Dissertations

Spring 5-2017

Population Structure, Connectivity, and Phylogeography of Two Balistidae with High Potential for Larval Dispersal: *Balistes capriscus* and *Balistes vetula*

Luca Antoni
University of Southern Mississippi

Recommended Citation

POPULATION STRUCTURE, CONNECTIVITY, AND PHYLOGEOGRAPHY OF

TWO BALISTIDAE WITH HIGH POTENTIAL FOR LARVAL DISPERSAL

BALISTES CAPRISCUS AND *BALISTES VETULA*

by

Luca Antoni

A Dissertation
Submitted to the Graduate School
and the School of Ocean Science and Technology
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved:

_____
Dr. Eric Saillant, Committee Chair
Associate Professor, Ocean Science and Technology

_____
Dr. Chet Rakocinski, Committee Member
Professor, Ocean Science and Technology

_____
Dr. Donald Johnson, Committee Member
Senior Research Scientist, Fisheries Research and Development

_____
James Franks, Committee Member
Senior Research Scientist, Fisheries Research and Development

_____
Dr. Karen S. Coats
Dean of the Graduate School

May 2017

Published by the Graduate School

THE UNIVERSITY OF
SOUTHERN
MISSISSIPPI.

ABSTRACT

POPULATION STRUCTURE, CONNECTIVITY, AND PHYLOGEOGRAPHY OF

TWO BALISTIDAEWITH HIGH POTENTIAL FOR LARVAL DISPERSAL

BALISTES CAPRISCUS AND *BALISTES VETULA*

by Luca Antoni

May 2017

The gray triggerfish (*Balistes capriscus*) and the queen triggerfish (*Balistes vetula*) are two exploited reef fish distributed in tropical and temperate shelf waters of the Atlantic Ocean and the Mediterranean Sea. Both species are highly sedentary as adults but disperse pelagic larvae for extended periods of time potentially allowing connectivity across long distances under the action of oceanic currents. In this work population structure, phylogeography, and migration patterns were examined in the two species and contrasted with predictions of larval transport based on surface circulation data. A total of 1,017 gray triggerfish from twelve sampling localities spanning the species distribution range were assayed at 17 homologous microsatellite markers and sequence variation at the ND4 mitochondrial gene. Four genetically distinct populations were detected including (i) a North Atlantic group that comprised the North American, European, and Northwest African populations, (ii) a Mediterranean group that was inferred to result from a recent colonization of the Mediterranean Sea by a small number of migrants of North Atlantic origin, (iii) a southeastern Atlantic group that included populations from the Gulf of Guinea and Southwest Africa, and (iv) a southwestern Atlantic group recently diverged from the southeastern group. Analysis of phylogeography supported long-term historical isolation of the South Atlantic and North Atlantic groups. Assignment tests and

ii

isolation-by-distance analysis supported the hypothesis of long-distance connectivity with evidence for transatlantic migrations and estimates of the mean dispersal distance of 740 km or greater. The high estimates of contemporaneous migration rates (up to 36.7%) may reflect increased larval transport in connection with the recent development of new *Sargassum* in the equatorial region. Analysis of high-density genome scans revealed homogeneous distributions of genetic variants among queen triggerfish from the French Antilles, the U.S. Virgin-Islands, and South Florida, suggesting high connectivity is occurring across the region.

These findings suggest that, in both species, local recruitment depends largely on the output of spawning populations located hundreds or thousands of kilometers away from a given stock, highlighting the need to conserve populations across each species' range in particular in areas where circulation patterns predict a low likelihood of incoming migrants.

DEDICATION

This dissertation is dedicated to my parents, Alberto and Morena, my sisters, Arianna and Viola, and my family who have been the most supportive during this long process. A special shout out to my amazing grandma Tina, who I am sure will always be looking over me, for all the recipes which made me survive this long away from Italian food.

Additionally, a big thank to all my friends (too many to name) and roommates, particularly Jennifer Frey Andres and Michael Andres, who were my family away from home and made this stay in Mississippi much more enjoyable and full of great memories.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I - INTRODUCTION

Understanding the factors that affect the abundance and distribution of marine organisms is essential in order to develop effective conservation strategies. While the dispersal capability and the availability of suitable habitats account for the potential range of occurrence of a species, the degree of connectivity among geographic populations is determined by the spatial distribution of migration events and the sizes of local stocks. Information on connectivity is required for practical management in many situations, including for example when assessing (i) whether local stocks can be replenished from other geographic populations and at what pace (Waples 1998), (ii) whether Marine Protected Areas can effectively contribute to recruitment in neighboring populations (Harrison et al. 2012), and/or (iii) whether local stocks could become depleted if fishing pressure is not regulated locally and there is little or no connectivity with neighboring stocks (Carvalho and Hauser 1995). From a conservation genetics point of view, connectivity is important because the balance between gene flow and the other forces affecting genetic change determine genetic diversity and population long-term sustainability. In particular, migration rates determine the extent of local adaptation and the rate of spread of advantageous mutations (Lenormand 2002).

Marine organisms have often been expected to display high degrees of demographic (and genetic) connectivity among geographic populations because of the open and connected nature of marine habitats (Avise 1998), but little is known on the actual extent of this connectivity in species that differ in habitat use and dispersal capability. Many marine species, including several reef-associated fishes, have pelagic eggs and larvae for a period of time that can last several weeks, sometimes months

1

(Robertson 1988; Mora 2004; Ball et al. 2007; Simmons 2008). Pelagic larvae are transported by oceanic currents during that ecophase, promoting exchanges (migration) among populations even when older life stages are sedentary. In particular, there is increasing evidence to support the hypothesis that geographic populations of reef fishes are connected even when they are separated by long distances and adults are highly sedentary, which suggests that larval dispersal is an effective medium of connectivity in those species. However, little is known yet about the spatial scale of this connectivity and its magnitude (e.g. how far migrants are dispersing and what portions of stocks are being replenished from local gene pools and from migrants, respectively). Also, even though the marine environment is open, migrations are not expected to be uniform in all directions and some barriers to gene flow likely exist. Two of the important questions regarding the connectivity of marine populations are (i) what is the nature of these structuring factors and (ii) what are their actual effects on demography and gene flow. The pattern and velocity of oceanic currents are a priori two major candidate factors expected to impact gene flow in species where larval dispersal is the main driver of connectivity. Numerical circulation models of surface currents have been used to predict larval dispersal envelopes at a regional scale (e.g. Roberts 1997; Cowen et al. 2006; Johnson et al. 2009) but comprehensive data on genetic connectivity are still lacking in many marine reef fishes.

Reef fishes that are highly resident as adults but have a prolonged pelagic larval dispersal phase are very valuable models in order to study larval mediated connectivity (Leis 2002). In these species, limited adult movement is expected to result in negligible adult migration, favoring the hypothesis that pelagic larval dispersal is the primary driver

of gene flow. The extended duration of larval transport provides increased power to demonstrate the role of oceanic currents in promoting gene flow, for example, if populations separated by large sections of unsuitable adult habitats, but connected by oceanic currents, do appear connected by gene flow (e.g. if population on the two sides of the Atlantic Ocean appear connected genetically through larval transport across the Atlantic via transatlantic oceanic currents) or if populations that are relatively close to one another but are not connected by currents appear isolated reproductively and genetically.

Connectivity from a genetic point of view is the exchange of genes between populations and formally occurs when individuals reproduce and spread their genes in another population than the population where they were born, thereby contributing to the gene pool of the recipient population (Lowe and Allendorf 2010). Studying larval mediated connectivity requires the deployment of population genetic models in order to account for other evolutionary forces impacting genetic variation in populations (namely genetic drift, mutation, and selection) when measuring gene flow. Phylogeography and population historical demography also need to be considered when interpreting genetic patterns in terms of gene flow. For example, populations recently expanding their geographic range may not yet show the signal of restricted gene flow due to insufficient time to accumulate genetic differences among geographic locations (Gold and Richardson 1998). Alternatively, large populations isolated for a long time may show levels of divergence similar to those in smaller populations connected by residual gene flow; distinguishing the latter two alternative scenarios effectively is essential in order to assess gene flow and genetic connectivity.

1.1 Experimental Approaches to Study Connectivity

Many different approaches have been used to study connectivity in marine

populations. One of the most straightforward methods is the analysis of the results of

mark-recapture experiments. Both artificial and natural tags can be used in these studies

(Thorrold et al. 2002). Physical tagging with artificial tags has been used in marine

species for a long time. However, as discussed earlier, many reef fishes disperse

primarily during early life stages when individuals are generally too small to be

physically tagged. Other artificial tags include fluorescent compounds, elemental tags,

and radioactive isotopes. Fluorescent labels and elemental tags (e.g. rare earth elements)

can be incorporated into calcified tissues of larvae by immersion. The former are detected

under lights of specific wavelength (Thorrold et al. 2002) while the latter can be detected

using spectrometric instruments. The use of radioactive isotopes is impractical in many

cases due to safety issues, but gamma-emitters like ^{85}Sr or ^{65}Zn have been used

successfully to tag embryos or yolk-sac larvae and subsequently detect tagged individuals

through gamma-ray detectors. These tags, particularly fluorescent compounds, and

elemental tags can be applied and detected on large numbers of individuals, and do not

affect individual growth and mortality rate which makes them very useful to study

connectivity, in particular when specimens cannot be tagged using traditional methods.

However, all these approaches require collecting large numbers of larvae in their natural

habitat and perform tagging in situ which can be challenging in practice. The most widely

used natural tags are genetic markers, which will be presented in detail later in this

chapter, and geochemical signatures in calcified structures such as the otoliths. The latter

approach is based on the fact that calcified structures "record" information on the

chemistry of the water where the fish is living. If enough differences in elemental composition and/or levels of stables isotopes occur between localities, these differences can be used to develop signatures of regional nurseries. When fish migrate from one region to another, the signature of the new region is incorporated in the outer rings of the otoliths while the origin of the fish can be determined using the signature present in the core (Thorrold et al. 1998). A limitation to the methods discussed above is that they provide information on fish movements but do not allow determining if fish that moved to a new region successfully spawned in the recipient population and contributed to its gene pool.

For this reason, inference on gene flow using genetic data and population genetics models is essential to develop a reliable assessment of connectivity. As discussed earlier, genetic patterns are influenced not only by gene flow but also by genetic drift, natural selection, and mutation. Historical events also influence patterns of genetic variation and incorrect account for population demographic history can lead to misinterpretation of genetic patterns when inferring gene flow.

1.2 Genetic Approaches to Study Population Structure and Gene Flow

Studies of population structure historically relied on the simple island model, initially developed by Fisher and Wright (Fisher 1930; Wright 1931, 1943). This model makes a lot of unrealistic assumptions such as considering an infinite number of populations with equal (and constant through time) effective size for all demes, identical and constant migration rates between all pairs of populations, no effects of mutation and natural selection on gene frequencies, migration-drift equilibrium, and equal reproductive success for migrants and resident individuals. Because these assumptions are highly

violated in real populations, this model cannot be used to make reliable inferences on migrations (Whitlock and McCauley 1999). The island model was extended to the continent island model (Haldane 1930; Nagylaki 1992), which assumes that the gene pools of small populations (islands) in the proximity of a large population (mainland) are largely determined by the migration from the mainland to the islands, but this derived model still carries most of the unrealistic assumptions of the classical island model.

Significant progress was made when the concept of metapopulation was introduced by Levins (1969, 1970). Levins defined a metapopulation as "a population of populations which go extinct locally and recolonize". One of the most important aspects of Levins' metapopulation model is that it considers the fact that species can go extinct locally and fluctuate demographically due to the patchiness of the habitat whereas, according to the continent island model, island populations reach an equilibrium state and never go extinct (Sale et al. 2006). This flexibility makes the metapopulation theory a better framework to study real natural populations.

Adapting the metapopulation concept from terrestrial to marine systems needs some additional specifications. Levins' model is appropriate for species with small discrete local populations but it is unrealistic for large populations which can vary in size but very likely will never go extinct (Kritzer and Sale 2004; Sale et al. 2006). Also, patterns of connectivity need to account realistically for the life history of the organism considered. In reef fishes that broadcast pelagic eggs, dispersal occurs primarily via passive larval transport and has been typically described to be limited in time corresponding to the completion of early developmental stages (e.g. Gaylord and Gaines 1999; Siegel et al. 2003), meaning that dispersal is limited to the neighborhood of the

individual dispersing gametes. Accordingly, a stepping stone model (Kimura and Weiss 1964) where larvae are more likely to reach habitat patches close to their origin is likely in these species. A continuous model where dispersal distance follows a continuous probability distribution may also be appropriate to describe these metapopulations. The expected pattern of population structure, in that case, is an isolation-by-distance model where genetic distance between populations increases as a function of geographic distance (Rousset 1997). Portions of the species range may also be isolated due to discrete barriers to gene flow, such as sections of unsuitable habitats or unfavorable current direction that would prevent larval exchange between specific sub-populations. In such a situation, genetic discontinuities would be expected. Thus, the overall pattern of population structure may be a combination of discrete units separated by genetic discontinuities due to barriers to gene flow and isolation-by-distance within units. Parameters of the landscape (e.g. temperature, salinity, or other landscape features) may also explain genetic structure if subpopulations have evolved to adapt to different habitat characteristics. Understanding marine metapopulations requires distinguishing between these patterns. Spatially explicit models (e.g. Dupanloup et al. 2002; Manni et al. 2004; François et al. 2006) and analyses accounting for features of the landscape (Hansen and Hemmer-Hansen 2007) are therefore required to elucidate genetic structure and appropriately investigate gene flow.

Historical events such as population bottlenecks, demographic expansion, or episodes of isolation followed by restricted migration or secondary contacts all impact genetic variation observed in population samples (Templeton 1998) and the signature of these historical events may impact inferences on gene flow based on genetic data due to

confounding effects on genetic patterns. In consequence, once population structure is elucidated, assessing the demographic history of the metapopulation is an essential step in order to analyze gene flow. Coalescent approaches studying the genealogy of alleles at genetic markers (Kingman 1982) are widely used to infer historical demography of populations based on genetic data. Among the available analytical methods, the isolation with migration model (Hey and Nielsen 2004) and approaches allowing evaluating the likelihood of alternative historical scenarios using approximate Bayesian computation algorithms (Cornuet et al. 2008; Wegmann et al. 2010) are most suited to account for the complexity of potential alternative hypotheses. Along with inferring historical scenarios for populations of the studied species, these coalescent methods allow estimating long-term historical gene flow jointly with other population genetic parameters such as the effective size of demes or population growth rates.

Contemporaneous gene flow can be estimated using assignment methods (e.g. Bayesian clustering, Cornuet et al. 1999; Pritchard et al. 2000; Paetkau et al. 2004; François et al. 2006), which allow detecting F0 migrants with genotypes assigned to one population but sampled in another one as well as admixed individuals or putative hybrid descendants from migrants. These assignment methods perform well when populations are sufficiently differentiated and all the potential sources of migrants are accounted for (i.e. when the species range is comprehensively sampled). A robust initial description of population structure is, therefore, essential in order to infer contemporaneous gene flow with confidence using assignment approaches. Gene flow can also be estimated by deriving the dispersal parameters of an isolation-by-distance relationship when genetic patterns are consistent with this model (Rousset 1997). Inferences under isolation-by-

distance are now improved by the maximum likelihood approach that was recently developed to estimate jointly effective population size and dispersal under this model (Rousset and Leblois 2012).

The outcome of the approaches described above is the identification of divergent units separated by genetic discontinuities and of isolation-by-distance within those units when these patterns are present. Coalescent analyses and assignment methods also provide information on historical and current rates and patterns of gene flow and on the ancestry and demography of populations. These inferences all rely on the assumption that natural selection is not impacting genetic structure and genetic patterns only reflect the neutral processes of genetic drift and gene flow.

However, the degree of divergence among populations can be affected by natural selection and local adaptation. This source of genetic variation affects specific regions of the genome and its analysis, therefore, requires the development of genetic markers reflecting the signature of selection (non-neutral markers). The affected genomic regions (and associated non-neutral markers) are expected to display increased divergence as compared to the rest of the genome and thus this divergent selection can be studied during an outlier analysis seeking to identify loci or genomic regions showing significantly higher levels of divergence among populations than neutral regions (Beaumont and Nichols 1996; Foll and Gaggiotti 2008; Lotterhos and Whitlock 2014). Identification of genomic regions and genetic loci affected by selection is essential in order to describe and study the effects of local adaptation but also to separate neutral loci unaffected by selection for use during the study of the neutral processes of gene flow and genetic drift.

9

The combination of the analyses of neutral and non-neutral variation thus allows for a comprehensive assessment of a metapopulation system, in particular, drawing inferences on historical and contemporaneous connectivity. The data required are genetic surveys of geographic populations at molecular markers. Therefore, adequate sets of molecular markers are needed in order to deploy these methods successfully and provide reliable estimates of gene flow and other dispersal parameters.

1.3 Molecular Approaches to Characterize Gene Flow

Population structure has historically been assessed using a variety of approaches including comparisons of phenotypic and meristic characters among geographic populations (Swain and Foote 1999). However, this method provides limited information on genetic stock structure because phenotypic differences can reflect, for a large part, effects of environmental differences among regions such as varying habitat quality or exploitation rates. For these reasons, identification of genetic stocks and assessment of gene flow need to be based on data collected at molecular markers and the deployment of population genetics models (Allendorf and Luikart 2007).

DNA markers assessing directly variation of the DNA sequence at selected genetic loci most accurately characterize genetic diversity and are best suited for population genetic studies (Wright and Bentzen 1994; Sunnucks 2000; Allendorf and Luikart 2007). Mitochondrial DNA (mtDNA) markers have been widely used for population genetics and phylogeography studies. The maternal inheritance and haploid nature of mtDNA lead to increased effects of genetic drift thereby potentially increasing the resolution of mtDNA datasets relative to those based on nuclear markers; mtDNA markers also potentially allow examining differences between sexes in dispersal and

genetic demography (Prugnolle and De Meeus 2002). In addition, genetic assay of the mitochondrial genome can easily be developed using readily available universal primers (Sunnucks 2000), which allows selecting regions with different evolutionary rates and refining inferences on the status of clades showing various levels of divergence (Avise 2004). A major drawback of mtDNA is that it represents only one locus, which limits inference power. Also, while mtDNA has been widely used to describe phylogeography under the assumption of neutrality, there is increasing evidence that the mitochondrial genome is under selection, which potentially compromises phylogeographic inference (Nachman et al. 1994; Rand 2001; Ballard and Whitlock 2004). Microsatellites are arrays of short motifs of DNA composed of 2 to 6 nucleotides aligned in tandem (Detrich et al. 1998). Advantages of microsatellites over other markers for population genetics studies include high polymorphism owing to elevated mutation rates, straightforward assay and scoring, and Mendelian codominant inheritance. These qualities result in high power for the resolution of fine-scale patterns of population genetic structure, assessment of migrations and gene flow, estimation of effective population size, and inferences on demographic history (Frankham et al. 2002). Microsatellites can be multiplexed for simultaneous assay in the same reaction allowing efficient high throughput assay of large numbers of individuals. This allows examining large sample sizes during genetic studies, which is preferable in species with high gene flow and low levels of divergence among populations as is expected in the selected species (see below). These qualities have led to a high popularity of microsatellites for population genetic studies until the costs of methods based on Next Generation Sequencing (NGS) decreased sharply in recent years. The assumed neutrality of microsatellite markers is, however, limiting, considering that

11

the low marker density that can be achieved for large-scale surveys (15-20 loci) makes it

very unlikely that any of the loci is linked to regions of the genome undergoing selection,

so that microsatellite datasets are inappropriate to characterize variation under selection

and local adaptation although accounting for this source of variation is essential for

conservation (Gold et al. 2001; Funk et al. 2012). In recent years the progress of NGS

technologies has led to the development of new methods of high throughput genotyping.

Among those, the methods derived from the Restriction Associated DNA (RAD)

sequencing protocol (Baird et al. 2008; Peterson et al. 2012) have become extremely

popular because the costs of analysis per sample is now compatible with relatively large-

scale studies and the large numbers of loci that can be recovered result in high-density

genome scans (several thousands of loci in typical studies, Avise 2010; Helyar et al.

2011) allowing studying both neutral and selected regions of the genome. These

genotyping methods can be applied directly *de novo* in non-model species, although

inferences are improved if a reference genome is available through implementation of

sliding window analysis (Hohenlohe et al. 2010). While a variety of polymorphisms can

be studied through NGS and genotyping by sequencing approaches, Single Nucleotide

Polymorphisms (SNPs) have been the most widely used in molecular ecology so far

(Payseur and Jing 2009; Helyar et al. 2011). SNPs are distributed throughout the genome

in both neutral and selected regions such that their analysis during genome scans obtained

via sequencing approaches allows for a comprehensive assessment of genome-wide

genetic variation. Analysis of SNPs in non-model organisms, where a reference genome

is not available, is limited by the lack of information on physical linkage and genomic

position of specific SNPs, but inferences can still be made on neutrality using outlier

analyses (Foll and Gaggioti 2008; Whitlock and Lotterhos 2015). Another advantage of SNPs is that they can be described using a simple mutation model (Helyar et al. 2011).

1.4 Considerations on Sampling Designs

It is very difficult to evaluate a priori the power of a dataset to infer population genetics parameters because of the very large number and multidimensional aspect of alternative scenarios that need to be evaluated. However, a comprehensive coverage of the distribution area of a species, when possible, is valuable in order to infer correctly parameters, in particular, accounting for gene flow from all possible sources of migrants in the estimation (Wang and Whitlock 2003; Beerli 2004; Slatkin 2005).

Analyses of spatial genetic variation based on 50 to 100 individuals per geographic location usually allows resolving population structure and estimating genetic distance among populations, assuming a large number of loci is used (Ruzzante et al. 1998; Cornuet et al. 1999). Temporal replication of sampling is also recommended in order to ensure that observed differences are not due to a sampling artifact or temporal fluctuations in local populations (Waples 1998). This can be achieved for example by surveying distinct cohorts. Regarding inferences on historical demography, simulation studies of coalescent-based approaches indicate that 25 to 30 individuals per population are sufficient to recover demographic signals in a wide range of situations (Beerli and Felsenstein 1999; Cornuet et al. 2008). Thus, the sample sizes required to describe spatial genetic variation are more than sufficient to conduct phylogeographic analyses and estimate historical demography parameters and gene flow.

1.5 Model Species Chosen

The two species chosen for this project are the gray triggerfish (*Balistes capriscus*) and the queen triggerfish (*Balistes vetula*). These two species are exploited by fisheries in various parts of their range yet they both lack the essential demographic data, including information on population structure and connectivity, needed to develop management plans. Both species settle on benthic habitats as juveniles and remain highly sedentary on these habitats for the rest of their life. In consequence, most of the connectivity can be hypothesized to be due to the passive transport of pelagic larvae and young juveniles under the action of oceanic currents, making the two species very well suited to study the connectivity that results from pelagic dispersal. In addition, the extremely long duration of larval transport in these species (> 60 days for *B. vetula*, Robertson 1988 corrected by Lindeman et al. 2000; up to 4-7 months for *B. capriscus*, Simmons 2008) and their broad geographic distributions allow making and testing predictions on migrations between regional populations separated by large sections of unsuitable adult habitats but connected by oceanic currents, or, alternatively, restrictions to gene flow between populations that are proximal geographically but not well connected via oceanic currents. A comprehensive analysis of population structure and gene flow will provide useful data in order to develop management strategies and also evaluate the importance of pelagic transport-mediated gene flow.

1.6 Research Objectives

The primary objective of this work was to investigate population structure and gene flow in the two studied species and determine whether patterns were consistent with predictions of larval transport by predominant surface currents. A comprehensive

analysis of spatial genetic variation was conducted in order to identify units separated by genetic discontinuities, when they occurred, and isolation-by-distance. Homologous microsatellites for gray triggerfish were developed as an initial step and are presented in Chapter II. Microsatellites were used as the primary markers for the study of population structure in gray triggerfish in Chapter III and patterns were contrasted to those obtained using mtDNA markers, in particular, to assist during phylogeographic analysis. Contemporaneous gene flow was also analyzed using assignment approaches introduced in this chapter and considering the main units identified during population structure analysis. A preliminary phylogeographic analysis was conducted in order to investigate historical processes shaping population structure. The parameters of dispersal were also estimated in Chapter IV taking advantage of the continuous sampling available for the U.S. population and exploiting isolation-by-distance models. Queen triggerfish stock structure and connectivity were investigated using genome scans generated using the RAD-sequencing methodology in Chapter V. The analysis also allowed examining non-neutral variation in the species through an outlier analysis. The overall analysis provided information on stock structure, rates of migrations, and the spatial scale of connectivity in the two species. The patterns and rates of gene flow were overlaid with predictions of larval transport based on oceanic currents in order to evaluate concordance. Because both species display extreme life histories, characterized by extended larval pelagic periods and limited adult movements, oceanic current patterns were hypothesized to constrain gene flow among localities leading to specific hypotheses that were tested in different parts of the species' range as outlined in Chapters III and V.

The implications of findings on connectivity and gene flow for the assessment and management of U.S. and foreign populations of these two species are discussed in concluding remarks.

CHAPTER II – DEVELOPMENT OF HOMOLOGOUS MICROSATELLITE

MARKERS FOR GRAY TRIGGERFISH

2.1 Introduction

The acquisition of suitable panels of molecular markers is critical in order to

achieve reliable inference during conservation genetic studies. Microsatellites or Single

Sequence Repeats (SSR) were discovered in 1981 by Miesfield et al. (1981) and were the

most widely used markers in population genetic studies during the following two decades

(Jarne and Lagoda 1996; Zane et al 2002; Oliveira et al. 2006). Microsatellites are arrays

of short DNA sequence motifs (2-6 bp) repeated multiple times in a head to tail fashion.

They are widely distributed throughout the genome and are generally assumed to belong

to the non-coding fraction of DNA. The development of a specific assay for a target

microsatellite is achieved by designing Polymerase Chain Reaction (PCR) primers in the

specific DNA region flanking the repeat array which allows for the specific amplification

of microsatellite alleles and the determination of their sequence. The obtained assay can

then be used to survey sequence variation among alleles at the microsatellite in different

individuals and populations. Microsatellites possess numerous advantageous

characteristics over other classes of genetic markers for population genetic inference.

They are assumed to be non-affected by natural selection and are therefore appropriate to

study neutral processes including genetic drift and gene flow. In addition, microsatellites

are characterized by a high degree of polymorphism due to elevated mutation rates, and a

Mendelian codominant inheritance, both contributing to their high power of inference for

population genetics and parentage studies. Sequence variation at microsatellite alleles can

be easily assayed by simple sizing of PCR amplicons on an electrophoresis gel. Multiple

17

microsatellites can also be assayed in a single reaction in a multiplex protocol (Renshaw et al. 2006), allowing cost-effective genotyping of large numbers of samples.

While microsatellites have been successfully transferred among species (e.g. Gold et al. 2009; Karlsson et al. 2009; Renshaw et al. 2012), the transfer is usually restricted to closely related taxa, and heterologous assays tend to be more prone to scoring errors and null alleles such that the development of homologous microsatellites is usually preferable in order to achieve reliable interpretation of electropherograms and scoring of genotypes. Development of homologous microsatellites *de novo* has become much more efficient thanks to recent protocol developments including the enrichment of genomic libraries (Zane et al. 2002) and the application of next-generation sequencing technologies (Malausa et al. 2011; Castoe et al. 2012). The enrichment procedure involves the capture of microsatellite-rich fragments of DNA by hybridization prior to cloning and screening of genomic libraries for SSRs and has been shown to significantly increase the proportion of positive clones actually carrying a microsatellite (Kijas et al. 1994).

The development of specific assays for new microsatellites discovered during screening of genomic libraries is followed by testing of the obtained assays and selection of markers that show desirable characteristics for population genetic inference. Loci showing high levels of polymorphism, consistent amplification success across individuals and populations, reproducible assay and scoring, and easy-to-interpret electropherograms are preferred. Loci, where scoring artifacts such as null alleles, large allele dropout, or stuttering are detected, are prone to genotyping errors and need to be omitted for inference (Van Oosterhout et al. 2004). Null alleles are alleles that fail to amplify during PCR leading to the false inference of a homozygote genotype for individuals carrying a

18

null allele. Large allele dropout occurs when amplification of alleles that contain large numbers of repeats is weaker than that of shorter alleles, ultimately leading to the non-detection of the large alleles in heterozygotes carrying both a large and a short allele. Large allele dropout leads to an apparent (and erroneous) deficit in heterozygotes for larger alleles. Stuttering is due to slippage of the DNA polymerase during amplification and leads to multiple peaks resulting from the amplification of a single microsatellite allele. This phenomenon leads to difficulties distinguishing heterozygotes for alleles differing by one repeat from homozygotes showing stutter bands. Because all these artifacts lead to errors in estimation of allele and genotype frequencies and would compromise population genetic inference, an initial survey of candidate microsatellites in a collection of specimens from one population is necessary in order to establish the occurrence of such artifacts and the suitability of loci for population genetic inference.

Panels of 15 or more microsatellites are usually needed to achieve reliable inference on population structure (e.g. Koskinen et al. 2004) and demographic history. While greater numbers of microsatellites would be expected to improve the power of datasets, the number of microsatellites that can be deployed is limited by the technical feasibility and cost-effectiveness of genotyping of large numbers of samples and the availability of specific primers to amplify microsatellite loci. Fifteen to 20 loci can usually be combined in a moderate number (4 or 5) of multiplex panels for high throughput assay compatible with genotyping of moderately large numbers of samples (usually several hundreds) needed to study population structure in species such as marine fishes where spatial divergence is low.

To date, no molecular markers have been developed for the gray triggerfish or any closely related species. The objective of this chapter was to develop a panel of homologous microsatellites markers from gray triggerfish enriched genomic libraries and select a minimum of 15 loci showing consistent amplification success and reliable scoring across samples. Multiplex assays were also optimized for the selected loci.

2.2 Material and Methods

2.2.1 Library Construction and Sequencing

Microsatellite-enriched genomic libraries were developed following the protocol described by Bloor et al. (2001) incorporating modifications by Zane et al. (2002) and John and Quinn (2008). Muscle, liver, and spleen tissue samples from one gray triggerfish specimen were collected immediately after death and snap frozen by direct immersion in liquid nitrogen. DNA was isolated from the three tissues using the DNA-easy (Qiagen Inc., Valencia, California) extraction kit following protocols from the manufacturer. The degree of degradation of the DNA obtained from each tissue was evaluated on an agarose electrophoresis gel and DNA concentration was determined using a NanoDrop Spectrophotometer. The highest quality DNA was obtained from the spleen tissue sample and was selected to prepare the genomic libraries. DNA was digested with the restriction enzymes HaeIII, AluI, and RsaI, and ligated with blunt-end adaptors described in Bloor et al. (2001) for cloning. The obtained library was size-selected to retain fragments ranging in size between 400 and 1000 bp. Genomic sequences obtained from shorter fragments have a low probability to include both the microsatellite arrays and portions of their flanking regions as necessary to design PCR primers, while fragments longer than 1,000 base pairs cannot be sequenced entirely using

20

forward and reverse sequencing which prevents determination of the microsatellite array

sequence and ultimately the reliable evaluation of the suitability of the locus as a genetic

marker. In order to size-select the library, the product of the digestion reaction was run on

an agarose gel containing Ethidium Bromide and a size standard. The DNA present in the

agarose gel after electrophoresis was visualized on a U.V. Transilluminator which

allowed identifying the portion of the gel containing fragments between 400 and 1000 bp,

based on the location of the corresponding size standards, and excising it from the gel.

DNA fragments were isolated from the excised fraction using the QIAquick gel

extraction kit (Qiagen). A first library was enriched with microsatellite arrays featuring a

CA dinucleotide motif. To do so, the DNA was hybridized to a $(CA)_{12}$ probe to select

fragments containing CA microsatellite arrays. Similarly, a second library targeting

TAGA microsatellites was generated by hybridizing the digested and size selected DNA

to a $(TAGA)_9$ probe. Subsequently, libraries were mixed with streptavadin-coated

magnetic beads (Dynabeads M-280, Life Technologies) in order to capture the

microsatellite-containing DNA fragments. Hybridization temperature was 50°C for both

probes and followed the protocol described in Bloor et al. (2001). The enriched library

was amplified through 20 PCR cycles, and amplified fragments were ligated into pGEM-

T cloning vector, and transformed into *Escherichia coli* competent cells (JM-109 strain,

Promega Inc., Madison, Wisconsin) as described in Bloor et al. (2001). Colonies of *E.

coli* were then grown on LB-Ampicillin X-Gal agar plates. Colonies with plasmids are

ampicillin-resistant and grow on the selective medium with ampicillin. Colonies with

inserts display a white color indicating successful ligation of an insert within the gene

coding for the β-galactosidase enzyme (colonies with no insert express the gene and

metabolize X-Gal resulting in a blue color). White colonies were hand-picked with a

sterile toothpick and incubated separately in LB medium buffer. Clones inserts were

amplified through PCR using the universal primers M13 (-21) Forward and M13 -22mer

Reverse that frame the cloning site. Amplicons were then sequenced using the same

forward and reverse primers on a 3730XL automated sequencer (Applied Biosystems,

Foster City, California) at the High Throughput Genomic Unit of Seattle (WA).

2.2.2 Primer Design and Optimization

The sequences obtained from the clones were aligned and edited using the

software SEQUENCHER v.4.10.1 (Gene Codes Corporation, Ann Arbor, Michigan). Edited

clone sequences were screened for microsatellite arrays featuring multiple repeats of the

probe sequences used to enrich the libraries (CA or TAGA). PCR primers were designed

in priority for both perfect and imperfect microsatellite arrays containing 12 to 25 repeats.

Loci with less than 12 repeats are not likely to be polymorphic while those containing

more than 25 repeats are more difficult to amplify consistently and are more prone to

scoring errors. Primers were designed using the software PRIMER3 v.0.4.0 (Rozen and

Skaletsky 2000) available at http://frodo.wi.mit.edu/ and used to amplify the target

microsatellites through PCR. PCR reactions were conducted in a 10 µL reaction volume

containing 10–20 ng of genomic DNA, 4 pmol of each primer, 0.5 U of TAQ polymerase

(Promega), 2 nmol of deoxynucleotide triphosphates (dNTPs), 15 nmol of MgCl2, and 1

×buffer (Promega). Amplification by PCR consisted in an initial denaturation at 95°C for

3 min, 35 cycles of 95°C for 30 s, T_a (annealing temperature) for 30 s, 72°C for 45 s, and

a final extension of 15 min at 72°C. T_a was decreased during the amplification reaction

according to the touchdown principle: T_a was 60°C for the first 7 cycles, 57°C for the

next 7 cycles, and 53°C for the remaining 21 cycles. Six to 8 samples were used in these

amplification tests. Markers showing consistent amplification and polymorphic PCR

products as evaluated on a high-resolution NuSieve-GTG agarose gel (Lonza) were

selected for further testing. The 5' end of one of the two PCR primers was labeled with

one of the three fluorescent dyes 6-Fam, Hex, or Ned, to allow detection of labeled PCR

products on an automated sequencer ABI-377XXL (Applied Biosystems). Primers

producing PCR products in the same size range were labeled with different fluorescent

dyes to allow simultaneous assay and detection in the same multiplex assay. Labeled

primers were test-amplified at different annealing temperatures ranging from 54°C to

62°C to determine the optimal annealing temperatures for each microsatellite. PCR

products were run on a ABI-377XXL sequencer to verify polymorphism and evaluate

consistency of amplification and interpretability of electropherograms.

Electropherograms were analyzed in the software GENESCAN and allele calling was

performed in the software GENOTYPER.

2.2.3 Assessment of the Suitability of Markers for Population Genetic Inference

The selected microsatellites were tested by genotyping 35 individuals collected

offshore the Louisiana coastline during the 2008 and 2009 fall SEAMAP surveys of the

U.S. National Marine Fisheries Service. DNA was extracted using a phenol-chlorophorm

method (Sambrook et al. 1989) and microsatellites were amplified by PCR using the

optimal temperature cycling protocol developed as described above. The obtained PCR

products were assayed on an automated sequencer ABI-377XXL. Conformance of

genotypic proportions to Hardy-Weinberg expectations was tested for each microsatellite

using exact tests in the software GENEPOP v.4.2 (Raymond and Rousset 1995; Rousset 2008a). GENEPOP was also used to perform Linkage Disequilibrium (LD) tests of the null hypothesis that alleles at different loci were associated at random (Slatkin 2008). The exact probability value for each test was estimated using a Monte Carlo approach based on 5,000 dememorizations, 500 batches, and 5,000 iterations per batch. The False Discovery Rate (FDR, Benjamini and Hochberg 1995) procedure was used to correct significance thresholds for probability values in order to control for multiple testing. The number of alleles (A), and the observed (H_o) and expected heterozygosity (H_e) statistics were calculated using the software ARLEQUIN v.3.01 (Excoffier et al. 2005). The presence of scoring artifacts such as null alleles, large allele dropout, or stuttering was tested using the software MICROCHECKER v.2.2.3 (Van Oosterhout et al. 2004). Loci showing evidence of scoring artifacts were removed from the panel of markers.

2.2.4 Multiplex Assay Optimization

Multiplex panels for the simultaneous assay of multiple microsatellites in the same reaction were optimized using a procedure similar to that described in Renshaw et al. (2006). Candidate panels of microsatellites compatible for multiplex amplification were first identified. In order to ensure that all microsatellites incorporated in a panel can be scored unambiguously on an electrophoresis gel, loci combined in the reaction need to be labeled with different fluorescent dyes unless they display non-overlapping allele distributions. In the latter case, PCR primers may be labeled with the same fluorescent dye and microsatellite alleles for the two loci labeled with the same dye are scored in their respective allele size range. Potentially compatible sets of microsatellites identified based on these principles were used in test amplification reactions conducted on the DNA

24

of 6 to 8 individuals. When the markers selected in a multiplex differed in their annealing temperature, amplification tests employed the touchdown protocol where annealing temperature is progressively decreased during consecutive amplification cycles in order to include cycles at annealing temperatures compatible with all loci present in the reaction. Optimization aimed to develop panels of up to 4 or 5 microsatellites in order to increase the cost-effectiveness of assays for large numbers of samples.

When the optimal panel of markers for each multiplex was identified, the amount of primers needed in each PCR reaction cocktail was optimized so that each marker in a multiplex would show similar intensity on acrylamide gels for simultaneous scoring. Final reactions were run in a total volume of 5.6 µL with amount of reagents adjusted to reflect proportions presented in section 2.2.2 for a reaction volume of 10 µL.

2.3 Results (Antoni and Saillant 2012)

Screening of the enriched libraries revealed 206 colonies with inserts among a total of 501 transformed colonies. The inserts of the positive clones were amplified through PCR and all obtained amplicons were sequenced. A total of 125 of the sequenced clones (122 for the CA library and 3 for the TAGA library) contained microsatellites bearing more than 9 repeats. PCR primers were designed and tested for 50 of those markers. Twenty-five primer pairs successfully amplified a PCR product that appeared variable in size among individuals when run on an agarose gel. One of the two primers for each of the 25 pairs was then labeled with a fluorescent dye for further evaluation of the scorability of amplicons on an acrylamide gel. Consistent amplification of a scorable PCR product was achieved for 21 of the primer pairs. The characteristics of the 21 microsatellites based on a sample of 35 gray triggerfish are presented in Table 2.1. Per

locus estimates of diversity, indices ranged between 4 and 27 for *A*, and 0.502 and 0.967

for H_e respectively. Genotypic proportions did not depart significantly from H–W

equilibrium expectations except for loci BC20 and BC36. Six out of 210 tests of linkage

disequilibrium were significant before FDR correction of critical probability values.

None of the tests remained significant after FDR correction.

Table 2.1

Microsatellite Markers Characteristics

Locus	Repeat motif	Primer sequence 5'-3' (label)	T_a (°C)	A	Allele-size range	H_o	H_e	P_{HW}	GenBank accession no.
BC1	$(TAGA)_{14}$	F: TGCCTGGAGGAGAAAAGAGA R: AATTCAAAACCAGGCCACAC (NED)	62	23	138-266	0.8	0.88	0.116	JQ285974
BC2	$(GT)_{26}$	F: TACATGTGGGCAATGTGGTT (FAM) R: CTGGAATCTCCAGCATCTCA	62	22	150-200	0.89	0.95	0.308	JQ285975
BC3	$(CA)_3CGCTGGC$ $G(CA)_{12}CT(CA)_3$	F: CATTTGGTCTCCTTGCAAAA (FAM) R: ACAGACCCTGCTTTGCTTGT	60	5	146-156	0.51	0.6	0.702	JQ285976
BC9	$(CA)_{13}$	F: GACGGATGAAACGTGAAGGT (FAM) R: CCTGGACCTGCTAAGTCTGC	60	22	204-256	0.97	0.95	0.149	JQ285977
BC13	$(CA)_{15}$	F: GCGCCATTTTATTCTTGGAA (HEX) R: TGGTATGTGGCGGTCAATAA	60	21	227-277	0.86	0.89	0.799	JQ285978
BC14	$(GT)_{32}GC(GT)_4$	F: CAGAGGAATTCCACCTGATGA (FAM) R: CAGTCTCCAGATGGACCACA	60	18	101-197	0.83	0.92	0.053	JQ285979
BC16	$(GT)_{12}$	F: CAGCAATGTGGGGGTAAAGT (FAM) R: GCTTATGCGCTCAAAAGTCC	58	4	249-257	0.49	0.5	0.508	JQ285980
BC17	$(CA)_{18}$	F: TGGCTGGAATCAATGAACAA (NED) R: TTCACTCTGCTCATCCCACA	56	12	205-253	0.63	0.61	0.513	JQ285981
BC19	$(GT)_{14}CT(GT)_2$	F: TTGTTCTCACCTCCCCTCTG (FAM) R: CTCTTTGACATGCCCACAAA	62	10	201-225	0.71	0.72	0.64	JQ285982
BC20	$(CA)_{19}$	F: TCTGTGTTTCGGAACGTTTG (FAM) R: TCCCGCCTTTTGTGTAGTTC	62	13	165-207	0.44	0.9	0	JQ285983
BC25	$(GT)_{22}$	F: TCCCGCCTTTTGTGTAGTTC (HEX) R: TTCAGAAATGCTGCTGGATG	60	13	143-175	0.8	0.88	0.186	JQ285984

Table 2.1 (continued).

Locus	Repeat motif	Primer sequence 5'-3' (label)	T_a (°C)	A	Allele-size range	H_o	H_e	P_{HW}	GenBank accession no.
BC26	$(CA)_{21}$	F: GGCGTGAAATCATCTAGGACA (HEX) R: TCAGCAAGCGCTTCTCACTA	62	16	188-244	0.89	0.88	0.985	JQ285985
BC27	$(GT)_{13}$	F: TAAGTAACCGCAGCCTCACC (HEX) R: ATACTTCCACAGCCGAGACG	62	6	165-177	0.71	0.74	0.65	JQ285986
BC34	$(GT)_{19}$	F: GCAGGTCACACACTCCAGTC (HEX) R: TGTGCTCCCATCACACAGTT	62	14	186-272	0.83	0.87	0.336	JQ285987
BC36	$(GT)_{12}$	F: GGGGAACAAGGAGAGAGGAG (HEX) R: CTGGTGCGTCTGTGGTCTTA	62	7	264-278	0.63	0.73	0.003	JQ285988
BC41	$(GT)_{12}$	F: TGGAATTGCTCGAGGCTAAG R: TGATATCAGCGGACCTGTCA (HEX)	62	7	128-148	0.54	0.51	0.241	JQ285989
BC44	$(CA)_4CG(CA)_{13}$	F: TTTGTCCTGCAGAACCTCCT R: ACATTTGCAATCAGCAGCAC (FAM)	58	18	249-299	0.89	0.93	0.233	JQ285990
BC45	$(CA)_{16}CG(CA)_8$	F: GCCTTCAAGACCAGGAAGTG (FAM) R: TGTTTTGATCAGAGGCGTTG	54	10	251-271	0.83	0.81	0.553	JQ285991
BC46	$(GT)_{28}$	F: TGAGAGTGAAATGGGTCACG R: TGAGCACGCTGGTTTGTTTA (FAM)	58	27	112-184	0.91	0.97	0.227	JQ285992
BC47	$(CA)_{28}TA(CA)_4$	F: ATCAGCTGCCTCCGTTTG (HEX) R: ATTAACCACCAGCCAACTGC	56	26	108-174	0.89	0.96	0.066	JQ285993
BC49	$(GT)_{14}$	F: TCATCAACTGTCGGCTCTCA (FAM) R: CAACAGTCAGCCGTCTAGCA	56	7	244-270	0.62	0.5	0.933	JQ285994

Characteristics of 21 microsatellite markers developed for gray triggerfish. Summary statistics are based on 35 specimens caught offshore the Louisiana coast (United States). T_a=specific annealing temperature, A=number of alleles, H_o=observed heterozygosity, H_e=expected heterozygosity, P_{HW}=probability of departure from Hardy–Weinberg equilibrium.

Analyses in MICROCHECKER indicated possible occurrence of null alleles and/or stuttering at locus BC20. There was no evidence for artifact impacting scoring at locus BC36. In addition, locus BC9 displayed alleles differing in size by one base pair only, thus departing from the pattern of variation in number of repeats (2 or 4 bp difference between consecutive alleles) expected for microsatellites. Loci BC20 and BC9 were not included in multiplex assay development.

All microsatellites but BC1 and BC16, for a total of 17 markers, were successfully incorporated in 4 multiplex panels. The composition of the 4 multiplex panels and optimum annealing temperatures used in multiplex PCR are presented in Table 2.2. PCR reactions for BC16 were performed in simplex and combined with multiplex 40 for electrophoresis on automated sequencer. BC1 could not be incorporated in the final panel because its range in allele size and fluorescent label were incompatible with electrophoresis within any of the multiplex.

Table 2.2

Multiplex Polymerase Chain Reaction Protocols

Multiplex #	Marker ID	Fluorescent dye	Primer quantity (pmol)	T_a
37	BC44	FAM	0.49	
	BC46	FAM	0.49	7 cycles T_a = 62 °C
	BC27	HEX	0.38	7 cycles T_a = 60 °C
	BC34	HEX	0.44	21 cycles T_a = 58 °C
	BC19	FAM	0.41	
39	BC47	HEX	0.57	
	BC13	HEX	0.40	
	BC17	NED	0.40	35 cycles T_a = 56 °C
	BC3	FAM	0.52	
	BC45	FAM	0.32	
40	BC14	FAM	0.73	
	BC36	HEX	0.73	35 cycles T_a = 60 °C
	BC25	HEX	0.73	
44	BC26	HEX	0.60	
	BC41	HEX	0.60	7 cycles T_a = 62 °C
	BC2	FAM	0.60	7 cycles T_a = 60 °C
	BC49	FAM	0.40	21 cycles T_a = 58 °C
Simplex	BC16	FAM	2.20	35 cycles T_a = 58 °C

Multiplex PCR protocols for 18 homologous microsatellites developed for gray triggerfish. Primer quantities (pmol) are given for a 5.6 µL total reaction volume. T_a: specific annealing temperature.

2.4 Discussion

In this work, specific assays for 21 new homologous microsatellites for gray triggerfish were developed. Testing of assays on a sample of 35 wild triggerfish revealed that 19 of the loci could be reliably amplified and scored across samples and showed no signs of scoring artifacts. One locus showed 1 bp intervals between consecutive alleles inconsistent with the expected variation at microsatellites. This locus would be difficult

to score reliably due to the limited resolution of acrylamide gels for peaks differing by one bp only. A second locus showed evidence for null alleles segregating in the sampled populations and was rejected from the final panel. The remaining loci showed polymorphism levels similar to those reported for marine species by DeWoody and Avise (2000) and multiplex panels could be optimized for the assay of all but one of the 19 markers. The 18 markers could be amplified in 5 PCR amplification reactions and assayed in only 4 different electrophoresis gels allowing cost effective assay of large numbers of samples as needed in the following chapters of this dissertation. The pairwise test of independence of genotypes between loci did not reveal evidence of significant linkage between any of the microsatellites. The test is limited when a large frequency of double heterozygote genotypes is present in a dataset as is expected in highly polymorphic microsatellites (Waples 2015). Thus, weak levels of linkage might not have been detected during the analysis. However, the non-significant outcomes of all the pairwise comparisons performed indicates that none of the loci is closely linked and that the 18 microsatellites can be treated as independent during population genetic analysis, further suggesting that the obtained marker system is informative on several of the 22 balistidae chromosomes (Sá-Gabriel and Molina 2005).

The protocol used in this study based on enrichment of a genomic library followed by cloning and screening of transformants (Bloor et al. 2001; Zane et al. 2002; John and Quinn 2008) is more time-consuming than newer techniques utilizing next generation sequencing (Castoe et al. 2012) but has the advantage of accessing the complete repeat array of a microsatellite and its flanking regions during sequencing which enables improved prediction of the size of the PCR products and overall a lower

rate of rejection during primer' testing. A better control of the expected size of each

microsatellite marker at the stage of primer design allows evening the proportion of

microsatellites assays yielding small-, medium-, and large-size amplicon respectively,

thereby avoiding excessive overlaps in size range and improved multiplexing

opportunities.

The 18 markers were used in this work to investigate spatial genetic variation in

Chapters III and IV of this dissertation.

CHAPTER III – POPULATION STRUCTURE, CONNECTIVITY, AND
DEMOGRAPHIC HISTORY OF GRAY TRIGGERFISH IN THE ATLANTIC OCEAN

3.1 Introduction

Understanding the structure and connectivity of marine metapopulations is
essential in order to develop effective conservation strategies. Marine fish populations
have often been assumed to show high levels of connectivity because of the open nature
of marine habitats and the high dispersal potential of many species (Avise 1998). This
idea has however been challenged by findings in several studies where population
structure and bio-complexity were evidenced in relation to habitat characteristics,
geographic distance, or other factors of the sea landscape (Hauser and Carvalho 2008). A
more general framework to assess marine metapopulations was proposed by Kritzer and
Sale (2004) and considers demes partially isolated and independent but connected to each
other via gene flow. Under this model, extinction and recolonization events are possible,
but not as essential as in earlier models (e.g. Levins 1970), recognizing that marine
populations rarely go extinct and also that they often maintain some degree of
connectivity even if migration events are rare and episodic. Metapopulations are usually
not structured according to simple models such as the island model (Whitlock and
McCauley 1999) but rather feature structures involving unequal migration rates among
demes and variable deme sizes and growth rates. Therefore, a full understanding of
metapopulation structure requires assessing demes across the species range in order to
account for all sources of migrations when inferring gene flow and demographic
dynamics of local demes. This is particularly important when defining conservation units
in cases where local stocks are sustained by migrations from geographically distant

populations, as failure to account for these sources of migrants could lead to ineffective management efforts (Beerli 2004; Slatkin 2005). The genetic structure of a metapopulation also often reflects non-equilibrium situations and the effects of historical events. Studies of the phylogeography and historical demography of populations, therefore, provide critical information in order to interpret current patterns of genetic variation and assess the status of subpopulations. Studies of metapopulations of many reef fishes are facilitated by the highly sedentary behavior of adults leading to the prediction that dispersal and connectivity is due for a large part to the (passive) dispersal of pelagic eggs and larvae under the action of oceanic currents. This dispersal process allows developing hypotheses on migration routes based on current circulation patterns and the distribution of suitable habitats for adults. These hypotheses are difficult to test in practice because of challenges involved in tracking larvae in their natural pelagic environment (Thorrold et al. 2002). Genetic methods provide an alternative measure of migrations through the detection of migrants based on their genotypes at molecular markers (Pritchard et al. 2000; Anderson and Thompson 2002). Another critical population parameter influencing the demographic dynamic and genetic structure of a metapopulation is the effective size of demes. The effective population size determines the capacity of a population to adapt to changing environmental conditions and recover from depletion (Hauser et al. 2002) and, in combination with the migration rates, its capacity to develop local adaptations and avoid genetic swamping (Lenormand 2002; Aitken and Whitlock 2013).

In this chapter, the population structure, connectivity, and demographic history of a reef fish with high potential for dispersal, the gray triggerfish, were investigated.

34

Populations of gray triggerfish are found in various coastal regions surrounding the Atlantic basin and in basins connected to it, leading to the prediction of a potentially complex metapopulation structure shaped by historical events and involving demes of variable size and variable degrees of connectivity.

3.1.1 Distribution of Gray Triggerfish

The gray triggerfish (*Balistes capriscus*) is a reef fish widely distributed in temperate and tropical offshore shelf waters of the Atlantic basin (Figure 3.1). Adults are found associated with benthic structures on the continental shelf at depth varying from 0 to 100 m (Harmelin-Vivien and Quéro 1990). The species has been reported from Canada to Argentina in the West Atlantic and from the United Kingdom to South Africa, including the Mediterranean Sea, in the East Atlantic (Robins and Ray 1986; Sazonov and Galaktionova 1987).
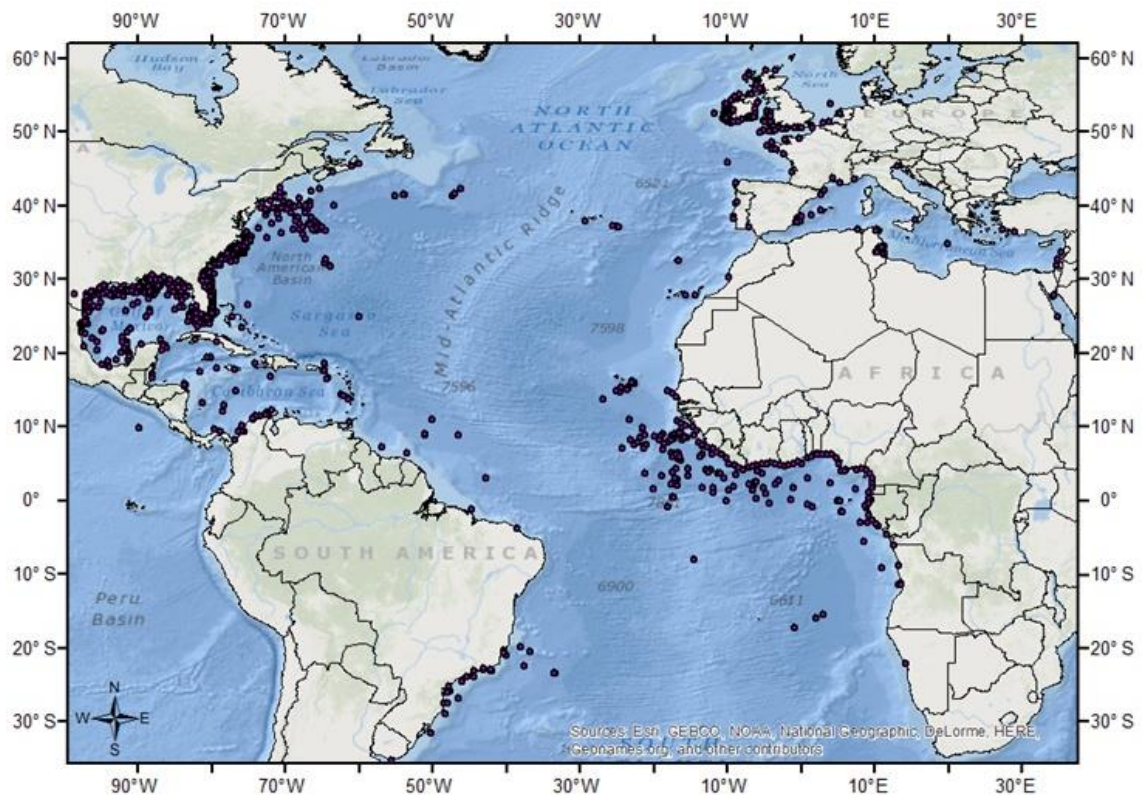
Figure 3.1 Gray Triggerfish Distribution Map

Gray triggerfish (*Balistes capriscus*) distribution range based on reviewed point observations. www.aquamaps.org, version of Aug. 2010. Updated 02/21/13.

In the West Atlantic, gray triggerfish are abundant in the northern Gulf of Mexico and in offshore waters along the U.S. East coast up to the Carolinas (Personal communication of the National Marine Fisheries Service, Fisheries Statistics Division) and off Bermuda Island (J. Pitt, personal communication). In South America, the species is reported along the central coast of Brazil (Bernardes 2002; A. Martin, personal communication), and become scarce moving South to Argentina (I. Masson, personal communication). Further North, they are rare in French Guyana (F. Blanchard, personal communication) and are also very infrequent in Venezuela (F. Arocha, personal communication), the French Antilles (L. Reynal, personal communication), and the U.S. Caribbean (A. Rosario, personal communication). They have been reported in Colombia (Garcìa et al. 1998; Lopez-Pena and Orlando-Duarte 2012) and may occur in other parts of the western Caribbean. In the East Atlantic, gray triggerfish have been reported off the western coasts of England (G. Baker, personal communication) and France (D. Milly, personal communication). The species occurs in the Mediterranean Sea where it is exploited by Turkish fisheries (Ismen et al. 2004). Further South, gray triggerfish are relatively abundant off of the Canary Islands (J. Castro, personal communication). They are not reported off of Mauritania, Senegal, and Liberia, but were historically abundant off of Guinea Bissau and Guinea, Ghana, and Angola (Stromme 1984).

3.1.2 Fisheries

In the Mediterranean Sea, gray triggerfish are exploited commercially along the Turkish coasts although these fisheries are considered of moderate importance (Ismen et

al. 2004). An average of 77 tons was caught in Tunisia between 1993 and 2009, and Libya had the highest catch recorded in 2009 with 432 tons (FAO 2014).

According to Mensah and Quaatey (2002), gray triggerfish were very abundant in the Gulf of Guinea in the early 1970's and represented the most common demersal species in the 1980's but, at the time the paper was published (2002), gray triggerfish had declined sharply and become very infrequent in the region. More recently, Aggrey-Fynn (2009) reported that, even though the species had been declining for more than two decades, it was still present in the Gulf of Guinea and its growth parameters and geographic distribution had not changed. Stromme (1984) identified differences in the size distribution of gray triggerfish found off of Ghana (showing a mode of 5.7 inches in fork length) as compared to that of populations located off of Guinea Bissau where the mode was 7.5 in. These differences may reflect in part different age structures and/or different growth conditions in these two regions, but tentatively suggest the possible occurrence of different demographic stocks.

Gray triggerfish have been commercially targeted in Brazil for the past two decades becoming one of the most commonly caught species at least up to the late 1990s (Bernardes 1988; Castro 2000), but landings have been decreasing (Ataliba et al. 2009) and the species has been declared commercially extinct due to overexploitation in the Espirito Santo State (Netto and Di Beneditto 2010). In U.S. waters, gray triggerfish was not considered a desirable catch until the 1980s when the decreased abundance of other reef fishes such as the red snapper and increased harvest restrictions on fisheries harvesting them, diverted fishing effort towards alternative species (Valle et al. 2001). Gray triggerfish are currently harvested by both recreational and commercial fisheries,

and young specimens (generally 0-1-year-old) are captured as bycatch by the shrimp fishery. To our knowledge, the United States is the only country in which gray triggerfish fisheries are regulated. The assessment conducted in 2006 revealed that the species was overfished and undergoing overfishing (SEDAR-9 2006). The updated assessment conducted in 2015 (SEDAR-43 2015) concluded that the stock was not undergoing overfishing anymore but was still overfished.

In 2015 gray triggerfish was listed in the IUCN Red list of threatened species under the "Vulnerable" category due to the fact that the species had experienced an overall decline across its distribution range of at least 30% over the past three generations. Declines by as much as 63-68% were reported for the Gulf of Mexico, Gulf of Guinea, and Brazil (Liu et al. 2015).

3.1.3 Life History

Gray triggerfish can reach 23 inches in total length (TL) (Harmelin-Vivien and Quéro 1990) and weighs up to 13.5 lb (IGFA 2001). Their longevity is medium with a maximum reported age of 16 years (SEDAR-9 2006). Adults are generally found associated with reefs and hard structures, such as oil rigs (Ingram 2001), where they maintain a highly sedentary lifestyle. During mark and recapture studies conducted by Ingram (2001), tagged triggerfish moved by less than 9 km from artificial reefs and less than 23 km from natural reefs over periods of up to 5.5 months. Movements increased, yet remained relatively limited, during hurricanes. These data are consistent with those obtained by Beaumariage (1964) who recorded all recaptures (36.9% of tagged fish, some recaptured after being at large for 16 months) in the vicinity of release sites. One study suggested significant, yet temporary, adult movements of gray triggerfish off the Ghana

coasts (FAO 1980). However, the hypothesized migrations were seasonal and from nearshore to offshore areas in connection with a seasonal upwelling.

Direct observations during diving surveys using fixed underwater cameras showed that this species exhibits territorial behavior; dominant males defend nesting areas potentially harboring multiple females (Simmons 2008). Eggs are laid on plants in areas where the bottom substrate is hard (Ismen et al. 2004) or on nests dug in the sand. Both sexes defend the nest until the eggs hatch (MacKichan and Szedlmayer 2007). The spawning season in the northern hemisphere extends from May to July in the Gulf of Mexico (Ingram 2001; MacKichan and Szedlmayer 2007), and from April to July in the Mediterranean Sea although fish in spawning condition can be occasionally found as late as August (Ismen et al. 2004). In the southern hemisphere, spawning occurs from October to December along Ghanaian coasts (Ofori-Danson 1990), and from November to February in South Brazil (Bernardes and Dias 2000).

Gray triggerfish larvae have been shown to utilize the pelagic environment (Richards and Lindeman 1987; Leis 1991) and juveniles are commonly found in association with floating *Sargassum* (Aiken 1983), where they represent a very high fraction of the icthyofauna (32% in the study of Hoffmayer et al. 2005; 4th most abundant species in the study of Casazza and Ross 2008). An aging survey of newly settled gray triggerfish in the northern Gulf of Mexico revealed that the pelagic phase can last up to 4-7 months in that region (Simmons 2008).

Gray triggerfish larvae and juveniles can thus be transported across long distances during the extended period of their pelagic ecophase as the *Sargassum* beds they use as habitat move under the action of ocean currents. This pelagic dispersal could potentially

39

promote connectivity among populations separated by long distances. Patches of floating *Sargassum* are known to move extensively across the Northwest Atlantic [e.g. Sargasso Sea water gets entrained in the Florida current and the Gulf Stream and transported to the eastern Atlantic (Casazza and Ross 2008)] or from the Gulf of Mexico to the U.S. East coast through the Loop current and the Gulf Stream (Gower and King 2008). In addition, in recent years, pulses of *Sargassum* deriving from the North Equatorial Recirculation Region (NERR), i.e. formed in the South Atlantic, have also been observed to travel along the North Brazil current up to the Caribbean, or eastward toward western Africa (Johnson et al. 2013).

3.1.4 Major Oceanic Circulation Patterns in the Atlantic Ocean and Working Hypotheses on Connectivity

Connectivity among gray triggerfish populations is hypothesized to derive primarily from movements of larvae and juveniles during their pelagic ecophase. During that phase, gray triggerfish larvae follow *Sargassum* patches or other floating materials until they settle on suitable benthic habitats where they are assumed to remain sedentary for the rest of their life. *Sargassum* beds circulate primarily between the 20° and 40°N of latitude and between the American coast and 30°W of longitude (Weis 1968). Because *Sargassum* can float indefinitely (Parr 1939; Woodstock 1950), gray triggerfish larvae and juveniles can potentially remain in association with it for a very long time resulting in transport across long distances. Such a scenario could occur if fish delay settlement on benthic habitat because *Sargassum* patches get entrained offshore where no suitable bottom structure are present. Pelagic durations of up to 7 months were indeed inferred by Simmons (2008) for newly settled juveniles in the Gulf of Mexico as discussed above.

Data on oceanographic currents in the Atlantic were obtained from the

Cooperative Institute for Marine and Atmospheric Sciences (CIMAS) website and are

presented in Figure 3.2a and 3.2b. Surface current data were obtained from the U.S.

Coast Guard's Mariano Global Surface Velocity Analysis (MGSVA) 1.0 (Mariano et al.

1995). Sea Surface Velocity was determined from ship-drift measurements and drifting

buoy trajectories. All the maps available for the Atlantic Ocean were connected to obtain

a global image of general current routes. The main surface current patterns during peaks

of the spawning season for gray triggerfish in both northern and southern Atlantic are

shown below (Figures 3.2a and 3.2b) and were used to formulate hypotheses on the

connectivity among regional populations.
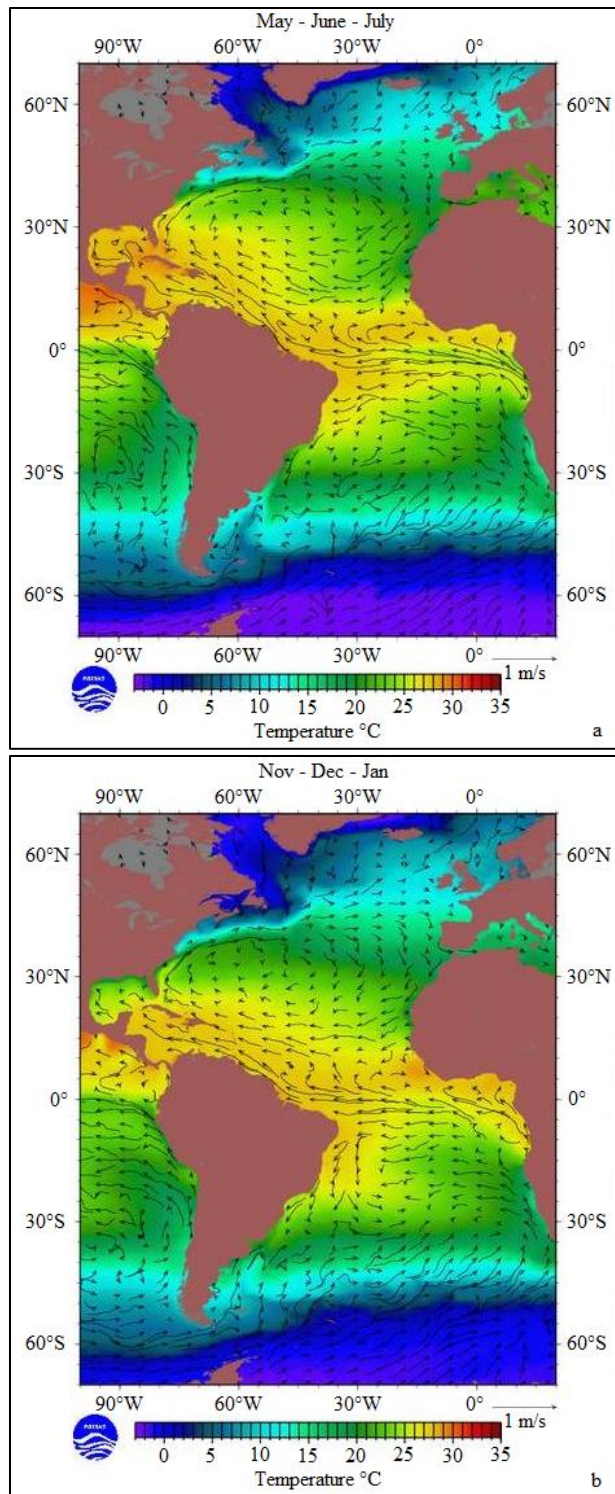
Figure 3.2 Main Surface Currents During Gray Triggerfish Spawning Season

42

Diagram depicting the main surface currents during the peaks of spawning for the northern hemisphere (a) and southern hemisphere (b) in the Atlantic basin and the Gulf of Mexico. (source: http://oceancurrents.rsmas.miami.edu/mgsva/slideshow.html?imgnum=6&basin=atlantic).

The following hypotheses on genetic connectivity between regional stocks in the Atlantic Ocean were developed under the assumption that most of the movements of larvae and juveniles occur following free-floating material under the action of oceanic circulation and accounting for the velocity and direction of main known surface currents.

a. Connectivity of northern Brazil, Central America, and U.S. East coast populations: Connectivity in that region could be mediated by the Caribbean current that originates from the North Brazil and Guyana currents and flows to the West entering the Gulf of Mexico as the Loop current. Regions connected through this current include North Brazil, Venezuela, Nicaragua-Honduras, then Belize, and southern Yucatan. The Loop current could further connect the above-mentioned regions to the northern Gulf of Mexico and then the East coast of Florida. An alternative potential route is through the Antilles although currents are slower in that region which is expected to lead to increased divergence among geographic populations. As discussed in section 3.1.1, gray triggerfish are infrequent in the Antilles and Bahamas. Because currents are generally unidirectional in the western Atlantic along a South-North axis, an Isolation-by-Distance (IBD) pattern with asymmetric gene flow (primarily South to North) is expected. These currents can run at relatively high velocities (60-200 cm/s), with the exception of the Caribbean current which is relatively slow (30-40 cm/s), overall suggesting high levels

of connectivity among populations may occur considering the 4-7 months larval dispersal period hypothesized for gray triggerfish.

b. Connectivity between the U.S. East coast and northwestern Europe: The Gulf Stream and the North Atlantic currents could transport gray triggerfish juveniles, associated with *Sargassum* patches, to northern Europe. Buoy data showed that drifting materials or seaweeds can move about 15°E and 5°N in a month following the Gulf Stream (Rossby et al. 1985). If *Sargassum* patches follow the Gulf Stream and considering an average current velocity of 66 cm/s reported by Rossby et al. (1985), gray triggerfish larvae could drift from Florida to northern Europe in a minimum of 5 months assuming they do not get caught in gyres. However, gyres are pretty common in this area. Examination of data from 232 drifters deployed in the North Atlantic from 2000 to 2012 (available at http://www.aoml.noaa.gov/phod/dac/index.php) revealed that only 8 successfully crossed from the U.S. waters to coastal areas in the East Atlantic; transport between the two sides of the basin, separated by approximately 60° in longitude, took a minimum of 11 months (for a drifter travelling 64°E). While this duration is much longer than the 4 to 7 months pelagic period reported in the northern Gulf of Mexico by Simmons (2008), the decreasing water temperature expected as the Gulf Stream current proceeds North may result in a reduction in growth rate and an increase of the duration of the early development phases that precede settlement (Pankhurst and Munday 2011). Indeed, a recent study of larval coral reef fishes (Rankin and Sponaugle 2011) showed that larvae grow slower in colder water and

44

tend to settle later. Also, gray triggerfish may delay settlement until suitable shelf waters become available. For all these reasons, and although restricted gene flow is expected between the U.S. East coast and northwestern Europe, the possibility that some sporadic migrants are able to reach the European Coast cannot be excluded.

c. Northern Europe to southern Europe and Mediterranean Sea: Surface currents along the Atlantic coast of Europe show slow velocity and no clear directional pattern. Larvae from the Gulf Stream seem more likely to reach the UK but may also on occasion be transported directly to central or southern Europe. Migration between northern and southern Europe may be limited, possibly following a pattern of IBD, although genetic homogeneity could be maintained if migrants from the Gulf Stream are dispersed broadly across the European coasts. If migrants reach Portugal they could enter the Mediterranean Sea or move southward along the African coasts. Transport in the Mediterranean Sea may be restricted due to the Alboran Seafront which has been shown to be associated with both ecological and genetic discontinuities in many marine species with pelagic larvae, from mussels (e.g. Quesada et al. 1995) to fish (e.g. Naciri et al. 1999; Lo Brutto et al. 2004).

d. Connectivity of northwestern African populations: populations located in Northwest Africa could receive migrants from Europe as described in Hypothesis C, but also West Atlantic immigrants incoming through the Azores current that is formed by the southern branch of the Gulf Stream. Triggerfish larvae spawned in Morocco and Canary Islands are expected to be

45

transported in a Southwest direction by shelf surface currents and to move off

of African coasts in the North Equatorial current. This mechanism would be

consistent with the apparent absence of gray triggerfish in Mauritania and

Senegal. Reproductive isolation and a genetic discontinuity may, therefore,

occur between North African populations and the stock located in Guinea

(Off Guinea, Sierra Leone, and Guinea Bissau). However, it is also possible

that some migrants could reach the Guinea current, which originates from the

Canary current (Gyory et al. 2005) and reach the Gulf of Guinea thereby

connecting northern populations to those located in the Gulf of Guinea.

Further South and East, the Guinean stock could be connected to those in the

western Gulf of Guinea (e.g. Benin, Ghana) through transport along the

Guinea current that flows along the coast West to East. Although the Guinea

current is close to its minimum during the spawning season of gray triggerfish

(Gyory et al. 2005), the occasional reversal of the direction of the flow could

favor mixing across the area. Interestingly, Stromme (1984) identified two

separate stocks in that part of the species range: The eastern stock was located

off Ghana and the western stock off Guinea Bissau and Guinea. The two

stocks were considered distinct because they were separated by an area where

gray triggerfish was absent (the species was not reported in Liberia) and fish

caught in the western stock were larger than those from the eastern stock.

However, no genetic data were provided in support of this hypothesis and

these demographic differences may simply reflect different habitat

characteristics and age structure between the two putative stocks.

e. Connectivity of southeastern African populations: The Angola current (flowing northward), and the Guinea current (flowing mostly eastward), converge off of Nigeria potentially isolating populations of the western Gulf of Guinea and those from Southwest Africa (Angola, Namibia). However, connectivity could occur if the Guinea current changes directions, as explained above, or through westward transport of fish from Southwest Africa by the South Equatorial current and successive mixing with gray triggerfish populations off of the Ghana coast.

f. Connectivity between East and West Atlantic through the Equatorial currents: Such connectivity needs to be examined because the geographic distance between the two continents along the Equatorial current axis is shorter than that between North America and Europe. However, Equatorial currents are much slower than the Gulf Stream and only reach 25 cm/s. In addition, floating materials moving along the North Equatorial current are likely to be caught by gyres that are frequent in that region which would further slow-down transport of *Sargassum* patches to suitable habitats on the other side of the Atlantic basin. To formally evaluate the hypothesis of larval transport through the Equatorial currents, data from 413 drifters moving along the South Equatorial System were downloaded from the Global Drifter Program (available at http://www.aoml.noaa.gov/phod/dac/index.php). For each drifter, coordinates at the time of release and pick-up, and drifting time were recorded. Sixty-one drifters crossed the ocean from West Africa to the American continental shelf, with an average calculated travel time between

47

the two sides of 9.1 months. The faster drifter covered 39.7° in longitude in 3.9 months. Considering that the distance that separates the two continental shelves corresponds approximately to 35-40° in longitude, it is not impossible that some *Sargassum* patches or flotsam offering shelter to gray triggerfish juveniles reach West Atlantic adult habitats in a reasonable amount of time compatible with fish survival. Migrations from South America to West Africa seem less likely due to the weakness or absence of the North Equatorial counter current during the gray triggerfish spawning season. In summary, gene flow is hypothesized to be restricted across the Equatorial Atlantic and migration events appear more likely to occur from southeastern Atlantic populations toward southwestern ones.

g. Connectivity of South Brazilian populations: The main surface current that flows along the central and southern coasts of Brazil is the Brazilian current that originates at the eastern tip of Brazil (Figure 3.2b) and flows southward. An IBD pattern is therefore hypothesized along this axis. Because of the divergence of the North Brazil (flowing North) and Brazilian (flowing South) currents, Brazilian populations under the influence of the Brazilian current are hypothesized to be isolated from those further North discussed in Hypothesis A.

Very few studies of reef fish population structure have been conducted at the scale of an entire Ocean basin. Ball et al. (2007) detected three distinct populations of red porgy (*Pagrus pagrus*) in the Atlantic (northwestern, northeastern, and southwestern Atlantic). In addition, samples from the Azores in the eastern North Atlantic diverged

from those obtained from Greece and Madeira indicating some level of structure within

the northeastern Atlantic. Red porgy life history presents similarities with that of the gray

triggerfish although the pelagic phase is much shorter (up to 30 days). While life history

differences between Gulf of Mexico and Atlantic coast of U.S populations suggested at

least partial isolation of the two stocks, genetic data indicated homogeneity within the

area suggesting that the short pelagic duration was sufficient to maintain gene flow. The

separation between Northwest and Southwest Atlantic stocks of red porgy was stronger

($F_{ST} > 0.02$) and a genetic discontinuity between these two regions has also been reported

for other reef fish species, including two species of surgeonfishes (Rocha et al. 2002), red

snapper (Norrell 2016), and yellowtail snapper (Vasconcellos et al. 2008). The

discontinuity in the study of Rocha et al. (2002) corresponded to the Amazon freshwater

plume. The third species of surgeonfish surveyed in that study did not show a

discontinuity in the region possibly because of its ability to use deep water corridors and

get around the plume. Since gray triggerfish disperse in surface waters occupied by

floating material, they may not be able to overcome the physical barrier created by a

freshwater plume. Surprisingly, gene flow between the populations located in the mid-

Atlantic Islands and those off the coast of Brazil seemed to be occurring as inferred from

genetic homogeneity observed for all 3 surgeonfish species studied by Rocha et al.

(2002). Thus, considering the long duration of the pelagic phase of gray triggerfish,

connectivity between West African populations and South American ones cannot be

excluded as discussed in Hypothesis F.

In this chapter, gray triggerfish from various geographic populations were characterized using the microsatellite markers developed in Chapter II in order to test the hypotheses developed above.

3.2 Materials and Methods

3.2.1 Sampling Design

Sampling targeted gray triggerfish populations from various locations within the distribution range of the species in order to address the hypotheses outlined in introduction of this chapter. The geographic populations surveyed thus included the Northwest Atlantic (Gulf of Mexico and U.S. East coast), the Northeast Atlantic (including the European Atlantic coast, the Atlantic coast of North Africa, and the Mediterranean Sea basin), the Southeast Atlantic (Gulf of Guinea and Southwest Africa), and the Southwest Atlantic (Brazil). Tissue samples were requested from fisheries surveys (U.S. populations) and provided by collaborators to the project in other parts of the range.

The Northwest Atlantic (Gulf of Mexico and U.S. East coast) appears to be the center of abundance of the species with annual catches reported by the recreational and commercial fisheries exceeding 203,382 specimens (personal communication of the fisheries statistics division of NOAA fisheries). This part of the range was, therefore, more intensively surveyed to achieve a dense sampling reflecting the near continuous distribution of gray triggerfish and estimate dispersal parameters using isolation-by-distance theory in Chapter IV. Samples were collected during the summer and fall of 2008, 2009, and 2010. The geographic locations of collecting sites within U.S. waters and the aggregation of individual samples in regions for analysis are presented in Figure 3.3.

Specimens from the north-central and northwestern Gulf of Mexico (South Texas to the

Alabama-Florida border) were collected by trawling in conjunction with the groundfish

SEAMAP surveys conducted by the National Marine Fisheries Service (NOAA-

Fisheries). The survey employs a stratified randomized design to sample benthic shelf

habitats used by triggerfish juveniles and adults (5-50 fathoms depth) by trawling from

Pensacola to the U.S./Mexico border (Nichols 2004). Additional samples from

Mississippi were provided by anglers. Samples from the vicinity of Panama City, Florida

were collected at recreational fishing docks or using traps during fishery independent reef

fish monitoring programs. Samples from Southwest Florida (SWF) were obtained in

conjunction with the reef fish monitoring program and the SEAMAP surveys

implemented by the Florida Fish and Wildlife Research Institute. Samples from Southeast

Florida (SEF) were obtained via hook and line fishing conducted by Capt. W. Taylor in

the context of his professional activity. Finally, samples from offshore South Carolina

(SC) waters were collected by trapping in conjunction with the MARMAP survey

implemented by the South Carolina Department of Natural Resources. Samples from the

north-central and northwestern Gulf were initially aggregated in 3 regions (South Texas:

STX, East Texas-Louisiana: ETX-LA, Mississippi-West Florida: MS-WF, Figure 3.3) in

order to account for gaps in the distribution of available samples along the coastline.
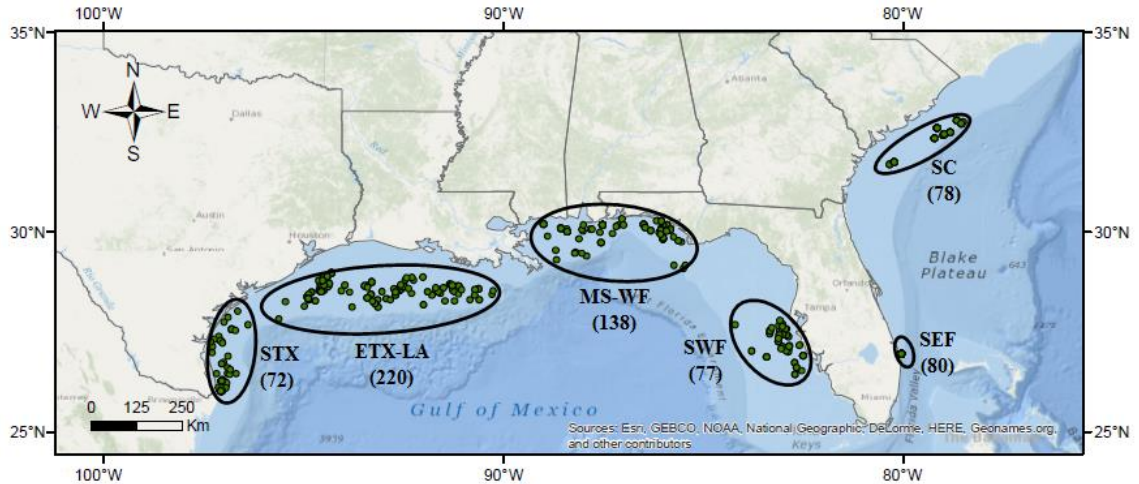
Figure 3.3 Sampling Localities for Gray Triggerfish in U.S. Waters

Sample sizes per location are showed between parentheses. STX: South Texas; ETX-LA: East Texas-Louisiana; MS-WF: Mississippi-West Florida; SWF: Southwest Florida; SEF: Southeast Florida; SC: South Carolina. Green circles represent individual sampling localities within a region.

Gray triggerfish tissue samples from the rest of the distribution range (Figure 3.4) were collected between 2010 and 2014 by collaborators and provided for this study. Individuals from the Bay of Biscaye (FR) were obtained as bycatch of Mr. Dominique Dirassar commercial fishing activity and provided by Mr. David Milly (OP CAPSUD, France). Sampling in the Mediterranean region (MED) was coordinated by Dr. Stefanos Kalogirou (IMBC, Rhodes) and included tissue samples provided by anglers operating in the vicinity of the Island of Rhodes. Specimens from the Canary Islands (CA) were obtained during fishery dependent sampling and provided by Dr. Jose Castro (Universidad de las Palmas de Gran Canaria). Samples from Benin (BE) were collected at fishing docks by personnel at the Centre de Recherches Halieutiques et Océanologiques du Bénin (CRHOB) coordinated by Dr. Roger Djiman. Specimens from Angola (AN) offshore waters were collected during the 2012 and 2013 Nansen fisheries surveys conducted by the Institute of Marine Research (Norway) and provided by Mrs. Diana

Zaera-Perez and Dr. Kathrine Michalsen. Finally, samples from Brazil (BR) were obtained by Drs. Cristiano Albuquerque and Agnaldo Martins (Universidade Federal do Espírito Santo) in the context of fisheries dependent sampling and provided for this study.

Tissue samples (finclips and/or muscle tissue) were collected post-mortem on fresh or frozen specimens and preserved at room temperature in 95% ethanol or in a DMSO salt-saturated fixative (0.5 M EDTA, 20% Dimethyl sulfoxide, NaCl, ddH$_2$O) until DNA isolation except for the samples from South Carolina that were preserved in a sarkosyl urea lysis buffer (1% n-lauroylsarcosinate, 20 mM NaPO4, 8 M urea, 1 mM EDTA).

DNA extraction was performed using the phenol-chloroform protocol (Sambrook et al. 1989). All specimens were assayed at 17 microsatellite markers developed during the project (see Chapter II for detailed assay methods). Locus BC36 was initially included in the survey but was eventually removed from the panel of markers during the course of the study because several sets of samples failed to provide interpretable electropherograms at this marker.

A sub-sample of 352 specimens (18-39 per locality, Table 3.1) was also assayed at a portion of the ND4 subunit of the NADH dehydrogenase encoded by mitochondrial DNA. This coding gene was successfully used in previous population genetic studies of a variety of organisms including marine reef fishes in the Gulf of Mexico and Caribbean regions (Pruett et al. 2005; Saillant et al. 2012). PCR amplification and sequencing of PCR products employed the universal primers NAP-2 (Arevalo et al. 1994) and ND4LB (Bielawski and Gold 2002). PCRs for the mtDNA marker were conducted in a 25 μL volume solution containing 25-50 ng of genomic DNA, 10 pmol of each primer, 1.25 U

of TAQ polymerase (Promega Inc., Madison, Wisconsin), 5 nmol of dNTPs, 37.5 nmol of MgCl2, and 1X buffer (Promega). Amplification by PCR consisted in an initial denaturation at 95°C for 3 min, 35 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 1 min and 30 s, and a final extension of 15 min at 72°C.

The PCR products were purified using the Exo-SAP-IT PCR clean-up kit (GE Healthcare, Piscataway, New Jersey) and sequenced using the Big-Dye Terminator v.3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, California) following instructions from the manufacturer. Sequencing reaction products were run on a ABI-3730XL capillary sequencer (Applied Biosystems) at the High-Throughput Genomics Center in Seattle or on an ABI-3130 capillary sequencer at the USM-Gulf Coast Research Laboratory.

Sequences were aligned and edited in the software SEQUENCHER v.4.10.1 (Gene Codes Corporation, Ann Arbor, Michigan). Unique haplotypes were re-sequenced to confirm their sequence and reduce the risk of erroneously introducing new genetic variants in the dataset.

Sample sizes available for each sampling locality at the two types of markers are presented in Table 3.1.

Table 3.1

Gray Triggerfish Sample Sizes per Locality

Locality	# samples microsatellites dataset	# samples mtDNA dataset
STX	72	30
ETX-LA	220	31
MS-WF	138	27
SWF	77	-
SEF	80	32
SC	78	30
FR	64	37
CA	76	35
MED	17	18
BE	72	39
AN	70	38
BR	52	35

Numbers of gray triggerfish specimens analyzed for each sampling locality as defined in section 3.2.1.
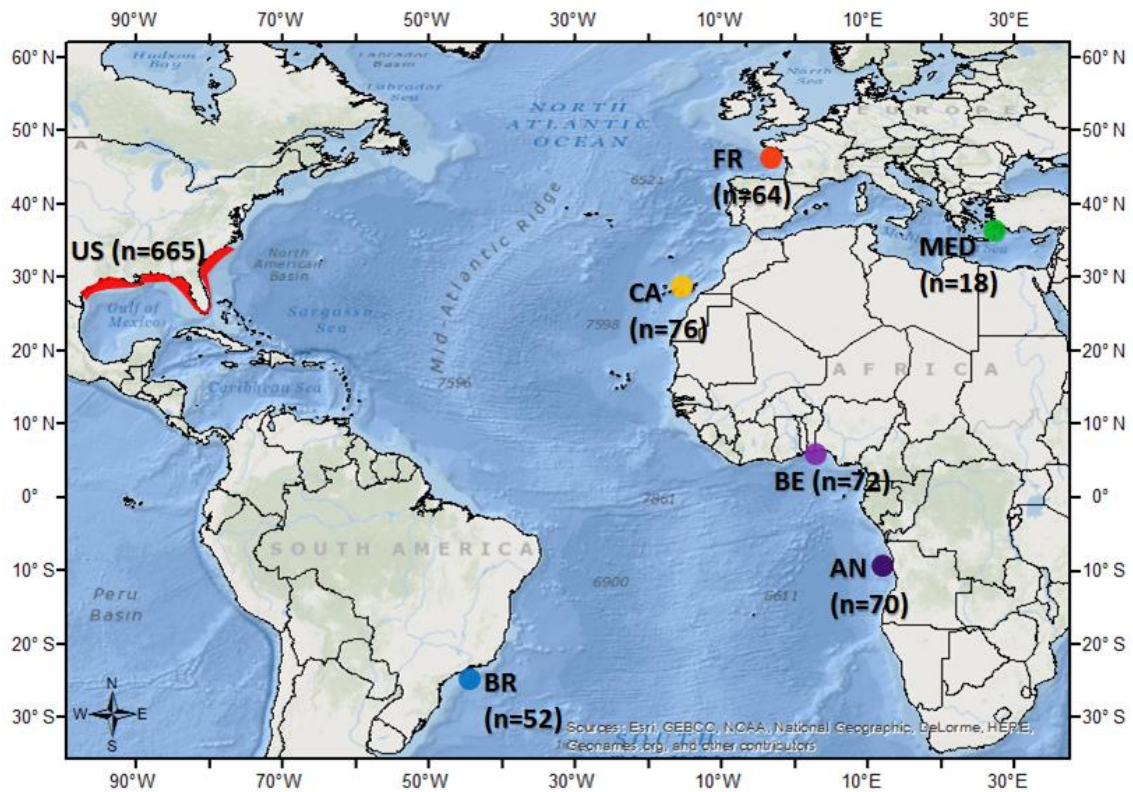
Figure 3.4 Sampling Localities for Gray Triggerfish

Sample sizes are showed between parentheses. Sampling localities in U.S. waters of the Gulf of Mexico and South Atlantic regions are detailed in Figure 3.3. US: United States; FR: France; CA: Canary Islands; MED: Mediterranean Sea; BE: Benin; AN: Angola; BR: Brazil.

3.2.2 Data Analysis

The conformance of genotypic proportions to Hardy-Weinberg equilibrium expectations for each locus within each population was tested using exact tests as implemented in GENEPOP v.4.2 (Raymond and Rousset 1995). The exact probability for each test was estimated using a Monte Carlo Markov Chain approach as per Rousset and Raymond (1995) and based on 10,000 dememorizations, 500 batches, and 5,000 iterations per batch. The software MICROCHECKER v.2.2.3 (Van Oosterhout et al. 2004) was used to test for the occurrence of scoring errors due to null alleles, stuttering bands, and large allele dropout in each of the sampling localities. The inbreeding coefficient (F_{IS}) measured as Weir and Cockerham's (1984) *f*, the number of alleles, allelic richness, and gene diversity were calculated for each regional sample in FSTAT v.2.9.3 (Goudet 1995). Allelic richness is a measure of the number of alleles independent of sample sizes that is based on the rarefaction method (El Mousadik and Petit 1996) and allows comparing allelic diversity in groups with different sample sizes. Gene diversity was calculated as described in Nei (1987). Homogeneity in allelic richness and gene diversity among samples, or groups of samples, was tested using the Wilcoxon ranks test, as implemented in SPSS v.20 (IBM Corp., Armonk, NY, USA).

The mitochondrial DNA data were aligned to confirm consistency of base calls across the gene sequence in SEQUENCHER, and the validity of variant calls was checked

by cross-examining all sequences at each variable site and by verifying that no stop codons had been introduced by substitution calls within the coding fragment analyzed.

Summary statistics, including number of haplotypes and haplotype diversity, were computed in DNASP v.5.10 (Rozas et al. 2003). Haplotype richness was estimated using CONTRIB v.1.0.2 (Petit et al. 1998). Homogeneity in haplotype richness and diversity among groups was tested via a bootstrap (random) resampling approach wherein the probability that the number of different haplotypes or the haplotype diversity observed in a group would be observed in a random sample of the same size taken from the other group was estimated. The program POPTOOLS (a free add-in software for EXCEL, Hood 2010) was used to generate random bootstrap samples with replacement for all 4 groups. The bootstrap sample size was determined as the size of the smallest group for each comparison performed. Random sampling was performed 10,000 times, and the average number of haplotypes, average haplotype diversity, and their upper (0.975) and lower (0.025) percentiles were recorded. Significant differences between groups were inferred when the observed value for one group lied outside the bounds of the obtained confidence interval for the group it was compared to (i.e. pairwise comparisons were performed).

3.2.2.1 Analysis of Spatial Genetic Variation

The magnitude of divergence among geographic samples was assessed using Weir and Cockerham's (1984) unbiased estimate of F_{ST} (θ). Estimates of θ were generated using FSTAT, and the probability that $\theta = 0$ was determined via exact tests of genic and genotypic differentiation (Raymond and Rousset 1995) in GENEPOP.

Pairwise comparisons were performed by computing estimates of pairwise θ between individual regions and performing associated pairwise exact homogeneity tests.

Markov Chain parameters during exact homogeneity tests were the same as above (Exact tests of H-W equilibrium). The False Discovery Rate (FDR, Benjamini and Hochberg 1995) procedure was used to determine the significance threshold for P-values when multiple independent tests were conducted simultaneously.

The occurrence of barriers to gene flow within the sampling surface was assessed using a modified version of Monmonier's (1973) maximum difference algorithm as implemented in the software BARRIER v.2.2 (Manni et al. 2004). BARRIER seeks to identify boundaries, areas where differences between pairs of sample localities are largest, within a genetic landscape. To do so, a Voronoi diagram is constructed that defines the boundaries of each sampling locality neighborhood by enclosing it in a polygonal cell. Barriers are initiated by the edge of the Voronoi diagram that corresponds to the highest pairwise genetic distance estimate across the entire dataset and continues through adjacent edges according to the Monmonier's algorithm until the border of the sampled area is reached or the barrier closes around a set of localities. The pairwise genetic distance between localities used in computations was the weighted average of F_{ST} (across the 17 loci) calculated as described by Weir and Cockerham (1984). The support of each barrier was determined by resampling loci in the software POPTOOLS. One thousand boostrap datasets were generated and the matrix of pairwise F_{ST} values was recalculated. The obtained 1,000 matrices were used as input to BARRIER to determine the support of the inferred barriers.

Population structure was examined using Spatial Analysis of Molecular Variance (SAMOVA, Dupanloup et al. 2002) using the software SAMOVA 1.0 available at http://web.unife.it/progetti/genetica/Isabelle/samova.html. SAMOVA employs a

58

simulated annealing algorithm to optimize the allocation of N geographic populations into K groups (2 < K < N). Allocation is optimized by maximizing the proportion of total genetic variance due to genetic variation among the inferred groups. A total of 100 simulated annealing processes were used to determine the optimal allocation of the 12 geographic samples into groups. SAMOVA was performed using both the mitochondrial DNA and the microsatellite datasets in separate analyses.

Population structure was also inferred using the model-based Bayesian clustering method implemented in the software STRUCTURE v.2.3.4 (Pritchard et al. 2000). STRUCTURE optimizes the allocation of individuals into putative populations (clusters) that minimize departure from Hardy-Weinberg and linkage equilibrium in the overall dataset. Another outcome of analysis in STRUCTURE is the ancestry of sampled individuals and the potential inference of migrants (individuals showing ancestry in one cluster but sampled in a geographic region showing a majority of ancestry in another cluster) and admixed individuals showing shared ancestry in multiple clusters through a probability vector of admixture proportions. The number of subpopulations K is *a priori* unknown and is determined by performing replicate runs of structure for different values of K and comparing the posterior probability of the data under the optimum model as described by Evanno et al. (2005). Three replicate runs were performed for each tested value of K. Each run consisted of 10^7 Monte Carlo steps and a burn-in period of 10^6 steps. Calculations were performed considering sampling localities as prior and using the correlated allele frequency model described in Falush et al. (2003). The logarithm likelihood probabilities of the data were averaged among replicate runs for comparison and determination of the optimum value of K.

3.2.2.2 Contemporaneous Gene Flow

Inferences on contemporaneous gene flow relied on the assignment of genotypes (or fractions of the genotypes) in the dataset to the geographic groups identified during analysis of population structure. Direct migrants (F0) or progeny of migrants (F1 hybrids) were inferred when pure or admixed genotypes involving one cluster were inferred in geographic locations dominated by another cluster (see section 3.3.2 of "Results"). Two approaches were used to assign individuals as putative pure genotypes from one of the 4 groups or first generation F1s (admixed individuals) involving parents from two of the groups within the dataset.

The first approach is based on the inferred ancestry of individuals given by STRUCTURE. STRUCTURE optimizes for each individual a vector of admixture proportions *q* that describes the proportion of ancestry to each of the inferred K clusters. Individuals showing ancestry of at least 90% in a cluster were considered as pure. Pure individuals were then used to simulate hybrid genotypes between clusters using the software HYBRIDLAB v.1.1 (Nielsen et al. 2006). Considering hybridization between two clusters (cluster 1 and cluster 2), simulations generated first generation hybrids (F1s), second generation hybrids (F2s), and both types of F1 backcrosses (F1 crossed with either a pure genotype from the cluster 1 or with a pure genotype from cluster 2). One hundred genotypes were simulated for each cluster pair and each hybrid type. All simulated genotypes were then added to the real dataset and a new STRUCTURE run was implemented using the parameters described above. The ancestry coefficients were obtained for the 100 simulated genotypes during STRUCTURE runs using all 17 loci. Analyses were also conducted using partial datasets that included only 16 of the loci in

order to generate 95% CI for the estimated individual ancestry proportions to each cluster according to the Jackknife procedure (Quenouille 1949, 1956). The mean and range of the proportion of ancestry to each cluster were calculated based on the data obtained for the 100 simulated individuals in each pure and hybrid category. The means were compared with theoretical expectation and the range was used to determine thresholds for the assignment of sampled genotypes to each pure or hybrid categories.

A second approach to infer migrants and hybrids employed the Bayesian method implemented in the software NEWHYBRIDS v.1.1 (Anderson and Thompson 2002). NEWHYBRIDS uses a Gibbs sampler to estimate the posterior probability that sampled individuals fall into each of a set of user-defined hybrid categories. The method assumes that source reference populations are known and sampled individuals include pure individuals and recent hybrids. The hybrid categories considered in this study were F1, F2, and both backcrosses as defined by Anderson (2003). Since NEWHYBRIDS can only assign pure and admixed individuals from two populations, analysis focused on assessing migrations and admixture between each pair of populations. Pure reference individuals for each cluster were selected as described above for the initial analyses in HYBRIDLAB and used as priors in the analysis. Two runs were performed to infer hybrid categories for each cluster pair: The first run only included the sampled genotypes from the original dataset and was used to assign individuals to a pure category (cluster 1 or cluster 2) or to one of the hybrid classes (F1, F2, or one of the two backcrosses). The second run included both the original dataset and simulated genotypes for each hybrid class in order to test the power of NEWHYBRIDS for the assignment to each pure and hybrid class. An individual was assigned unambiguously to a class when the probability of assignment to

that class was at least 3 times greater than the probability value obtained for the second

most supported class for that individual (Odd Ratio, OR, criterion). When the difference

between the probability values for the two most likely classes was less than 3 fold, the

class with the highest probability was reported as highest support but not significant

(Relaxed criterion).

Migration rates between pairs of localities (proportions of F0 migrants or F1

hybrids descended from migrants) were calculated based on the proportion of F0 or F1

individuals inferred with significant statistical support (OR criterion defined above).

3.2.2.3 Phylogeography and Historical Demography

The historical relationships among geographic populations were examined based

on similarity. Phylogenetic trees were generated from allele frequency data using the

software POPTREE2 (Takezaki et al. 2010). A first analysis was conducted considering all

12 geographic samples. Trees were also constructed based on allele frequencies generated

for 4 main groups identified during the analysis of population structure (see section

3.2.2.1 of Results). Clustering employed the neighbor-joining (NJ) algorithm (Saitou and

Nei 1987) and the $(\delta\mu)^2$ distance of Goldstein et al. (1995) or the F_{ST} distance of Latter

(1972) corrected by the sample size. Trees obtained using the two distances were

compared in order to assess the relative role of mutations and genetic drift in generating

the observed divergence (Hardy et al. 2003). The support for each topology was inferred

by bootstrapping over loci (10,000 bootstraps) as per Felsenstein (1985).

A statistical parsimony network of mtDNA haplotypes was generated in TCS

v.1.21 (Clement et al. 2000).

Historical gene flow and effective population size in the population units

identified during analysis of population structure were estimated using the Bayesian

coalescent approach in MIGRATE-N v.3.6.11 (Beerli and Felsenstein 2001; Beerli 2006).

Because coalescent methods are computationally demanding, particularly when several

markers are used as in this study, this analysis was performed using a reduced dataset

obtained by subsampling at random from the complete dataset. Thirty

genotypes/haplotypes were subsampled per population (see section 3.2.1) except for the

Mediterranean region where only 17 genotypes and 18 mtDNA haplotypes were available

respectively. The reduced dataset thus included a total of 107 samples for microsatellites

and 108 samples for mtDNA. Starting parameters for each run were generated from F_{ST}

estimates and a uniform prior distribution was used for all parameters. Minimum,

maximum, and delta priors for the parameter Θ ($\Theta = 4N_e\mu$ for nuclear genes and $N_e\mu$ for

mitochondrial genes) were determined after a series of test runs, in order to narrow down

the range of possible values for each parameter. The range of migration rates allowed

under the prior was from zero to one. Average mutation rates used to derive N_e values

from Θ estimates were 5×10^{-4} (Leblois et al. 2004) and 10^{-8} (Bermingham et al. 1997) for

microsatellites and the mtDNA marker respectively. The parameters of the final run used

to calculate posterior distribution for each parameter are presented in Table 3.2. For the

microsatellite dataset, a Brownian approximation of the stepwise mutation model was

used and the rate was allowed to vary among loci. A mutation rate modifier was deduced

directly from the data based on the ratio of the Watterson's (1975) estimate of theta for

the locus to the average theta over all loci. The modifier was used to scale the mutation

rate for an individual locus. For the mtDNA dataset, a Kimura 2 parameters mutation

model was used. Monte Carlo searches employed 4 long chains consisting of 1.5×10^7 steps with parameters recorded every 1,000 iterations and the first 10,000 trees discarded as burn-in. An adaptive heating scheme with initial temperatures of 1, 1.5, 3, and 1×10^6 was used to ensure mixing of the chains. Convergence was assumed when the plots of the posterior distribution had a single peak and MCMC effective sample size was very large (> 1,000) as recommended by Beerli (2012).

Table 3.2

Prior Distribution Parameters Used in MIGRATE Runs

	mtDNA			microsatellites		
	Min	Max	Delta	Min	Max	Delta
Θ	0	6.0	0.6	0	100	10
M	0	10^8	10^7	0	10,000	1,000

Final priors used for mtDNA and microsatellites datasets during MIGRATE runs. Θ: theta; M: migration rate (scaled by mutation rate).

3.3 Results

Summary statistics for both mtDNA and the 17 microsatellites in the 12 geographic samples are presented in Appendix Table C.1.

Significant departure from H-W expectations were detected during 15 out of 221 uncorrected tests but only 1 test remained significant after adjusting probability values to control the False Discovery Rate (Locus BC49 for BE sample). MICROCHECKER analyses indicated the presence of null alleles during 8 of the 221 (3.6%) locus by region combination tests (BC13 in BR, BC14 in ETX-LA and MED, BC17 in SEF, BC3 in SWF, BC34 in CA, and BC45 and BC46 in BR). Samples from SWF also showed the presence of stuttering at marker BC3. Because significant MICROCHECKER test outcomes occurred in only one or at most two populations and did not lead to significant departure from H-W equilibrium expectations, all markers were kept for further analysis.

3.3.1 Analysis of Spatial Genetic Variation

Exact tests of population differentiation indicated occurrence of significant heterogeneity in allele frequencies among samples (P < 0.0001). The estimate of $\theta\,(F_{ST})$ was 0.017 (95% CI: 0.009-0.029). Pairwise comparisons were conducted to assess occurrence of geographic patterns and are summarized in Table 3.3.

Table 3.3

Pairwise F_{ST} Estimates and Associated P-Values Obtained During Comparisons of Gray Triggerfish Locality Samples

	NA								MED	SEA		SWA
	STX	ETX-LA	MS-WF	SWF	SEF	SC	FR	CA	MED	BE	AN	BR
STX	*	0.1815	0.5426	0.1629	0.2197	0.5144	0.0479	0.0520	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
ETX-LA	0.0007	*	0.0477	**0.0177**	0.2029	0.6179	0.2119	**0.0002**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
MS-WF	0.0003	0.0000	*	0.2594	0.2965	0.3225	**0.0203**	**0.0054**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
SWF	0.0004	0.0008	0.0004	*	**0.0032**	0.2649	**0.0015**	**0.0032**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
SEF	0.0018	0.0006	0.0012	0.0018	*	0.0979	0.0726	**0.0018**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
SC	-0.0002	-0.0006	-0.0006	0.0003	0.0015	*	0.5170	**0.0189**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
FR	0.0008	-0.0007	-0.0002	0.0019	0.0008	-0.0010	*	0.0693	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
CA	0.0000	0.0013	0.0016	0.0014	0.0029	0.0003	0.0001	*	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
MED	0.0665	0.0704	0.0729	0.0721	0.0710	0.0699	0.0798	0.0677	*	**<0.0001**	**<0.0001**	**<0.0001**
BE	0.0324	0.0388	0.0398	0.0406	0.0375	0.0368	0.0367	0.0259	0.0765	*	0.7717	**<0.0001**
AN	0.0322	0.0378	0.0390	0.0397	0.0360	0.0377	0.0377	0.0269	0.0726	0.0004	*	**<0.0001**
BR	0.0449	0.0538	0.0551	0.0563	0.0505	0.0501	0.0508	0.0377	0.0932	0.0120	0.0186	*

Estimates of F_{ST} (Weir and Cockerham θ) (lower diagonal) and probability that $F_{ST} = 0$ (upper diagonal) for pairwise comparisons of microsatellite allele distributions between gray triggerfish geographic samples. Probability values that differed significantly from zero following correction for multiple tests are in bold.

All exact homogeneity tests involving pairwise comparisons between North Atlantic, Mediterranean, and South Atlantic samples were highly significant after FDR correction indicating significant divergence between these 3 groups. Corresponding F_{ST} estimates were all greater than 0.026 (0.048 on average).

Divergence among samples from the North Atlantic, including all U.S. sampling localities, FR, and CA, was very low (average F_{ST} estimate 0.0006, range -0.0010-0.0029) and did not reveal a clear geographic pattern within that area, although two tests comparing FR to U.S. populations were significant (MS-WF and SWF comparisons) and 5 out of 6 comparisons between CA and U.S. samples were significant. While the latter results suggest divergence of the CA sample from the U.S. group, differentiation was minimal with F_{ST} estimates averaging 0.0013 and ranging between 0 and 0.0029 (Table 3.3).

In the South Atlantic, the samples from Benin and Angola did not differ significantly in allele frequencies from one another but both samples differed significantly from the Brazil sample (BE vs BR: $F_{ST} = 0.0120$, P < 0.0001; AN vs BR: $F_{ST} = 0.0186$, P < 0.0001).

The most supported barrier detected in the software BARRIER isolated the North Atlantic and South Atlantic groups with 100% bootstrap support (Figure 3.5). Secondary barriers between the North Atlantic and Mediterranean regions and between the Southeast and Southwest Atlantic were detected but with weaker support (36.8% for the North Atlantic-Mediterranean Sea barrier and 39.6% or less for the Southeast-Southwest Atlantic barrier). Other barriers received less than 15% bootstrap support and are not discussed further.
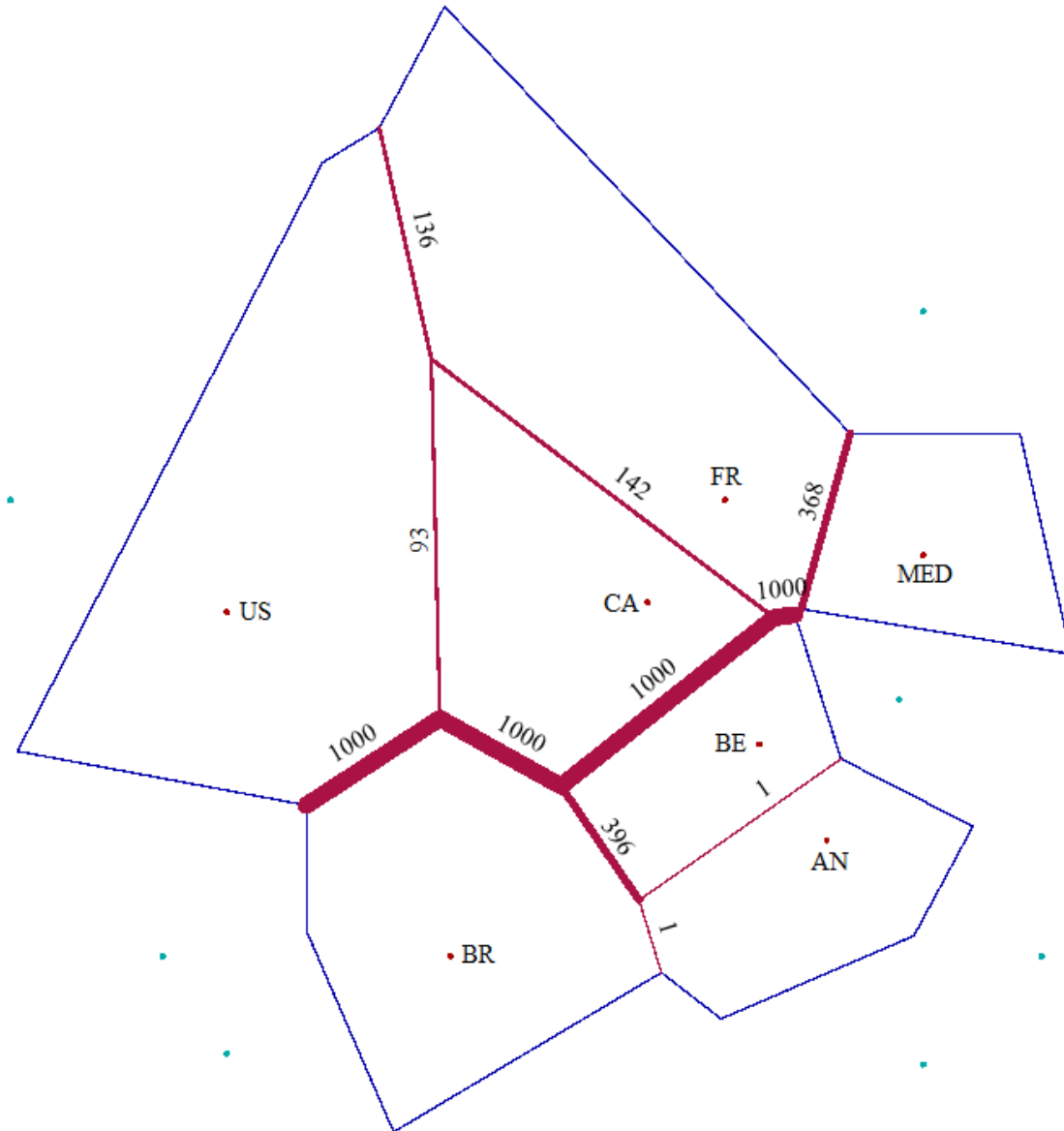
Figure 3.5 Voronoi Diagram Delimiting the Neighborhood of Gray Triggerfish Sampling Localities and Featuring Detected Barriers to Gene Flow

Voronoi diagram with detected barriers and associated bootstrap support between sampling locations in the Atlantic Ocean. The six

U.S. sampling locations are grouped in the figure. US: United States; CA: Canary Islands; FR: France; MED: Mediterranean; BE:

Benin; AN: Angola; BR: Brazil.

The SAMOVA model that led to the highest amount of genetic variance among

groups (5.5% of the total variance) isolated the Mediterranean Sea sample in one group

from all the other locations in a second group. However, this partition did not receive

significant statistical support (F_{CT} = 0.055, df = 1, P-value = 0.08). The model that

yielded the second highest among groups component of molecular variance (4.17%)

isolated 4 groups, matching those discussed above based on results of exact tests and

BARRIER (North Atlantic, Mediterranean Sea, Southeast Atlantic, and Southwest

Atlantic), and the associated among-group component of molecular variance was

significant (F_{CT} = 0.042, df = 3, P = 0.001).

Analysis of the mtDNA sequence dataset in SAMOVA only provided support for

two groups: a North Atlantic group, including the Mediterranean samples, and a South

Atlantic group with eastern and western populations combined. This partition gave the

highest and very strong among groups component of molecular variance (46.6% of the

total variance, F_{CT} = 0.466, df = 1, P = 0.005).

The logarithm of the probability of the data obtained during Bayesian clustering

runs in STRUCTURE increased until K = 3 and then stabilized. The delta K method of

Evanno et al. (2005) confirmed the choice of K = 3. The 3 groups reflected a clear

geographic pattern consistent with previous analyses with most individuals from the

North Atlantic showing a high percentage of ancestry in the first cluster, the second

cluster including individuals from the Mediterranean, and most of the samples from the

South Atlantic featuring close to 100% ancestry in the third cluster (Figure 3.6a). The

occurrence of structure within the North and South Atlantic regions was assessed by re-

running STRUCTURE within these two groups (i.e. using partial datasets consisting of

North Atlantic or South Atlantic samples only) as recommended by Pritchard et al.

(2010). No additional subdivision was detected within the North Atlantic group (Figure

3.6c), but the South Atlantic group converged to a two cluster model, one cluster

corresponding to the Southeast Atlantic (all sampled individuals showed ancestry in that

cluster) while the second one (Southwest cluster) was only detected in samples from the

Southwest Atlantic; Individuals in that region had either a mixed ancestry or pure

ancestry to the Southwest cluster (Figure 3.6b).
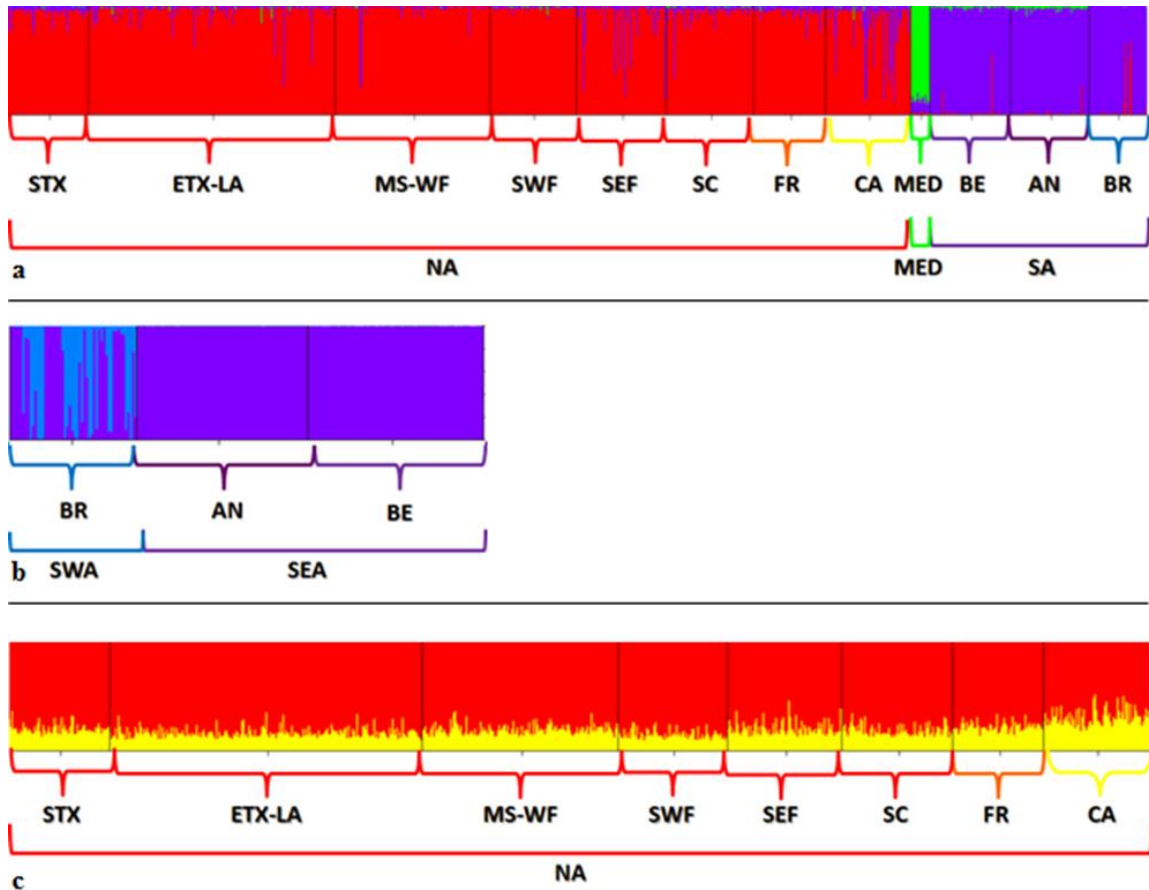
Figure 3.6 Individual Ancestry Bar Plots Generated During Bayesian Clustering

Summary plots representing the results of Bayesian clustering in STRUCTURE for the full dataset (a), South Atlantic region partial

dataset (b), and North Atlantic region partial dataset(c). Each individual is represented by a single vertical line, with the proportion of

assignment to each of the K clusters depicted in a different color. Population labels match the denomination as described in section

3.2.1. NA: North Atlantic; MED: Mediterranean Sea; SEA: Southeast Atlantic; SWA: Southwest Atlantic.

Allelic richness was significantly lower in the South Atlantic than in the North

Atlantic samples and substantially lower in the Mediterranean group (Table 3.4).

Heterozygosity did not differ significantly between the North and South Atlantic groups,

70

but the Mediterranean sample had a lower heterozygosity than the other three groups
(0.66 versus 0.76-0.78, Table 3.4).

Haplotype richness was significantly higher in the North Atlantic group than in
either southern group. The Mediterranean group had an intermediate value and the
Confidence Interval for this group overlapped with all the CI obtained in other groups.
However, only 4.3% of the resampled data sets from North Atlantic gave values lower
than the observed A_R for the Mediterranean Sea group. The lowest haplotype richness
was in the Southwest Atlantic although observed values did not differ significantly from
those in the Mediterranean Sea or Southeast Atlantic groups. Haplotype diversity in the
North Atlantic group was significantly higher than in the other regions. H_D in the
Mediterranean region was higher than that in Southwest Atlantic but not than that
recorded in the Southeast Atlantic. The haplotype diversity in the Southeast and
Southwest Atlantic did not differ significantly.

Table 3.4

Comparison of Genetic Diversity Statistics Among Geographic Regions

	NA	MED	SEA	SWA
A_R	10.40[a]	6.98[c]	9.60[b]	9.44[b]
	(3.65 – 19.32)	(1.00 – 15.79)	(2.11 – 17.80)	(1.93 – 17.36)
He	0.78[a]	0.66[b]	0.76[a]	0.77[a]
	(0.39 - 0.96)	(0.00 - 0.93)	(0.14 – 0.96)	(0.08 – 0.96)
H_R	8.746[a]	5.000[ab]	4.373[b]	2.571[b]
H_D	0.8694[a]	0.7255[b]	0.4525[bc]	0.2689[c]

A_R: allelic richness; H_e: expected heterozygosity; H_R: haplotype richness; H_D: haplotype diversity; superscript letters show

homogeneous groups as inferred during pairwise comparisons using the Wilcoxon test (microsatellites) or overlaps of 95% bootstrap

resampling confidence intervals (mtDNA, see section 3.2.2 for details).

3.3.2 Contemporaneous Gene Flow

The very small sample sizes available for the Mediterranean region (n = 17) prevented estimating reliably allele frequencies in this population. Because inferences on ancestry and hybridization involving that population would be compromised due to this small sample size, the Mediterranean cluster was not considered in statistical analyses aiming to formally estimate migrations and hybridization rates; these analyses focused on inferences of ancestry in the remaining 3 groups. Bayesian clustering in STRUCTURE did not reveal any significant ancestry from the Mediterranean Sea cluster in any other geographic region (Figure 3.6a) and all individuals sampled in MED had very similar ancestry patterns with a large majority (> 79%) of ancestry in the MED cluster suggesting they were all purebred individuals from that population. Thus, while interactions between MED and the other regions cannot be ruled out, these results suggest that this population is effectively isolated from the rest of the range and the impact of omitting MED during inferences on migrations and hybridization in other populations discussed below is minimal.

The first part of this analysis consisted in assigning individuals to one of the three parental stocks or one of the hybrid categories using the procedure described in section 3.2.2.2 based on the identification of reference putative pure genotypes for each of the three clusters in STRUCTURE and their use to calculate the probability of assignment of all individuals in the dataset to purebred or hybrid categories in NEWHYBRIDS. Detailed assignments results are presented in Appendix Table A.1. The reliability of assignments to parental stocks and hybrid categories was assessed by simulating the genotypes of hybrid and pure individuals in HYBRIDLAB and examining the proportion of correctly and

incorrectly classified genotypes during analysis in NEWHYBRIDS as well as the 95% CI of ancestry proportions inferred in STRUCTURE for these genotypes. The results of this analysis are presented in Table 3.5.

Examination of the ancestry proportions of simulated pure individuals from each of the three groups and their hybrids (F1, F2, and two types of backcrosses) indicated that ancestry proportions were consistent with expectations: pure individuals averaging 96.3% in their origin cluster, F1s and F2s hybrids displaying close to 50% of ancestry in each of their two parental clusters on average (Table 3.5), and backcrosses displaying ancestry in their two parental clusters close to the expected 25%/75% ratio. However, 95% Confidence Intervals overlapped largely between categories preventing assignment as pure or hybrid types with confidence using this method.

The power of the dataset to assign individuals to pure or hybrid categories was further examined based on the comparison of posterior odds in NEWHYBRIDS. Pure individuals from all three regions were correctly reclassified with high confidence (> 99.7%) in NEWHYBRIDS and the percentages of F1 correctly assigned varied between 31% and 64% (Table 3.5). On average only 3.3% of simulated second generation hybrids and backcrosses were correctly assigned indicating that the dataset does not allow drawing robust inference on F2 and backcross hybrids. Thus, results presented below only focus on inferred individuals of pure ancestry and F1 hybrids. Assignment decisions were considered using both the odd ratio and relaxed criteria presented in section 3.2.2.2. While the percentage of recovered pure and F1 hybrids was slightly improved when using the relaxed criterion, the percentage of false positive also increased, sometimes in

high proportions (Table 3.5). Therefore, assignment of individuals to categories for the

analysis of migrations below focused on results obtained based on the odd ratio criterion.

Table 3.5

Results of Assignment Tests Conducted on Purebred from Each Population and Simulated Two-Way Hybrids

| Parental pair | Hybrid category | STRUCTURE | | | NEWHYBRIDS | | | | | |
| | | NA | SEA | SWA | Correct assignment | | Incorrect assignment | | False positive | |
					OR	Relaxed	OR	Relaxed	OR	Relaxed
NA x SEA	pureNA	0.925-0.991	0.008-0.069	0.001-0.014	100	100	0	0	16.3	20.5
	pureSEA	0.024-0.086	0.805-0.933	0.035-0.139	100	100	0	0	14	25
	F1	0.457-0.508	0.489-0.540	0.002-0.004	64	88	36	12	12.5	26.3
	F2	0.432-0.525	0.473-0.566	0.002-0.003	0	1	100	99	0	0
	BxNA	0.703-0.779	0.219-0.295	0.001-0.003	0	0	100	100	0	0
	BxSEA	0.225-0.300	0.699-0.773	0.001-0.003	1	8	99	92	0.3	5.8
NA x SWA	pureNA	0.925-0.991	0.008-0.069	0.001-0.014	100	100	0	0	9.5	12.3
	pureSWA	0.006-0.069	0.036-0.272	0.720-0.958	92	96	8	4	8	15.8
	F1	0.460-0.512	0.002-0.003	0.486-0.537	31	69	69	31	7.8	20
	F2	0.418-0.539	0.001-0.002	0.459-0.580	1	2	99	98	0	0.5
	BxNA	0.688-0.771	0.002-0.004	0.226-0.309	0	0	100	100	0	0
	BxSWA	0.197-0.272	0.002-0.003	0.726-0.800	8	38	92	62	6.8	24.5
SEA x SWA	pureSEA	0.024-0.086	0.805-0.933	0.035-0.139	100	100	0	0	12.5	20.8
	pureSWA	0.006-0.069	0.036-0.272	0.720-0.958	92	96	8	4	15.3	27
	F1	0.001-0.001	0.472-0.559	0.441-0.528	52	71	48	29	18.3	34.5
	F2	0.001-0.001	0.452-0.547	0.452-0.547	0	0	100	100	0	0
	BxSEA	0.001-0.002	0.725-0.813	0.186-0.274	0	0	100	100	0	0
	BxSWA	0.001-0.001	0.270-0.468	0.530-0.729	0	0	100	100	0	0

Results of assignment of purebred and simulated hybrids to hybrid classes in NEWHYBRIDS for the three crosses considered (NA x SEA, NA x SWA, and SEA x SWA). STRUCTURE results feature the percentages of ancestry in each of the three clusters, while results of NEWHYBRIDS include the percentage of simulated individuals correctly assigned to their category (correct) or to another category (incorrect) and the percentage of false positive (individuals from other categories assigned to the category under consideration by error). F1: first generation hybrid; F2: second generation hybrid; Bx: backcross; OR: odd ratio in favor of the selected assignment greater than 3 fold; Relaxed: assignment to the category for which the posterior probability is the highest.

Individuals assigned as pure genotypes from a parental population other than that hosted in the geographic region where the sample was collected were considered direct (F0) migrants and their numbers were used to calculate contemporaneous migration rates among the four geographic regions (Table 3.6). Migration rates were generally low except for SWA where 40.8% of the individuals were assigned as immigrants. Immigrants to SWA were mostly from the SEA region (36.7%) although a small fraction was assigned to NA (4.1%). No emigrant from this region was detected. No migrant to and from the MED population was detected, although statistical evaluation was not conducted in NEWHYBRIDS as discussed above. Exchanges of migrants between the SEA and NA were moderate (1.4% of migration from NA to SEA and 2.4% migration from SEA to NA).

Table 3.6

Estimates of Current Migration Rates Between Geographic Regions

| | | To | | | |
		NA	MED	SEA	SWA
From	NA	97.6	0	1.4	4.1
	MED	0	100.0	0	0
	SEA	2.4	0	98.6	36.7
	SWA	0	0	0	59.2
	n	791	17	138	49

Migration matrix representing the pairwise migration rates between geographic populations. Migration rates are calculated as the frequency of F0 migrants inferred in NEWHYBRIDS.

Migrations routes among regions are summarized in Figure 3.7.
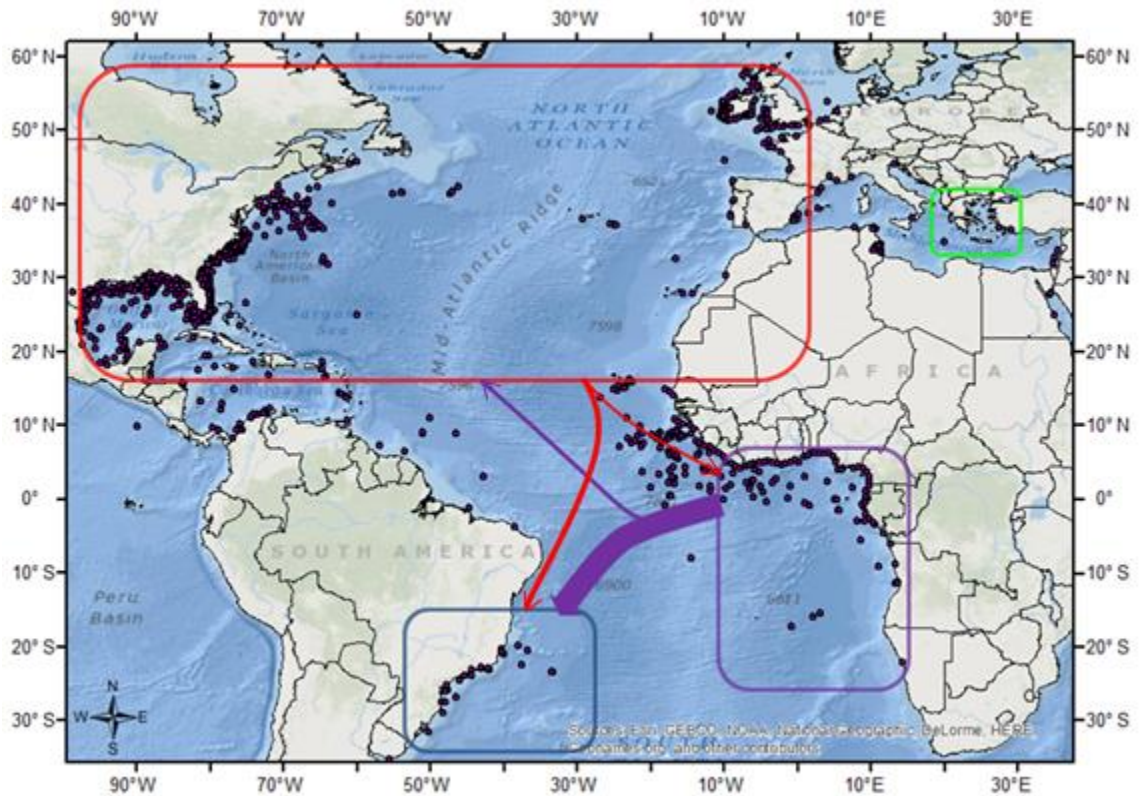
Figure 3.7 Migration Routes for Gray Triggerfish Inferred from the Results of Assignment Tests

Diagram representing contemporaneous migrations rates among gray triggerfish populations from various regions within the Atlantic Ocean. Arrows represent migration routes supported by assignment results in NEWHYBRIDS presented in Table 3.6. Red: migration from NA; Purple: from SEA; Blue: from SWA. The thickness of the arrow is proportional to the inferred migration rate.

F1 hybrids were only detected between SEA and NA and were found in much lower abundance than the F0 migrants, with a total of 0.3% of F1s in the overall dataset versus 4.1% of F0 migrants. The percentage of F1s inferred in the SEA, SWA, and NA regions were 0.7%, 0%, and 0.2%, respectively (Table 3.7).

Table 3.7

Frequencies of First Generation Hybrids F1 Inferred During NEWHYBRIDS Assignments

		Location of capture			
		NA	MED	SEA	SWA
	NA		0.0	0.7	0.0
Origin of	MED	0.0		0.0	0.0
parent #2	SEA	0.2	0.0		0.0
	SWA	0.0	0.0	0.0	
	n	805	17	142	52

3.3.3 Phylogeography and Historical Demography

The phylogenetic trees generated from the microsatellite dataset considering $(\delta\mu)^2$ and the F_{ST} distance are presented in Figure 3.8 and 3.9 respectively. Trees based on all 12 sampling localities are presented in Figure 3.8a and 3.9a while those generated using the 4 groups inferred during analysis of spatial genetic variation are presented in Figures 3.8b and 3.9b. All the populations from the North Atlantic group clustered in the same branch consistent with findings of the population structure study (Figures 3.8a and 3.9a). Similarly, the 3 populations from the South Atlantic group (Brazil, Benin, and Angola) clustered in a separate branch. The distinction of the South Atlantic and North Atlantic regional groups respectively was recovered with high support ($\geq 84\%$) in both analyses (F_{ST} and $(\delta\mu)^2$ based) consistent with the ranking of F_{ST} estimates between these two regions (Table 3.3). The status of the Mediterranean Sea group differed depending on the molecular distance used to generate the tree. Trees obtained based on F_{ST} isolated MED in a separate branch consistent with the results of pairwise comparisons of MED with all the other populations that yielded the highest F_{ST} estimates (Table 3.3). However, the

$(\delta\mu)^2$ distance between MED and the North Atlantic was smaller than the distance

between MED (or North Atlantic) and the South Atlantic groups which led to the

clustering of NA and MED populations in the same branch with high support ($> 87\%$) in

this analysis.
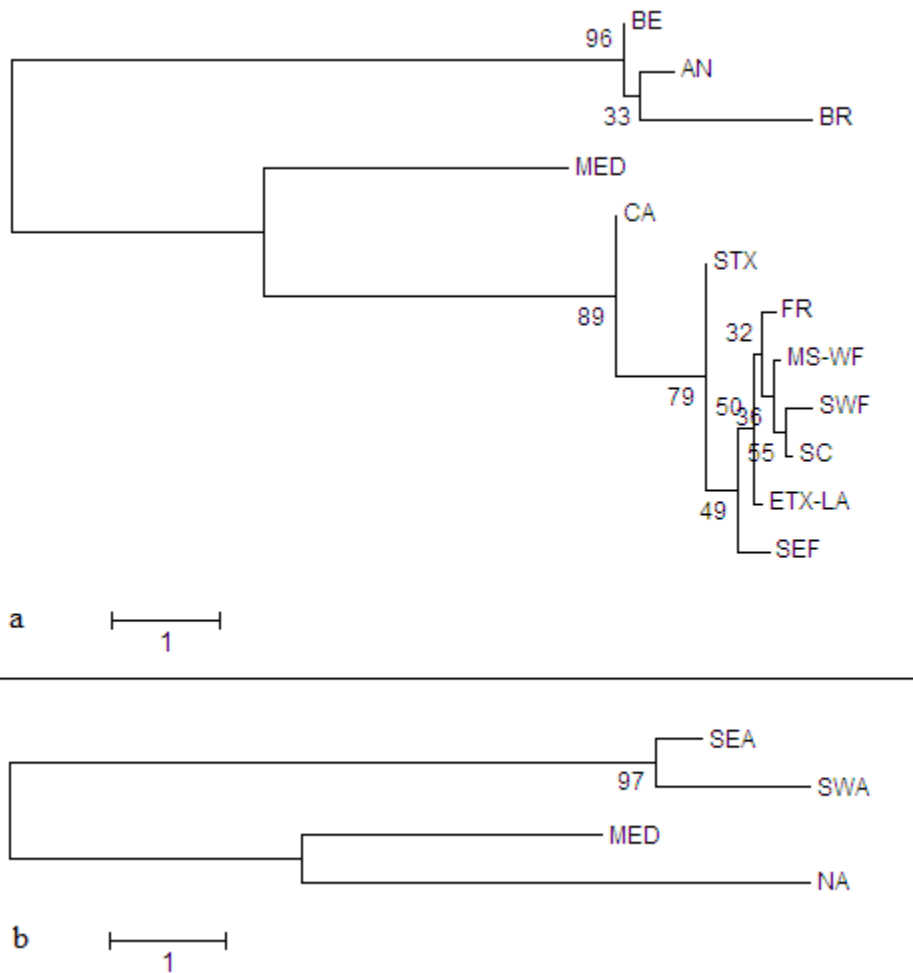
Figure 3.8 Neighbor-Joining Tree of Gray Triggerfish Populations Based on the $(\delta\mu)^2$
Distance

Neighbor-joining tree constructed based on the $(\delta\mu)^2$ distance for the microsatellite dataset accounting for the 12 sampling localities

(a) or 4 geographic regions (b). The scale bar under the trees represents one unit $(\delta\mu)^2$ distance.
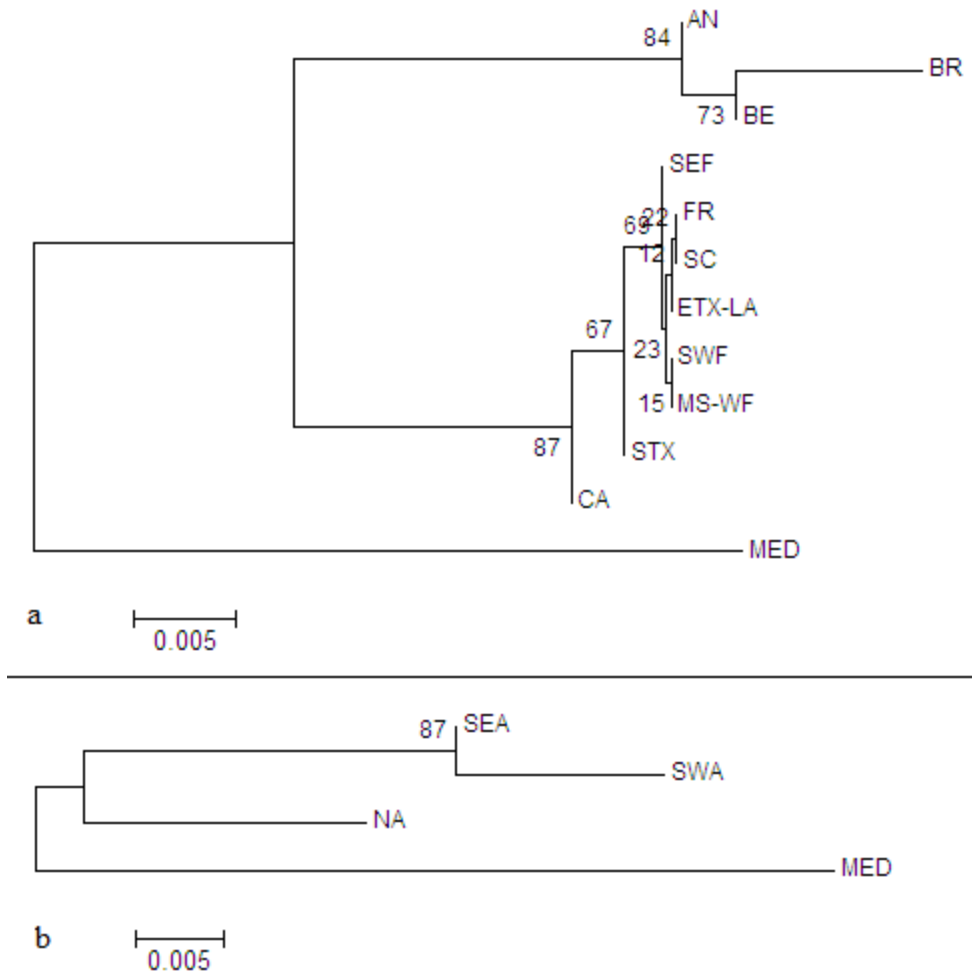
Figure 3.9 Neighbor-Joining Tree of Gray Triggerfish Populations Based on the F_{ST} Distance

Neighbor-joining tree constructed based on the F_{ST} distance for the microsatellite dataset accounting for the 12 sampling localities (a) or 4 geographic regions (b). The scale bar under the trees represents units of F_{ST} distance.

The demographic history of gray triggerfish populations was also examined based on the mitochondrial DNA dataset (Figure 3.10). The statistical parsimony network of haplotypes revealed a clear divergence between the North (NA) and South Atlantic (SA) groups with most of the South Atlantic samples bearing haplotypes clustered in a distinct 'South Atlantic' clade. The South Atlantic clade displays a star like phylogeny derived from a haplotype that has a very low frequency in the North Atlantic group. This clade

has very low frequency of occurrence in the North Atlantic group (4%) although its

frequency is slightly higher in the Mediterranean Sea (MED, 17%). The phylogeny of

haplotypes in the North Atlantic group is more complex with three different haplotypes

found at high frequencies (18.7-43.2%) all showing numerous derived haplotypes found

at low frequencies (< 4%). This clade has a very low frequency in the South Atlantic

group (7%).
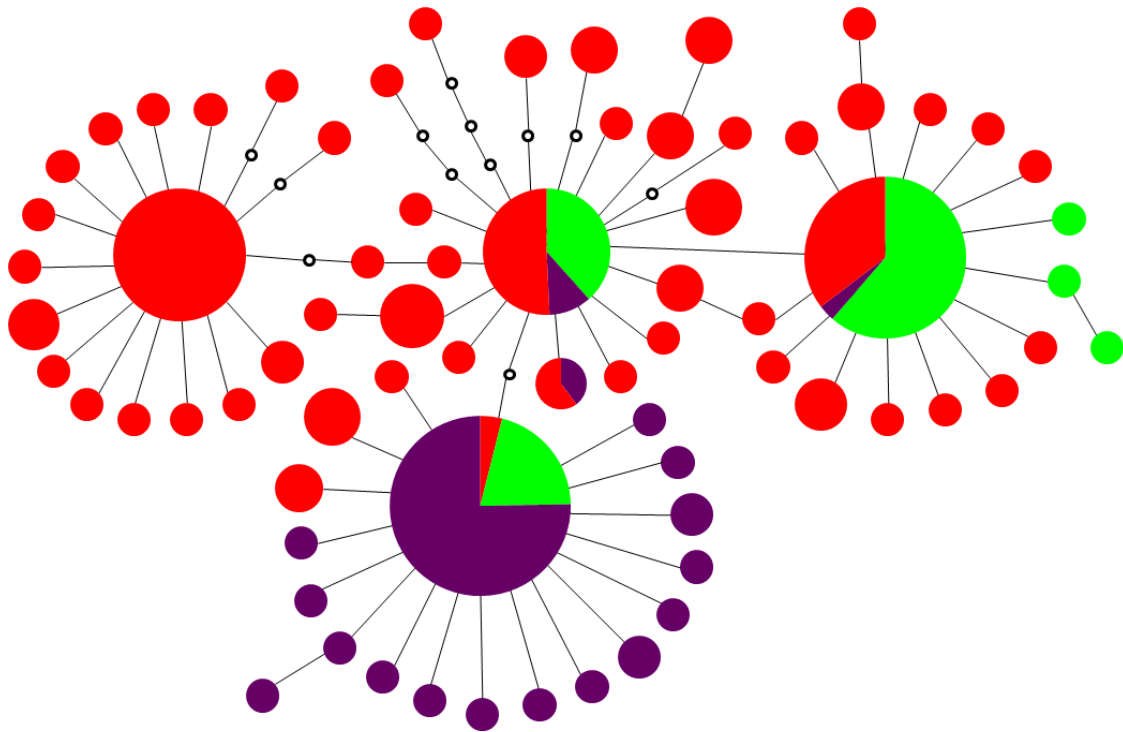
Figure 3.10 Statistical Parsimony Network of Gray Triggerfish MtDNA Haplotypes

Each circle represents a unique haplotype with size proportional to its frequency in the dataset. Black hollow circles represent

unsampled haplotypes. Pie charts show the relative frequencies of each of the three groups identified during analysis of spatial genetic

variation. Red circles: NA; Green circles: MED; Purple circles: SA.

Bayesian coalescent estimates of historical migration rates (*m*) and effective

population size (N_e) obtained from the microsatellite dataset are presented in Table 3.8.

Based on 95% highest posterior density, the North Atlantic group showed a significantly

larger estimate of N_e than the Mediterranean Sea and Southwest Atlantic groups and the

Mediterranean group had a significantly smaller N_e than any other location. These results

parallel the outcome of comparisons of summary statistics for the same dataset

summarized in Table 3.4. Migration rates between regions did not differ significantly

from each other.

Table 3.8

Bayesian Coalescent Estimates of Gray Triggerfish Effective Population Size and

Historical Migration Rates in the Atlantic

		NA	MED	SEA	SWA
N_e		3,150 (1,700 - 4,600)	17 (0 - 833)	883 (33 - 1,767)	850 (0 - 1,633)
m *from*	NA		0.012 (0 - 0.093)	0.008 (0 - 0.09)	0.008 (0 - 0.09)
	MED	0.008 (0 - 0.09)		0.008 (0 - 0.09)	0.008 (0 - 0.09)
	SEA	0.012 (0 - 0.093)	0.008 (0 - 0.09)		0.008 (0 - 0.09)
	SWA	0.008 (0 - 0.09)	0.008 (0 - 0.09)	0.008 (0 - 0.09)	

Estimates of N_e and pairwise *m* with 95% confidence intervals in 4 gray triggerfish populations within the Atlantic basin. Estimates are based on 17 microsatellite markers and assuming a mutation rate of 5×10^{-4}. Source and sink for migrants are represented in rows and columns respectively.

Bayesian coalescent estimation of migration rates and effective population size

based on the mtDNA dataset was evaluated during several test runs performed using a

broad range of priors and various chain lengths. All searches attempted failed to

converge, and these estimates are therefore not presented.

3.4 Discussion

3.4.1 Population Structure

The analysis of genetic variation among the 12 sampled geographic populations of gray triggerfish spanning across the entire species range revealed the occurrence of significant spatial heterogeneity in microsatellite allele and mitochondrial DNA haplotype frequencies. Pairwise comparisons of allele distributions, Bayesian clustering in STRUCTURE, and spatial analysis of molecular variance based on the microsatellite dataset revealed the occurrence of four differentiated groups. The first group encompassed the North Atlantic and included the samples collected in U.S waters for the Northwest Atlantic (Gulf of Mexico and U.S. East coast) and those collected in France and the Canary Islands for the Northeast Atlantic. Pairwise F_{ST} estimates within this group were all very low (less than 0.003), a finding consistent with the hypothesis that high levels of genetic connectivity are maintained across the area. Homogeneity in the Northwest Atlantic was hypothesized to occur as a result of transport of *Sargassum* across the Gulf of Mexico and then along the Florida current and the Gulf Stream, potentially connecting the different parts of the Gulf of Mexico and the U.S. East coast. While an isolation-by-distance pattern of population structure was proposed, the high velocity of some of the currents involved, in particular, the Loop current – Florida current – Gulf Stream system, was expected to promote effective transport of gray triggerfish across the entire region during the course of their extended pelagic ecophase. Genetic heterogeneity within that part of the range was, in fact, minimal with F_{ST} estimates all lower than 0.0019. Other species occupying offshore benthic habitats similar to those

83

used by adult gray triggerfish (e.g. the red snapper, Hollenbeck et al. 2015; Norrell 2016, or the red Porgy, Ball et al. 2007) and dispersing pelagic larvae under the influence of the same currents also display homogeneous allele frequencies across the Gulf of Mexico and South Atlantic regions despite much shorter pelagic transport durations than that reported for the gray triggerfish.

Connectivity between the Northwest Atlantic and the Northeast Atlantic was hypothesized to result from pelagic transport along the Gulf Stream. The homogeneity in allele frequency between U.S. and eastern Atlantic stocks indicated by very low F_{ST} estimates (less than 0.002) suggests that connectivity is occurring between the two regions and that some gray triggerfish larvae/juveniles do cross the Atlantic via the Gulf Stream and contribute to populations reported in Europe (England and France). Data from mixed-layer Ocean drifters were used to simulate the transport of surface waters (and *Sargassum* patches) between the U.S. East coast and the Northeast Atlantic. The predicted duration of the transport involved in connecting the two regions outlasted the 4 to 7-month duration of the pelagic phase of gray triggerfish estimated by Simmons (2008) in the Gulf of Mexico. However, estimates of the time required to cross the Atlantic only accounted for a limited number of drifters and didn't consider the potentially important effect of storms in causing episodic connectivity. In addition, gray triggerfish post-larvae trapped in patches of *Sargassum* drifting across the Atlantic may delay settlement for as long as they remain in the open Atlantic until a benthic habitat suitable for settlement is reached, leading to a longer pelagic transport period than that reported in the Gulf of Mexico where benthic habitats are expected to be encountered

more frequently by juveniles. Finally, the lower surface temperatures expected in the

Northeast Atlantic would be expected to slow down the growth and development of

triggerfish juveniles as the Gulf Stream proceeds North towards that region (Pankhurst

and Munday 2011; Rankin and Sponaugle 2011), further contributing to favor a delayed

settlement.

Connectivity between the Northwest Atlantic and the Northwest Africa was

hypothesized to occur through transport via the Azores current that arises from the

branching out of the Gulf Stream as it reaches the East Atlantic. Patches of *Sargassum*

and triggerfish larvae coming from the West Atlantic through the Gulf Stream could be

transported towards the Northwest coast of Africa through that route, leading to the

genetic similarity of Canary Islands and U.S. populations. The Canary Islands sample

was slightly differentiated from U.S. samples (5 out of 6 pairwise comparisons of allele

distributions between these two stocks were significant and the average F_{ST} during those

comparisons was 0.0013) suggesting that the Canary Islands stock may maintain some

degree of demographic independence. However, these slight differences in allele

frequencies were in part due to the high proportion of F0 migrants from the South

Atlantic in the samples collected for this study as discussed below. Indeed, the average

F_{ST} between Canary Islands and U.S. populations was only 0.0007, and only two

comparisons (instead of 5) between Canary Islands and U.S. samples remained

significant when F0 migrants and F1 hybrids were excluded from the dataset. Individuals

that were not identified as F0 or F1 hybrids also showed a clear ancestry in the northern

group during Bayesian clustering in STRUCTURE. Overall, these results are consistent with

the hypothesis that the Canary Islands have historically maintained genetic connectivity with the rest of the North Atlantic group. The observation of F0 migrants and F1 hybrids thus suggests that connectivity between this geographic population and southern stocks in the South Atlantic is a recent event. If sustained, this new connectivity may lead to progressive divergence of Canary Islands gray triggerfish from the northern group, if migrants from southern stocks successfully interbreed with local fish and contribute to the gene pool. Genetic monitoring of the gray triggerfish in this area is therefore warranted in order to determine if migrations from the southern group continue and delineate their geographic range along the Northwest African coasts.

The F_{ST} estimates generated during comparisons of microsatellite allele distributions in the Mediterranean Sea sample to those in the North Atlantic samples were the highest in magnitude among those obtained from the entire dataset (> 0.066, Table 3.3). All the samples obtained from the Mediterranean Sea sampling locality showed high levels of ancestry in an independent cluster in STRUCTURE and the MED locality was assigned to a distinct group in SAMOVA. While the sample size for the MED location is small, the lack of evidence for significant ancestry from the MED cluster in any other locality sample suggests that no migration from this group to the North Atlantic region has occurred in recent generations. Similarly, there was no sign of migration in the MED region from the North Atlantic group based on STRUCTURE results. Overall these results suggest that the MED population is not currently exchanging migrants with other groups, in particular, the North Atlantic group. Reduced connectivity between Mediterranean gray triggerfish and the rest of the range was hypothesized to result from the Alboran Sea

front that has been shown to isolate populations in the Mediterranean Sea from those in the Atlantic for several species (Quesada et al. 1995; Naciri et al. 1999; Lo Brutto et al. 2004) and seems to be effective at preventing entry of North Atlantic gray triggerfish into the Mediterranean Sea. Further study of the Mediterranean population employing larger sample sizes and additional cohorts is needed to confirm this assessment.

The two remaining groups identified during the analysis of population structure were located in the South Atlantic, with a southeastern group that included samples from the Gulf of Guinea and Angola, and a southwestern group that included samples from the South coast of Brazil. The divide between the North Atlantic and South Atlantic regions was strongly supported by differences at both microsatellite allele and mtDNA haplotype frequencies. The separation of these two regions received 100% support in BARRIER and, while F_{ST} at microsatellites was moderate (0.04 on average), the divergence was very strong at mtDNA with estimates of F_{ST} between northern and southern localities averaging 0.55 and a near complete sorting of mtDNA haplotype lineages between the two regions inferred during statistical parsimony clustering. The lack of connectivity between the North and South Atlantic on the East side of the basin may be due to slow and diverging currents between West Africa and the Gulf of Guinea. Stromme (1984) hypothesized that stocks off of Guinea Bissau and the Gulf of Guinea were independent based on the observation of differences in life history traits, a subdivision consistent with the observed genetic discontinuity between the Canary Islands and the Gulf of Guinea locality in this study. On the western side of the basin, the South Brazilian and North Brazilian currents flow in opposite directions and are expected to isolate populations off

87

Central and South Brazil from those along the northern part of South America and the

North Atlantic group as discussed in Hypothesis G described in section 3.1.4. The finding

of genetic differences between the Southwest Atlantic and the North Atlantic is,

therefore, consistent with the proposed hypothesis.

Isolation between the Southeast and Southwest Atlantic was hypothesized to

result from the very slow velocity of the South Equatorial current, the main surface

current capable of transporting floating *Sargassum* and gray triggerfish larvae from

Southwest Africa to South America. A specific cluster was identified in the Southwest

region in STRUCTURE but divergence between the Southeast and Southwest groups was

moderate at microsatellites (BR vs BE: $F_{ST} = 0.012$; BR vs AN: $F_{ST} = 0.019$), and non-

significant at mtDNA (P-value > 0.69). The inferred barrier to gene flow in BARRIER had

very weak support suggesting that connectivity is occurring, yet is limited considering the

significant differences in microsatellite allele frequencies between the two groups. The

sample from southwestern Atlantic included a high fraction (36.7%) of F0 from the

Southeast group. When only individuals assigned as pure genotypes from this cluster

were used to characterize allele frequencies in the group, the divergence was more

pronounced (BR vs BE: $F_{ST} = 0.029$; BR vs AN: $F_{ST} = 0.034$). Although the small

number of samples from the southwestern group prevented estimating accurately F_{ST} and

quantifying the divergence between the two groups, these results suggest that past

connectivity was limited and led to the accumulation of allele frequency differences.

Further monitoring of this population in the context of new and possibly continued

migrations from the other groups, in particular, the southeastern group, due to new

Sargassum influx in the equatorial region is therefore warranted.

3.4.2 Contemporaneous Migration and Connectivity Among Geographic

Populations of Gray Triggerfish

Based on the results of the analysis of population structure, the four groups

described above were distinguished during a study of contemporaneous migration

patterns. The Mediterranean cluster was omitted from this analysis due to the sample size

in this population that was too small to infer reliably allele frequencies and ancestry

proportions. However, all individuals sampled in the Mediterranean location had a very

high ancestry proportions in that cluster suggesting that no F0 migrants or F1 hybrids

were present in the dataset. While the occurrence of immigrants from other groups, in

particular from the North Atlantic group, cannot be ruled out, the present results indicate

that their frequency is likely moderate at best. In addition, the proportion of ancestry

attributed to the Mediterranean cluster was less than 5% in the genotypes of 98.6% of the

individuals sampled in other regions. The remaining individuals collected in non-

Mediterranean Sea locations (1.4%) had proportion of ancestry in the MED cluster

between 5% and 16%, effectively ruling out the possibility of F0 migrants or F1 hybrids

from the Mediterranean population in other parts of the range. Further analysis, therefore,

focused on the relationships among the remaining 3 groups (North Atlantic, Southeast

Atlantic, and Southwest Atlantic).

The ancestry of individuals was inferred in STRUCTURE and the first step of this

work focused on determining the power of the dataset for the detection of individuals of

pure ancestry from each stock, F1 and F2 hybrids, or backcross of F1 hybrids to parental

stocks. Reference pure individuals from each of the 3 groups were selected and used to

simulate hybrid genotypes that were subjected to inference of ancestry in STRUCTURE and

formal assignment to pure parental or hybrid categories in NEWHYBRIDS. The mean

ancestry proportions obtained from pure and simulated individuals were generally

consistent with expectations in that pure individuals from each of the 3 groups had a high

ancestry proportion in their cluster during analyses in STRUCTURE ($\geq 72\%$), simulated F1

and F2 hybrids had intermediate values, and simulated backcross hybrids were

approaching the expected 25%/75% of ancestry in the migrant and backcrossed

population respectively (Table 3.5). However, the ranges of ancestry proportions inferred

in simulated individuals overlapped between groups preventing reliable classification of

individuals in one of the hybrid or pure categories. The identification of pure and hybrid

genotypes was therefore conducted using the Bayesian approach in NEWHYBRIDS which

revealed that pure individuals were correctly classified in 99.7% of the cases and a

reasonable percentage of F1s were identified (31 to 64%) as well, while F2 and backcross

hybrids were poorly recovered (only 3% were identified) and mostly confounded with

pure individuals. Challenges in resolving F1, F2, and backcross hybrids, in this case, are

likely due for a large part to the low levels of divergence among the 3 groups assessed. A

simulation study conducted by Vähä et al. (2006) revealed that 12 to 24 microsatellites

were needed to accurately distinguish F1 and purebred individuals when the F_{ST} values

between the parental populations were 0.24 and 0.12 respectively. The F_{ST} estimates

obtained in this study are lower than these values and the recovery of only a proportion of

90

F1 and pure individuals is, therefore, consistent with the study of Vähä et al. (2006).

Distinguishing backcross hybrids from pure individuals was more challenging and would

have involved a heavy genotyping effort (48 microsatellites or more were required in the

conditions of the simulations conducted by Vähä et al. 2006) inconceivable in a study

such as ours. Thus, the low power for detection of backcross hybrids in the present study

is also consistent with findings of Vähä et al. (2006). Considering the high level of

divergence at mtDNA between the northern and southern groups recorded in this study,

incorporating the mtDNA haplotypes in future inferences may be useful in order to assess

maternal ancestry in these two groups. The occurrence of separate lineages largely sorted

between the northern group and the two southern groups would be expected to help

resolve the ancestry of some of the F0 migrants and some of the F1 or backcross hybrids,

although the occurrence of a few haplotypes from the southern clade in the northern

localities and of haplotypes from the northern clades in the southern localities indicates

that the rejection of northern (or southern) maternal origin would not be possible with

high confidence and would require including other genetic loci. Alternatively, high-

density genome scans such as those used in Chapter V of this dissertation could be used

to characterize individuals and improve the statistical power to distinguish the various

categories of hybrids. Genome scans, if used in the context of a linkage map, allow

assessing independently the ancestry at the level of each chromosome or known portions

of chromosomes accounting for linkage, which would be expected to improve the

accuracy of the estimation of ancestry. Using this approach, Pritchard et al. (2016) were

indeed able to accurately distinguish hybrid classes up to the third hybrid generation in

the context of hybridizing salmon populations. Based on the results of simulations conducted in the present study, inferences on migrations focused on pure and F1 individuals only, considering that only few F1s were expected to be misclassified as pure individuals (6.3%) and that the probability of pure individuals being classified as F1 was less than 1%. A caveat that needs to be recognized is that a fraction of the inferred pure and F1 individuals (up to 31.3% for the pure and 17.1% for the F1s) could be backcross or F2 hybrids misclassified in NEWHYBRIDS. However, considering the very low proportion of F1 hybrids detected in the study, the proportions of F2 and backcross hybrids in the dataset was likely negligible and in consequence, the proportion of genotypes misclassified as F1 or pure was also expected to be negligible.

Direct F0 migrants were inferred as individuals with a pure genotype assigned to one of the 3 clusters but were captured in a geographic locality housing one of the other clusters. As discussed above, all 17 individuals sampled in the Mediterranean Sea population showed very similar ancestry patterns in STRUCTURE with a large majority in the Mediterranean Sea cluster, suggesting that they were purebred from this cluster or at least did not have a recent ancestor in another cluster. Similarly, the proportion of ancestry attributed to the Mediterranean Sea cluster in individuals collected in the other regions was very low (0.7% on average) as discussed above, indicating that no F0 migrant from MED or F1 hybrid descending from a F0 Mediterranean migrant was found in the other 3 regions. Overall, and although this inference needs to be confirmed using larger sample sizes, these observations suggest that current migrations between the Mediterranean Sea and the rest of the range are very infrequent and that the isolation of

92

the MED population inferred during population structure analysis was still effective in the generations sampled during the study. Isolation is hypothesized to result from the Alboran Sea front as discussed above and the low genetic diversity in the MED sample suggests a genetic bottleneck may have occurred in this population, possibly signaling a founder effect where colonization of the Mediterranean basin resulted from a rare intrusion and successful establishment of a few gray triggerfish.

Migration rate estimates between the North Atlantic and the two southern stocks were relatively low but multiple migrants were detected. The estimated migration rate from NA to SEA was 1.4%, but it was more than 4% to the SWA and a 2.4% migration rate from SEA to NA was inferred. No migration from SWA to SEA was detected as supported by the odd ratio criterion, but a high proportion (36.7%) of genotypes obtained from individuals collected in SWA were classified as pure SEA origin (F0 migrants). Considering the long-term estimates of effective population size for the NA, SEA, and SWA, estimates of the effective number of migrants per generation would be 12 from NA to SEA, 49 from NA to SWA, 79 from SEA to NA, and 315 from SEA to SWA. Under an island model of migration, the effective number of migrants exchanged between regions being greater than 50 for all pairs, F_{ST} would be expected to be in the order of 0.005, close to one order of magnitude lower than the actual estimates of F_{ST} obtained based on the microsatellite dataset. The migration rates inferred here are therefore surprisingly high and *a priori* incompatible with the magnitude of F_{ST} estimates, in particular with the high divergence recorded at mtDNA between the southern and northern groups. Two hypotheses can be offered to explain this result. A first hypothesis

93

is that the migration rates inferred from the percentages of F0 in the present dataset do not reflect the average migration rate per generation that led to the accumulation of the observed divergence among populations. Until recently, pelagic *Sargassum* was hypothesized to form in the Gulf of Mexico and circulate around the Florida peninsula to join the Sargasso Sea (Gower and King 2008, 2011). However, since 2011, multiple massive *Sargassum* landings in the Caribbean Sea and West Africa have occurred and the investigation conducted on these events suggested the occurrence of new *Sargassum* sources North of the Amazon evidenced by Gower et al. (2013) and backtracked to the NERR by Johnson et al. (2013) and Franks et al. (2016) using numerical circulation models. The *Sargassum* produced in this area was hypothesized to circulate in the NERR before being entrained in the North Brazil current and reaching the Caribbean region. This new source of *Sargassum* and the circulation pattern associated with the NERR could have promoted connectivity between southern populations of gray triggerfish located on the East and West side of the Atlantic basin and also connected those populations with northern stocks, part of the NA group. The occurrence of migrations from NA to southern populations likely does not involve the Northwest Atlantic populations but rather the genetically similar stocks in the Northeast Atlantic, in particular, those in the Canary Islands region. Gray triggerfish spawned in the Canary Islands region could be caught in the NERR and transported to southeastern or southwestern geographic populations of gray triggerfish through the NERR circulation. Similarly, gray triggerfish larvae spawned in the Gulf of Guinea could be collected in the NERR and further entrained in the North Brazil current which could transport them to the

Gulf of Mexico or alternatively these larvae could connect with the South Brazilian current and be transported to the SWA population. The duration of the pelagic transport involved in reaching the Brazilian coasts is expected to be very long as discussed in section 3.1.4. Examination of the geographic distribution of migrants from SEA to NA indicates that 42% (8 out of 19) were found in the Canary Islands and represented 10.5% of collected individuals in that locality versus only 1.5% in the rest of the North Atlantic group, indicating that a substantial part of the connectivity between NA and SEA occurs in Northwest Africa. However, some SEA genotypes were also found in the Gulf of Mexico and in Southeast Florida and, as discussed above, the direct transport from the Gulf of Guinea seems very unlikely considering the extremely long duration of transport involved in reaching South America. Most triggerfish larvae and juveniles reaching the South American shelf would be more likely to attempt to settle on suitable habitats in that region rather than remaining for extended periods of additional time in the *Sargassum* to reach the Gulf of Mexico or the U.S. East coast. Alternatively, the observation of SEA genotypes in U.S. geographic populations may be explained by the occurrence of new settlements of gray triggerfish of SEA origin on the North Brazilian coast following transport by the NERR during recent *Sargassum* events. If these populations only include SEA genotypes, their larvae can be entrained in the North Brazil current and distribute migrants to U.S. waters as discussed in section 3.1.4.

Transport from SEA to SWA could follow the above pattern involving the NERR, although direct transport seems difficult considering the slow velocity of the NERR system. Recent settlement of SEA genotypes at the tip of Brazil (e.g. state of Rio Grande

95

Do Norte) could originate from triggerfish larvae that would be transported South by the South Brazilian current. The apparent rate of migration from SEA to SWA was very large (36.7%). This could reflect that the Brazilian population is small (a hypothesis consistent with the reported decline of the species in that region, Ataliba et al. 2009; Netto and Di Beneditto 2010) and/or that it received a substantial amount of direct migrants from the SEA in the sampled generation following the recent *Sargassum* events.

Finally, migrations from SWA to NA or to SEA were not detected, a finding consistent with the lack of surface currents connecting central Brazil to the North Brazil current or the NERR.

Further insights into contemporaneous gene flow among gray triggerfish regional populations are provided by the analysis of F1 hybrids. While the simulation study indicated that only about half of the F1 hybrids were expected to be detected, the very low proportion of F1 hybrids detected throughout the range suggests that the frequency of F1 hybrids in the different groups is much smaller than that of F0 migrants and is more compatible with the observed divergence (F_{ST} estimates) between populations in allele and haplotype distributions. This result could be due to a much lower proportion of F0 in the previous generation, which is a possibility considering that the new *Sargassum* events and associated larval transport through the NERR and between southern and northern stocks discussed above were only detected since 2011 and would not have affected (promoted) gene flow in the previous generation. An alternative hypothesis explaining the low abundance of F1 hybrids is that only a small proportion of F0 migrants successfully transmit their genes in the recipient population due to some form of selection

96

against migrant genotypes. Monitoring of the frequency of F1 hybrids in the next gray triggerfish generation would allow further evaluation of this second hypothesis.

3.4.3 Phylogeography and Demographic History

Phylogeography was evaluated based on both the mtDNA and the microsatellite datasets. Neighbor-joining clustering was conducted using the 12 geographic samples or the 4 homogeneous groups inferred during analysis of population structure. Clustering using F_{ST} and the $(\delta\mu)^2$ distance resulted in the grouping of all populations of the North Atlantic in the same branch with high degree of support although there was some support ($\geq 67\%$) for clustering of Canary Islands in a separate branch. Clustering of the 4 main groups using F_{ST} as a measure of genetic distance isolated the MED group and reflected the higher F_{ST} estimate between the MED group and the other 3 groups (average 0.074 while pairwise F_{ST} comparing NA, SWA and SEA were all less than 0.057). In contrast, clustering using the $(\delta\mu)^2$ distance placed MED in the same branch as the NA group and the two South Atlantic groups in another branch. The $(\delta\mu)^2$ distance accounts for the evolutionary distance between microsatellite alleles and puts a higher weight on the mutation process, while the F_{ST} distance is expected to reflect primarily the effect of genetic drift and migrations. The $(\delta\mu)^2$ distance yielded highest values for comparisons between the southern and North Atlantic groups. The higher value of the North-South distance during comparisons using the $(\delta\mu)^2$ measure thus indicates that mutations explained a significant fraction of the divergence between the two groups (Hardy et al. 2003). This observation suggests that the isolation between northern and southern groups is more ancient than the separation between NA and MED and SWA and SEA

respectively as it led to the accumulation of different mutations in the two latitudinal groups. This inference is consistent with the mtDNA statistical parsimony network that indicates that samples from southern and northern populations are essentially on separate clades. In contrast, while the MED sample differed in allele frequencies from NA stocks, the $(\delta\mu)^2$ distance was smaller indicating a highest affinity of MED with NA, its likely origin. Similarly, mtDNA haplotypes in MED were all clustered within the NA clades, a result consistent with the hypothesis of a recent isolation of MED and insufficient time for new mutations at mtDNA to accumulate. The low levels of genetic diversity in MED also suggest that this population experienced a recent genetic bottleneck. Considering the apparent lack of current migration between NA and MED, we hypothesize that a relatively recent colonization by a small number of founders resulted in the establishment of the MED population. Similarly, the lack of divergence between SWA and SEA at mtDNA suggests that the isolation between these two groups is relatively recent.

The long-term migration rates between stocks could not be estimated with confidence, but are expected to be limited considering the F_{ST} estimates and the effective population size estimates in each population, which were of several hundreds or thousands. The samples from FR or CA had estimates of N_e similar to those obtained for the U.S. populations (Appendix Table B.1). Gray triggerfish are not very frequent in western Europe, yet the high N_e estimates comparable to the large value obtained for the U.S. population suggest that migration rates are substantial between the two groups. Migrations would be expected to be asymmetric (unidirectional from U.S. to France or

Canary Islands) and may contribute a high fraction of the gray triggerfish present in those waters.

In summary, the gray triggerfish metapopulation in the Atlantic is structured, complex and features 4 main groups. The data are consistent with an ancient isolation between the North and South Atlantic populations and a more recent divergence of Mediterranean populations derived from the North Atlantic stock and a partial isolation of the southwestern and southeastern Atlantic. Analysis of contemporaneous migrations revealed high rates of migrants in current generations possibly reflecting the recent expansion of *Sargassum* in the Equatorial region. If this new *Sargassum* dynamic is maintained in the future, some changes in the population structure of gray triggerfish are expected to occur including an increased connectivity between the North and South populations and also between the East and West Atlantic. Further genetic monitoring of gray triggerfish populations under the influence of the Equatorial circulation is therefore warranted.

CHAPTER IV – CHARACTERIZATION OF AN ISOLATION-BY-DISTANCE

MODEL FOR THE GRAY TRIGGERFISH: APPLICATION TO THE ESTIMATION

OF DISPERSAL PARAMETERS

4.1 Introduction

Characterizing genetic and demographic connectivity among geographic

populations is essential to design effective conservation strategies (Lowe and Allendorf

2010). The marine environment is *a priori* open to migrations (Avise 1998) and many

marine species display a continuous geographic distribution across large portions of their

range, sometimes encompassing several thousand kilometers, leading to the assumption

that connectivity occurs across these large geographic areas, as promoted by unrestricted

dispersal. The spatial scale of the actual genetic connectivity is however influenced by

several factors including the dispersal capability of organisms, the density of populations

and the strength of local adaptation. The dispersal itself is determined by several factors

including the occurrence of barriers to gene flow resulting from discontinuities of suitable

habitat, the limited duration of the physical transport of eggs and larvae, the velocity of

currents utilized for species with pelagic planktonic phases, and the movement

capabilities and behaviors of adults. In reef fishes, movements of adults are often limited,

and when this is the case, the larval transport processes are assumed to be the major

determinants of dispersal (Leis and McCormick 2002; Jones et al. 2009; Shanks 2009).

Genetic connectivity can be maintained even when only a few effective migrants

are exchanged per generation (Waples 1998), which is often enough to rapidly achieve

the spread of advantageous mutations across large distances (Lowe and Allendorf 2010).

However, management is also concerned with local demographic change of populations, in particular, the relative role of local recruitment and migration in determining local demographic dynamics, or the potential for local replenishment through migration from external populations (Kritzer and Sale 2004). The spatial scale of this demographic connectivity is often different from that of the genetic connectivity and is also more difficult to determine because it requires estimating migration rates. Obtaining direct estimates of the spatial scale of demographic connectivity requires data on local recruitment as well as quantitative estimates of migrations from and to other demes (Lowe and Allendorf 2010). This information is particularly challenging to obtain when boundaries between demes are not clearly defined, as is the case in many marine species that are structured in large continuous metapopulations. In such cases, tag and recapture studies or studies of elemental signatures in otoliths can provide information on juvenile and adult movements, but these methods are not adapted to measure dispersal in most reef fishes that are sedentary as adults but broadcast planktonic eggs and larvae that cannot easily be tagged (Thorrold et al. 2002). Particle tracking may be used to predict larval envelopes (e.g. Roberts 1997; Cowen et al. 2006; Johnson et al. 2009), but this approach can also be challenging in species that cannot be modeled as a simple particle including, for example, those utilizing pelagic habitats that fluctuate over time in terms of size and shape, such as floating *Sargassum* beds.

Genetic methods provide an alternative approach to assess demographic connectivity. Genetic estimation of contemporaneous rates of gene flow through assignment tests has been used in several species (Lowe and Allendorf 2010), but this

method is only effective when migrants are exchanged between discrete and differentiated populations. This approach is irrelevant when there is isolation-by-distance in a continuous population. However, in that case, inferences on dispersal can be made using the isolation-by-distance theoretical framework (Rousset 1997; Puebla et al. 2009). Recent developments of individual-based models and maximum likelihood (ML) algorithms to estimate isolation-by-distance parameters (Rousset and Leblois 2007, 2012; Watts et al. 2007) have allowed assessing dispersal in a variety of demographic scenarios, including in metapopulations showing high degrees of genetic connectivity (and homogeneity) across large geographic areas (e.g. Puebla et al. 2012).

Studies in reef fishes to date have revealed relatively small (less than 100 km in most cases) larval dispersal envelopes (Roberts 1997; Cowen et al. 2006; Shanks et al. 2009; Puebla et al. 2012), but the species considered were characterized by short dispersal durations, usually less than a month. On the other hand, data on the spatial scale of demographic connectivity are lacking for species where larval dispersal lasts longer. In those species, rare successful long distance dispersal events could maintain genetic connectivity across long distances even if the majority of dispersal events are restricted to local areas; in that situation, the local spawning biomass would retain a strong influence on recruitment. Alternatively, longer larval transport could result in high proportion of dispersal events at long distances and a reduced importance of local spawning stocks for recruitment. Distinguishing between these two alternative demographic scenarios is essential in order to determine effective conservation and management strategies.

The gray triggerfish is a reef fish that shows highly sedentary behavior as adult. Dispersal is thought to occur primarily during the larval and juvenile stages (Wells and Rooker 2004; Franks et al. 2007) when the species is pelagic. This pelagic phase (4-7 months, Simmons 2008) lasts longer than in most other reef fishes, and, during that period, larvae and juveniles are found associated with floating seaweeds and flotsam (mostly *Sargassum sp.*) until they settle on hard benthic structures. Gray triggerfish have a center of abundance in the Southeast United States (Gulf of Mexico and Southeast U.S. coast) where they approach a continuous distribution along shelf habitats. The life history features of gray triggerfish predict structuring according to an isolation-by-distance model, as discussed in section 3.1.4 of this dissertation, where dispersal is limited by the spatial scale of the pelagic larval transport. The results of the population structure analysis across the species' range conducted in Chapter III revealed that long distance movements do occur, but it has also been hypothesized that larvae could be retained in local eddies and recruit close to their spawning location (NMFS 2006), and the relative importance of local or regional retention versus long distance dispersal is not known. The availability of a large continuously distributed populations in the southeastern U.S. provides the opportunity to describe the isolation-by-distance model and assess quantitatively the spatial scale of demographic connectivity resulting from larval dispersal in this species.

In this chapter, estimates of dispersal parameters in gray triggerfish were generated by exploiting the comprehensive sampling and survey of genetic variation in U.S. waters where the species shows a near continuous distribution.

103

4.2 Materials and Methods

4.2.1 Sampling

Samples of gray triggerfish were obtained during the summer and fall of 2008, 2009, and 2010 from South Texas to South Carolina in conjunction with the groundfish SEAMAP surveys conducted by the National Marine Fisheries Service (NOAA-Fisheries) and by the Florida Fish and Wildlife Research Institute, the MARMAP survey implemented by the South Carolina Department of Natural Resources, and additional sampling by local fishermen.

The SEAMAP groundfish surveys operate from Pensacola to the U.S./Mexico border and employ a stratified randomized design to sample benthic shelf habitats up to ~100 m depth by trawling (Nichols 2004). Sampling in the northern Gulf (1,400 km of coastline, Figure 3.3) yielded 430 specimens and resulted in minimal gaps in this section of the studied range except for the shelf nearing the Mississippi estuary delta and a small portion of the Texas shelf North of Corpus Christi.

Two hundred and thirty-five additional samples from Southwest Florida, Southeast Florida, and South Carolina waters were also analyzed. Specifications on sample size per location and detailed collection methods are presented in section 3.2.1.

Specimens were preserved frozen on board (SEAMAP samples) or kept on ice until fish were landed. Muscle tissue and fin clips were collected and stored in 95% alcohol, Dimethyl Sulfoxide (DMSO) salt-saturated storage buffer (0.25 M EDTA, 20% DMSO, 30% H_2O, and NaCl), or in a Sarkosyl urea lysis buffer (1% *N*-lauroylsarcosinate, 20 mM NaPO4, 8 M urea, 1 mM EDTA) prior to DNA extraction.

4.2.2 Laboratory Assays

DNA extraction was performed following a phenol-chloroform protocol (Sambrook et al. 1989). The fish were genotyped at 17 microsatellite markers, assayed in four multiplexes panels, as described in Chapters II and III.

4.2.3 Data Analysis

Samples were initially grouped in six regional populations based on gaps in sampling (Figure 3.3). The occurrence of scoring errors due to null alleles, stuttering bands, and large allele dropout in each regional population sample was tested in MICROCHECKER v.2.2.3 (Van Oosterhout et al. 2004). The conformance of genotype proportions to Hardy-Weinberg (H-W) equilibrium expectations was tested using exact tests in GENEPOP v.4.2 (Raymond and Rousset 1995; Rousset 2008a). Probability-value estimates were based on 10,000 dememorizations, 500 batches, and 5,000 iterations per batch. Departure from H-W equilibrium (F_{IS}) measured as Weir and Cockerham's (1984) f, the number of alleles, allelic richness, and gene diversity were calculated for each regional sample in FSTAT v.2.9.3 (Goudet 1995). Allelic richness is a measure of the number of alleles independent of sample sizes that allows comparing groups with different sample sizes (El Mousadik and Petit 1996). Gene diversity was calculated as described in Nei (1987).

4.2.3.1 Analysis of Spatial Genetic Variation

Homogeneity in allelic richness and gene diversity among samples was tested using the Friedman ranks test, as implemented in SPSS v.20 (IBM Corp., Armonk, NY, USA). The degree of population differentiation (F_{ST}) among regions was estimated as

Weir and Cockerham (1984) θ as calculated in FSTAT and homogeneity of allele

distributions among regional samples was tested using exact tests in GENEPOP. Pairwise

comparisons were performed by computing estimates of pairwise θ between individual

regions and performing associated pairwise exact homogeneity tests. Markov Chain

parameters during exact homogeneity tests were the same as above (Exact tests of H-W

equilibrium). The False Discovery Rate (FDR, Benjamini and Hochberg 1995) procedure

was used to determine the significance threshold for P-values when multiple independent

tests were conducted simultaneously.

Isolation-by-distance due to limited dispersal potential and barriers to gene flow

(genetic discontinuities) may both account for divergence among geographic samples.

Spatial genetic variation within the region was therefore further explored using the

Bayesian clustering approach implemented in the software TESS v.2.3.1 (Chen et al.

2007; Durand et al. 2009a). TESS aims to detect genetic discontinuities within

continuously distributed populations of a species based on the distribution of multilocus

genotypes. This approach accounts for the decay of spatial autocorrelation that occurs

due to isolation by distance and is therefore well suited for populations displaying

spatially restricted dispersal and a predicted isolation-by-distance pattern. One hundred

runs were performed using a conditional autoregression (CAR) admixture model,

allowing for correlated allele frequencies among populations. Each Monte Carlo

simulation included 250,000 sweeps with the first 50,000 sweeps discarded as burn-in.

The 20 runs showing the lowest Deviance Information Criteria (Spiegelhalter *et al.* 2002)

were retained to make inferences, as recommended by Durand et al. (2009b).

Structuring according to an isolation-by-distance mechanism was examined

within ranges where no evidence of genetic discontinuity was found. Both spatial

autocorrelation analysis (Smouse and Peakall 1999) and the method developed by

Rousset (2000) and Leblois et al. (2004) revealed weak structuring consistent with

isolation-by-distance. Only results of the second approach are presented here as this

method also allows estimating dispersal parameters based on existing theory of isolation-

by-distance (Rousset 1997).

The genetic distance between pairs of individuals was estimated as the ê statistics

(Watts et al. 2007) computed in the software GENEPOP. The ê statistics is more powerful

in cases where the spatial pattern of population structure is weak (Watts et al. 2007), as is

the case in the present study (see section 3.3.1 and section 4.3 of this chapter).

Considering the shelf habitat used by gray triggerfish, two approaches were used to

compute geographic distances between individuals and isolation-by-distance statistics

from individual capture location coordinates. In a first approach, a one-dimensional axis

(along the coastline and approximated using mid-shelf transects) was used, thus assuming

dispersal in a one-dimensional linear habitat. In a second approach, a two-dimensional

habitat consisting of a 10 km-wide strip surrounding the transect line was considered. The

analysis of isolation-by-distance focused on data obtained on specimens (n = 430)

collected between South Texas and West Florida because this portion of our sampling

design approached best a continuous sampling along the coastline as recommended to

infer parameters of the model (Leblois et al. 2004). Because estimation of the parameters

of the isolation-by-distance model is biased when the geographic distance between

samples being compared is greater than $0.56\sigma/\sqrt{2\mu}$, where σ is the standard deviation of

parental position relative to offspring position and μ is the mutation rate (Rousset 1997),

a bootstrap resampling approach was used to investigate the effect of the spatial scale of

sampling on estimates of the slope of the isolation-by-distance relationship and σ.

Subsamples were drawn by resampling sets of 100 individuals located within subsections

of the lattice of various lengths using the software POPTOOLS v.3.2.5 (Hood 2010) and the

slope of the linear regression between genetic and geographic distance (b) was estimated

for each resampled dataset.

This slope was then used to calculate σ, given the effective population density

(D), using the relationship (Rousset 1997):

$$\sigma = \sqrt{\frac{1}{4Db}} \qquad\qquad \text{equation 1}$$

Inferences on σ thus require information on population density. Two approaches

were taken to obtain values for D and discuss values of σ and the distribution of dispersal

distances. An upper bound for D is given by the census population density (D_c). The

census density of gray triggerfish in U.S. waters was estimated based on average landing

data in the Gulf of Mexico during the sampling period obtained from the recreational

fisheries statistics database of the Fisheries Statistics Division of the National Marine

Fisheries Service (personal communication, database accessed 08 January 2016) and

accounting for estimates of fishing mortality rates ranging between 0.435 and 0.53

(NMFS 2011). The census number for the Gulf of Mexico was divided by the length of

the corresponding portion of linear coastline to derive an estimate of census density.

Effective density (D_e) was also directly estimated using genetic data. Considering the observed homogeneity in allele frequencies across the sampling surface (see "results"), an estimate of the effective size for the overall metapopulation was generated using the ML coalescent approach in the software MIGRAINE v.0.4.1 (Rousset and Leblois 2007, 2012; Leblois et al. 2014). The OnePopVarSize demographic model allowing accounting for historical change in population size was used in the estimation. Runs employed the Importance Sampling (IS) algorithm statistic and assumed a Generalized Stepwise Mutation (GSM) model. Test runs were performed initially in order to find the proper range of surveyed parameters as recommended by R. Leblois (personal communication). The final run included 5,000 sampled points and 100,000 runs per point. The parameter N that represents an estimate of the current effective population size was calculated assuming an average mutation rate across microsatellites of 5×10^{-4} (Leblois et al. 2014). N was also calculated considering mutation rates of 10^{-3} and 10^{-4} in order to evaluate the sensitivity of parameter estimates to the mutation rate. The obtained estimates of N were applied to the entire section of coastline surveyed (3,100 km) to derive estimates of effective density in the one-dimensional model and to the entire two-dimensional shelf area (123,331 km^2) for the two-dimensional model. Other parameters estimated in MIGRAINE included the ratio between the current and ancestral effective population sizes (N_{ratio}), the time in generations since the initiation of the population expansion/contraction phase scaled by the effective population size ($D_g/2N$), and the distribution parameter of the generalized stepwise mutation model ($pGSM$).

Estimates of contemporaneous N_e by the linkage disequilibrium method were also generated aggregating samples across different distance classes as recommended by Neel et al. (2013). All the obtained estimates were infinite or very large (greater than 1,000). These values are considered incompatible with reliable estimation using the linkage disequilibrium method (Waples and Do 2010) and estimates by this method are therefore not discussed further.

Because the genetic consequences of dispersal depend on the shape of the distribution of dispersal distance (Rousset 2008b), a simulation approach after Puebla et al. (2012) was taken to determine the parameters of dispersal distance distributions yielding isolation-by-distance slopes consistent with that estimated from the empirical dataset. Coalescent simulations were implemented in the software IBDSIM v.2.0 (Leblois et al. 2009) considering various distribution functions (Geometric, Pareto, and Sichel). Simulations employed a one-dimensional lattice of 10,000 km with absorbing boundaries; 1,400 nodes were sampled on the lattice, corresponding to the studied portion of the northern Gulf of Mexico (South Texas to West Florida, approximately 1,400 km of coastline) where available samples best reflected the continuous distribution of gray triggerfish. Simulated datasets included 17 unlinked loci following a GSM mutation model with a mean mutation rate of 5×10^{-4} and a geometric variance of multi-step mutations with parameter estimated during ML coalescent analysis of the dataset in MIGRAINE above. The simulated datasets were processed for isolation-by-distance analysis as described above. Parameters for each of the dispersal distribution functions were adjusted during initial exploratory simulations to determine ranges of parameter

values leading to isolation-by-distance slopes *b* similar to those obtained with the

empirical dataset. Series of simulations were then conducted in triplicates within this

range to identify the parameter values (or combination of parameter values) that led to

isolation-by-distance slopes closest to the estimates from the empirical dataset. The

influence of the mutation rate on the dispersal distribution parameters was evaluated by

simulating dispersal distributions using the parameter combination optimized above for

each distribution function but varying the mutation rate. Mutation rates of 10^{-3} and 10^{-4}

were considered as an upper and lower bound of the mutation rate for the panel of

microsatellites used in the study, respectively.

Finally, ML estimates of σ were generated using both the linearIBD and the

planarIBD demographic models implemented in MIGRAINE. These methods provide an

estimate of the neighborhood size parameter (*Nb*) from which an estimate of σ can be

derived. The planarIBD model accounts for a two-dimensional habitat while the

linearIBD model assumes dispersal along a one dimension (linear) lattice. Computations

for the linearIBD and planarIBD models were performed accounting for the one-

dimensional lattice and two-dimensional habitat described above for estimation of the $\hat{e}$

statistics. Estimates of the neighborhood size were generated during three replicate runs

employing the Product of Approximate Conditional (PAC) likelihoods algorithm and

based on 2,000 points with 100 runs per point.

Estimates of the parameter σ were derived from *Nb* using the relationships

$Nb = 2D\sigma^2$ (equation 2) and $Nb = 2D\pi\sigma^2$ (equation 3) for the linear and the two

dimensional model respectively.

111

Values of *D* applied for the one and two-dimensional models were the same as those used above to derive estimates from the slope of the isolation by distance regression.

4.3 Results

Four out of 102 tests of Hardy-Weinberg equilibrium were significant before FDR correction for multiple tests performed simultaneously. None of the tests remained significant after correction. MICROCHECKER analyses indicated possible occurrence of null alleles at locus BC14 in the ETX-LA region, locus BC17 in the SEF region, and stuttering and/or null alleles at locus BC3 in the SWF region. Because the scoring artifacts at these three loci were found in one region (out of 6) only and did not lead to significant departure from Hardy-Weinberg expectation, all 17 markers were kept for further analysis.

Summary statistics per locus and per region including number of alleles, allelic richness, gene diversity, inbreeding coefficient, and probability of significance of tests of Hardy-Weinberg equilibrium are presented in Appendix Table C.1. The number of alleles (*A*) per locus averaged 25.6 and ranged between 9 (locus BC16) and 45 (locus BC46). Gene diversity ranged between 0.27 (locus BC16 in the SEF region) and 0.969 (locus BC46 in the SWF region). Allelic richness and gene diversity did not differ significantly among localities ($P = 0.240$ and $P = 0.083$ respectively).

The estimate of θ was very low (0.0004, 95% bootstrapping Confidence Interval CI: 0-0.001) and the probably that θ was different from zero from exact homogeneity tests was 0.031. Homogeneity tests at individual loci did not reveal significant

heterogeneity in allele frequencies among regions except for one locus (BC46) that

showed significant heterogeneity ($P = 0.042$) before FDR correction, but not after

correction. All pairwise θ values between individual regions were very low (average

0.0006, range -0.0006-0.0018, Table 3.3) and only two pairwise exact homogeneity tests

(across loci) were significant after FDR correction for multiple tests performed

simultaneously (SWF versus ETX-LA comparison: $P = 0.0177$, estimate of $\theta = 0.0008$;

SWF versus SEF comparison: $P = 0.0032$, estimate of $\theta = 0.0018$). Bayesian clustering

runs in TESS all converged towards a single population unit with isolation-by-distance

but with no genetic discontinuity within the sampled range. Further analysis of isolation-

by-distance proceeded under this assumption.

The ML coalescent estimates of demographic parameters for the metapopulation

generated in MIGRAINE are reported in Table 4.1. The estimate of the current effective

size (N) derived assuming an average mutation rate of 5×10^{-4} was 29,940 (95% CI:

18,570-62,630). The ML estimate of the ratio between current and ancestral population

size was an 8.9 fold increase of effective population size. Applying the ML value of N to

the estimates of $Dg/2N$, the population expansion would have occurred during the past

1,377 generations (95% CI: 305-21,557). The genetic estimate of D_e was generated by

applying the estimate of N to the entire one-dimensional lattice (from South Texas to

South Carolina) yielding a value of 9.66 (95% CI: 5.99-20.20) for D_e for the linear

model. N was then applied to the entire approximated shelf area (123,331 km^2) for the

two-dimensional model yielding a value of 0.24 (95% CI: 0.15-0.51) for D_e. The census

density D_c for the section of the Gulf of Mexico used to generate estimates of the

113

isolation-by-distance slope was 175 (linear model) or 2.89 (two-dimensional model) giving a ratio of effective to census density of 0.055 (95% CI: 0.034-0.115) for the linear model and 0.083 (95% CI: 0.052-0.176) for the two-dimensional model.

Table 4.1

Maximum Likelihood Coalescent Estimates of Gray Triggerfish Demographic Parameters

Parameter	*MLE*	*95%-*	*95%+*
$N*$	29,940	18,570	62,630
N_{ratio}	8.9	3.7	2.1×10^5
$Dg/2N$	2.3×10^{-2}	5.1×10^{-3}	3.6×10^{-1}
pGSM	0.65	0.59	0.70

Maximum likelihood coalescent estimates (*MLE*) and 95% Confidence Intervals for the parameters N, N_{ratio}, $D_g/2N$, and *pGSM* based on 17 microsatellite loci * Estimates generated assuming an average mutation rate of 5×10^{-4}.

Estimates of the slope of the isolation-by-distance model for the one-dimensional model using subsets of the data encompassing increasing distance ranges revealed a high variance among slopes when resampled datasets were generated using genotypes found within short distance ranges (< 1,100 km, Figure 4.1). The mean and standard error of slopes from resampled datasets stabilized between 3.4×10^{-8} and 4.4×10^{-7} when the sampled range was between 1,400 and 1,700 km (Figure 4.1). Accordingly, final estimates were generated based on all available data for the area between South Texas and West Florida where the high density of sampling locations with minimal gaps best reflected the near-continuous distribution of gray triggerfish along the continental shelf. The obtained estimate of the slope was 3.1×10^{-8} (lower and upper bounds -5.24×10^{-7} and 4.61×10^{-7}). Point estimates generated using greater portions of the dataset (i.e. including

localities in South Florida and South Carolina) were all included within the bounds of the confidence interval described above.
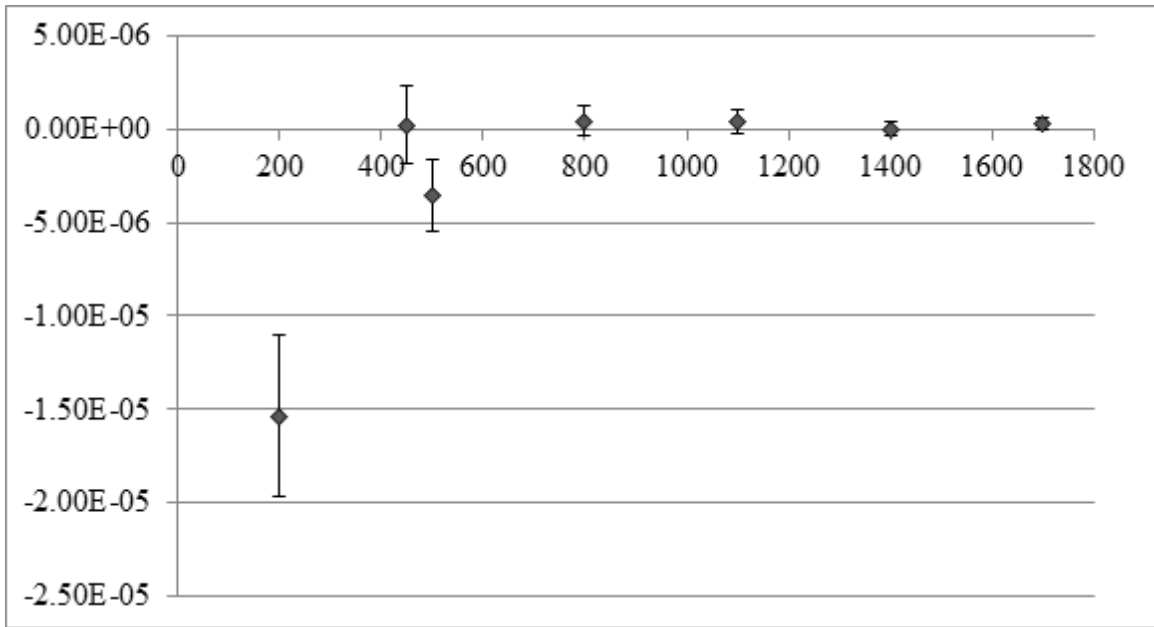
Figure 4.1 Slope of the IBD Model at Various Spatial Scales of Observation

Average (± standard error) of the point estimate of the slope (y-axis) of the isolation-by-distance model in replicate subsamples of the dataset (n = 100) at increasing spatial scale of observation (x-axis).

Considering the genetic estimate of effective density, the corresponding values of σ derived using equation 1 was 914 (95% CI: 237 - +∞).

Estimates of σ derived from Nb values obtained from the ML approach in MIGRAINE using equations 2 and 3 were 780 (95% CI: 255-2,517) for the linear model and 740 (95% CI: NA-7,330) for the two-dimensional model. Because dispersal along a coastline one-dimensional axis can be approximated more easily and the ML estimates of dispersal using the one dimensional and two-dimensional dispersal models were similar, further analysis of dispersal distributions via simulations focused on the one-dimensional model.

115

Because the genetic effects of dispersal depend largely on the shape of the distribution of dispersal distances, simulations were conducted to generate dispersal distributions, based on various families of distribution functions yielding isolation-by-distance slopes similar to that obtained with the empirical dataset. The obtained dispersal distribution functions, their parameters, and the slope of the associated isolation-by-distance relationships are presented in Table 4.2. All compatible distributions involved mean dispersal distances greater than 123 km and at least 10% of dispersal events at distances greater than 286 km. The average 90% percentile across simulated distributions was 1,809 km (Table 4.3) indicating that 10% of dispersal events were predicted at distances greater than 1,809 km from origin.

Table 4.2

Parameters of Simulated Distributions Yielding Isolation-by-Distance Slopes Comparable to That of the Empirical Dataset

Model	μ_x	σ (sim./est.)	IBD Slope	95%-	95%+	*P*-value
$D_e = 10$/km						
Pareto (M=0.995 ; n=1.16)	482	1,509/293	3.02E-07	1.25E-07	4.99E-07	0.0000-0.0002
Pareto (M=0.97 ; n=0.92)	1323	2,666/938	2.94E-08	-9.93E-08	1.88E-07	0.0248-0.0670
Geometric (m=0.98 ; g=0.993)	140	203/242	4.41E-07	2.50E-07	6.48E-07	0.0000
Geometric (m=0.95 ; g=0.999)	950	1,400/1,031	2.43E-08	-1.02E-07	1.74E-07	0.0000
Sichel (γ=-0.0005 ; ξ=15000 ; Ω=0.002)	420	1,037/318	2.56E-07	4.89E-08	6.64E-07	0.0000-0.0004
Sichel (γ=-0.002 ; ξ=15000 ; Ω=0.001)	505	1,263/806	3.98E-08	-8.92E-08	2.05E-07	0.0113-0.0794
$D_c = 175$/km						
Pareto (M = 0.95; n = 0.98)	1047	2,300/219	2.99E-08	-8.65E-08	1.51E-07	0.0143-0.1863
Geometric (m=0.98 ; g=0.992)	123	175/213	3.14E-08	-6.28E-08	1.45E-07	0.0085-0.0522
Sichel (γ=-0.001 ; ξ=10000 ; Ω=0.004)	278	660/212	3.19E-08	-9.57E-08	1.69E-07	0.0011-0.1525

Parameters of simulated distributions yielding isolation-by-distance slopes comparable to that of the empirical dataset (point estimate and upper bound). D_e/D_c: effective/census population density; μ_x: mean (simulated) dispersal distance; σ: standard deviation of parental position relative to offspring position; sim.: simulated; est.: estimated.

Table 4.3

Percentile Distribution of the Simulated Functions Compatible with the Isolation-by-Distance Slope Estimated During the Study

Distribution parameter	Percentile			
	25	50	75	90
Pareto (M=0.995 ; n=1.16)	3	12	137	1,247
Pareto (M=0.97 ; n=0.92)	12	166	1,519	4,866
Geometric (m=0.98 ; g=0.993)	40	97	196	326
Geometric (m=0.95 ; g=0.999)	238	643	1,336	2,251
Sichel (γ=-0.0005 ; ξ=15,000 ; Ω=0.002)	14	67	356	1,205
Sichel (γ=-0.002 ; ξ=15,000 ; Ω=0.001)	12	65	403	1,485
Pareto (M = 0.95; n = 0.98)	6	75	912	3,889
Geometric (m=0.98 ; g=0.992)	35	85	172	286
Sichel (γ=-0.001 ; ξ=10,000 ; Ω=0.004)	13	54	241	726

Estimates of σ and simulated distributions for the various functions were generated accounting for a mutation rates of 10^{-4} and 10^{-3} respectively to evaluate the sensitivity of dispersal distribution characteristics to this parameter (Table 4.4). The ML estimates of the standard deviation of the parent-offspring dispersal distance remained large for both mutation rates with a point estimate at 349 km (95% CI: 114-1,126) when a low average mutation rate (10^{-4}) was considered, while the estimate using a high mutation rate scenario (average 10^{-3}) yielded substantially larger values for sigma (point estimate 1,103, 95% CI: 361-3,559).

Simulated dispersal distributions compatible with the observed IBD slope and accounting for the two mutation rates in IBDSim all yielded an estimate of σ greater than 123 km ($\mu = 10^{-4}$) or 141 km ($\mu = 10^{-3}$) when the census density was used in calculations, or 231 ($\mu = 10^{-4}$) and 259 ($\mu = 10^{-3}$) when the estimate of effective density was used.

118

Table 4.4

Effect of the Mutation Rate on Estimates of the Parameter Sigma

Model	$\mu = 1.E-04$		$\mu = 5.E-04$		$\mu = 1.E-03$	
	σ	IBD Slope	σ	IBD Slope	σ	IBD Slope
$D_e = 10$/km						
Pareto (M=0.995 ; n=1.16)	∞	-2.56E-08	293	3.02E-07	345	4.35E-07
Pareto (M=0.97 ; n=0.92)	428	2.83E-08	938	2.94E-08	946	5.78E-08
Geometric (m=0.98 ; g=0.993)	231	9.70E-08	242	4.41E-07	259	7.74E-07
Geometric (m=0.95 ; g=0.999)	343	4.39E-08	1,031	2.43E-08	∞	-1.09E-08
Sichel (γ=-0.0005 ; ξ=15,000 ; Ω=0.002)	466	2.38E-08	318	2.56E-07	451	2.54E-07
Sichel (γ=-0.002 ; ξ=15,000 ; Ω=0.001)	2,738	6.91E-10	806	3.98E-08	928	6.01E-08
$D_c = 175$/km						
Pareto (M = 0.95; n = 0.98)	∞	-3.71E-08	219	2.99E-08	∞	-3.57E-08
Geometric (m=0.98 ; g=0.992)	148	6.54E-08	213	3.14E-08	141	7.21E-08
Sichel (γ=-0.001 ; ξ=10,000 ; Ω=0.004)	123	9.39E-08	212	3.19E-08	∞	-3.19E-08

Comparison of simulated distributions obtained using three different mutation rates. D: population density; σ: standard deviation of parent-offspring dispersal distance.

4.4 Discussion

Allele frequencies at the 17 microsatellites were homogeneous across the sampled area as indicated by the very low estimate of F_{ST} (0.0004) and the lack of significance of exact homogeneity tests at all individual loci. Pairwise F_{ST} estimates did not exceed 0.0018 and the two significant pairwise comparisons during pairwise exact tests involved samples collected along Southwest Florida coast compared with the Southeast Florida and East Texas/Louisiana samples respectively. These three geographic samples did not differ significantly in allele frequencies from any other regional samples in the remaining portions of the range in the Gulf of Mexico or along the east coast, leading to the interpretation that the marginal difference between these localities did not correspond to true barriers to gene flow. In addition, Bayesian clustering using a spatially explicit approach in TESS converged toward a single unit with isolation-by-distance and no discontinuity. Altogether, these results are consistent with the inference of a lack of detectable barrier to gene flow within the area and the occurrence of genetic connectivity among geographic populations across the sampled portion of the species range.

Genetic discontinuities within the sampled area have been evidenced in a variety of other marine and coastal species, in particular between the Gulf of Mexico and the U.S. East coast (Avise 1992), or between populations East and West of Mobile Bay (Karlsson et al. 2009; Portnoy and Gold 2012). These reported genetic breaks involved species occupying coastal or estuarine habitats, or species using offshore habitats but displaying characteristics prone to maintaining geographic structure such as limited dispersal abilities. In contrast, species occupying outer shelf habitats similar to those used

120

by the gray triggerfish and dispersing pelagic larvae did not display clear genetic discontinuities across the same geographic area (e.g. red porgy, *Pagrus pagrus*, Ball et al. 2007, or the red snapper, *Lutjanus campechanus*, Saillant et al. 2010; Hollenbeck et al. 2015), suggesting that no major barrier to the dispersal of larvae is present.

As pointed out by Lowe and Allendorf (2010), weak or lack of divergence among geographic populations is not synonym for panmixia across the region but may instead indicate that gene flow is sufficient to maintain a high level of genetic connectivity. Considering the large distance over which allele frequencies appear homogenous (see Chapter III) and the highly sedentary behavior of gray triggerfish adults (Ingram 2001), the hypothesis of panmixia appears unrealistic. Instead, a metapopulation model as defined by Kritzer and Sale (2004) where geographic populations are connected by gene flow can serve as a framework to assess gray triggerfish in the region. In reef species showing sedentary adult behavior and pelagic larvae such as the gray triggerfish, gene flow occurs primarily through dispersal of pelagic eggs and larvae (Shanks et al. 2009). This dispersal phase is finite in duration and thus, structuring is expected to follow an isolation-by-distance pattern where genetic relatedness decreases as a function of geographic distance (Puebla et al. 2009). Under isolation-by-distance, both population density and dispersal distance contribute to determine the spatial scale of differentiation (Rousset 1997). Apparent genetic homogeneity may, therefore, occur across large distances even when dispersal and demographic connectivity does not (e.g. Puebla et al. 2012).

The spatial scale of demographic connectivity in gray triggerfish was explored through estimating the parameters of the isolation-by-distance model. Both the moment estimator of Watts et al. (2007) and the ML estimate in MIGRAINE (Rousset and Leblois 2007, 2012) yielded large estimates of neighborhood sizes with estimates of the parameter σ approaching 800 km. This result held when the two-dimensional dispersal model was applied. Dispersal in two dimensions in this study was considered by approximating the shelf area in the Gulf of Mexico as a 10 km-wide strip framing the mid-shelf transect line used in the one-dimensional model and allowing dispersal in two dimensions within the delimited geometric area. This approximation is underestimating the shelf area in some parts of the Gulf (in particular the western Gulf) and also does not account for dispersal across the Gulf. Considering dispersal across sections of the open Gulf of Mexico (e.g. from South Texas to West Florida) is challenging because gray triggerfish juveniles cannot settle in the middle of the Gulf thus violating assumptions of the dispersal model. Such dispersal events, if they occur, would lead to a moderate overestimation of dispersal distances when they occur considering the shape of the continental shelf in the northern Gulf. On another hand, the underestimation of the shelf habitat due to a broader shelf in the western Gulf and the non-inclusion of the open Gulf as a potential habitat for dispersal leads to a potential substantial overestimation of population effective density during the estimation which would result in an underestimation of σ. While further developments of isolation-by-distance models to allow accounting for the specific characteristics of habitats and dispersal in gray triggerfish in the Gulf would be needed, the inference of large neighborhood sizes and

122

long distance dispersal in this study seems largely supported by the two models. Also, considering the limitations of the two-dimensional model discussed above, inferences focused on the one-dimensional model. While this model is an approximation, it seems to more realistically reflect the dispersal of gray triggerfish among geographic regions than applicable two-dimensional models and it also provides interpretations that can be more easily be incorporated into spatial management of populations in the Gulf of Mexico. Simulated distributions of dispersal distances using different families of functions and different mutation rates corresponded to average dispersal distances between 123 and 1,323 km. Moreover, examination of the simulated distributions of dispersal distances indicated that 10% of dispersal events resulted in migrations across very long distances from origin (on average greater than 1,809 km). Interestingly, the high frequency of long distance dispersal events was observed in all simulations, including those where the census population density (which can be considered as an upper bound of effective density) was used in simulations, which indicates that the inference that demographic connectivity occurs across long distances is not affected by uncertainties on the value of effective population density. A fraction of immigrants of 10% is usually considered as a threshold below which connected populations transition from demographic dependence to independence (Hastings 1993; Waples and Gagiotti 2006). While gene flow cannot be easily quantified in terms of a percentage of immigrants in the case of isolation-by-distance, the long distances traveled by a substantial fraction of gray triggerfish before recruiting to benthic habitats and subsequently to breeding populations is consistent with a large degree of demographic dependency of local recruitment from non-local spawning

123

stocks, including those located several hundreds of km from a given recipient benthic habitat. This result contrasts with finding in studies of the demographic connectivity of various reef fishes (e.g. Roberts 1997; Cowen et al. 2006; Puebla et al. 2012) that concluded that dispersal of ecological significance was occurring within short distances (less than 100 km in most cases). The species considered in these studies dispersed larvae over a period limited to a few weeks and usually less than 40 days. Gray triggerfish, similar to most reef fishes, are highly sedentary as adults as demonstrated by tagging experiments (Ingram 2001), but larvae and juveniles remain in the *Sargassum* habitat for 4 to 7 months (Simmons 2008). Although local spawners could contribute to recruitment in the same region if larvae are caught in local eddies (NMFS 2006), the present results indicate that such local retention, if it occurs, is limited and local recruitment is dependent for a large part on the output of spawning populations located at long distances from recipient habitats. An important consequence for management of gray triggerfish populations and fisheries targeting them is that recruitment cannot be predicted from local spawning biomass since it depends for a large part on non-local spawning populations. Instead, recruitment indices may need to be based on the abundance of newly settled juveniles in order to maintain healthy local populations.

Inferences based on the isolation-by-distance relationship imply that dispersal was symmetrical along a one-dimensional axis. Information on the movement and dynamics of *Sargassum* patches used by gray triggerfish larvae and juveniles is still limited. The peak of the gray triggerfish spawning season occurs in June and July (Simmons and Szedlmayer 2011). During these months *Sargassum* is found in abundance in the Gulf of

Mexico and tends to move off the Florida coast and along the Gulf Stream in September (Gower and King 2008). This could favor asymmetric dispersal rates from the Gulf to the Atlantic, a hypothesis that cannot be formally tested within the framework of currently available methods to analyze isolation-by-distance. Improved data on the accumulation and movement of *Sargassum* would be helpful in order to develop more accurate dispersal models for gray triggerfish in the region. Another underlying assumption made during inferences on connectivity based on population genetics models is that the population has reached an equilibrium situation. Simulations conducted by Hardy and Vekemans (1999) showed that, in the range of scenarios considered in that study, equilibrium was reached within a few generations when high migration rates were considered, suggesting that with the high rates of migrations inferred in this study for gray triggerfish, a quasi-equilibrium situation would be reached rapidly. Repeated temporal sampling would, however, be useful in order to refine demographic parameter estimates and test the temporal stability of patterns of population structure described in this study.

The analysis conducted in this work also implicitly neglected the effects of immigration from geographic populations in other portions of the species' range. Gray triggerfish are reported in Central and South America, in Europe and the Mediterranean Sea, and in western Africa (Robins and Ray 1986; Sazonov and Galaktionova 1987). The main circulation patterns in the Atlantic Ocean during the summer period when immigrant larvae could potentially be transported to U.S. populations from other regions of the Atlantic are summarized on the Cooperative Institute for Marine and Atmospheric

Sciences (CIMAS) Ocean reference website (available from

http://oceancurrents.rsmas.miami.edu/atlantic/atlantic.html) and presented in section

3.1.4. Examination of these data indicates that the main source of immigrants from the

East Atlantic would be the North Equatorial Current. This hypothesis was confirmed by

analysis of contemporaneous gene flow as described in Chapter III which revealed that

1.6% of individuals samples in U.S. waters were F0 migrants from the Southeast Atlantic

and 1.4% were F1 hybrids between the two groups. Removing genotypes identified as

possible migrant or hybrid from the dataset resulted in similar dispersal distribution

parameters (σ estimate: 1,014, 95% CI: 229-+∞) suggesting that the effect of

immigration was minimal on estimates generated during this study.

In addition to the potential influence of migrants from Southeast Atlantic, the

Caribbean current that originates from the North Brazil and Guyana currents and flows to

the West, entering the Gulf of Mexico as the Loop current, could transport immigrants

from North Brazil, Venezuela, Nicaragua-Honduras, Belize, and/or southern Yucatan to

habitats of the Gulf of Mexico or the U.S. East coast; the Antilles current could also

transport migrants through the Caribbean to the U.S. East coast. While gray triggerfish

are described in the Caribbean, Central America, and along the North coasts of South

America, attempts to obtain specimens for genetic analysis in this study were

unsuccessful due the rarity of gray triggerfish in fisheries catch recently reported in the

different Caribbean and South American countries, suggesting that the populations are

small and the demographic impacts of immigrants from these regions would be

tentatively limited at best. This assumption would, however, deserve further examination

by characterizing gray triggerfish populations in the Caribbean and Central American regions if samples can be obtained.

Gray triggerfish are also present in the southern Gulf of Mexico (e.g. the Bay of Campeche) although samples from that region could not be obtained in this study. While the abundance of gray triggerfish in that area and their connectivity with U.S. populations could not be established in this study, populations from the southern Gulf would be expected to be connected to the studied populations and follow the isolation-by-distance pattern described in this study with the additional implication that the axis used to characterize dispersal would be longer by approximately 43% to incorporate the corresponding section of shelf habitat. Data on the abundance of gray triggerfish in Mexican waters of the Gulf of Mexico would, therefore, be useful for incorporation in future assessments of the gray triggerfish regional metapopulation.

The ratio of effective to census population density was approximately 5.5×10^{-2} for the one-dimensional model and 8.3×10^{-2} for the two-dimensional model. These values are intermediate between the extremely low ratios (10^{-3} to 10^{-5}) reported in studies of other marine fishes (Hauser et al. 2002; Turner et al. 2002; Saillant and Gold 2006) and the range (> 0.1) expected in most situations based on demographic models (Nunney and Elam 1994). Estimating effective population size/density is particularly challenging in marine species structured in large connected populations, as is the case for gray triggerfish (Hare et al. 2011). Methods based on coalescent simulations such as the model used in the present study tend to estimate the size of the overall metapopulation that includes all demes connected to one another by migrations as long as migration is not too

127

low (Hare et al. 2011). These methods also integrate the various historical events experienced by the metapopulation over time meaning that it is difficult to determine an appropriate census number that can be matched with the obtained estimates of N_e. The model used in the present study accounted for the historical population growth rate of gray triggerfish and thus the estimate of N generated is expected to reflect current/recent N_e, after the detected recent change in population size event. The method employs a generalized stepwise mutation model that has been shown to be robust, in particular in order to avoid false signals of population reduction in size (Leblois et al. 2014). However, very recent changes in population size may not be reflected in the coalescent estimate and the ratio D_e/D_c may be biased if the estimates of census and effective size respectively correspond to different time periods. The spawning biomass of gray triggerfish in the studied region is estimated to have declined by 43-58% in the past 3 generations (Liu et al. 2015), a reduction in population size that was not detected in the analysis of effective population size. The species was not targeted by direct fisheries until the early 1990s and changes in population size prior to that period might have been minor in comparison to recent changes due to fisheries harvests, suggesting tentatively that the estimated ratio may only be moderately biased. Methods to estimate contemporaneous effective size such as the linkage disequilibrium (Waples 2006; Waples and Do 2010) would have been preferable to match directly census and effective numbers for the same cohorts (Hare et al. 2011), but these methods are very imprecise when N_e is greater than 1,000. When there is isolation-by-distance, estimates of N_e by the linkage disequilibrium based on samples collected within a breeding window tend to reflect the neighborhood

size (Neel et al. 2013). The ineffectiveness of the linkage disequilibrium method in the present case, with infinite or very large estimates, is thus consistent with the very large neighborhood size inferred during isolation-by-distance analysis (lower bound of ML estimates 1.2×10^6). Alternative approaches to evaluate the current effective population density are based on life history data (Nunney and Elam 1994). The acquisition of these parameters is, unfortunately, challenging for marine species such as the gray triggerfish where little information is available on early mortality rates and quantitative data on reproductive behaviors and fecundity are still unreliable. The census density estimate was derived based on catch data available from the NOAA Office of Science and Technology database for the period that matched genetic sampling and approximates the density of adults present on benthic habitats. This value was uncorrected for potential factors likely to lower N_e, such as biased sex ratio and variance in reproductive success, and accordingly, can be considered an upper bound for population density.

In conclusion, estimates of dispersal parameters among geographic populations of gray triggerfish obtained from genetic data suggested the presence of large neighborhoods and dispersal events involving long sections of the shelf habitat used by the species. In contrast to what was proposed for other reef fishes with pelagic larvae characterized by shorter pelagic durations, local recruitment seems to depend substantially on non-local spawning stocks, including some located hundreds or even thousands of kilometers away. This suggests that management procedures should be reevaluated to consider the reduced role of local spawning biomass in determining recruitment in this species and allocating fisheries harvests using alternative metrics such

as recruitment indices. This finding also highlights the need to consider management over very broad scales encompassing multiple countries to ensure long-term sustainability of this species.

CHAPTER V POPULATION STRUCTURE AND CONNECTIVITY OF QUEEN

TRIGGERFISH IN THE ANTILLES AND SOUTHEASTERN U.S.

5.1 Introduction

This chapter focuses on a second Balistid with high potential for larval dispersal,

the queen triggerfish (*Balistes vetula*). The life history of the queen triggerfish is similar

to that of the gray triggerfish studied in Chapter III and IV in that it features a highly

sedentary adult behavior and a long pelagic larval phase. As in the gray triggerfish, these

features lead to the prediction that connectivity among geographic populations results for

a large part from the process of passive dispersal of larvae and juveniles such that

patterns of gene flow are determined by the combination of the duration of the pelagic

phase, the direction of surface currents, and the velocity of these currents. In this chapter,

genome-wide genetic variation among queen triggerfish populations was assessed using a

high-density genome scan which allowed analyzing neutral processes related to gene

flow, dispersal, and genetic drift, but also non-neutral variation and the potential role of

natural selection and local adaptation in shaping the genetic structure of this species.

5.1.1 Distribution and Life History of Queen Triggerfish

The queen triggerfish is a member of the Balistidae family found on tropical and

subtropical reef habitats of the Atlantic basin. Reports of the species in the eastern

Atlantic are sparse (K. Michalsen, Institute of Marine Research, personal

communication) and mostly consist of country records that are not accompanied by

voucher specimens or detailed information on criteria used to confirm species

identification diagnoses. In the western Atlantic, queen triggerfish are most abundant in

the Caribbean Sea, off the Southeast coast of Florida, and off the central coast of Brazil. In the latter region, triggerfish species observed a major decline between 2001 and 2006 (IBAMA 2006) and the current abundance of queen triggerfish is very low (C. Albuquerque, Universidade Federal do Espírito Santo, Brazil Personal communication). The species is also very infrequent in U.S. and Mexican waters of the Gulf of Mexico (National Marine Fisheries Service, Fisheries Statistics division, A. Aguilar-Perera, Universidad Autónoma de Yucatán personal communication) (Figure 5.1) and in Central America (Belize: M. Gongora, Belize Fisheries Department, Personal Communication; Colombia: species listed as endangered, Mejía and Acero 2002; Panama: C. Vergara, Universidad Tecnológica de Panamá Personal communication; Venezuela: F. Arocha, Universidad de Oriente Personal communication). Along the East coast of North America, although the species has been reported as far North as Canada (Scott and Scott 1988) or Massachusetts (Robins and Ray 1986), catches are anecdotal North of Florida (National Marine Fisheries Service, Fisheries Statistics Division, Figure 5.2). Accordingly, the geographic area where queen triggerfish can be found in abundance compatible with sampling for population genetic studies appears to be restricted to the Antilles and the southern tip of Florida.
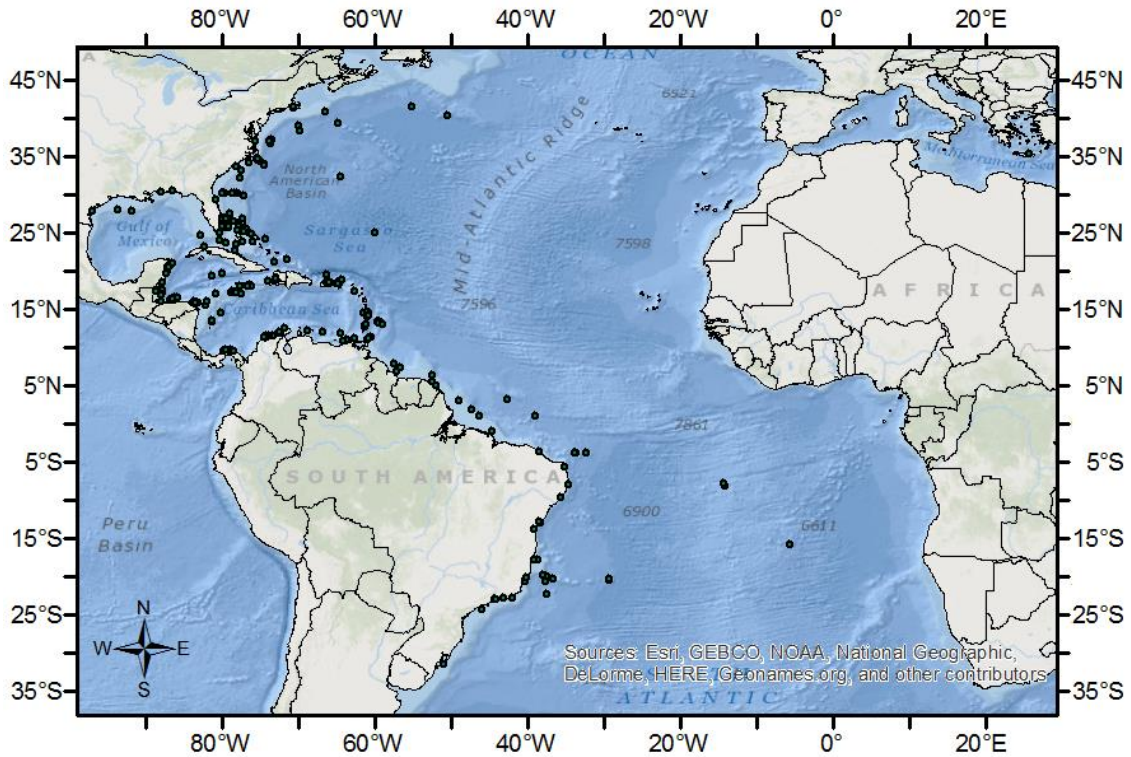
Figure 5.1 Queen Triggerfish Distribution Map

Queen triggerfish (*Balistes vetula*) distribution range based on reviewed point observations. www.aquamaps.org, version of Aug. 2010. Updated 02/21/13.

Queen triggerfish maximum reported length and weight are 57 cm FL (Randall 1968) and 5.4 kg (IGFA 1991) respectively. The species has a medium longevity with maximum reported ages of 7 years old (Manooch and Drennon 1987) and 14 years old (Albuquerque et al. 2011) in the Caribbean and Brazil respectively. According to Aiken (1983) males and females reach maturity at a mean size of 26.5 cm and 23.5 cm respectively. The spawning season of queen triggerfish is protracted as commonly observed in tropical fishes. Available data in the Caribbean indicate occurrence of fish in spawning condition throughout most of the year although peaks of spawning activity are suggested to occur in winter (January-February) and in the late summer/early fall

133

(August-October) based on the presence of high proportions of spawning-capable fish in samples during these periods (Aiken 1983). Queen triggerfish have been reported to form small spawning aggregations during the periods of peak spawning in the Netherland Antilles, Puerto Rico, and St. Croix (Heyman et al. 2014), and more recently near St. Thomas (R. Nemeth, University of the Virgin Islands Personal communication). Little is known yet regarding these aggregations although the sedentary behavior of queen triggerfish adults and the relatively small number of individuals observed in these aggregations (R. Nemeth, University of the Virgin Islands, personal communication) suggest that long-distance migrations are not involved.

Tagging studies indicate limited movement of queen triggerfish once they settle on reefs although the size distributions observed in catches at different depths suggest that larger fish tend to move to deeper waters (Aiken 1983). Larvae and juveniles are planktonic and can remain in the water column for 63 to 83 days (Robertson 1988 corrected by Lindeman et al. 2000). The distribution of early larvae is not documented but Munro et al. (1973) reports that juveniles are found in mid-water and settle on reefs when they reach a size of 4.8-6.9 cm (Robertson 1988). Thus, the extended duration of the pelagic transport experienced by queen triggerfish larvae could promote connectivity and gene flow across large geographic distances, despite of the sedentary behavior of adults.

5.1.2 Fisheries Exploitation and Conservation Status

Queen triggerfish is one of the predominant species in current landings of commercial and recreational fisheries in the U.S. Caribbean. Catches in that region have

increased substantially in the last few decades due to the decline of other reef fishes

(SEDAR-30). The queen triggerfish is also an important game fish in South Florida and

is harvested for the ornamental fish market throughout its range. Available data suggest

that the species has experienced a pronounced decline in abundance since the 1980s in

continental U.S. waters (Figure 5.2; Personal communication of the National Marine

Fisheries Service at the Fisheries Statistics Division. Database accessed 23 September

2016): As illustrated by Figure 5.2, estimates of yearly recreational catches for the 1981-

1990 period averaged 44,915 fish harvested, but only 8,478 from 1991 to 2015. Signs of

decline of queen triggerfish in other parts of its range led to its listing in 1996 as

"Vulnerable" (Roberts 1996) on the IUCN Red List of Endangered Species. More

recently the species status was updated to the category "Near Threatened" on the basis of

new estimates of populations decline over the distribution range which were lower than

the previous ones and did not meet the 30% reduction in abundance threshold used to list

species in the Vulnerable category (Liu et al. 2016).
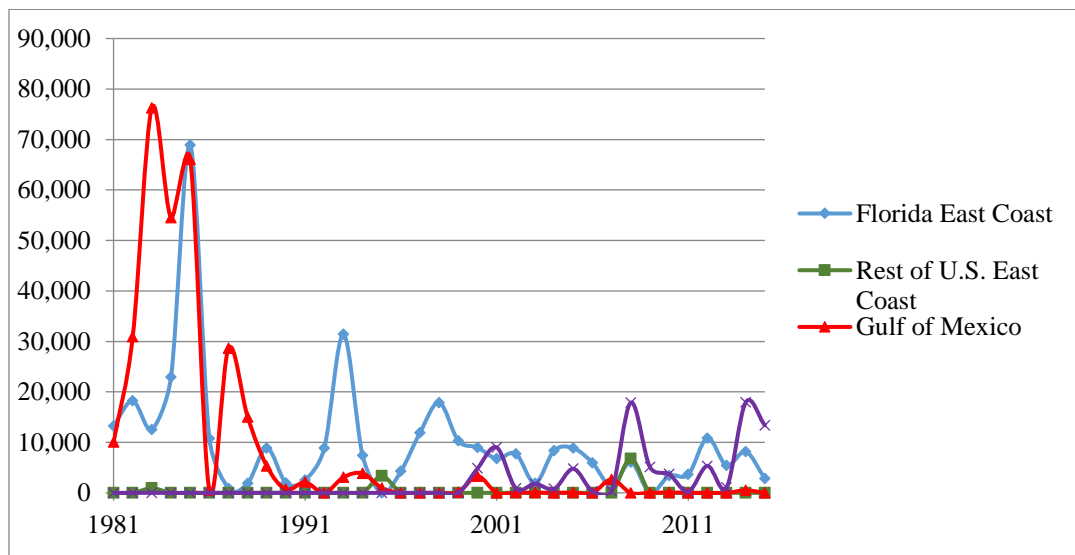
Figure 5.2 Queen Triggerfish Landings in U.S. Waters

Number of queen triggerfish landed per year from 1981 to 2015 in U.S. waters.

Despite of these signs of decline, management plans for queen triggerfish are still undeveloped. In the U.S., the species is currently managed within the filefish and triggerfish Fishery Management Unit under the Reef Fish Fishery Management Plan with no stock subdivision. The absence of specific management for queen triggerfish is largely due to the lack of essential data on population structure and stock abundance and biomass which prevents reliable assessment of regional stocks (SEDAR-30 - U.S. Caribbean Queen Triggerfish Stock Assessment Report 2013). Catches in the U.S. Caribbean showed an increasing trend in the numbers of larger individuals caught in recent years, particularly along the East coast of Puerto Rico but these trends may also reflect increased fishing effort. Indicators of the abundance, including indirect measures such as effective population size inferred from genetic data, would be valuable for evaluating the status of stocks along with improved data collection. Understanding stock structure, in particular within the U.S., and the connectivity of U.S. populations with other geographic stocks is also needed to predict recruitment and design management plans.

5.1.3 Major Oceanic Circulation Patterns in the Western Atlantic Ocean and Working Hypotheses on Connectivity

Because of the sedentary behavior of adults discussed above, population connectivity is hypothesized to depend primarily on the passive dispersal of pelagic larvae and juveniles. Since these phases have limited swimming capacity, their transport is expected to follow oceanic surface currents. In order to formulate hypotheses on the direction and distance of migrations, surface velocity data were gathered from the

136

Cooperative Institute for Marine and Atmospheric Studies hosted by the University of Miami RSMAS (http://oceancurrents.rsmas.miami.edu/index.html). The current patterns and velocity during and following the peak spawning season for queen triggerfish (February to May) are shown in Figures 5.3 and 5.4.

The patterns observed during the spring peak of spawning shown on these two figures are similar to those reported for the second period of high spawning activity (October, data not shown), so that hypotheses on the geographic scope of connectivity formulated below are expected to apply to the two peaks of spawning activity of queen triggerfish.
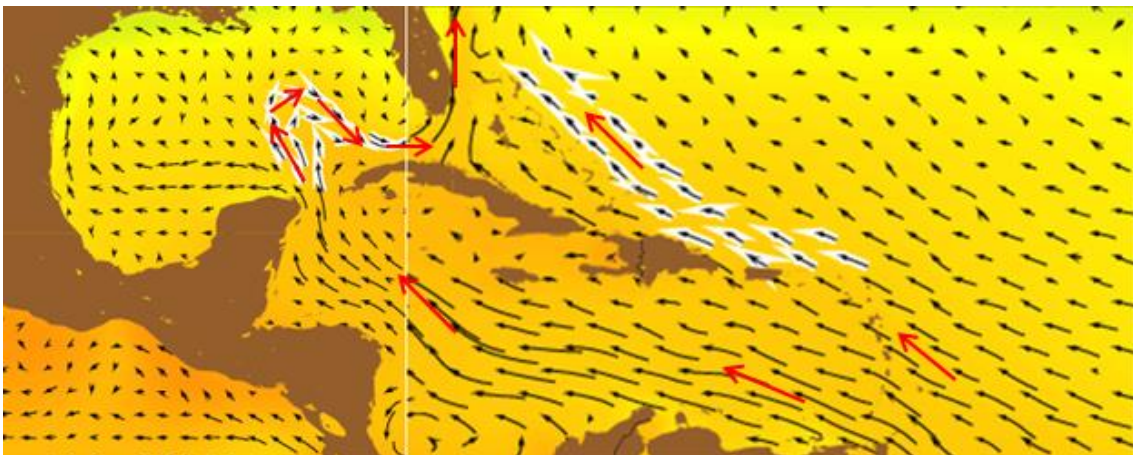
Figure 5.3 Prevailing Surface Currents in the Caribbean and the Gulf of Mexico during the Queen Triggerfish Spawning Season

Diagram depicting the main surface currents in the Caribbean Sea and Gulf of Mexico regions

(http://oceancurrents.rsmas.miami.edu/atlantic/atlantic.html).
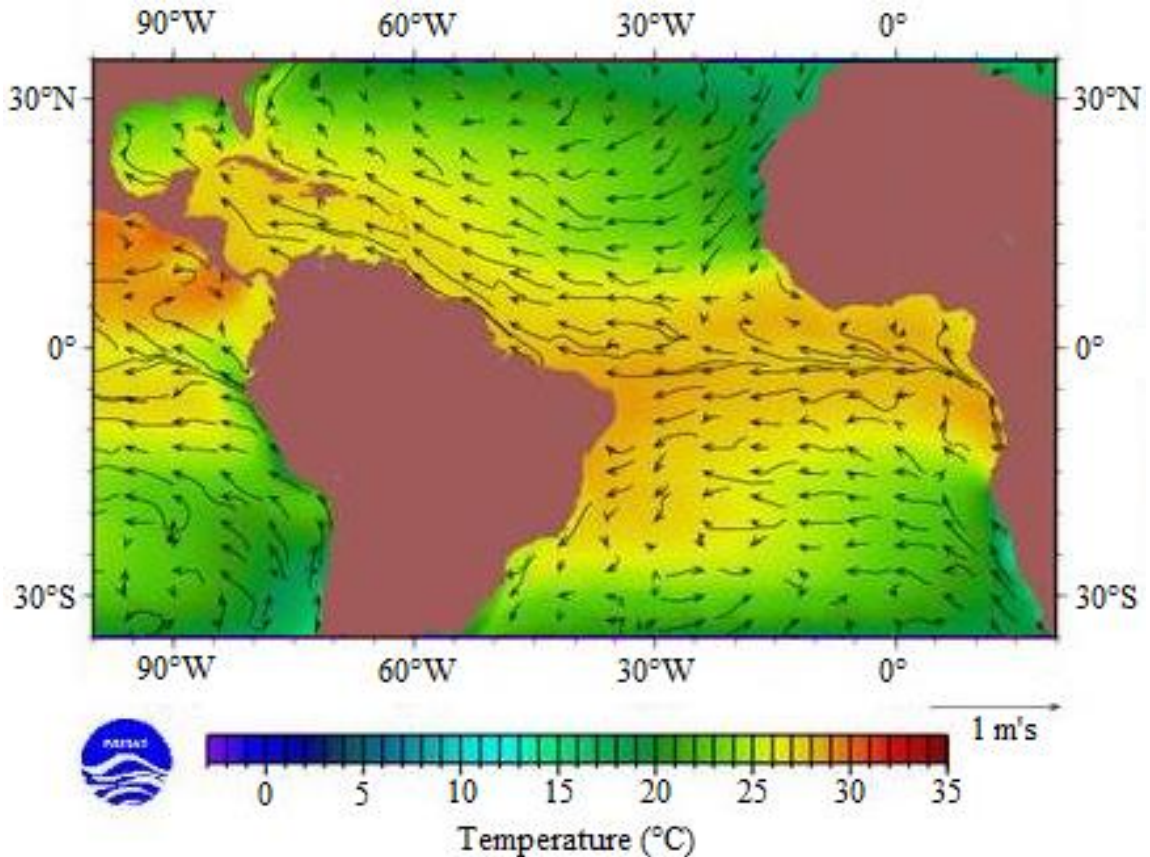
Figure 5.4 Prevailing Surface Currents During Queen Triggerfish Spawning Season Across the Species Distribution Range

Diagram depicting the main surface currents across queen triggerfish distribution range.

(http://oceancurrents.rsmas.miami.edu/mgsva/slideshow.html?imgnum=4&basin=atlantic).

Examination of the current patterns in Figure 5.3 and 5.4 leads to the following hypotheses predicting connectivity of geographic populations of queen triggerfish:

a. The Antilles current flowing from the Lesser Antilles to the Bahamas is hypothesized to connect island populations located in its path. Dispersal of larvae through pelagic transport along this current is predicted to lead to a unidirectional (Southeast to Northwest) pattern of migration and isolation-by-distance along this axis. This current is very slow (4-9 cm/s) and

138

discontinuous (Lee et al. 1996), which is expected to result in moderate dispersal distances (average dispersal would range from 218 km to 645 km considering a 63-83 days duration of the pelagic phase respectively) and increased demographic independence and divergence among geographically distant populations.

b. The Caribbean current and the Loop current are hypothesized to connect the northern countries of South America (e.g. Venezuela and Colombia) to Belize and South Florida although the Panama-Colombian Gyre may contribute to isolate populations in that region from those in the rest of the Caribbean (Figure 5.3) as suggested for other reef species (Cowen et al. 2006). As in the case of the Antilles current, a unidirectional (South to North) pattern along the current axis is expected.

c. The Bahamas and Florida peninsula are separated by unsuitable adult habitat (water depth reaches 6,000 ft) unlikely to be crossed by adult queen triggerfish. The fast-flowing Florida current runs through this trench, in an orientation parallel to the coastline and would therefore not promote exchange of larvae between the two stocks (Figure 5.3), leading to a predicted barrier to migration between the Bahamas and the Florida East coast.

d. Queen triggerfish, as gray triggerfish studied in Chapter III, occur along the central coast of Brazil. As discussed in Chapter III, populations of queen triggerfish in Brazilian coastal waters are expected to be isolated from Caribbean stocks due to the divergence of surface currents along the

northeastern region of Brazil (exposed to the North Brazil current flowing

Northwest) and those along the central coast of Brazil (exposed to the Brazil

current flowing South) (Figure 5.4).

e. The exchange of migrants between the Southeast and Southwest Atlantic via

transport by the South Equatorial current is hypothesized to be negligible:

The minimum time required to cross the Atlantic along this current, as

estimated from drifters' data in Chapter III, is significantly longer than the

pelagic phase duration (3.9 months vs 2.8 months). In addition, as described

in section 5.1.1, the presence of queen triggerfish in those areas is uncertain

and the abundance is very low such that even if residual gene flow was

occurring, the small populations from West Africa, would only be able to

provide very small numbers of effective migrants to western populations.

Additional hypotheses on fine scale stock structure among regional populations of

queen triggerfish in the U.S. Caribbean and South Florida can also be formulated. The

South Atlantic region has been treated as an independent stock in assessments of

Lutjanids with pelagic larvae (e.g. the yellowtail snapper, SEDAR-8 2005) considering

the large geographic distance between South Florida and the U.S. Caribbean which was

assumed to be incompatible with direct larval transport between the two regions.

In the U.S. Caribbean, two stocks have also been distinguished in assessment

efforts, one corresponding to the northern Puerto Rican Platform that includes the islands

of Puerto Rico, St. Thomas, St. John, and the British Virgin Islands and a second stock

consisting of the southern St. Croix Platform that includes the island of St. Croix. These

two geologic platforms are separated by a ~4,000 meters deep trench (CFMC 2004) that likely prevents adult queen triggerfish migration. Connectivity between the two platforms could, however, be maintained owing to larval pelagic transport as discussed above. Roberts (1997) examined surface current patterns in the Caribbean and inferred potential dispersal routes and envelopes for pelagic larvae. Considering a two-month period of larval drift (a lower bound for queen triggerfish larvae as discussed above), areas off of Puerto Rico and the U.S. Virgin Islands could be connected via direct larval transport. There are however a number of caveats to these predictions. Indeed, Shulman and Bermingham (1995) in a review of Caribbean reef fishes showed that the length of the larval planktonic phase was not necessarily an effective predictor of gene flow. More recently, Cowen et al. (2000, 2006) using advanced biophysical models showed that effective larval dispersal in a variety of reef fishes was restricted to distances in the order of 10-100 km. In addition, the direction of dominant currents may lead to unidirectional gene flow (e.g. East to West in the U.S. Caribbean) and promote local larval retention on the leeward side of islands while windward sides would be more prone to receiving migrants (Swearer et al. 1999).

Evidence of partially restricted gene flow between populations in the region was indeed recently documented in genetic studies conducted in three reef-associated Lutjanids (Carson et al. 2011; Gold et al. 2011; Saillant et al. 2012). These studies were consistent with demographic independence of the Caribbean and South Atlantic populations and also with a possible demographic independence between St. Croix and

141

the northern platform and between the eastern (East Coast of Puerto Rico, St. Thomas) and western sides of the northern platform (West coast of Puerto Rico).

Because of the unavailability of samples of queen triggerfish in the lower Caribbean and South America, due to low and declining abundance of the species as discussed in section 5.1.1, this project focused on the axis Antilles – South Florida, where multiple locations could be sampled. The study examined the hypothesis that populations are structured according to an isolation-by-distance model, and also evaluated the potential occurrence of finer patterns of population structure involving demographic independence of populations occupying the two Puerto Rican platforms or partial isolation of populations located on the leeward side of the U.S. Caribbean through local retention of larvae.

Studies of genetic variation and stock structure during the past few decades have largely focused on DNA markers (Sunnuck 2000) and have been dominated by microsatellite loci. Although datasets generated with microsatellites generally reach higher power of inference than those resulting from older marker systems thanks to the high polymorphism of loci (see Chapter II of this dissertation), their ability to detect population structure is still limited for species with large population size and high gene flow where levels of divergence among populations at neutral loci are expected to be very low (Waples 1998; Allendorf et al. 2008). In addition, microsatellites are generally assumed to be neutral to natural selection and therefore are not suitable to detect local adaptation of regional populations, a most important source of divergence to consider when designing conservation units. Loci undergoing divergent selection show greater

142

divergence and were instrumental at resolving stock structure in marine species such as cod (Pampoulie et al. 2006) or salmon (Vasemägi et al. 2005). Loci undergoing divergent selection are a priori unknown when performing population genetic studies and therefore assessment of variation under selection needs to rely on high-density genomic scans including large number of genetic markers; loci under divergent selection are detected during an outlier analysis (Whitlock and Lotterhos 2015). Such high-density genome scans were cost prohibitive and incompatible with large-scale population surveys until recently. However, with the recent development of next-generation sequencing techniques and genotyping by sequencing protocols (e.g. Baird et al. 2008; Peterson et al. 2012), such genome scans can now be performed with sample sizes compatible with population genetic studies. This study, therefore, employed high-density genome scans based on Single Nucleotide Polymorphism markers generated using the double digest RAD sequencing technology to study both neutral and non-neutral genetic variation among geographic populations of the queen triggerfish.

5.2 Material and Methods

5.2.1 Sampling

Sampling aimed to assess fine-scale structure in the U.S. Caribbean (Puerto Rico and U.S. Virgin Islands) where the species is particularly abundant and can easily be sampled. Samples from four locations were taken in this area in order to assess genetic variation in the region: East and West coasts of Puerto Rico and St. Thomas Island on the northern Puerto Rican platform and St. Croix Island on the southern platform. Sampling also aimed to describe genetic variation along the Antilles axis (Lesser Antilles to Greater

143

Antilles) and evaluate a possible isolation-by-distance pattern. Samples from the Lesser

Antilles were pursued and obtained from La Martinique Island. Samples from further

South (Trinidad, Venezuela, Panama, and Colombia) were pursued without success due

to the rarity of queen triggerfish in these waters. Samples were pursued from Bahamas

without success but were obtained from the Southeast of Florida (Jupiter) and included in

the study.

A total of 470 specimens were collected between 2012 and 2015 (Figure 5.5).

Sampling focused on the winter peak of the spawning season to describe the breeding

structure. Individuals from Jupiter, Florida (JP, n = 79) were collected by Capt. W.

Taylor via hook and line. Specimens from East Puerto Rico, near Fajardo (FA, n = 79),

were obtained with the help of Capt. M. Hanke and caught via hook and line and using

traps. Samples from West Puerto Rico, off Mayaguez (MAY, n = 75), were caught via

spearfishing by Capt. Benigno Rodriguez in collaboration with the Fisheries Research

Laboratory of the Department of Natural Resources. Individuals from Saint Thomas (ST,

n = 87) were collected by Capt. Daryl Bryan. Saint Croix (SCR, n = 79) specimens were

collected by trapping in collaboration with Capt. Thomas Daley. Finally, individuals from

La Martinique Island (MA, n = 66) were provided by the institute IFREMER. Fin clips

were collected from fresh fish at the dock or directly on the boat and preserved in a

DMSO salt-saturated fixative (0.5 M EDTA, 20% Dimethyl sulfoxide, NaCl, ddH$_2$O).
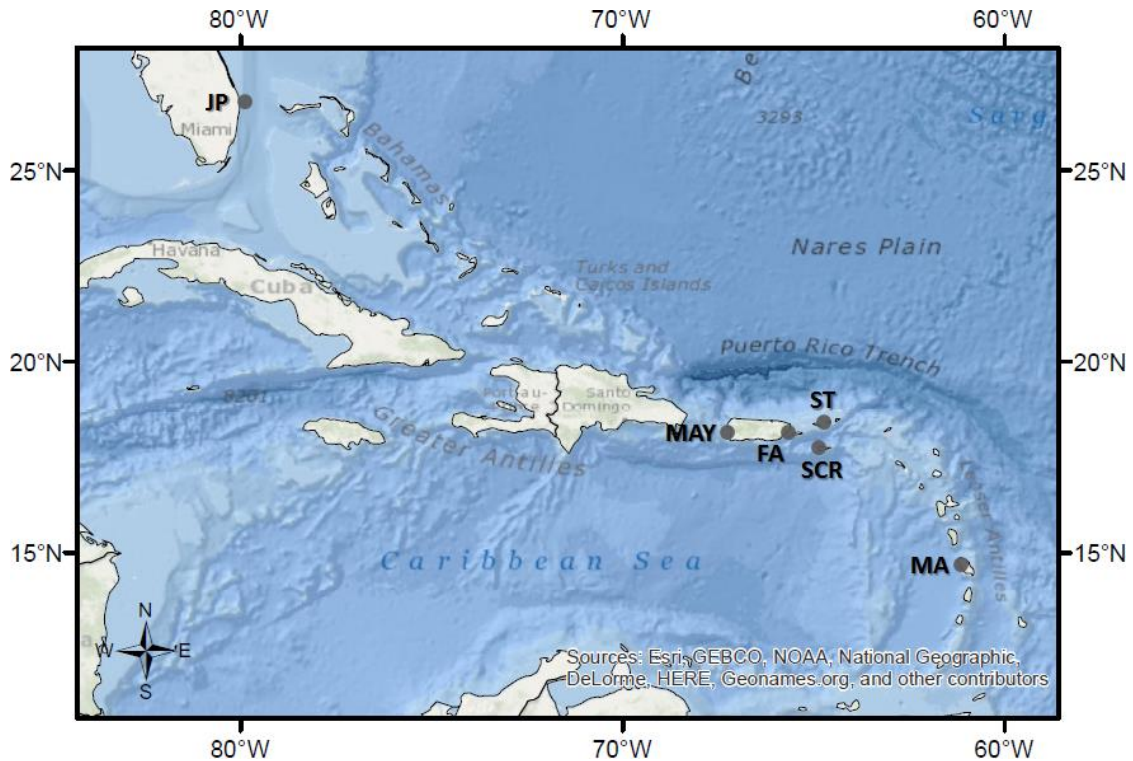
Figure 5.5 Queen Triggerfish Sampling Localities

Sampling localities for the study of population structure in queen triggerfish. JP: Jupiter; MAY: Mayaguez, PR; FA: Fajardo, PR; SCR: Saint Croix; ST: Saint Thomas; MA: La Martinique.

5.2.2 Library Preparation and Sequencing

Genomic DNA was extracted using the phenol-chloroform method (Sambrook et al. 1989). DNA concentration for each individual sample was determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and the quality of the DNA was also evaluated during agarose electrophoresis. DNA solutions were concentrated or diluted in order to achieve a final concentration of 40-100 ng/µL. Double-digest RAD-Tag (ddRAD) sequencing libraries were prepared from each individual DNA using protocols modified from Baird et al. (2008) and Peterson et al. (2012). Each library was made from 0.52-1.3 µg of genomic DNA that was simultaneously digested with the restriction enzymes Sau3AI (3.75 units, New England

Biolabs Inc. [NEB], Ipswitch, MA, USA) and SPEI (7.5 units, NEB) at 37°C for 1.5 h.

Enzymes were then heat-inactivated at 65°C for 15 min and the digested fragments were

ligated at 16°C for 30 min to sample-specific Illumina adapters using 400 units of T4

Ligase (NEB). In order to identify sequencing reads originating from a specific individual

sample, a unique 6 bp barcode was included in each adapter pair. Adapters also included

a degenerate 8 bp unique molecular identifier (UMI) which was used to identify and

remove PCR duplicates during processing of sequencing reads as described by Schweyen

et al. (2014). Ligation reactions were terminated by diluting 5 folds with ddH2O. The

products of the ligation reaction were then purified and size-selected to retain fragments

larger than 250 bp using Agencourt AMPure XP magnetic beads (Beckman Coulter Inc.,

Brea, CA, USA). Beads were applied at a 0.65x ratio and unbound DNA was removed

during two washes with 70% ethanol. Cleaned DNA products (2 µL in volume) were then

amplified through PCR in 25 µL reactions consisting of 5 pmol of forward and reverse

primers and 12.5 µL of 2X Master Mix, containing Taq polymerase (50 units/mL),

dNTPs (0.4 mM), MgCl2 (3 mM), and 2X buffer, (NEB). The thermal cycle consisted of

an initial denaturation step of 3 min at 95°C, 30 cycles of 95°C for 30 s, 55°C for 30 s,

and 72°C for 30 s, and a final extension at 72°C for 7 min. PCR products from individual

libraries were pooled for multiplex sequencing (75-100 individuals were pooled in each

Illumina Hi-seq sequencing lane). The pool of libraries was cleaned with Agencourt

AMPure XP beads as described above and libraries were size-selected on a PippinPrep

(2% agarose gel cartridge, Sage Science, Beverly, MA, USA) to isolate fragments

ranging in size between 300 and 500 bp. The products of the elution step were purified

using a 1.8x ratio of Agencourt AMPure XP beads and submitted to the University of

Colorado, Denver School of Medicine (Aurora, CO, USA) for 2x125 paired-end

sequencing on the Illumina HiSeq2000 platform.

5.2.3 Bioinformatics Analysis of Sequencing Data and SNP Discovery

An initial quality check was performed on the Illumina raw reads to reduce the

probability of base-calling errors: Paired-end reads were trimmed by removing portions

of the sequence with Phred-score below 30. The resulting sequences were discarded from

the dataset if either the forward or the reverse read was shorter than 75 bp. The dataset

was then screened for PCR duplicates. Duplicates were inferred when multiple reads

showed identical sequences and shared the same UMI. Because PCR duplicates lead to an

increase in homozygosity and an inflated confidence in genotype calls (Andrews et al.

2014), only one read from each set of detected duplicates was kept for further analysis.

Filtered data were processed using the dDocent pipeline (Puritz et al. 2014) using

default parameters. Because there is no reference genome for queen triggerfish, a *de novo*

assembly was performed to generate a reference directly from the dataset. During this

process, unique reads with low frequency of occurrence in the dataset were not

considered as these reads are generally the result of sequencing errors or are only shared

by a very small number of individuals which makes them uninformative for SNP

genotyping. The threshold frequency of occurrence used to decide if reads were retained

in the reference assembly or discarded was determined as described below. Reads were

assembled using the program RAINBOW (Chong et al. 2012) to form reference clustered

sequences for each locus. These reference sequences obtained were further compared and

clusters were merged based on their overall sequence similarity, represented as the percentage of sequence identity, using the software CD-HIT (Li and Godzik 2006; Fu et al. 2012) in order to establish a reference genome consisting of putatively unique contigs. In order to find optimal values for the threshold frequency of unique reads and the minimum degree of sequence similarity to use for merging, test assemblies were performed allowing threshold minimum read frequencies of 10, 12, and 15, and similarity scores ranging from 80 to 99%.

The software FREEBAYES (Garrison and Marth 2012) was used to locate and genotype Single Nucleotide Polymorphisms (SNPs) and other variants such as insertions/deletions (INDELSs), multi-nucleotide polymorphisms (MNPs), and complex events (composite insertions and substitutions). Complex variants (INDELSs, MNPs and complex events) were excluded and only SNPs with depth (the number of reads found for each SNP locus) between 10 and 75 and a quality score greater than 30 were included in the genotype of an individual for further analyses. Genotypes with likelihood lower than 0.99 were coded as missing data. Individuals showing a high percentage of missing genotypes at the filtered loci within each population were excluded from the dataset. An inherent characteristic of SNP datasets generated from ddRAD sequencing reads is a high percentage of missing data in all samples. In order to evaluate the impact of filtering individuals with high percentages of missing data on the structure of the final dataset, datasets were generated using various threshold values (25, 35, 40, or 50% missing data) and compared. The maximum percentage of missing data within an individual retained in the final dataset was chosen accounting for the number of loci retained and the retained

sample size (number of individuals per sampling locality). SNPs, where genotype

frequencies departed significantly from Hardy-Weinberg equilibrium expectations during

exact tests performed in VCFtools (Danacek et al. 2010), were discarded. Tests were

performed within all 6 sampling localities and SNPs departing significantly from

expectations in at least one population after FDR correction were removed from the

dataset. A variant call file (VCF) including all the SNPs shared by at least 75% of the

individuals in the whole dataset was created and used for further analysis.

5.2.4 Data Analysis

The structure of the final dataset was first examined using a Principal Component

Analysis (PCA). Principal components were calculated using the software PLINK v.1.9

(Purcell et al. 2007, available at http://pngu.mgh.harvard.edu/purcell/plink/) and

individual genotypes labeled by sampling site were plotted on principal components in

Excel.

Data analysis then proceeded with a search for loci showing significant deviation

from neutral patterns of genetic variation among localities in an outlier analysis

performed with the software BAYESCAN v.2.1 (Foll and Gaggiotti 2008). In BAYESCAN,

locus-specific F_{ST} coefficients are decomposed into a population-specific component (β)

shared by all loci, and a locus-specific component (α). Departure from neutrality of a

locus is inferred when α is significantly different from 0. A positive value of α indicates

diversifying selection, while negative values are expected to reflect balancing or

purifying selection. For each locus, a reversible-jump Monte Carlo Markov Chain

(MCMC) evaluates the posterior probability of models with and without selection (α

149

component being either present or absent, respectively) to determine the significance of α. The VCF file generated after the last step of the filtration process was converted to the BAYESCAN input file format using the software PGDSPIDER v.2.0.9.0 (Lischer and Excoffier 2012). Default parameters were used for the run which consisted in 20 pilot runs of 5,000 iterations each, followed by a longer run of 100,000 iterations with the first 50,000 iterations discarded as burn-in. Various values were evaluated for the prior odds (the ratio between the probability of a model without selection and that of a model with selection) input. Higher values for the prior odds lead to reduced type I errors (i.e. the probability to eliminate false-positive outliers is higher) but also reduce the power to detect loci under selection (type II error, the probability of rejecting true candidate selected loci). Three separate analyses were performed considering prior odds parameter of 5, 10, and 50. The three values of the prior odds chosen were determined based on recommendations of Foll (2010) and considering the number of markers in the dataset.

Summary statistics per sampling locality, including the number of polymorphic loci and gene diversity, were calculated using ARLECORE v.3.5.2 (Excoffier et al. 2005). Gene diversity and its standard deviation were calculated as described in Nei (1987).

Homogeneity of gene diversity among sampling locations was tested using a Wilcoxon signed rank test in SPSS v.20 (IBM Corp., Armonk, NY, USA).

5.2.4.1 Spatial Genetic Variation and Contemporaneous Gene Flow

The fixation index F_{ST} among sampling locations was estimated as Weir and Cockerham (1984) θ using VCFTOOLS and exact homogeneity tests in GENEPOP v.4.2 (Raymond and Rousset 1995) were used to test the significance of differences among

localities in allele distributions. Pairwise exact tests were also performed using GENEPOP

for comparison of individual localities. When the exact probability could not be

computed, it was estimated using a MCMC method. Markov chain parameters were

10,000 dememorizations, 500 batches, and 5,000 iterations per batch.

Analyses of Molecular Variance (AMOVA) were performed with the software

ARLECORE. In all analyses, significance of molecular variance components and fixation

indices was determined using 10,000 permutations of pseudo-haplotypes, and utilizing

the number of pairwise differences as the molecular distance between haplotypes. Input

files were generated from VCF files using PGDSPIDER.

Population structure was further investigated using the Bayesian clustering

approach in FASTSTRUCTURE v.1.0 (Raj et al. 2014). FASTSTRUCTURE, as the

program STRUCTURE (Pritchard et al. 2000; Falush et al. 2003; Hubisz et al. 2009)

utilizes a Bayesian method to identify differentiated subpopulations and infer the ancestry

of sampled individuals. The allocation of genotypes to subpopulation clusters is

optimized in an iterative process that minimizes the departure from Hardy-Weinberg

equilibrium and the linkage disequilibrium in the obtained partition of the dataset.

FASTSTRUCTURE implements a variational Bayesian inference algorithm that allows

approximating the marginal likelihood of the data with much reduced computational

requirements than older versions of structure, making this software suitable for inference

of structure even with very large datasets such as the one analyzed in the present study.

The criteria used to determine the number of subpopulations (K) included K_ε^*, which is

the value that maximizes the marginal likelihood of the model, and $K_{\emptyset_C}^*$, which is the

151

smallest value of *K* that explains at least 99.99% of the ancestry in the sample. One hundred runs were performed using a logistic prior and compared for each candidate value of *K*; *K* values between 1 and 6 were considered.

A Mantel test of the correlation between genetic and geographic distance was computed in GENALEX v.6.5 (Peakall and Smouse 2006, 2012) and the significance of the correlation was tested using 1,000 permutations. The genetic distance between individual localities used in the test was computed as $F_{ST}/(1-F_{ST})$ as recommended by Rousset (1997) where F_{ST} was the Weir and Cockerham (1984) θ. In this analysis, dispersal of queen triggerfish was assumed to follow a stepping stone model along the Antilles archipelagos where gene flow results from the transport of juveniles and larvae along the Antilles current in a one-dimensional dispersal process. The geographic distance was calculated as a linear distance among sampling localities, in kilometers.

The slope of the regression between genetic and geographic distances (*b*) was used to calculate the parameter σ (the standard deviation of parental position relative to offspring position) using the formula $\sigma = \sqrt{\dfrac{1}{4D_e b}}$ (equation 1) where D_e is the effective population density.

The effective population density used to estimate σ was derived as $D_e = \dfrac{N_{eH}}{L_T}$, where L_T is the length of the transect, expressed in kilometers, across the entire sampling range following the Antilles archipelago axis until Puerto Rico and then continuing to Florida along the Antilles current (Figure 5.6), and N_{eH} is the harmonic mean of the effective population size (N_e) estimates in the different localities.

152

Effective population size was estimated using both the linkage/gametic disequilibrium (LD) method (Hill 1981; Waples 2006) and the heterozygote excess (HE) method (Pudovkin et al. 1996) for the 6 sampling locations. Computations were performed using the software NEESTIMATOR v.2.0.1 (Do et al. 2014). Markers with minor allele frequencies below 0.02 were discarded in the calculation as recommended by Waples and Do (2010) considering the available sample sizes.
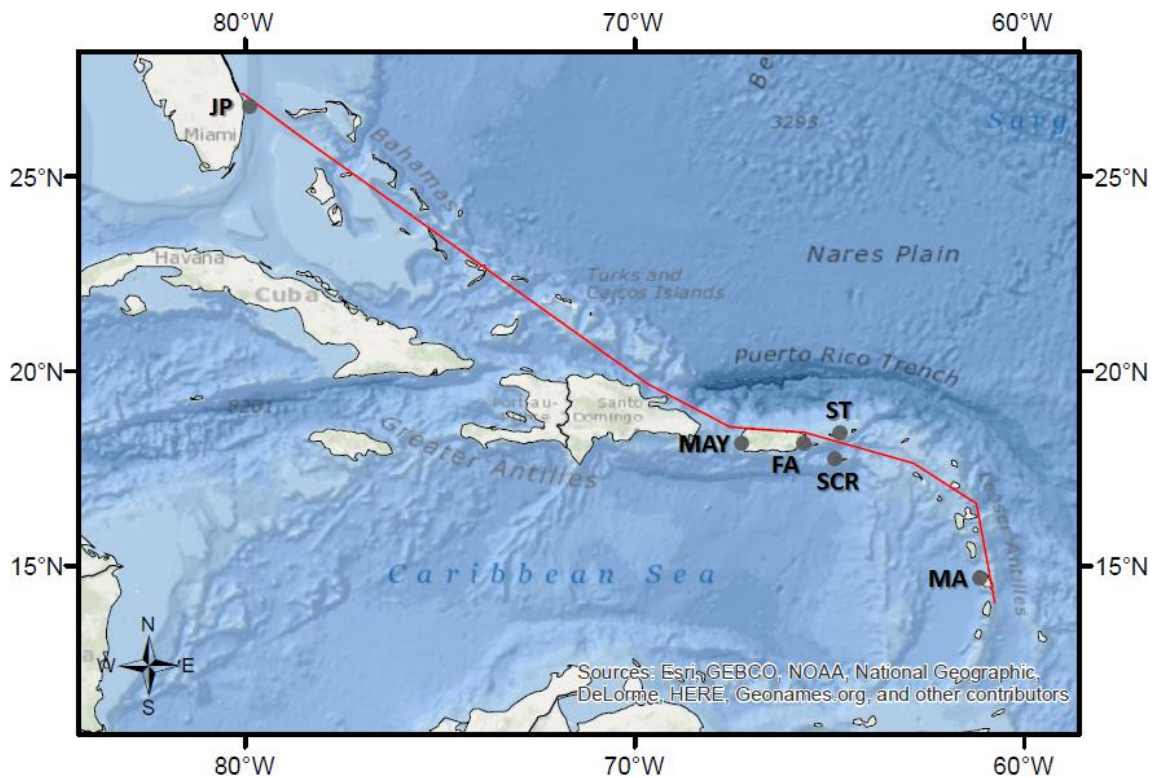
Figure 5.6 Transect Line Used to Approximate the Shortest Distance Between Queen Triggerfish Geographic Populations

Sampling locations for the study of population structure of queen triggerfish. The red line represents the transect used to describe the isolation-by-distance model.

5.3 Results

Sequencing of the libraries on the Illumina platform yielded an average of 1,502,928 raw reads per sample. A total of 229,506,813 raw reads (on average 32% of the

reads obtained for an individual) were identified as PCR duplicates and removed from the

dataset reducing the final average number of usable raw reads per sample to 1,014,097.

The minimum and maximum numbers of reads per sample were 5,502 and 2,501,012

respectively; the minimum was 24,796 when excluding the two samples with the lowest

coverage. Within these reads a total of 108,890 SNPs were identified in FREEBAYES, half

of which (49,027) were shared among all localities. The slope of the plot of the number

of clusters as a function of similarity in CD-HIT sharply increased when similarity

reached values greater than 91% and this pattern was unaffected by the choice of the

minimum coverage for individual SNPs (10, 12 or 15). Therefore, the final dataset was

generated using a minimum coverage of SNPs of 10 and a similarity setting at 91% in

order to maximize the number of loci while minimizing the risk of erroneously splitting

clusters during assembly (Appendix Figure D1; Appendix Table D2).

The harmonic mean of the number of retained samples per location showed

minimal change ($< 4.2\%$) when the percentage of missing data per individual varied

(Appendix Table D3), but the number of loci shared among locations was greatly affected

by this parameter. Allowing 50% of missing data yielded 8,423 SNPs in the dataset after

filtration while allowing only 25% of missing data reduced the final number of SNPs to

5. When intermediate percentages were considered, a minimal increase in the percentage

of missing data (from 35% to 40%) resulted in nearly 3 times as many loci (3,177 vs

1,049) in the final dataset. Datasets allowing 35% and 40% missing data were both

considered during further analyses of population structure and yielded similar outcomes,

therefore only results obtained using the dataset allowing for 40% of missing data per

individual are presented below.

After filtration of the dataset, the sample size per locality averaged 71 individuals

(range 48-85) and the number of retained SNPs within a locality averaged 35,654 (range

16,811-44,868). A total of 1,547 SNPs (average of 258 per locality, range 101-418) were

discarded because genotype proportions departed significantly from Hardy-Weinberg

expectations potentially signaling genotyping errors (Table 5.1). The total number of

SNPs shared among all sampling localities with genotypes available for at least 75% of

the individuals and a minimum sample size of 48 per locality was 3,177 and served as the

final dataset.

Table 5.1

Number of Samples and SNP Loci Retained per Locality During Filtration of the Dataset

	JP		FA		MAY		ST		SCR		MA		TOT	
	BF	AF	BF	AF	BF	AF	BF	AF	BF	AF	BF	AF	BF	AF
# samples	79	78	79	76	75	72	87	85	79	68	66	48	470	427
# of SNPs	42,839	42,323	41,603	41,000	45,244	44,868	38,630	38,130	31,465	30,790	17,457	16,811	49,027	3,177
# SNPs out of HWE	418		271		251		286		220		101			

Sample sizes and number of SNPs retained per sampling locality before and after filtration of the dataset. BF = Before Filtration; AF = After Filtration. Locality abbreviations are defined in

Figure 5.5, TOT: overall dataset (all localities combined).

PCA revealed occurrence of 4 outliers in the samples from the lower Caribbean and 3 in the sample from South Florida. The coordinates of these outliers on the first two principal components departed from the centroid by at least 2.6 standard deviations (Figure 5.7). Summary statistics and results of analyses of population structure with or without these 7 outliers gave similar results but, in order to reduce the risk of artifacts impacting parameter estimation, the outliers were excluded from the dataset to generate the results presented below.
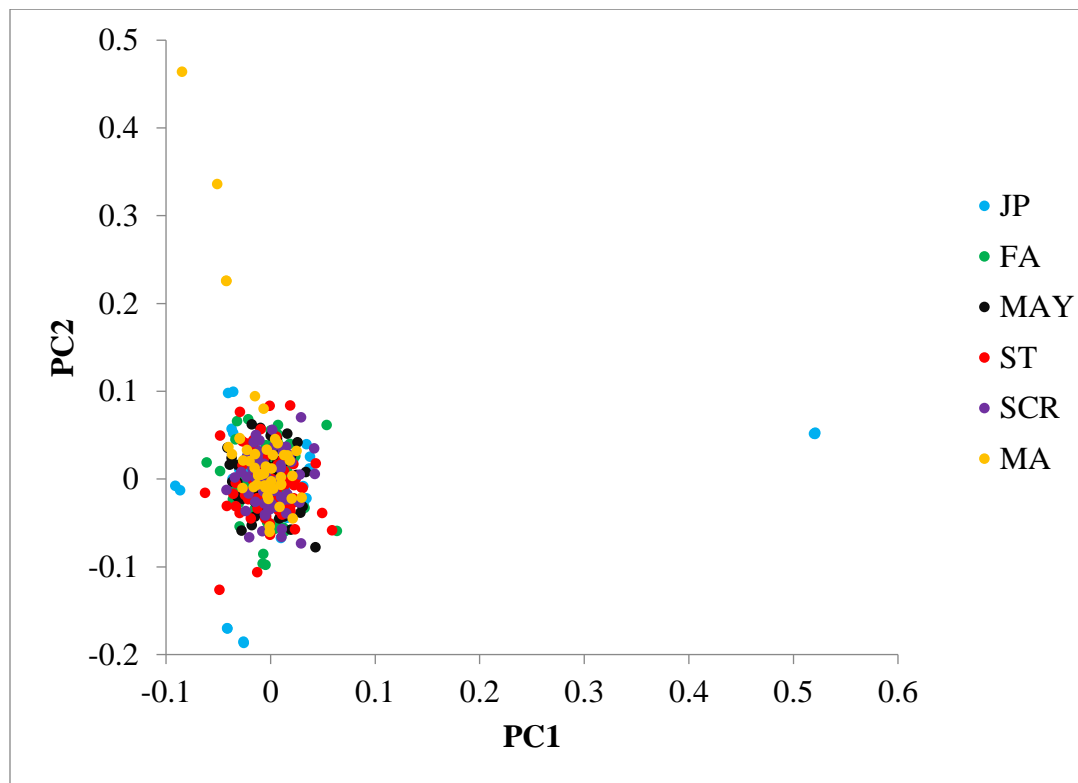
Figure 5.7 Principal Component Analysis of Queen Triggerfish Genotypes

Principal Component Analysis of genotypes from sampled individuals in each sampling location.

Analyses in BAYESCAN accounting for various values of the prior odds did not detect any outlier locus among the retained SNPs. The whole dataset was therefore kept

for further analysis of neutral spatial genetic variation and estimation of effective

population size.

Summary statistics per population are presented in Table 5.2. A minimum of

3,115 SNPs (out of 3,177 total SNPs surveyed) were polymorphic in each sampling

location. Gene diversity did not differ significantly among any pair of geographic

samples (P > 0.118).

Table 5.2

Genetic Diversity Statistics Estimates for Six Queen Triggerfish Geographic Populations

	JP	FA	MAY	ST	SCR	MA	All
# of individuals	75	76	72	85	68	44	427
# of polymorphic loci	3,157	3,163	3,165	3,161	3,167	3,115	3,177
gene diversity	0.203	0.156	0.161	0.145	0.161	0.170	0.164
gene diversity s.d.	0.096	0.074	0.077	0.069	0.077	0.081	0.078

Summary statistics per population overall loci. s.d.: standard deviation.

5.3.2 Spatial Genetic Variation and Contemporaneous Gene Flow

AMOVA and exact tests of population differentiation did not provide evidence of

significant differences in allele frequencies among geographic populations (Table 5.3), a

finding consistent with the very low estimate of F_{ST} overall populations (0.0007; 95% CI

0.0003-0.0011). Exact homogeneity tests across all loci were significant but none of the

tests performed at individual loci was significant after FDR correction. Bayesian

clustering runs in FASTSTRUCTURE accounting for different numbers of subpopulations

were compared to determine the model complexity (number of subpopulations K). All

test runs yielded an optimal value of one for K based on both the K_ε^* and $K_{\emptyset_C}^*$ criteria,

indicating that the structure in the dataset was best explained in a one population scenario

with no subdivision. The probability of significance of the correlation between genetic and geographic distance obtained during the Mantel test was 0.063 (Rxy = 0.848). The corresponding isolation by distance slope b was 5×10^{-7} (Figure 5.8).

Table 5.3

Pairwise F_{ST} Estimates and P-Values During Homogeneity Tests Comparing Samples from Queen Triggerfish Geographic Populations

JP	FA	MAY	ST	SCR	MA	
	0.986	0.844	0.214	0.999	0.999	JP
0.0014		1.000	1.000	1.000	1.000	FA
0.0011	0.0004		1.000	1.000	1.000	MAY
0.0013	0.0007	-0.0002		1.000	1.000	ST
0.0009	0.0007	0.0002	0.0001		1.000	SCR
0.0016	0.0007	0.0004	0.0006	0.0004		MA

Estimates of F_{ST} (lower diagonal) and exact probability (upper diagonal) obtained during pairwise homogeneity tests comparing queen triggerfish geographic samples.
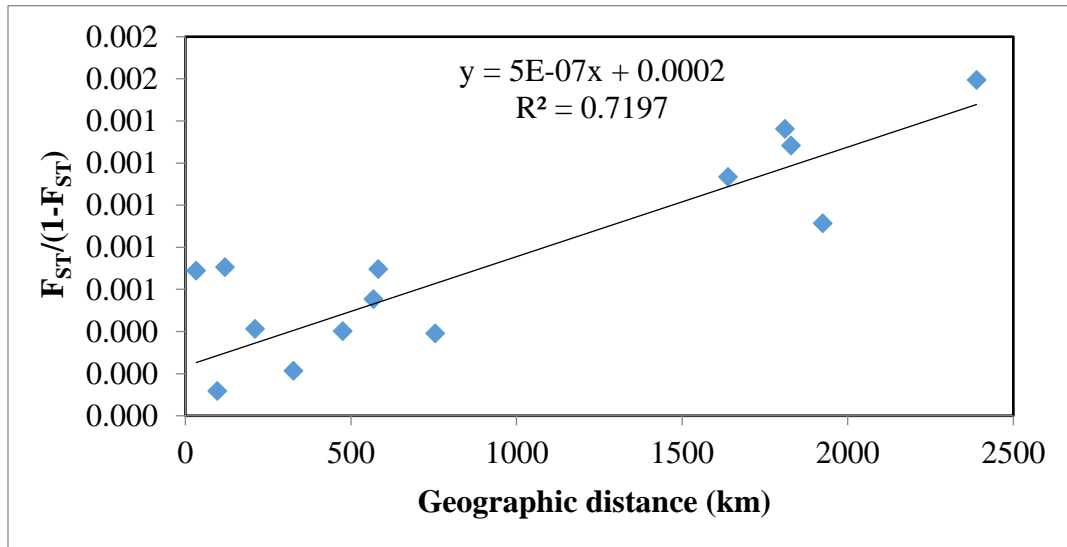
Figure 5.8 Relationship Between Genetic and Geographic Distance in Six Geographic Populations of Queen Triggerfish

Plot depicting the relationship between genetic and geographic distance in 6 geographic populations of queen triggerfish. The equation of the regression line and correlation coefficient R^2 are reported above the graph.

Estimates of effective population size by the heterozygote excess method were all

infinite. Estimates by the linkage disequilibrium method are presented in Table 5.4.

Estimates for the upper and lower Caribbean regions were significantly higher than the

one for South Florida (20.1 times higher on average). The harmonic mean of N_e

accounting for estimates obtained in all locations was 453 and corresponded to an

effective density D_e of 0.174 which yielded an estimate of 1,695 km for sigma based on

equation 1. Considering the extremely low effective size estimate for Florida and the long

geographic distance between South Florida and Caribbean locations, sigma was also

estimated accounting for the Caribbean samples only in order to evaluate dispersal

parameters in the Antilles region, where the density was higher. The harmonic mean of

the effective population size and the effective density, in this case, were 1,164 and 1.159

respectively, yielding an estimate of the dispersal parameter sigma of 657 km.

Table 5.4

Estimates of Effective Population Size by the Linkage Disequilibrium Method for Six

Queen Triggerfish Geographic Populations

Pop	N_e (LD)
JP	111.8 (110.9 - 112.7)
FA	818.3 (771.3 - 871.2)
MAY	4166.1 (3164.2 - 6089.7)
ST	3197.1 (2616.5 - 4106.1)
SCR	468.9 (451.0 - 488.2)
MA	2578.4 (1954.7 - 3782.0)

Effective population size estimates by the linkage disequilibrium method and 95% confidence interval in geographic populations of queen triggerfish.

5.4 Discussion

A total of 3,177 SNPs shared among the 6 sampling locations were generated through ddRAD sequencing and available to study spatial genetic variation in queen triggerfish. The deployment of a high density genome scan allowed examining whether some of the genetic loci were subjected to divergent selection and local adaptation through an outlier analysis. None of the loci under investigation were identified as significant outliers during analyses in BAYESCAN accounting for various values of the prior odds, suggesting that the 3,177 SNPs sampled were evolving neutrally. Considering the large number of loci surveyed in this genome scan and assuming conservative chromosome map lengths of 150 cM, the 22 chromosomes (Sá-Gabriel and Molina 2005)

of the queen triggerfish were expected to be covered by on average 144 SNPs such that

each locus under selection would be expected to be located within a centimorgan or less

of one of the SNP surveyed in this study. In this situation, the signature of even relatively

weak local selection would be expected to yield a F_{ST} signal at SNPs framing selected

loci in a broad range of demographic scenarios (Charlesworth et al. 1997; Storz 2005).

Thus, the lack of any significant outlier in the present study suggests that no genomic

region is undergoing strong divergent selection. Previous studies of marine species with

large dispersal capabilities using high density genomic scans revealed the occurrence of

outliers in association with significant neutral population structure (Nielsen et al. 2009;

Bradbury et al. 2010; Limborg et al. 2012; Laconcha et al. 2015), but also in cases where

there was no significant spatial structure at neutral loci (Lamichhaney et al. 2012; Grewe

et al. 2015). The present study on queen triggerfish thus contrasts with these studies in

that no significant outliers were detected. However, as pointed out by Lotterhos and

Whitlock (2014), several of the methods used to detect outliers have a high risk of false

positives, especially when no structure is present at neutral loci. In consequence, some of

the studies reporting outliers, in particular those where outliers were not found associated

with detectable neutral structure, may in fact have been impacted by high rates of false

positives and thus situations similar to the present study on queen triggerfish with no

significant support for outlier loci may be more common than currently apparent in the

literature. Considering that queen triggerfish have a high potential for gene flow, it is

possible that the lack of differentiation is the result of a selection-migration balance

where the differentiation caused by divergent selection and local adaptation is

counterbalanced by gene flow (gene swamping, Lenormand 2002; Conover et al. 2005; Cheviroz and Brumfield 2009). A first hypothesis would be that queen triggerfish are only found in habitats with similar characteristics leading to little or no local selection. A second hypothesis is that there is some degree of local selective pressure but migrations are sufficient to prevent divergence at impacted loci. As pointed out by Lenormand (2002), even though the potential for adaptation is greater in sparsely populated environments such as those occupied by queen triggerfish outside their center of abundance, the homogeneizing effects of gene flow is also stronger in those populations with a higher rate of effective immigration from the larger stocks in the center of abundance of the species. Such a scenario is plausible in the range surveyed in this study considering the high dispersal distance estimates compatible with direct transport of migrants across the sampling surface. Therefore, if local selective pressures are important determinants of the fitness of queen triggerfish populations, it would be important to rebuild healthy local stocks with large effective size so that the impact of migrants would be reduced and the selection of genotypes with higher fitness becomes more efficient.

Based on the negative outcome of outlier tests, the analysis of spatial genetic variation and demographic parameters continued assuming all the loci were evolving neutrally. Divergence among geographic samples was very low (F_{ST} of 0.0007) and a possible isolation-by-distance pattern was suggested by a positive correlation between genetic and geographic distance although the slope of the relationship was not significant (P = 0.06). Studies of population structure in the Southeast Florida and U.S. Caribbean were recently conducted in three Lutjanids, sharing the adult sedentary behavior and

163

pelagic larval lifestyle of the queen triggerfish. Although these species are characterized by a shorter pelagic phase (30-40 days, Lindenman et al. 2000) than that of the queen triggerfish, in all three species, divergence among samples was also very weak (-0.009 – 0.0095, Carson et al. 2011; Gold et al. 2011; Saillant et al. 2012).

The observation of very weak levels of divergence and lack of statistical significance of homogeneity tests in this study are thus consistent with the low levels of divergence reported in all three snapper species and the prediction of dispersal across longer distances based on the longer larval pelagic phase (almost twice the duration of that of snappers). Opportunities to overcome barriers such as the separation between the two Puerto Rican platforms or the longer distance between the U.S. Caribbean and South Florida are much greater in queen triggerfish. Homogeneity in the frequency of alleles was indeed observed even when the population from La Martinique was accounted for, bringing a total geographic distance covered by the sampling design to 2,604 km.

Divergence between North and South of the Puerto Rican platform and a possible isolation of populations located West of Puerto Rico were detected in yellowtail snapper (Saillant et al. 2012), and, for both yellowtail snapper and lane snapper, divergence between the U.S. Caribbean populations and South Florida was inferred (Gold et al. 2011; Saillant et al. 2012). These results possibly reflected an isolation-by-distance pattern as hypothesized in this study for queen triggerfish. Also, in the mutton snapper, differences in effective population size estimates among regions of the Caribbean suggested demographic independence of nearby populations, despite of the lack of significant divergence in allele frequencies in this species (Carson et al. 2012). These

164

results suggest that structure and demographic independence may occur on a small geographic scale, possibly due to physical factors such as the separation of the two Puerto Rican platforms, differing levels of local retention, or differing exploitation rates among regions, even when divergence in allele frequencies is very weak. The study of effective population size can provide further insights into the degree of demographic independence of local populations as in the case of mutton snapper (Carson et al. 2012).

In the queen triggerfish, even though gene diversity estimates did not differ significantly among sampling locations, comparisons of estimates of effective population size revealed significant differences between South Florida (N_e=111.8), St. Croix (N_e=468.9), Fajardo (N_e=818.3), and the rest of the Caribbean regions (N_e: 2,578.4 – 4,166.1) (See Table 5.4 in "Results"). According to the "50/500 rule" defined by Rieman and Allendorf (2001), an effective population size of at least 50 is sufficient to minimize inbreeding effects, while N_e greater than 500 would be necessary to maintain the equilibrium between the loss of adaptive genetic variance from genetic drift and its replacement by mutation. Accordingly, an effective size greater than 500 is a minimum target for management of populations in order to achieve long-term sustainability. Based on the results from this study, Caribbean sampling locations have high levels of genetic diversity and appear to maintain effective populations sizes greater than (or approaching) 500 (range 468.9 – 4,166.1) and would not be at immediate risk of extinction due to genetic factors.

The St. Croix sampling location was the only Caribbean location with effective size just below the upper threshold defined by Rieman and Allendorf (2001). Swearer and

165

collaborators (1999) found that bluehead wrasse (*Thalassoma bifasciatum*), a reef fish

with a long planktonic larval duration, exhibit recruitment around St. Croix dominated by

local retention. This is particularly true from June to August (Swearer et al. 1999). One of

the queen triggerfish main spawning periods occurs between August and October (Aiken

1983). Even though this only partially overlaps with the period during which local

recruitment seems to be favored, an increased local retention, at least at the beginning of

the spawning season could contribute to the lower effective population size estimate

around St. Croix. However, considering the geographic proximity with the other upper

Caribbean sampling locations where estimates of N_e were much larger, the low effective

size in St. Croix may also reflect a temporal artifact and this result needs to be confirmed

using additional samples.

The estimate of effective population size for the Florida region was very small

(N_e=111.8), on average 20.1 times smaller than those in other populations. Even though

this value is not below the threshold (N_e=50) under which the effect of inbreeding would

be strong, it is 4.2 times smaller than the minimum (500) recommended to ensure long-

term viability. Considering that genetic diversity was relatively high, one possible

hypothesis could be that the population was historically relatively large but underwent a

bottleneck in the sampled generations, which translated into a small effective size

estimate by the linkage disequilibrium method which estimates contemporaneous N_e (i.e.

the effective number of breeders in the generation that produced the sample). Thus, while

allele frequencies appear homogeneous across the area, effective population size

estimates vary among local populations suggesting some degree of demographic

166

independence, and the size of the Florida population appears very small when compared to that of Caribbean stocks.

Connectivity across the area was further examined by estimating the dispersal parameters using genetic data and the inferred isolation-by-distance relationship. This distribution is directly impacted by the population density and, in this study, this quantity was derived from the harmonic mean of the N_e estimates obtained for individual sampling localities. This calculation was strongly influenced by the small N_e obtained for the South Florida population.

The population density derived from U.S. Caribbean populations and La Martinique LDN$_E$ estimates yielded an average parent-offspring dispersal distance of 657 km.

Including the N_e from South Florida increased the dispersal distance parameter (by decreasing the inferred density) to a much larger value (1,695 km). The status of the South Florida population is unclear from this study. The small effective population size estimated in this work, if maintained over several generations, would be expected to lead to rapid divergence in allele frequencies between the Florida population and those in the U.S. Caribbean, unless there is a strong input of migrants from the more stable Caribbean metapopulation in the recruiting gene pool. Triggerfish in South Florida show highly variable recruitment among years (Figure 5.2), an observation inconsistent with the occurrence of a stable spawning population in that region. Insights on the occurrence and status of a breeding population in South Florida could be gained by looking for spawning aggregations or assessing if adults are spawning capable or actively spawning during the

spawning season. If South Florida lacks a spawning population, recruitment could be provided by spawning stocks in the Caribbean with occasional larval pulses settling in South Florida when pelagic transport conditions are favorable. This hypothesis would be consistent with the small effective population size estimate in that recruits would descend from one or a few spawning aggregations, whose offspring reached South Florida. Spawning aggregations of queen triggerfish appear to be small (Heyman et al. 2013; < 40 specimens, R. Nemeth, University of the Virgin Islands Personal communication) and would be expected to produce pools of migrants of limited effective size, a scenario consistent with the observation of a small N_e estimate in our sample. The potential contribution of immigrants to recruitment in the South Florida region can also be discussed based on the isolation-by-distance model.

For the sake of discussion, on Figure 5.9 the isolation-by-distance model was redrawn assuming a low population density for the Florida region, based on the Florida estimate of N_e, and the corresponding model was applied to predict genetic distances between Florida and Caribbean populations. All the points generated from the pairwise comparison of South Florida versus the rest of the sampling locations lie below the modified regression line suggesting that the observed genetic distance (F_{ST} estimates) are below expectations under the assumption of a low population density in the Florida region. The finding of a small N_e in Florida, therefore, warrants further investigation with additional temporal sampling. The lack of divergence of Florida samples from those in the Caribbean and the small effective size estimates obtained from this population, if

168

repeated over multiple sampling years, would support the hypothesis of a strong

dependency of recruitment in South Florida on migrants from the Caribbean.
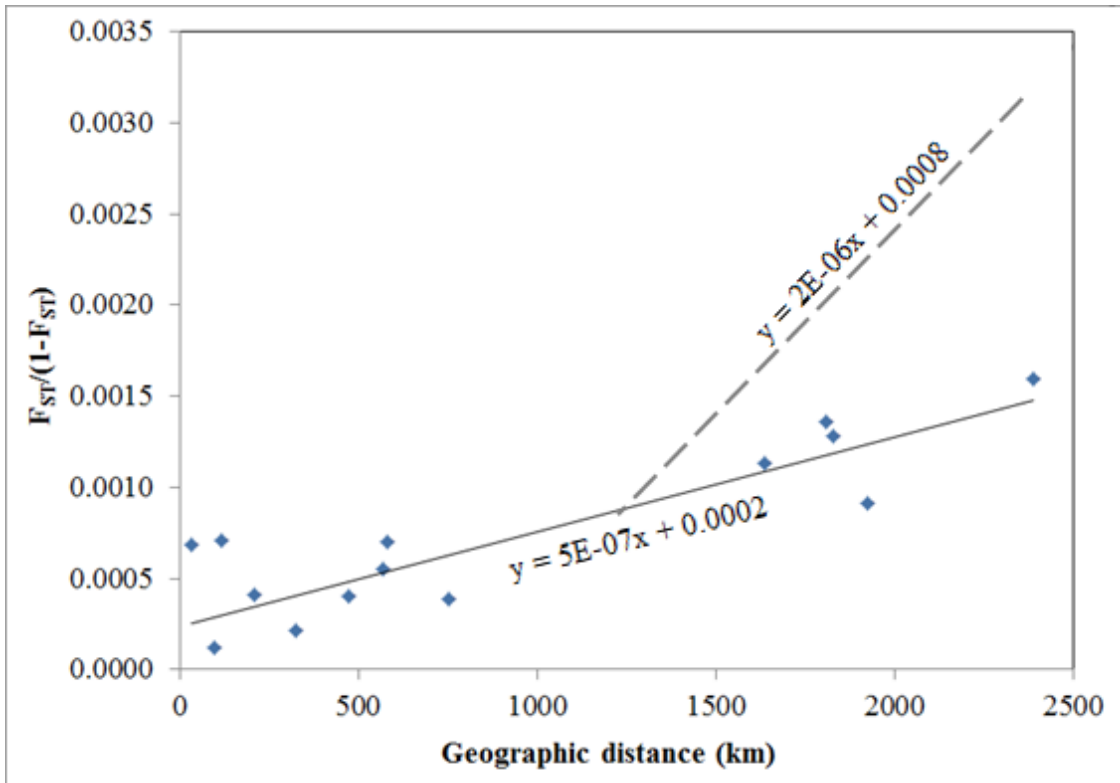
Figure 5.9 Plot Depicting the Relationship Between Genetic and Geographic Distance in Queen Triggerfish with Modified Regression Line

Plot depicting the relationship between genetic and geographic distance in 6 geographic samples of queen triggerfish. The continuous line represents the regression line obtained from the data while the dashed line was redrawn by modifying the slope to account for a possible smaller density in the northern part of the range (proximity of Florida). Equations for both lines are also presented on the figure.

Considering the estimates of isolation-by-distance parameters generated based on the high population density values obtained in the Caribbean localities, the slope of the IBD relationship was positive and yielded estimates of sigma of 657 km or higher, which corresponded to a mean dispersal distance of at least 524 km calculated assuming a normal dispersal distribution function in one dimension (as per Puebla et al. 2012).

According to calculations based on the average velocity of the Antilles current, and considering a pelagic phase lasting up to 63-83 days, an average dispersal distance along the axis of 218-645 km was estimated. Thus, this estimate is consistent with the genetic estimate of 524 km obtained in this study.

The sampling surface thus barely covered twice the average estimated dispersal distance. In consequence, individuals dispersing at long distances are not expected to be captured using the current sampling design, leading to a potential underestimation of sigma (Leblois et al. 2003). Sampling additional populations in order to increase the size of the lattice used to derive estimates would be beneficial to estimate more accurate dispersal parameters.

The large volume of catches in the U.S. Caribbean along with the rarity of queen triggerfish in other parts of the distribution range suggests that the northern Caribbean populations may represent the center of abundance of the species. Considering the range-wide decline of queen triggerfish (except in the northern Caribbean region), determining the connectivity of the northern Caribbean stock with other remaining populations is essential in order to identify geographic areas that are unlikely to receive migrants, as these populations would need to receive priority for the protection of local spawning stocks. Based on circulation patterns in the region, potential sources of migrants to the Caribbean would be located in South and Central America. Considering the low abundance of queen triggerfish in those regions, the contribution of these stocks to recruitment and genetic diversity in the Caribbean is likely very low. In addition, the orientation of surface currents mostly South to North predicts the lack of gene flow from

170

the northern Caribbean to southern populations in the lesser Antilles or South America. These southern populations, therefore, require specific conservation efforts, as they cannot rely on migrants to rebuild spawning biomass and genetic diversity. The northern Caribbean stock itself, as the possibly last remaining healthy stock of queen triggerfish, needs specific attention in order to maintain sufficient genetic diversity for the species to persist in the long term.

CHAPTER VI – GENERAL CONCLUSION

The objective of this work was to characterize the structure, connectivity, and demographic dynamics of the metapopulations formed by two species of triggerfish with high potential for larval dispersal. Both species are highly sedentary as adults and were predicted to maintain connectivity among geographic populations via the exchange of migrants dispersed as pelagic larvae. The duration of pelagic transport in these two species is in the upper range of that reported in marine reef fishes, leading to the prediction that connectivity is occurring across long distances, including between populations separated by large sections of habitats unsuitable for settlement.

The first part of this work focused on the gray triggerfish and employed homologous microsatellite markers developed for the purpose of this study and partial sequences of the mtDNA coding gene ND4. Both marker types were surveyed in 12 localities spanning most of the distribution range of the gray triggerfish and revealed the occurrence of 4 genetically distinct regional stocks. The genetic characterization of populations separated by wide areas of open ocean, unsuitable for benthic settlement of juveniles, provided an opportunity to assess the effectiveness of long distance dispersal in this species. Northeast and Northwest Atlantic populations had homogeneous microsatellite allele and mtDNA haplotype frequencies, indicating that connectivity between the two regions has been maintained historically and suggesting that the hypothesized transport of larvae across the Atlantic through the Gulf Stream system is effective. Similarly, while the Southeast and Southwest Atlantic appear to show some degree of historical isolation, as indicated by the detection of a differentiated

172

subpopulation in the Southwest Atlantic during Bayesian clustering of genotypes, a high

frequency of migrants from the Southeast was detected in the Southwest during analysis

of contemporaneous migrations. The latter result indicates that transatlantic connectivity

through the NERR can also occur at high rates. In both cases, larval transport covered

more than 3,800 km and durations were likely over 9 months. Altogether, these results

confirmed that substantial connectivity at the scale of the Atlantic is occurring in this

species, pending larvae and *Sargassum* patches can reach suitable surface currents for

transport. The parameters of dispersal estimated in the near-continuous population found

off of the southeastern United States also indicated that long-distance dispersal events are

frequent, several hundreds or even thousands of kilometers from the origin. An

implication of these findings is that recruitment in this species is relatively independent

from local spawning biomass, and may instead be determined by the output of spawning

stocks located outside the boundaries of regional or national fisheries management units.

The queen triggerfish shares the extended pelagic dispersal of the gray triggerfish

and, although only a reduced section of this species' range could be studied during this

work, estimates of dispersal distances from the isolation-by-distance relationship were

consistent with those obtained for gray triggerfish, indicating a substantial effect of long

distance migrations mediated by surface currents on recruitment dynamics as discussed

above.

Another prediction was that some areas would display little or no connectivity due

to the lack of favorable currents. This hypothesis was supported by findings in the gray

triggerfish with the isolation between northwestern and southwestern Atlantic stocks

173

consistent with the divergence of the Brazil and North Brazil currents. Similarly, a genetic discontinuity was found between the Gulf of Guinea and Northwest Africa, consistent with the scarcity and reduced intensity of currents connecting the two regions during the gray triggerfish spawning season. Overall these results are consistent with the hypothesis that surface currents are strong drivers of the actual connectivity among populations.

A third prediction was that gene flow would be asymmetric. The analysis of contemporaneous migrations in the South Atlantic revealed that all the dispersal events inferred from the data set occurred in a westward direction, a finding consistent with predictions based on surface circulation in that region. Asymmetric migrations across the North Atlantic, as mediated by the Gulf Stream, were also hypothesized. Although this prediction could not be formally tested during the analysis due to the inability to genetically distinguish Northwest Atlantic individuals from those collected in the Northeast Atlantic, the genetic homogeneity between the two regions is consistent with the hypothesis. The potential implications of these asymmetric migration patterns for conservation are significant: while some populations serve as 'sinks' for migrants and benefit from the outputs of both local spawning biomasses and those of foreign spawning stocks connected to them, others may serve as source populations but would get little or no immigrants due to the lack of adequate incoming currents. The latter populations would be at greater danger of extinction because they would not have the possibility to replenish their stocks from migrant pools following demographic bottlenecks. This is the case, for example, of the South American and western Caribbean populations of queen

174

triggerfish where the species has become rare and endangered in some countries. No healthy queen triggerfish stock appears connected to these regions, which may impair the recovery of populations if local spawning biomasses are severely depleted. Transplanting queen triggerfish for reintroduction in these regions may be a viable conservation option in this case, especially if local density has reached such a low level that the population is experiencing an Allee effect where population growth is not occurring because the remaining adults have reduced opportunities to find mates and breed (Gascoigne and Lipcius 2004). The concerns with transplantation have been related to the occurrence of local adaptation and the risks of outbreeding depression (Houde et al. 2011). However, the lack of evidence for adaptive variation and local adaptation in queen triggerfish suggests that this concern may be minimal in this species. In any case, the maintenance of healthy local stocks for populations that appear to show limited and infrequent connectivity with other stocks, as predicted based on oceanic currents circulation, seems essential and a priority for conservation.

Habitat availability was another factor hypothesized to structure populations and determine local abundance. The gray triggerfish, for example, seems absent from the Caribbean yet present in the Gulf of Mexico and South America. While the specific factor preventing the establishment of large gray triggerfish populations in the Caribbean is unclear, the high level of connectivity discussed above suggests that the very low abundance in the Caribbean region is related to ecological factors that will warrant investigation. Interestingly, the red snapper, a species that uses habitats very similar to

175

those used by gray triggerfish, is also nearly absent from the Caribbean region (Norrell 2016).

The occurrence of adaptive variation, where local populations would have been selected for genetic characteristics that maximize fitness in their home habitat, was examined in the queen triggerfish. The outlier analysis did not reveal occurrence of genomic regions undergoing divergent selection. While this analysis was limited by the lack of reference genome, the support for selection and local adaptation is tentatively weak in this species. The demographic dynamics of queen triggerfish, with potential strong effects of migrants on local gene pools, would be expected to counteract the effects of natural selection by homogenizing allele frequencies in populations, thereby preventing the selection of alleles and advantageous genetic combinations in local stocks (Lenormand 2002). Therefore, divergent selection and local adaptation may play a minor role in shaping the population structure in species such as the triggerfish studied in this work where recruitment seems to depend heavily on the migration of post-larvae from spawners located far from settling habitats. Natural selection is still expected to operate in these species but may favor mechanisms that maximize fitness across the range rather than in specific local geographic stocks.

6.1 Possible Changes in Connectivity Patterns Related to the North Equatorial Recirculation Region

The analysis of gray triggerfish in the Atlantic revealed the occurrence of recent transport through the NERR, concomitant with a new *Sargassum* source in the South Atlantic (Johnson et al. 2013; Franks et al. 2016). This new influx of *Sargassum* in the

176

equatorial region, if continued, is expected to keep promoting gene flow between the Southeast and the Southwest Atlantic and also between the South and the North Atlantic groups, as suggested by a few southeastern migrants identified in U.S. waters and in the Canary Islands. Accordingly, a change in the structure of gray triggerfish populations, as well as those of other organisms utilizing *Sargassum* as a habitat, is predicted and monitoring of population structure in conjunction with future formation and circulation of *Sargassum* in the NERR system is warranted.

6.2 Methods to Assay Genetic Variation

In this work, inferences on genetic variation employed microsatellites and mtDNA for gray triggerfish and high-density genome scans based on SNPs for queen triggerfish. When the gray triggerfish project began, next generation sequencing methods and the RAD-sequencing protocol were not available. The study was therefore conducted based on microsatellites and mtDNA focusing on describing neutral genetic variation and the effects of genetic drift and migrations. The dataset obtained for the queen triggerfish allowed examining both neutral and non-neutral genetic variation through the distinction of outlier loci potentially signaling regions of the genome undergoing selection in different part of the species' range. Divergent selection among geographic populations is a form of genetic structure of central importance for conservation because the extent of local adaptation is one of the main criteria used to design populations as Evolutionary Significant Units (Waples 1995). The results of the queen triggerfish study suggest that the role of divergent selection and local adaptation is reduced in this species, yet this finding will need confirmation by re-testing for the occurrence of outlier genomic regions

177

using a reference genome. The possibilities offered by high-density genome scans are vast (Fierst 2015) and their deployment has become achievable for costs similar to those of traditional markers such as microsatellites. Other than enabling the analysis of non-neutral genetic variation among populations as discussed above, high throughput genotyping projects using the RAD-sequencing methodology yield datasets showing higher power than those obtained using traditional methods for the assessment of neutral genetic structure and the estimation of demographic parameters such as the effective size of population thanks to the very large numbers of loci surveyed. This approach to assess genetic variation seems therefore recommended for future studies of marine populations. The development of reference genomic resources including a linkage map of the genome would allow mapping the position of genetic markers in the genome thereby improving the reliability of inferences on local adaptation and selection (Bourret et al. 2013) or that of estimates of effective population size (Larson et al. 2014). The acquisition of such resources to assist with the interpretation of the queen triggerfish genome scan data is warranted and in progress.

6.3 Conservation Implications

An important conservation implication of this study is related to the potential strong effect of migrations from distant populations on local recruitment. As suggested in Chapter IV, recruitment cannot be easily predicted from local spawning biomasses but instead may be more accurately predicted based on juvenile recruitment indices developed post settlement. Groundfish surveys targeting habitats used by triggerfish at

settlement could be designed to monitor incoming queen or gray triggerfish juveniles and generate the data needed for such indices.

Available information on queen triggerfish populations in various countries where they occurred historically indicate that the species is losing ground throughout most of its range. This suggests that the Antilles region may be the last center of abundance for queen triggerfish. Conservation of this population and maintenance or restoration of local stocks in other geographic populations are therefore priorities in order to ensure the sustainability of the metapopulation formed by the species to the long term. The gray triggerfish shares the extended larval dispersal and adult sedentary behavior life history traits of queen triggerfish and, while this species currently seems to maintain a larger number of relatively abundant populations, signs of decline for populations that lack predicted sources of immigrants, such as the Southeast and Southwest Atlantic, have already been reported and the overall status of gray triggerfish has been of concern even in the northern Gulf of Mexico, which is considered the center of abundance for this species (stocks were considered overfished in the U.S. since SEDAR-9 2006). Thus, the management and conservation implications suggested for queen triggerfish discussed above may apply to gray triggerfish as well.

Overall, the complexity and large geographic scale of the gray triggerfish and queen triggerfish metapopulations associated with a dependency of recruitment on the output of spawning events occurring far away from recipient habitats for grow-out, indicate that management and conservation of these species is difficult. In particular, populations exchanging migrants are located in foreign countries and jurisdictions, in

many cases, which impairs the coordination of conservation efforts. Some populations are potentially vulnerable to local extinction in regions that are less likely to receive immigrants due to unfavorable surface current patterns. These local extinction events pose the risk of a progressive destabilization of the metapopulations as some contributors to the migrant pools get depleted. Thus, monitoring these species in the various parts of their ranges with a focus on local spawning stocks is a priority.

Table A.1

Individual Assignment Test to Pure and Hybrid Categories*

*This table is uploaded as a supplementary file on the Aquila and ProQuest databases and is also available on a CD Rom in the bound

copy of this dissertation at the Gunter Library of the USM Gulf Coast Research Laboratory, Ocean Springs, MS.

APPENDIX B GRAY TRIGGERFISH EFFECTIVE POPULATION SIZE

Table B.1

Bayesian Coalescent Estimates of Gray Triggerfish Effective Size in Six Atlantic Regions

	US	FR	CA	MED	SEA	SWA
N_e	2,717	2,650	4,017	17	1,217	883
	(1,467 – 3,933)	(1,567 – 3,700)	(1,867 – 5,600)	(0 - 900)	(133 - 2233)	(0 – 1,700)

Estimates of N_e with 95% confidence intervals in 6 regions within the Atlantic basin obtained from microsatellite data assuming a mutation rate of 5×10^{-4}. US: United States; FR: France; CA: Canary Islands; MED: Mediterranean; SEA: Southeast Atlantic; SWA: Southwest Atlantic.

APPENDIX C SUMMARY STATISTICS

Table C.1

Microsatellite and MtDNA Summary Statistics per Geographic Sample

Locus	STX	ETX-LA	MS-WF	SWF	SEF	SC	FR	CA	GR	BE	AN	BR
BC13												
n	72	219	137	77	80	78	64	76	15	72	70	49
A	25	35	33	26	28	32	28	23	6	28	28	22
A_R	13.29	13.63	13.86	13.39	13.98	15.28	14.84	13.73	6.00	14.01	14.03	13.75
H_E	0.884	0.889	0.889	0.877	0.883	0.911	0.916	0.899	0.719	0.879	0.854	0.913
P_{HW}	0.155	0.050	0.802	0.258	0.972	0.010	0.657	0.240	0.959	0.104	0.286	0.070
F_{IS}	-0.021	0.003	-0.051	0.008	-0.048	0.071	0.028	0.078	-0.205	0.068	0.030	0.106
BC14												
n	72	220	137	77	80	77	64	76	17	71	70	52
A	23	30	30	21	27	26	25	28	7	26	22	22
A_R	14.66	14.81	14.55	13.38	15.72	14.79	14.77	15.96	6.74	12.45	11.92	12.92
H_E	0.922	0.922	0.913	0.917	0.930	0.919	0.910	0.942	0.732	0.849	0.847	0.889
P_{HW}	0.008	0.291	0.269	0.040	0.381	0.511	0.240	0.009	0.025	0.490	0.318	0.322
F_{IS}	0.051	0.048	0.049	0.065	0.006	-0.004	-0.030	0.037	0.357	0.005	-0.029	0.070
BC16												
n	72	220	138	77	80	78	64	76	17	72	70	52
A	5	6	6	5	5	6	4	6	1	3	3	3
A_R	3.73	3.67	3.67	3.87	3.32	3.63	3.21	3.95	1.00	2.09	2.15	1.93
H_E	0.369	0.422	0.409	0.413	0.270	0.345	0.420	0.442	0.000	0.131	0.159	0.075
P_{HW}	0.781	0.746	0.292	0.749	1.000	0.053	0.056	0.252	NA	0.285	1.000	1.000
F_{IS}	-0.054	0.019	-0.044	-0.037	-0.110	0.071	-0.189	0.048	NA	0.153	-0.079	-0.023

Table C 1 (continued).

Locus	STX	ETX-LA	MS-WF	SWF	SEF	SC	FR	CA	GR	BE	AN	BR
BC17												
n	72	220	138	77	80	78	64	75	17	72	70	52
A	18	23	22	18	16	15	15	19	6	19	22	22
A_R	8.68	16.3	8.54	7.41	7.62	7.26	8.64	9.44	5.65	11.29	11.45	12.42
H_E	0.748	0.663	0.679	0.639	0.644	0.641	0.695	0.766	0.454	0.872	0.871	0.866
P_{HW}	0.512	0.283	0.098	0.485	0.025	0.255	0.712	0.034	0.762	0.031	0.807	0.74
F_{IS}	0.108	-0.014	0.071	-0.017	0.127	0.02	0.011	0.077	-0.036	0.029	0.016	0.045
BC19												
n	72	220	138	77	80	78	64	70	15	71	70	51
A	10	11	13	12	10	12	9	12	6	10	12	9
A_R	6.87	9.9	7.31	7.02	7.4	6.99	6.38	8.16	6	7	7.92	6.57
H_E	0.744	0.737	0.759	0.755	0.759	0.742	0.719	0.772	0.802	0.763	0.788	0.752
P_{HW}	0.081	0.543	0.416	0.182	0.383	0.271	0.852	0.33	0.905	0.86	0.522	0.054
F_{IS}	-0.139	-0.011	-0.002	0.106	-0.005	0.015	-0.066	0.075	0.169	-0.016	0.021	0.061
BC2												
n	72	220	138	77	80	78	64	75	17	71	70	52
A	25	28	24	23	25	25	25	28	14	29	26	21
A_R	14.37	24.43	13.42	13.79	14.26	14.49	15.44	15.71	13.37	16.7	16.07	14.18
H_E	0.92	0.929	0.916	0.923	0.919	0.914	0.937	0.938	0.925	0.945	0.944	0.925
P_{HW}	0.359	0.051	0.535	0.378	0.393	0.138	0.642	0.125	0.242	0.783	0.955	0.343
F_{IS}	0.004	0.012	0.003	0.029	-0.007	0.018	0.032	-0.009	0.046	-0.043	-0.029	0.064

Table C 1 (continued).

Locus	STX	ETX-LA	MS-WF	SWF	SEF	SC	FR	CA	GR	BE	AN	BR
BC25												
n	72	220	138	77	80	78	64	75	17	72	70	52
A	18	17	18	15	18	13	13	17	7	14	12	13
A_R	11.26	14.66	10.91	10.85	10.94	9.99	10.09	10.89	6.87	8.66	8.7	9.58
H_E	0.896	0.886	0.897	0.857	0.885	0.886	0.886	0.882	0.838	0.84	0.848	0.843
P_{HW}	0.584	0.598	0.648	0.511	0.62	0.811	0.312	0.027	0.878	0.712	0.335	0.404
F_{IS}	0.023	0.025	-0.034	0	-0.073	-0.028	0.047	0.078	-0.123	0.057	-0.027	-0.072
BC26												
n	72	220	138	77	80	78	64	75	17	72	70	52
A	21	26	25	16	20	21	17	23	7	15	11	13
A_R	11.53	19.32	11.06	10.57	11.68	11.07	11.4	11.21	6.65	8.75	7.74	8.48
H_E	0.845	0.868	0.865	0.861	0.899	0.837	0.871	0.831	0.757	0.81	0.808	0.806
P_{HW}	0.755	0.331	0.292	0.172	0.748	0.884	0.576	0.62	0.042	0.166	0.18	0.461
F_{IS}	-0.035	0.005	0.028	0.065	-0.015	0.004	0.085	0.037	0.223	-0.046	0.046	0.045
BC27												
n	72	220	138	77	80	78	64	76	17	71	70	51
A	7	8	6	7	6	8	7	6	5	8	6	4
A_R	5.7	7	5.16	5.36	5.12	5.24	5.37	5.06	5	5.12	4.37	3.8
H_E	0.752	0.736	0.725	0.735	0.714	0.726	0.737	0.726	0.722	0.684	0.667	0.619
P_{HW}	0.734	0.833	0.252	0.714	0.475	0.284	0.776	0.893	0.464	0.978	0.261	0.511
F_{IS}	0.021	-0.044	0.03	0.028	-0.033	-0.059	-0.017	0.039	-0.059	-0.091	0.058	0.082

Table C 1 (continued).

Locus	STX	ETX-LA	MS-WF	SWF	SEF	SC	FR	CA	GR	BE	AN	BR
BC3												
n	72	219	138	77	80	77	64	75	15	72	70	52
A	5	9	8	8	7	7	7	6	3	8	4	5
A_R	4.19	5.62	4.65	4.72	4.43	4.52	4.64	4.31	3	4.53	3.52	3.55
H_E	0.654	0.613	0.633	0.606	0.655	0.604	0.642	0.582	0.693	0.589	0.591	0.582
P_{HW}	0.065	0.73	0.28	0.515	0.143	0.171	0.487	0.41	0.25	0.934	0.83	0.239
F_{IS}	0.044	0.009	-0.076	0.186	-0.011	-0.032	-0.095	-0.009	0.23	-0.013	0.009	-0.124
BC34												
n	72	220	138	77	80	78	64	76	17	72	70	50
A	19	27	20	17	21	19	16	23	9	20	21	21
A_R	11.21	18.31	10.06	9.7	12.26	10.38	10.41	12.02	8.64	10.52	12.08	12.01
H_E	0.893	0.879	0.879	0.871	0.907	0.887	0.893	0.906	0.857	0.823	0.88	0.84
P_{HW}	0.257	0.302	0.885	0.695	0.238	0.188	0.469	0.329	0.235	0.609	0.354	0.447
F_{IS}	0.02	0.028	0.019	0.016	-0.006	0.017	0.003	0.085	0.107	0.004	0.026	0.024
BC41												
n	72	220	138	77	80	78	64	75	17	72	70	52
A	8	9	10	9	9	9	9	9	2	4	4	5
A_R	5.7	7.68	5.69	5.45	5.87	5.8	6.26	5.75	1.99	3.2	3.32	4.26
H_E	0.571	0.534	0.54	0.569	0.534	0.569	0.631	0.588	0.114	0.54	0.434	0.696
P_{HW}	0.962	0.659	0.776	0.26	0.451	0.191	0.481	0.101	1	0.789	0.397	0.985
F_{IS}	-0.07	-0.03	0.02	0.018	0.016	-0.014	-0.015	-0.043	-0.032	0.049	0.013	-0.077

Table C 1 (continued).

Locus	STX	ETX-LA	MS-WF	SWF	SEF	SC	FR	CA	GR	BE	AN	BR
BC44												
n	72	220	138	77	80	78	64	72	17	72	70	48
A	22	28	29	21	24	23	23	24	11	23	21	16
A_R	14.78	22.42	15.01	13.95	14.52	13.73	14.53	14.72	10.4	12.94	12.12	11.46
H_E	0.937	0.928	0.939	0.931	0.935	0.92	0.937	0.935	0.853	0.92	0.914	0.905
P_{HW}	0.696	0.469	0.081	0.98	0.115	0.739	0.598	0.277	0.187	0.037	0.58	0.489
F_{IS}	0.007	0.006	0.028	-0.019	-0.03	-0.017	-0.034	0.049	-0.034	0.049	0	0.079
BC45												
n	72	220	138	76	80	78	64	75	17	72	70	52
A	12	14	15	12	14	13	13	11	4	12	13	11
A_R	7.87	12.28	8.97	8.56	8.62	8.8	9.08	8.33	3.77	6.95	8.19	8.7
H_E	0.795	0.816	0.834	0.827	0.833	0.807	0.826	0.843	0.395	0.774	0.806	0.829
P_{HW}	0.963	0.341	0.103	0.386	0.762	0.487	0.334	0.321	0.657	0.096	0.799	0.004
F_{IS}	-0.013	-0.036	0.035	0.03	-0.006	-0.016	0.073	0.083	0.107	0.103	-0.028	0.141
BC46												
n	72	220	137	77	80	78	64	74	17	72	70	52
A	38	42	41	41	33	38	33	38	17	33	32	30
A_R	19.55	36.28	19.29	20.1	18.04	19.28	18.4	19.03	15.79	17.65	17.92	17.36
H_E	0.965	0.965	0.963	0.969	0.959	0.96	0.959	0.961	0.932	0.952	0.958	0.956
P_{HW}	0.723	0.186	0.48	0.177	0.264	0.086	0.395	0.613	0.931	0.89	0.883	0.018
F_{IS}	-0.008	-0.008	0.014	0.008	0.036	0.039	0.023	0.001	-0.073	-0.007	-0.029	0.094

Table C 1 (continued).

Locus	STX	ETX-LA	MS-WF	SWF	SEF	SC	FR	CA	GR	BE	AN	BR
BC47												
n	71	219	137	76	80	78	64	75	17	72	70	52
A	37	39	39	34	32	35	34	35	15	29	31	24
A_R	19.58	32.27	18.98	17.67	17.92	18.32	19.4	19.12	13.94	17.1	17.3	15.63
H_E	0.967	0.962	0.962	0.953	0.957	0.959	0.966	0.965	0.915	0.951	0.953	0.944
P_{HW}	0.598	0.183	0.993	0.846	0.573	0.748	0.097	0.197	0.987	0.349	0.387	0.265
F_{IS}	-0.005	0.008	-0.017	-0.036	0.007	-0.002	-0.003	0.046	-0.092	-0.037	-0.005	-0.019
BC49												
n	72	219	138	77	80	77	64	76	17	72	70	51
A	8	15	11	10	9	11	9	9	4	10	7	5
A_R	5.57	9.92	5.49	5.51	5.2	5.95	5.14	5.33	3.87	5.22	3.95	3.92
H_E	0.593	0.541	0.5	0.527	0.529	0.571	0.455	0.602	0.439	0.657	0.562	0.6
P_{HW}	0.876	0.149	0.44	0.91	0.616	0.268	0.2	0.361	0.249	0.003	0.561	0.468
F_{IS}	-0.101	-0.004	-0.029	-0.01	-0.017	0	0.073	-0.072	0.197	0.133	-0.017	0.053
ND4												
n	30	31	27	-	32	30	37	35	18	39	38	35
H	14	11	14	-	16	12	15	20	6	9	10	6
H_R	9.19	6.64	9.33	-	9.78	8.22	7.8	11.11	5	4.2	4.52	2.57
H_D	0.892	0.815	0.9	-	0.907	0.887	0.826	0.921	0.725	0.449	0.461	0.269

Summary statistics for 17 nuclear-encoded microsatellites and mitochondrial DNA ND4 partial sequences in samples of gray triggerfish (*Balistes capriscus)*, collected from twelve localities in the Atlantic Ocean (see Figure 3.3 and 3.4). n: sample size; A: number of alleles; A_R: allelic richness; H_E: gene diversity (expected heterozygosity); P_{HW}: probability of conforming to expected Hardy-Weinberg genotypic proportions; F_{IS}: inbreeding coefficient measured as Weir and Cockerham's (1984) f; H: number of haplotypes; H_R: haplotype richness; H_D: haplotype diversity.
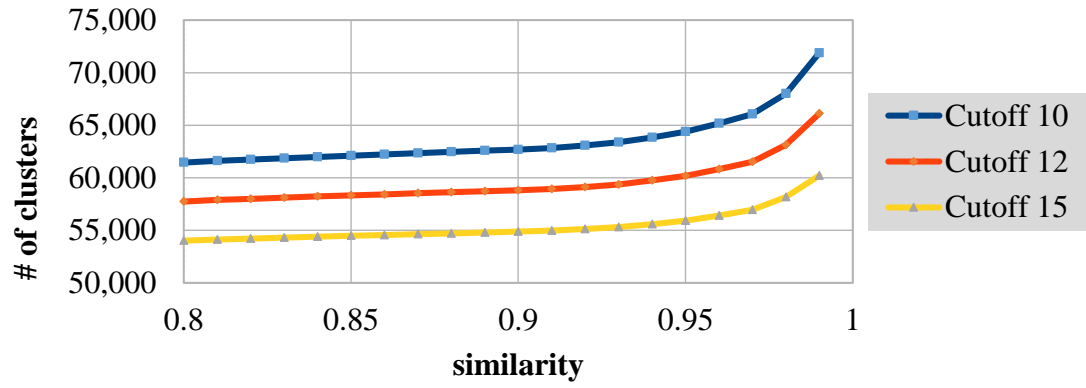
Figure D.1 Effect of Similarity and Cutoff Levels on the Number of Clusters Produced

Number of clusters obtained considering different similarity percentages for three cutoff levels, expressed as minimum frequency of unique reads.

Table D.2

Percentage of Increase in the Number of Clusters Generated at Various Similarity and Cutoff Levels

cutoff	similarity																		
	0.81	**0.82**	**0.83**	**0.84**	**0.85**	**0.86**	**0.87**	**0.88**	**0.89**	**0.9**	**0.91**	**0.92**	**0.93**	**0.94**	**0.95**	**0.96**	**0.97**	**0.98**	**0.99**
10	0.25	0.20	0.20	0.21	0.21	0.19	0.19	0.18	0.19	0.19	0.20	0.40	0.48	0.73	0.85	1.21	1.34	2.89	5.38
12	0.26	0.17	0.20	0.20	0.18	0.15	0.20	0.14	0.17	0.18	0.18	0.30	0.43	0.64	0.71	1.05	1.16	2.56	4.50
15	0.21	0.18	0.17	0.12	0.16	0.16	0.14	0.13	0.17	0.15	0.18	0.25	0.35	0.50	0.57	0.86	0.99	2.06	3.40

Percentages of increase in the number of clusters generated for each 0.01 increment in similarity and for the three cutoff thresholds considered.

Table D.3

Results of Testing for Different Percentages of Missing Data Allowed During Filtration

% missing data	shared SNPs	harmonic mean of retained samples
25	5	71.79
35	1,049	69.06
40	3,177	68.83
50	8,423	69.38

REFERENCES

Aggrey-Fynn, J. 2009. Distribution and Growth of Grey Triggerfish, *Balistes capriscus* (Family: Balistidae), in Western Gulf of Guinea. West African Journal of Applied Ecology 15:1-11.

Aiken, K. A. 1983. The biology, ecology, and bionomics of the triggerfishes, Balistidae. Pages 191-205 in J. L. Munro, editor. Caribbean coral reef fishery resources. ICLARM (International Center for Living Aquatic Resources Management) Studies and Reviews 7, Manila.

Aitken, S. N., and Whitlock M. C. 2013. Assisted gene flow to facilitate local adaptation to climate change. Annual Review of Ecology, Evolution, and Systematics 44:367-388.

Albuquerque, C. Q., A. S. Martins, N. de Oliveira Leite Junior, J. N. de Araújo, and A. M. Ribeiro 2011. Age and growth of the queen triggerfish *Balistes vetula* (tetraodontiformes, balistidae) of the central coast of Brazil. Brazilian Journal of Oceanography 59(3):231-239.

Allendorf, F. W., P. R. England, G. Luikart, P. A. Ritchie, and N. Ryman. 2008. Genetic effects of harvest on wild animal populations. Trends in Ecology & Evolution 23:327-337.

Allendorf, F. W., and G. Luikart. 2007. Conservation and the genetics of populations. Blackwell Publishing Ltd., Oxford, UK.

Anderson, E. C. 2003. User's Guide to the Program NewHybrids Version 1.1 beta. Available: http://ib.berkeley.edu/labs/slatkin/eriq/software/software.htm.

Anderson, E. C., and E. A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics 160:1217-1229.

Antoni, L., and E. Saillant. 2012. Development and characterization of Microsatellite markers in the gray triggerfish (*Balistes capriscus*). Conservation Genetic Resources 4:629-631.

Arevalo, E., S. K. Davis, and J. W. Sites Jr. 1994. Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosomes races of the *Sceloporus grammicus* complex (Phrynosomatidae) in Central Mexico. Systematic Biology 43:387-418.

Ataliba, C. C., P. M. G. de Castro, and M. H. Carneiro. 2009. Grey Triggerfish *Balistes capriscus* caught by the industrial fleet of the southeastern and southern Brazil, emphasis state of Sao Paulo. Boletim do Instituto de Pesca Sao Paulo 35(2):247-258.

Avise, J. C. 1992. Molecular Population Structure and the Biogeographic History of a Regional Fauna: A Case History with Lessons for Conservation Biology. Oikos 63(1):62-76.

Avise, J. C. 1998. Conservation genetics in the marine realm. The Journal of Heredity 89:377-382.

Avise, J. C. 2004. What is the field of biogeography, and where is it going? Taxon 53:893-898.

Avise, J. C. 2010. Perspective: conservation genetics enters the genomics era. Conservation Genetics 11:665-669.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3(10):e3376. doi:10.1371/journal.pone.0003376

Ball, A. O., M. G. Beal, and R. W. Chapman. 2007. Population structure of red porgy, *Pagrus pagrus*, in the Atlantic Ocean. Marine Biology 150:1321-1332.

Ballard, J. W., and M. C. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13(4):729-44.

Beaumariage, D. S. 1964. Returns from the 1963 Schlitz tagging program. Florida State Board of Conservation Technical Series 43:1-34.

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

Beerli, P. 2004. Effect of unsampled populations on the estimation of population sizes as migration rates between sampled populations. Molecular Ecology 13:827-836.

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

Beerli, P. 2012. Migrate user manual. Version 3.2.1. Available: http://popgen.sc.fsu.edu/migratedoc.pdf.

Beerli, P. and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152(2):763-773.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences USA 98:4563-4568.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B 57:289-300.

Bermingham, E., S. S. McCafferty, and A. P. Martin. 1997. Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. Pages 113-128 in T. D. Kocher, and C. A. Stepien, editors. Molecular systematics of fishes. Academic Press, San Diego, California.

Bernardes, R. A. 1988. Idade, Crescimento, Reprodução e Pesca do Peixe-Porco, *Balistes capriscus* (Gmelin,1788) capturado na costa sul do Estado de São Paulo (Brasil). Instituto Oceanográfico, Universidade de São Paulo.

Bernardes, R. A. 2002. Age, Growth, and longevity of the gray triggerfish, *Balistes capriscus*, from the Southeastern Brazilian Coast. Scientia Marina 66(2):167-173.

Bernardes, R. A. and J. F. Dias. 2000. Aspectos da reprodução do peixe-porco, *Balistes capriscus* (Gmelin) (Actinopetrygii, Tetraodontiformes, Balistidae) coletado na costa sul do Estado de São Paulo, Brasil. *Revista* Brasileira de Zoologia 17(3):687-696.

Bielawski, J. P., and J. R. Gold. 2002. Mutation patterns of mitochondrial H- and L-strand DNA in closely related cyprinid fishes. Genetics 161:1589-1597.

Bloor, P. A., F. S. Barker, P. C. Watts, H. A. Noyes, and S. J. Kemp. 2001. Microsatellite Libraries by Enrichment. Available: http://www.liv.ac.uk/~kempsj/genomics.html.

Bourret, V., M. P. Kent, C. R. Primmer, A. Vasemägi, S. Karlsson, K. Hindar, P. McGinnity, E. Verspoor, L. Bernatchez, and S. Lien. 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Molecular Ecology 22:532-551.

Bradbury I. R., S. Hubert, B. Higgins, T. Borza, S. Bowman, I. G. Paterson, P. V. R. Snelgrove, C. J. Morris, R. S. Gregory, D. C. Hardie, J. A. Hutchings, D. E. Ruzzante, C. T. Taggart, and P. Bentzen. 2010. Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. Proceedings of the Royal Society B: Biological Sciences 277:3725-3734.

Carson, E. W., E. Saillant, M. A. Renshaw, N. J. Cummings, and J. R. Gold. 2011. Population structure and long-term connectivity and effective size of mutton snapper (*Lutjanus analis*) in the -Caribbean Sea and Florida Keys. Fishery Bulletin 109(4):416-428.

Carvalho, G. R., and L. Hauser. 1995. Molecular genetics and the stock concept in fisheries. Pages 55-80 in G. R. Carvalho, and T. J. Pitcher, editors. Molecular genetics in fisheries. Chapman and Hall, London.

Casazza, T. L., and S. W. Ross. 2008. Fishes associated with pelagic Sargassum and open water lacking Sargassum in the Gulf Stream off North Carolina. Fishery Bulletin 106:348-363.

Castoe, T. A., A. W. Poole, A. P. J. De Koning, K. L. Jones, D. F. Tomback, S. J. Oyler-McCance, J. A. Fike, S. L. Lance, J. W. Stricher, E. N. Smith, and D. D. Pollock. 2012. Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. PLoS ONE 7:e30953. doi: 10.1371/journal.pone.0030953.

Castro, P. M. G. de. 2000. Estrutura e Dinâmica da Frota de Parelhas do Estado de São Paulo e Aspectos Biológicos dos Principais Recursos Pesqueiros Demersais Costeiros da Região Sudeste/Sul do Brasil (23°-29°S). Doctoral dissertation, Instituto Oceanográfico, Universidade de São Paulo, Brazil.

CFMC (Caribbean Fisheries Management Council). 2004. Caribbean SFA. Available: http://www.caribbeanfmc.com/fmp_sfa_amendment.html.

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetic Resources 70(2): 155-174.

Chen, C., E. Durand, F. Forbes, and O. François. 2007. Bayesian Clustering Algorithms Ascertaining Spatial Population Structure: A New Computer Program and a Comparison Study. Molecular Ecology Notes 7:747-756.

Chong, Z., J. Ruan, C.-I. Wu. 2012. Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. Bioinformatics 28:2732-2737.

Clement, M., D. Posada, and K. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Molecular Ecology 9:1657-1660.

Conover, D. O., S. A. Arnott, M. R. Walsh, and S. B. Munch. 2005. Darwinian fishery science: lessons from the Atlantic silverside. Canadian Journal of Fisheries and Aquatic Sciences 62:730-737.

Cornuet, J.-M., S. Piry, G. Luikart, A. Estoup, and M. Solignac. 1999. New Methods Employing Multilocus Genotypes to Select or Exclude Populations as Origins of Individuals. Genetics 153(4):1989-2000.

Cornuet, J.-M., F. Santos, M. A. Beaumont, C. R. Robert, J.-M. Marin, D. J. Balding, T. Guillemaud, and A. Estoup 2008. Inferring population history with DIY ABC: a user-friendly approach to Approximate Bayesian Computation. Bioinformatics 24:2713-2719.

Cowen, R. K., K. M. M. Lwiza, S. Sponaugle, C. B. Paris, and D. B. Olson. 2000. Connectivity of marine populations: open or closed? Science 287:857-859.

Cowen, R. K., C. B. Paris, and A. Srinivasan. 2006. Scaling of Connectivity in Marine Populations. Science 311:522-527.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156-2158.

Detrich, H. W. III, M. Westerfield, and L. I. Zon. 1998. Methods in Cell Biology. Volume 60. The Zebrafish: Genetics and Genomics.Academic Press, San Diego, California.

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

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

Dupanloup, I., S. Schneider, and L. Excoffier. 2002. A simulated annealing approach to define the genetic structure of populations. Molecular Ecology 11:2571-2581.

Durand, E., C. Chen, and O. François. 2009a. Tess version 2.1–reference manual. Available: http://membrestimc.imag.fr/Olivier.Francois/tess.html (July 2013).

Durand, E., F. Jay, O. E. Gaggiotti, and O. François. 2009b. Spatial inference of admixture proportions and secondary contact zones. Molecular Biology and Evolution 26(9):1963-1973.

El Mousadik, A., and R. J. Petit. 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theoretical and Applied Genetics 92:832-839.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of
    individuals using the software structure: a simulation study. Molecular Ecology
    14:2611-2620.

Excoffier, L., G. Laval, and S. Schneider. 2005. ARLEQUIN ver. 3.0: an integrated
    software package for population genetics data analysis. Evolutionary
    Bioinformatics Online 1:47-50.

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

FAO 1980. Eastern Central Atlantic Fisheries - Report of the ad hoc working group on
    Sardinella off the coast of Ivory Coast - Ghana - Togo. INT/79/019
    CECAF/ECAF/Series 80/21 (English). FAO. Available:
    http://www.fao.org/docrep/003/N8622E/n8622e00.htm#Contents. (May 2013).

FAO. 2014. FAO FishstatJ: Landings reports from 1950-2012. FAO. Available:
    http://www.fao.org/fishery/statistics/software/fishstatj/en.

Felsenstein, J. 1985. Confidence Limits on Phylogenies: An Approach Using the
    Bootstrap. Evolution 39(4): 783-791.

Fierst, J. L. 2015. Using linkage maps to correct and scaffold de novo genome
    assemblies: methods, challenges, and computational tools. Frontiers in Genetics
    6:1-8.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Clarendon, Oxford.

Foll, M., 2010. BayeScan v2.0 user manual. University of Bern Population Genetics.

    Available:

    http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.0_manual.pdf (24 May

    2016).

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci

    appropriate for both dominant and codominant markers: a bayesian perspective.

    Genetics 180:977-993.

François, O., S. Ancelet, and G. Guillot. 2006. Bayesian Clustering Using Hidden

    Markov Random Fields in Spatial Population Genetics. Genetics 174:805-816.

Frankham, R., J. D. Ballou, and D. A. Briscoe. 2002. Introduction to conservation

    genetics. Cambridge University Press, Cambridge, UK.

Franks, J. S., E. R. Hoffmayer, B. H. Comyns, J. R. Hendon, E. M. Blake, and D. P.

    Gibson. 2007. Investigations of fishes that utilize pelagic Sargassum and frontal

    zone habitats in Mississippi marine waters and the northcentral Gulf of Mexico.

    Final Report. Gulf Coast Research Laboratory, Ocean Springs, Mississippi and

    U.S. Fish and Wildlife Service, Atlanta, Georgia

Franks, J. S., D. R. Johnson, and D. S. Ko. 2016. Pelagic *Sargassum* in the Tropical

    North Atlantic. Gulf and Caribbean Research 27(1):SC6-SC11.

Funk, W. C., J. K. McKay, P. A. Hohenlohe, and F. W. Allendorf. 2012. Harnessing

    genomics for delineating conservation units. Trends in ecology & evolution

    27(9):489-496.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150-3152.

Garcìa, C. B., L. Orlando-Duarte, and D. von Schiller. 1998. Demersal fish assemblages of the Gulf of Salamanca, Colombia (southern Caribbean Sea). Marine Ecology Progress Series 174:13-25.

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. ArXiv:1207.3907 [q-Bio.GN]. Available: https://arxiv.org/abs/1207.3907.

Gascoigne, J., and R. N. Lipcius. 2004. Allee effects in marine systems. Marine Ecology Progress Series 269:49-59.

Gaylord, B., and S. D. Gaines. 1999. Temperature or Transport? Range Limits in Marine Species-Mediated Solely by Flow. The American Naturalist 155(6):769-789.

Gold, J. R., E. Park, and L. R. Richardson. 2001. Microsatellite Variation Among Red Snapper (*Lutjanus campechanus*) from the Gulf of Mexico. Marine Biotechnology 3:293-304.

Gold, J. R., and L. R. Richardson. 1998. Genetic Homogeneity among Geographic Samples of Snappers and Groupers: Evidence of Continuos Gene Flow? Gulf and Caribbean Fisheries Institute 50:709-726.

Gold, J. R., E. Saillant, N. J. Cummings, and M. A. Renshaw. 2011. Genetic Divergence and Effective Size among Lane Snapper in U.S. Waters of the Western Atlantic Ocean. North American Journal of Fisheries Management 31:209-223.

Gold, J. R., E. Saillant, N. D. Ebelt, and S. Lem. 2009. Conservation Genetics of Gray
Snapper (*Lutjanus griseus*) in U.S. Waters of the Northern Gulf of Mexico and
Western Atlantic Ocean. Copeia 2:277-286.

Goldstein, D. B., A. R. Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995. An
evaluation of genetic distances for use with microsatellite loci. Genetics 139:463-
471.

Goudet, J. 1995. F-STAT (Version 1.2): A computer program to calculate F-Statistics.
The Journal of Heredity 86:485-486.

Gower, J., and S. King. 2008. Satellite Images Show the Movement of -Floating
Sargassum in the Gulf of Mexico and Atlantic Ocean. Nature Proceedings:
hdl:10101/npre.2008.1894.1.

Gower, J., and S. King. 2011. Distribution of floating Sargassum in the Gulf of Mexico
and Atlantic Ocean mapped using MERIS. International Journal of Remote
Sensing 32:1917-1929.

Gower, J., E. Young, and S. King. 2013. Satellite images suggest a new *Sargassum*
source region in 2011. Remote Sensing Letters 4(8):764-773.

Grewe, P. M., P. Feutry, P. L. Hill, R. M. Gunasekera, K. M. Schaefer, D. G. Itano, D.
W. Fuller, S. D. Foster, and C. R. Davies. 2015. Evidence of discrete yellowfin
tuna (*Thunnus albacares*) populations demands rethink of management for this
globally important resource. Scientific Reports 5:16916.

Gyory, J., B. Bischof, A. J. Mariano, and E. H. Ryan. 2005. "The Guinea Current." Ocean Surface Currents. Available: http://oceancurrents.rsmas.miami.edu/atlantic/guinea.html.

Haldane, J. B. S.1930. A mathematical theory of natural and artificial selection. Part VI. Isolation. Proceedings of the Cambridge Philosophical Society 26:220-230.

Hansen M. M., and J. Hemmer-Hansen. 2007. Landscape genetics goes to sea. Journal of Biology 6:1-3.

Hardy, O. J., N. Charbonnel, H. Fréville, and M. Heuertz. 2003. Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163:1467-1482.

Hardy, O. J., and X. Vekemans. 1999. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. Heredity 83:145-154.

Hare, M. P., L. Nunney, M. K. Schwartz, D. E. Ruzzante, M. Burford, R. S. Waples, K. Ruegg, and F. Palstraa. 2011. Understanding and estimating effective population size for practical application in marine species management. Conservation Biology 25:438-449.

Harmelin-Vivien, M. L. and J.-C. Quéro, 1990. Balistidae. Pages 1055-1060. In J. C. Quero, J. C. Hureau, C. Karrer, A. Post, and L. Saldanha, editors. Check-list of the fishes of the eastern tropical Atlantic (CLOFETA). JNICT, Lisbon; SEI, Paris; and UNESCO, Paris. Vol. 2.

Harrison, H. B., D. H. Williamson, R. D. Evans, G. R. Almany, S. R. Thorrold, G. R. Russ, K. A. Feldheim, L. van Herwerden, S. Planes, M. Srinivasan, M. L. Berumen, and G. P. Jones. 2012. Larval export from marine reserves and the recruitment benefit for fish and fisheries. Current Biology 22:1023-1028.

Hastings, A. 1993. Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. Ecology 74:1362-1372.

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

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

Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, M. I. Taylor, R. Ogden, M. T. Limborg, A. Cariani, G. E. Maes, E. Diopere, G. R. Carvalho, and E. E. Nielsen. 2011. Application of SNPs for population genetics of non-model organisms: new opportunities and challenges. Molecular Ecology Resources 11(1):123-136.

Hey, J., and R. Nielsen. 2004. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747-760.

Heyman, W. D., S. Kobara, S. J. Pittman, and R. S. Nemeth. 2013. Caribbean Reef Fish Spawning Aggregations: Biogeography, Future Research, and Management Needs. Proceedings of the Gulf and Caribbean Fisheries Institute 66:401-407.

Hill, W. G. 1981. Estimation of effective population size from data on linkage disequilibrium. Genetics Research 38:209-216.

Hoffmayer, E. R., J. S. Franks, B. H. Comyns, J. R. Hendon, and R. S. Waller. 2005. Larval and Juveniles Fishes Associated with Pelagic *Sargassum* in the Northcentral Gulf of Mexico. Proceedings of the Gulf and Caribbean Fisheries Institute 56:259-269.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics 6:e1000862.

Hollenbeck, C. M., D. S. Portnoy, E. Saillant, and J. R. Gold. 2015. Population structure of red snapper (*Lutjanus campechanus*) in U.S. waters of the western Atlantic Ocean and the northeastern Gulf of Mexico. Fisheries Research 172:17-25.

Hood, G. M. 2010. PopTools version 3.2.5. Available: http://www.poptools.org (09 January 2016).

Houde, A. L. S., D. J. Fraser, P. O'Reilly, and J. A. Hutchings. 2011. Relative risks of inbreeding and outbreeding depression in the wild in endangered salmon. Evolutionary Applications 4(5):634-647.

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

IBAMA (Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais) 2006.

Estatisticas Pesqueiras (Fishery Statistics). Available:

http://www.ibama.gov.br/documentos-recursos-pesqueiros/estatistica-pesqueira.

IGFA (International Game Fish Association). 1991 .World record game fishes. IGFA,

Fort Lauderdale, Florida, USA.

IGFA (International Game Fish Association). 2001. Database of IGFA angling records

until 2001. IGFA, Fort Lauderdale, Florida, USA.

Ingram, G. W. Jr. 2001. Stock structure of gray triggerfish, *Balistes capriscus*, on

multiple spatial scales in the Gulf of Mexico. Doctoral dissertation. University of

South Alabama, Mobile.

Ismen, A., M. Turkog, and C. C. Yigin. 2004. The Age, Growth, and Reproduction of

Gray Triggerfish (*Balistes capriscus*, Gmelin, 1789) in Iskenderum Bay. Pakistan

Journal of Biological Sciences. 7(12):2135-2138.

Jarne, P., and P. J. L Lagoda. 1996. Microsatellites, from molecules to populations and

back. Trends in Ecology and Evolution 11:424-429.

John, J. St., and T. W. Quinn. 2008. Rapid capture of DNA targets. Bio Techniques

44:259-264.

Johnson, D. R., D. S. Ko, J. S. Franks, P. Moreno, and G. Sanchez-Rubio. 2013. The

Sargassum invasion of the Eastern Caribbean and dynamics of the equatorial

north Atlantic. Proceedings of the Gulf and Caribbean Fisheries Institute 65:102-

103.

Johnson, D. R., H. M. Perry, J. Lyczkowski-Shultz, and D. Hanisko. 2009. Red Snapper Larval Transport in the Northern Gulf of Mexico. Transactions of the American Fisheries Society 138:458-470.

Jones, G. P., M. J. Milicich, M. J. Emslie, and C. Lunow. 2009. Self-recruitment in a coral reef fish population. Nature 402:802-804.

Karlsson, S., E. Saillant, and J. R. Gold. 2009. Population structure and genetic variation of lane snapper (*Lutjanus synagris*) in the northern Gulf of Mexico. Marine Biology 156:1841-1855.

Kijas, J. M., J. C. Fowler, C. A. Garbett, and M. R. Thomas. 1994. Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Biotechniques 16(4):656-662.

Kimura, M., and G. Weiss. 1964. The stepping-stone model of population structure and the decrease of genetic correlation with distance. Genetics 49(4):561-576.

Kingman, J. F. C. 1982. The coalescent. Stochastic Processes and their Applications 13:235-248.

Koskinen, M. T., H. Hirvonen, P.-A. Landry, and C. R. Primmer. 2004. The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. Hereditas 141:61-67.

Kritzer, J. P., and P. F. Sale. 2004. Metapopulation ecology in the sea: from Levins' model to marine ecology and fisheries science. Fish and Fisheries 5:131-140.

Laconcha, U., M. Iriondo, H. Arrizabalaga, C. Manzano, P. Markaide, I. Montes, I. Zarraonaindia, I. Velado, E. Bilbao, N. Goñi, J. Santiago, A. Domingo, S. Karakulak I. K., Oray, and A. Estonba. 2015. New nuclear SNP markers unravel the genetic structure and effective population size of albacore tuna (*Thunnus alalunga*). PLoS ONE 10(6):e0128247. doi:10.1371/journal.pone.0128247.

Lamichhaney, S., A. Martinez Barrio, N. Rafati, G. Sundström, C.-J. Rubin, E. R. Gilbert, J. Berglund, A. Wetterbom, L. Laikre, M. T. Webster, M. Grabherr, N. Ryman, and L. Andersson. 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. Proceedings of the National Academy of Sciences USA 109:19345-19350.

Larson, W. A., L. W. Seeb, M. V. Everett, R. K. Waples, W. D. Templin, *and* J. E. Seeb. 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evolutionary Applications 7:355-369.

Latter, B. D. H. 1972. Selection in finite populations with multiple alleles. III. Genetic divergence with centripetal selection and mutation. Genetics 70:475-490.

Leblois, R., A. Estoup, and F. Rousset. 2003. Influence of Mutational and Sampling Factors on the Estimation of Demographic Parameters in a ''Continuous'' Population Under Isolation by Distance. Molecular Biology and Evolution 20(4):491-502.

Leblois, R., A. Estoup, and F. Rousset. 2009. IBDSim: a computer program to simulate genotypic data under isolation by distance. Molecular Ecology Resources 9:107-109.

Leblois, R., P. Pudlo, J. Néron, F. Bertaux, C. R. Beeravolu, R. Vitalis, and F. Rousset. 2014. Maximum-Likelihood Inference of Population Size Contractions from Microsatellite Data. Molecular Biology and Evolution 31:2805-2823.

Leblois, R., F. Rousset, and A. Estoup. 2004. Influence of spatial and temporal heterogeneities on the estimation of demographic parameters in a continuous population using individual microsatellite data. Genetics 166:1081-1092.

Lee, T. N., W. E. Johns, R. Zantopp, and E. R. Fillenbaum. 1996. Moored observations of western boundary current variability and thermohaline circulation 26.5°N in the subtropical North Atlantic. Journal of Physical Oceanography 26:962-963.

Leis, J. M. 1991. The pelagic stage of reef fishes: the larval biology of coral reef fishes. Pages 183-230 in P. F. Sale, editor. The ecology of fishes on coral reefs. Academic Press, San Diego, California.

Leis, J. M. 2002. Pacific coral-reef fishes: the implications of behavior and ecology of larvae for biodiversity and conservation, and a reassessment of the open population paradigm. Environmental Biology of Fishes 65:199-208.

Leis, J. M., and M. I. McCormick. 2002. The biology, behavior, and ecology of the pelagic larval stage of coral reef fishes. Pages 171-199 in P. F. Sale, editor. Coral Reef Fishes: Dynamics and Diversity in a Complex Ecosystem. Academic Press, San Diego, California, USA.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends in Ecology and Evolution 17(4):183-189.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America 15:237-240.

Levins, R. 1970. Extinction. Pages 77-107 in M. Desternhaber, editor. Some mathematical problems in biology. American Mathematical Society, Providence, Rhode Island.

Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-1659.

Limborg, M. T., S. J. Helyar, M. De Bruyn, M. I. Taylor, E. E. Nielsen, R. Ogden, G. R. Carvalho, F. P. T. Consortium, and D. Bekkevold. 2012. Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). Molecular Ecology 21:3686-3703.

Lindeman, K. C., R. Pugliese, G. T. Waugh, and J. S. Ault. 2000. Developmental patterns within a multispecies reed fishery: management applications for essential fish habitats and protected areas. Bulletin of Marine Science 66(3):929-956.

Lischer, H. E. L., and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299.

Liu, J., G. Zapfe, K.-T. Shao, J. L. Leis, K. Matsuura, G. Hardy, M. Liu, and J. Tyler. 2015. *Balistes capriscus*. The IUCN Red List of Threatened Species. Version 2015-4. Available: http:www.iucnredlist.org (08 January 2016).

Liu, J., G. Zapfe, K.-T. Shao, J. L. Leis, K. Matsuura, G. Hardy, M. Liu, and J. Tyler. 2016. *Balistes vetula*. The IUCN Red List of Threatened Species. Version 2016-2. Available: http:www.iucnredlist.org (08 November 2016).

Lo Brutto, S., M. Arculeo, and N. Parrinello. 2004. Congruence in genetic markers used to describe Mediterranean and Atlantic populations of European hake (*Merluccius merluccius* L. 1758). Journal of Applied Ichthyology 20:81-86.

Lopez-Pena, A., and L. Orlando-Duarte. 2012. Towards an identification of essential habitats in tropical coastal ecosystems: the case of demersal fishes in the Colombian Caribbean Sea. Proceedings of the Gulf and Caribbean Fisheries Institute 65.

Lotterhos, K. E., and M. C. Whitlock. 2014. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Molecular Ecology 23:2178-2192.

Lowe, W. H., and F. W. Allendorf. 2010. What can genetics tell us about population connectivity? Molecular Ecology 19:3038-3051.

MacKichan, C. A., and S. T. Szedlmayer 2007. Reproductive behavior of gray triggerfish, *Balistes capriscus*, in the Northeastern Gulf of Mexico. Proceedings of the Gulf and Caribbean Fisheries Institute 59:231-235.

Malausa, T., A. Gilles, E. Meglécz, H. Blanquart, S. Duthoy, C. Costedoat, V. Dubut, N. Pech, P. Castagnone-Sereno, C. Délve, N. Feau, P. Frey, P. Gauthier, T. Guillemaud, L. Hazard, V. Le Corre, B. Lung-Escarmant, P.-J. G. Malé, S. Ferreira, and J.-F. Martin. 2011. High throughput microsatellites isolation with 454 GS-FLX Titanium pyrosequencing. Molecular Ecology Resources 11:638-644.

Manni, F., E. Guerard, and E. Heyer. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". Human Biology 76(2):173-190.

Manooch, C. S., and C. L. Drennon. 1987. Age and Growth of Yellowtail Snapper and Queen Triggerfish Collected from the U.S. Virgin Islands and Puerto Rico. Fisheries Research 6:53-68.

Mariano, A. J., E. H. Ryan, B. D. Perkins, and S. Smithers. 1995. The Mariano Global Surface Velocity Analysis 1.0. USCG Report CG-D-34-95.

Mejía, L. S., and A. Acero, editors. 2002. Libro rojo de peces marinos de Colombia. INVEMAR, Instituto de Ciencias Naturales Universidad Nacional de Colombia, Ministerio del Medio Ambiente. La serie Libros Rojos de Especies Amenazadas de Colombia. Bogotá, Colombia.

Mensah, M. A., and S. N. K. Quaatey. 2002. An Overview of Fishery Resources and Fishery Research in the Gulf of Guinea. Pages 227-239 in J. M. McGlade, P. Cury, K. A. Koranteng, and N. J. Hardman-Mountford, editors. The Gulf of

Guinea Large Marine Ecosystem: Environmental Forcing and Sustainable Development of Marine Resources. Elsevier, Amsterdam.

Miesfeld, R., M. Krystal, and N. Arnheim. 1981. A member of a new repeated sequence family which is conserved throughout eukaryotic evolution is found between the human delta and beta globin genes. Nucleic Acids Research 9:5931-5947.

Monmonier, M. 1973. Maximum-difference barriers: an alternative numerical regionalization method. Geographical Analysis 3:245-261.

Mora, C. 2004. Importance of dispersal in coral reef fishes. Doctoral dissertation. University of Windsor, Windsor, Canada.

Munro, J. L., V. C. Gaut, R. Thompson and P. H. Reeson. 1973. The spawning seasons of Caribbean reef fishes. Journal of Fish Biology 5:69-84.

Nachman, M. W., S. N. Boyer, and C. F. Aquadro. 1994. Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proceedings of the National Academy of Sciences 91:6364-6368.

Naciri, M., C. Lemaire, P. Borsa, and F. Bonhomme. 1999. Genetic Study of the Atlantic/Mediterranean Transition in Sea Bass (*Dicentrarchus labrax*). The Journal of Heredity 90(6):591-596.

Nagylaki, T. 1992. Introduction to Theoretical Population Genetics. Springer-Verlag, Berlin, Germany.

Neel, M. C., K. McKelvey, N. Ryman, M. W. Lloyd, R. Short Bull, F. W. Allendorf, M. K. Schwartz, and R. S. Waples. 2013. Estimation of effective population size in

continuously distributed populations: there goes the neighborhood. Heredity 111:189-199.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia University Press, New York, NY, USA.

Netto, R. F., and A. P. M. Di Beneditto. 2010. Growth, mortality and exploitation rates of the Queen Triggerfish *Balistes vetula* (Tetraodontiformes: Balistidae) in the Brazilian East coast. Cahiers de Biologie Marine 51:93-99.

Nichols, S. 2004. Derivation of red snapper time series from SEAMAP and groundfish trawl surveys. SEDAR-7-DW-01. NMFS, Pascagoula, Mississippi.

Nielsen, E. E. G., L. Arvebach, and P. Kotlicki. 2006. Hybridlab (version 1.0): a program for generating simulated hybrids from population samples. Molecular Ecology Notes 6:971-973.

Nielsen, E. E., J. Hemmer-Hansen, N. A. Poulsen, V. Loeschcke, T. Moen, T. Johansen, C. Mittelholzer, G.-L. Taranger, R. Ogden, and G. R. Carvalho. 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). BMC Evolutionary Biology 9(276):1-11.

Norrell, A. E. 2016. Development of genomic resources for the evaluation of red snapper, an emerging species candidate for marine aquaculture and stock enhancement. Master's thesis. The University of Southern Mississippi, Ocean Springs, Mississippi.

NMFS (National Marine Fisheries Service). 2006. Stock Assessment Report of SEDAR-9: Gulf of Mexico Gray Triggerfish. Charleston, South Carolina.

NMFS (National Marine Fisheries Service). 2011. SEDAR 9 Update Stock Assessment
   Report Gulf of Mexico Gray Triggerfish. Tampa, Florida.

Nunney, L., and D. R. Elam. 1994. Estimating the effective population size of conserved
   populations. Conservation Biology 8:175-184.

Ofori-Danson, P.K. 1990. Reproductive ecology of the triggerfish, *Balistes capriscus*,
   from the Gahanian coastal waters. Tropical Ecology 31(1):1-11.

Oliveira, E. J., J. G. Pádua, M. I. Zucchi, R. Vencovsky, and M. L. C. Vieira. 2006.
   Origin, evolution and genome distribution of microsatellites. Genetics and
   Molecular Biology 29(2):294-307.

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

Pampoulie, C., D. E. Ruzzante, V. Chosson, T. D. Jörundsdóttir, L. Taylor, V.
   Thorsteinsson, A. K. Daníelsdóttir, and G. Marteinsdóttir. 2006. The genetic
   structure of Atlantic cod (*Gadus morhua*) around Iceland: insight from
   microsatellites, the Pan I locus, and tagging experiments. Canadian Journal of
   Fisheries and Aquatic Sciences 63:2660-2674.

Pankhurst, N. W., and P. L. Munday. 2011. Effects of climate change on fish
   reproduction and early life history stages. Marine and Freshwater Research
   62:1015-1026.

Parr, A. D. 1939. Quantitative observations on the pelagic Sargassum vegetation of the western North Atlantic. Bulletin of the Bingham Oceanographic Collection 6:1-94.

Payseur, B. A., and P. Jing. 2009. A Genomewide Comparison of Population Structure at STRPs and Nearby SNPs in Humans. Molecular Biology and Evolution 26(6):1369-1377.

Peakall, R., and P. E. Smouse. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

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

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, H. E. Hoekstra. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7(5):e37135. doi: 10.1371/journal.pone.0037135.

Petit, R. J., A. El-Mousadik, and O. Pons 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12:844-855.

Portnoy, D. S., and J. R. Gold. 2012. Evidence of multiple vicariance in a marine suture-zone in the Gulf of Mexico. Journal of Biogeography 39:1499-1507.

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

Pritchard, J. K., X. Wen, and D. Falush. 2010. Documentation for STRUCTURE software. Available: http://pritch.bsd.uchicago.edu/structure.html.

Pritchard, V. L., J. Erkinaro, M. P. Kent, E. Niemelä, P. Orell, S. Lien, and C. P. Primmer. 2016. Single nucleotide polymorphisms to discriminate different classes of hybrid between wild Atlantic salmon and aquaculture escapees. Evolutionary Applications 9(8):1017-1031.

Pruett, C. L., E. Saillant, and J. R. Gold. 2005. Historical population demography of red snapper (*Lutjanus campechanus*) from the northern Gulf of Mexico based on analysis of sequences of mitochondrial DNA. Marine Biology 147:593-602.

Prugnolle, F., and T. De Meeus 2002. Inferring Sex-biased dispersal from population genetic tools: a review. Heredity 88:161-165.

Pudovkin, A. I., D. V. Zaykin, and D. Hedgecock. 1996. On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. Genetics 144:383-387.

Puebla, O., E. Bermingham, and F. Guichard. 2009. Estimating dispersal from genetic isolation by distance in a coral reef fish (*Hypoplectrus puella*). Ecology 90:3087-3098.

Puebla, O., E. Bermingham, and O. McMillan. 2012. On the spatial scale of dispersal in coral reef fishes. Molecular Ecology 21:5675-5688.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: a tool set

for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81:559-575.

Puritz, J. B., C. M. Hollenbeck, and J. R. Gold. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431. doi:10.7717/peerj.431.

Quenouille, M. H. 1949. Problems in Plane Sampling. The Annals of Mathematical Statistics 20(3):355-375.

Quenouille, M. H. 1956. Notes on Bias in Estimation. Biometrika 43(3-4):353-360.

Quesada, H., C. M. Beynon, and D. O. F. Skibinski. 1995. A Mitochondrial DNA Discontinuity in the Mussel *Mytilus galloprovincialis* Lmk: Pleistocene Vicariance Biogeography and Secondary Intergradation. Molecular Biology and Evolution 12(3):521-524.

Raj, A., M. Stephens, and J. K. Pritchard. 2014. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Datasets. Genetics 197(2):573-89.

Rand, D. M. 2001. The Units of Selection on Mitochondrial DNA. Annual Review of Ecology and Systematics 32:415-448.

Randall, J. E. 1968. Caribbean reef fishes. T. F. H. Publications, Hong Kong.

Rankin, T. L., and S. Sponaugle. 2011. Temperature Influences Selective Mortality during the Early Life Stages of a Coral Reef Fish. PLoS ONE 6(5):e16814. doi:10.1371/journal.pone.0016814.

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

Renshaw, M. A., C. M. Hollenbeck, and J. R. Gold. 2012. Isolation of microsatellite markers from red drum, *Sciaenops ocellatus*, and characterization in red drum and spotted seatrout, *Cynoscion nebulosus*. Molecular Ecology Resources 12:570-572.

Renshaw, M. A., E. Saillant, S. C. Bradfield, and J. R. Gold. 2006. Microsatellite multiplex panels for genetic studies of three species of marine fishes: red drum (*Sciaenops ocellatus*), red snapper (*Lutjanus campechanus*), and cobia (*Rachycentron canadum*). Aquaculture 253:731-735.

Richards, W. J., and K. C. Lindeman. 1987. Recruitment dynamics of reef fishes: planktonic processes, settlement and demersal ecologies, and fishery analysis. Bulletin of Marine Science 41:392-410.

Rieman, B. E., and F. W. Allendorf. 2001. Effective population size and genetic conservation criteria for bull trout. Journal of North American Fisheries Management 21:756-764.

Roberts, C. 1996. *Balistes vetula*. IUCN Red List of Threatened Species. Version 2011-1. Available: www.iucnredlist.org.

Roberts, C. M. 1997. Connectivity and Management of Caribbean Coral Reefs. Science 278:1454-1457.

Robertson, D. R. 1988. Extreme variation in settlement of the Caribbean triggerfish, *Balistes vetula*, in Panama. Copeia 3:698-703.

Robins, C. R., and G. C. Ray. 1986. A field guide to Atlantic coast fishes of North

America. Houghton Mifflin Company, Boston, U.S.A. 354 p.

Rocha, L. A., A. L. Bass, D. R. Robertson, and B. W. Bowen. 2002. Adult habitat

preferences, larval dispersal, and the comparative phylogeography of three

Atlantic surgeonfishes (Teleostei: Acanthuridae). Molecular Ecology 11:243-252.

Rossby, T., A. S. Bower, and P.-T. Shaw. 1985. Particle Pathways in the Gulf Stream.

Bulletin American Meteorological Society 66(9):1106-1110.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics

under isolation by distance. Genetics 145:1219-1228.

Rousset, F. 2000. Genetic differentiation between individuals. Journal of Evolutionary

Biology 13:58-62.

Rousset, F. 2008a. Genepop'007: a complete reimplementation of the Genepop software

for Windows and Linux. Molecular Ecology Resources 8:103-106.

Rousset, F. 2008b. Dispersal estimation: Demystifying Moran's I. Heredity 100:231-232.

Rousset, F., and R. Leblois. 2007. Likelihood and approximate likelihood analyses of

genetic structure in a linear habitat: performance and robustness to model

misspecification. Molecular Biology and Evolution 24:2730-2745.

Rousset, F., and R. Leblois. 2012. Likelihood-based inferences under a coalescent model

of isolation by distance: two-dimensional habitats and confidence intervals.

Molecular Biology and Evolution 29(3):957-973.

Rousset, F., and M. Raymond. 1995. Testing heterozygote excess and deficiency.

Genetics 140:1413-1419.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and others methods. Bioinformatics 19:2496-2497.

Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. Pages 365-386 in S. Krawetz, and S. Misener, editors. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, New Jersey.

Ruzzante, D. E. 1998. A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. Canadian Journal of Fisheries and Aquatic Sciences 55:1-14.

Sá-Gabriel, L. G., and W. F. Molina. 2005. Karyotype diversification in fishes of the Balistidae, Diodontidae and Tetraodontidae (Tetraodontiformes). Caryologia 58:229-237.

Saillant, E., S. C. Bradfield, and J. R. Gold. 2010. Genetic variation and spatial autocorrelation among young-of-the-year red snapper (*Lutjanus campechanus*) in the northern Gulf of Mexico. ICES Journal of Marine Science 67:1240-1250.

Saillant, E., and J. R. Gold. 2006. Population structure and variance effective size of red snapper (*Lutjanus campechanus*) in the northern Gulf of Mexico. Fishery Bulletin 104:136-148.

Saillant, E., M. A. Renshaw, N. J. Cummings, and J. R. Gold. 2012. Conservation genetics and management of yellowtail snapper (*Ocyurus chrysurus*) in the U.S. Caribbean and South Florida. Fisheries Management and Ecology 12:301-312.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for

    reconstruction of phylogenetic trees. Molecular Biology and Evolution 4:406-425.

Sale, P. F., I. Hanski, and J. P. Kritzer. 2006. Pages 3-28 in J. Kritzer, and P. F. Sale,

    editors. Marine Metapopulations. Academic, San Diego, California.

Sambrook, J. E., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory

    manual, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor,

    New York.

Sazonov, Y. G., and A. I. Galaktionova. 1987. Some data on the morphometrics of the

    grey triggerfish, *Balistes carolinensis*, of the central-eastern Atlantic. Journal of

    Ichthyology 27:173-176.

Schweyen, H., A. Rozenberg, and F. Leese. 2014. Detection and removal of PCR

    duplicates in population genomic ddRAD studies by addition of a degenerate base

    region (DBR) in sequencing adapters. Biological Bulletin 227:146-160.

Scott, W. B., and M. G. Scott. 1988. Atlantic fishes of Canada. Canadian Bulletin of

    Fisheries and Aquatic Sciences 219. University of Toronto Press, Toronto,

    Canada.

SEDAR-8 (Southeast Data, Assessment, and Review). 2005. Caribbean yellowtail

    snapper, *Ocyurus chrysurus*. (a) Stock assessment report, (b) Stock assessment

    workshop report, and (c) Consensus summary report. SEDAR/NMFS, Report 1,

    Charleston, South Carolina.

SEDAR-9 (Southeast Data, Assessment, and Review). 2006. Stock Assessment Report of
    SEDAR-9: Gulf of Mexico Gray Triggerfish. Report 1, Charleston, South
    Carolina.

SEDAR-30 (Southeast Data, Assessment, and Review). 2013. U.S. Caribbean Queen
    Triggerfish. SEDAR, North Charleston, South Carolina.

SEDAR-43 (Southeast Data, Assessment, and Review). 2015. Gulf of Mexico Gray
    Triggerfish. SEDAR, North Charleston, South Carolina.

Shanks, A. L. 2009. Pelagic Larval Duration and Dispersal Distance Revisited. Biological
    Bulletin 216:373-385.

Shulman, M. J., and E. Bermingham. 1995. Early life histories, ocean currents, and the
    population genetics of Caribbean reef fishes. Evolution 49:897-910.

Siegel, D. A., B. P. Kinlan, B. Gaylord, and S. D. Gaines. 2003. Lagrangian descriptions
    of marine larval dispersion. Marine Ecology Progress Series 260:83-96.

Simmons, C. M. 2008. Gray triggerfish, *Balistes capriscus*, reproductive behavior, early
    life history, and competitive interactions between red snapper, *Lutjanus
    campechanus*, in the northern Gulf of Mexico. Doctoral dissertation. Auburn
    University, Auburn, Alabama.

Simmons, C. M., and S. T. Szedlmeyer. 2011. Recruitment of Age-0 Gray Triggerfish to
    Benthic Structured Habitat in the Northern Gulf of Mexico. Transactions of the
    American Fisheries Society 140:14-20.

Slatkin, M. 2005. Seeing ghosts: the effect of unsampled populations on migration rates
    estimated for sampled populations. Molecular Ecology 14:67-73.

Slatkin, M. 2008. Linkage disequilibrium - understanding the evolutionary past and
mapping the medical future. Nature Reviews Genetics 9:477-485.

Smouse, P. E., and R. Peakall. 1999. Spatial autocorrelation analysis of individual
multiallele and multilocus genetic structure. Heredity 82:561-573.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian
measures of model complexity and fit (with discussion). Journal of the Royal
Statistical Society B 64:191-232.

Storz, J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive
population divergence. Molecular Ecology 14: 671-688.

Stromme, T. 1984. Report on the R/V Dr. Fridtjof Nansen Fish Resource Surveys off
West Africa: Morocco to Ghana and Cape Verde. CECAF/ECAF SERIES 84/29.

Sunnucks, P. 2000. Efficient genetic markers for population biology. Trends in Ecology
& Evolution 15(5):199-203.

Swain, D. P., and C. J. Foote. 1999. Stocks and chameleons: the use of phenotypic
variation in stock identification. Fisheries Research 43:1123-1128.

Swearer, S. E., J. E. Caselle, D. W. Lea, and R. R. Warner. 1999. Larval retention and
recruitment in an island population of a coral-reef fish. Nature 402:799-802.

Takezaki, N., M. Nei, and K. Tamura. 2010. POPTREE2: Software for constructing
population trees from allele frequency data and computing other population
statistics with windows interface. Molecular Biology and Evolution 27:747-752

Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Molecular Ecology 7:381-397.

Thorrold, S. R., G. P. Jones, M. E. Hellberg, R. S. Burton, S. E. Swearer, J. E. Neigel, S. G. Morgan, and R. R. Warner. 2002. Quantifying larval retention and connectivity in marine populations with artificial and natural markers. Bulletin of Marine Science 70(1):291-308.

Thorrold, S. R., C. M. Jones, P. K. Swart, and T. E. Targett. 1998. Accurate classification of nursery areas of juvenile weakfish (*Cynoscion regalis*) based on chemical signatures in otoliths. Marine Ecology Progress Series 173:253-265.

Turner, T. F., J. P. Wares, and J. R. Gold. 2002. Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). Genetics 162:1329-1339.

Vähä, J.-P., and C. R. Primmer. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Molecular Ecology 15:63-72.

Valle, M., C. M. Legault, and M. Ortiz. 2001. A stock assessment for gray triggerfish, *Balistes capriscus*, in the Gulf of Mexico. National Oceanic and Atmospheric Administration–Fisheries, Sustainable Fisheries Division Contribution SFD-00/01-124, Miami, Florida.

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

Vasconcellos, A. V., P. Vianna, P. C. Paiva, R. Schama, and S. Sole`-Cava. 2008. Genetic and morphometric differences between yellowtail snapper (*Ocyurus chrysurus*, Lutjanidae) populations of the tropical West Atlantic. Genetics and Molecular Biology 31:308-316.

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

Wang, J. L., and M. C. Whitlock. 2003. Estimating effective population size and migration rates from genetic samples over space and time. Genetics 163:429-446.

Waples, R. S. 1995. Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. Pages 8-27 in J. L. Nielson, editor. Evolution and the aquatic ecosystem: defining units in population conservation. American Fisheries Society, Monterey, California.

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

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics 7:167-184.

Waples R. S. 2015. Testing for Hardy–Weinberg proportions: have we lost the plot? Journal of Heredity 106:1-19.

Waples, R. S., and C. Do. 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evolutionary Applications 3:244-262.

Waples, R. S., and O. Gaggiotti. 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 15:1419-1439.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theoretical Population Biology 7(2):256-276.

Watts, P. C., F. Rousset, I. J. Saccheri, R. Leblois, S. J. Kemp, and D. J. Thompson. 2007. Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion mercuriale*: Odonata: Zygoptera) populations: analysis of 'neighborhood size' using a more precise estimator. Molecular Ecology 16:737-751.

Wegmann, D., C. Leuenberger, S. Neuenschwander, and L. Excoffier. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. BMC Bioinformatics 11:116.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358-1370.

Weis, J. S. 1968. Fauna Associated with Pelagic *Sargassum* in the Gulf Stream. American Midland Naturalist 80(2):554-558.

Wells, R. J. D., and J. R. Rooker. 2004. Spatial and temporal patterns of habitat use by fishes associated with *Sargassum* mats in the northwestern Gulf of Mexico. Bulletin of Marine Science 74:81-99.

Whitlock, M. C., and D. E. McCauley. 1999. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. Heredity 82:117-125.

Whitlock, M. C., and K. E. Lotterhos. 2015. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F_{ST}. The American Naturalist 186(S1):S24-S36.

Woodstock, A. H. 1950. Subsurface pelagic *Sargassum*. Journal of Marine Resources 9:77-92.

Wright, S. 1931. Evolution in Mendelian population. Genetics 16:97-159.

Wright, S. 1943. Isolation by distance. Genetics 28:114-138.

Wright, J. M., and P. Bentzen. 1994. Microsatellites: genetic markers for the future. Reviews in Fish Biology and Fisheries 4:384-388.

Zane, L., L. Bargelloni, and T. Patarnello. 2002. Strategies for microsatellite isolation: a review. Molecular Ecology 11:1-16.