POPULATION GENOMICS ANALYSIS OF YELLOWFIN TUNA *THUNNUS ALBACARES* OFF SOUTH AFRICA REVEALS NEED FOR A SHIFTED MANAGEMENT BOUNDARY

A thesis submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in Ichthyology and Fisheries Science

of

RHODES UNIVERSITY

By

RACHEL BRENNA MULLINS

April 2017

DECLARATION

I am aware of Rhodes University's policy on plagiarism. I have not included ideas, phrases, passages or illustrations from another person's work without acknowledging their authorship, and declare that all the work presented here is my own, with the exception of the development of mtDNA Control Region primers, and the extraction of DNA and library preparation for tGBS®, which was performed by Dr Niall McKeown of the Institute of Biological, Environmental and Rural Sciences at Aberystwyth University.

_____

Rachel Brenna Mullins

9 February 2017

ABSTRACT

Yellowfin tuna *Thunnus albacares* is a commercially and economically important fisheries species, which comprises the second largest component of South Africa's catch of tuna and tuna-like species. Catches of the species off South Africa are treated as two discrete stocks by the two tuna Regional Fisheries Management Organisations (tRFMOs) under whose jurisdictions they fall. Individuals caught off the Western Cape, west of the boundary between the tRFMOs at 20°E, are included in assessment and management of the Atlantic Ocean yellowfin tuna stock by the International Commission for the Conservation of Atlantic Tunas (ICCAT), and those caught east of this boundary are assessed and managed as part of the Indian Ocean stock by the Indian Ocean Tuna Commission (IOTC). The boundary between these stocks is based on the confluence of the two oceans in this region and does not incorporate the population structure of species. For sustainable exploitation of fisheries resources, it is important that the definition of management stocks reflects species' biological population structure; the fine-scale stock structure of yellowfin tuna off South Africa is therefore a research priority which this study aimed to address by means of population genomics analyses.

Yellowfin tuna exhibit shallow genetic differentiation over wide geographic areas, and as such traditional population genetic approaches have limited power in resolving fishery significant population structure in the species. Herein, a population genomic approach was employed, specifically, genome-wide analysis of single nucleotide polymorphisms (SNPs) discovered using a next-generation DNA sequencing approach, to confer (i) increased statistical power to detect neutral structuring reflecting population connectivity patterns and (ii) signatures of local adaptation. The mitochondrial Control Region (mtDNA CR) was also sequenced to compare the resolving power of different approaches and to permit coalescent based analyses of the species evolutionary history in the region.

Neutral SNP loci revealed significant structure within the dataset (F_{ST}=0.0043; P<0.0001); partitioning of this differentiation within the dataset indicated significant differentiation between yellowfin tuna from the Western Cape and the Gulf of Guinea in the eastern Atlantic Ocean, with no significant differentiation between individuals from the Western Cape and Western Indian Ocean regions. This indicates two population units wherein there is a separation of the Gulf of Guinea from the remaining samples (Indian Ocean including

Western Cape) that are largely derived from a single genetic population. This pattern was also supported by assignment tests. Positive outlier SNPs, exhibiting signatures of diversifying selection, suggest that individuals from these regions may be locally adapted, as well as demographically isolated. The mtDNA CR did not reveal any significant genetic structure among samples (F_{ST}=0.0030; P=0.309), demonstrating the increased resolving power provided by population genomics approaches, but revealed signatures of historical demographic fluctuations associated with glacial cycles.

Based on the findings of this study, it is suggested that yellowfin tuna caught off the Western Cape of South Africa are migrants from the Indian Ocean population, exhibiting significant genetic differentiation from the Atlantic Ocean Gulf of Guinea individuals, and should thus be included in the assessment and management of the Indian Ocean stock. It is therefore recommended that the boundary between the Atlantic and Indian Ocean yellowfin tuna stocks, under the mandates of ICCAT and the IOTC respectively, should be shifted to approximately 13.35°E to include all individuals caught in South African waters in the Indian Ocean stock.

ACKNOWLEDGEMENTS

My greatest thanks go to my supervisors, Professors Paul Shaw and Warwick Sauer for giving me the opportunity to complete this fascinating project and learn all that I have through it, and to Paul for taking me on with almost no knowledge of genetics. I have gained an enormous amount of knowledge and experience, in population genetics and genomics, and research in general, and for that I am extremely grateful. Thank you very much to Dr Niall McKeown for your help and guidance with all things genetics; for teaching me lab techniques, for doing a large amount of my lab work for me before I came to Aberystwyth, and for the help with the analyses of overwhelming datasets. Your help was invaluable in the writing of this thesis.

Thank you very much to everyone who collected fin clippings and muscle samples for me! Without your willingness to assist me this project could not have gone ahead. Thank you to Wendy West from DAFF for providing samples from the Western Cape, as well as for giving me insight into South Africa's tuna fisheries. Thank you to Professor Colin Attwood and Stewart Norman from UCT for providing additional Western Cape samples. Thank you to John Filmalter for collection of the Gulf of Guinea sample; this was an invaluable addition to the study. Thank you to Ryan Daly for the collection of samples from Mozambique. Thank you to Jonathan Booysen, Janeva Grimmick and the Shelly Beach Ski Boat Club, Chris and Cameron Gemmel and JP Jordaan for collection of samples from KwaZulu-Natal. Many thanks also to the Port Elizabeth Deep Sea Angling Club and the Port St Francis Ski Boat and Yacht Club for your co-operation in my collection of samples at your competitions.

Thank you to all the funders who allowed me to carry out this research. The Western Indian Ocean Marine Science Association (WIOMSA) provided me with a MARG-I research grant for personal and project-related funding. Personal funding was additionally provided by the National Research Foundation (NRF) of South Africa through an Innovation Masters Scholarship, and the Rhodes University Research Committee through an Ian Mackenzie Scholarship and a Henderson Masters Bursary. The Gulls Project, funded by the Belmont Forum, provided funding for the next-generation sequencing. Last but certainly not least, thank you to the Marine Stewardship Council (MSC) for awarding me a Research Scholarship which allowed me to travel to Aberystwyth University to undertake work in the Shaw Laboratory and make this project happen.

Thank you to all the friends I have made at Rhodes, and to everyone at DIFS for your friendship and tea-time conversations over the past three years. Thank you especially to everyone who has come and gone from Office S5 over the past two years for the emotional support throughout the whole Masters process!

Thank you to everyone at IBERS at Aberystwyth University for making my time in the UK so enjoyable; thank you especially to Amy Healey, Max Blake and Niall for making me feel so welcome and for helping me out with numerous genetics questions. Thank you Amy for joining me in my swim in the Irish Sea!

Finally, thank you Mommy and Daddy, for all your love and support, and for encouraging me to study my passion. Thank you Leah for all your encouragement, and for being a constant source of entertainment in my life. Thank you Granny and Lionel for all your interest in my project and all the newspaper clippings about anything ocean-related. Thank you Oli for always cheering me up when I got overwhelmed by it all, and for your encouragement when I thought I couldn't do it. And thank you Jasper, for being my sidekick, and for cheering me up after many long days in the office.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 *Yellowfin tuna* Thunnus albacares *(Bonnaterre, 1788)*

Yellowfin tuna *Thunnus albacares* (Bonnaterre, 1788) is a member of the marine Scombridae family (the tunas, mackerels, Spanish mackerels and bonitos), the Scombrinae subfamily and the tribe Thunnini (the tunas) (Collette & Nauen 1983; Collette *et al.* 2001). Yellowfin tuna is a tropical tuna species, and is characterised by a relatively fast and opportunistic life history strategy relative to tropical tunas, with a fast growth rate reaching a maximum size of 231 cm, early maturation, high fecundity and protracted spawning season, and relatively short longevity of approximately seven years (Fromentin & Fonteneau 2001; Collette *et al.* 2011; Juan-Jordá *et al.* 2013a, b). They are epipelagic predators and feed on a wide variety of prey species, including fish, crustaceans and cephalopods, and can adjust their main prey species depending on availability (Collette *et al.* 2011; Juan-Jordá *et al.* 2013b; Potier *et al.* 2004, 2007; Pecoraro *et al.* 2016b).

The tunas have counter-current heat exchangers, allowing them to increase their body temperature above ambient water temperatures, however these are less developed in tropical tuna species (Block & Finnerty 1994; Brill & Bushnell 2001; Graham & Dickson 2001); yellowfin tuna thus preferentially inhabit tropical and subtropical waters with temperatures of 18 to 31°C, although they can withstand colder water for shorter periods of time, for example to reach prey (Brill & Lutcavage 2001; Dagorn *et al.* 2006; Galli *et al.* 2009; Juan-Jordá *et al.* 2013b). Yellowfin tuna is highly migratory, and has a cosmopolitan oceanic distribution, and occupies tropical and subtropical waters of the Atlantic, Indian and Pacific Oceans and their adjacent seas (excluding the Mediterranean Sea), between the latitudes of approximately 40°N and 40°S (Collette *et al.* 2011).

1.2 *Global yellowfin tuna fisheries*

1.2.1 History and nature of yellowfin tuna fisheries

Yellowfin tuna is a commercially and economically important fisheries species, with a high demand due to its large size, accessibility and high quality of meat resulting in its suitability for the sashimi market (Majkowski 2007; Miyake *et al.* 2010; Juan-Jordá *et al.* 2013a;

Pecoraro *et al.* 2016b). Yellowfin tuna is currently the second most important of the seven principal market tuna species in global fisheries for tuna and tuna-like species, after skipjack tuna *Katsuwonus pelamis*, in terms of both global catch volume and weight (Miyake *et al.* 2010; Juan-Jordá *et al.* 2013a; FAO 2017). Prior to the 1950s, all yellowfin tuna catches were taken by small-scale coastal fishers, however following the increased demand for canned tuna in the 1950s, industrial fishing of tuna by longliners, purse-seiners and baitboats developed and fisheries for this species expanded further offshore and throughout the species' distribution (Fonteneau 1997; Miyake *et al.* 2010; Pecoraro *et al.* 2016b). Catches of yellowfin tuna thus increased from less than 300,000 metric tons (MT) per annum in the 1960s, to in excess of 1 million MT in recent years, with an average annual catch of ~1.25 million MT over the past decade (Pecoraro *et al.* 2016b). Global yellowfin tuna catches peaked in 2003 at ~1.5 MT (Miyake *et al.* 2010; FAO 2017), with the most recent available catch statistics showing a catch of ~1.4 MT in 2014, the greatest proportion of which is taken by the purse-seine fishery in the western-central Pacific Ocean, accounting for ~40% in 2014 (Pecoraro *et al.* 2016b). The species however supports important fisheries throughout its distribution. South African yellowfin tuna fisheries are associated with the Atlantic and Indian Ocean stocks of yellowfin tuna, and thus the yellowfin tuna fisheries in these oceans will be discussed here.

1.2.1.1 Atlantic Ocean fisheries

Industrial-scale fishing for yellowfin tuna in the Atlantic Ocean expanded in the mid-1950s, with the development of industrial longline and baitboat fishing in the region. The purse-seine fishery developed later, and is currently the most important fishing gear in this ocean followed by longliners. Yellowfin tuna catches increased to an average of 150,000 MT in the 1980s, and peaked at ~193,000 MT in 1990 (Figure 1.1) (ICCAT 2016b). Yellowfin tuna was the dominant species in Atlantic Ocean tuna fisheries until the 1990s, after which skipjack tuna became dominant due to shifted targeting effort (Pecoraro *et al.* 2016b). Yellowfin tuna catches still remain important however, with a total reported catch of 104,000 MT for this stock in 2014 (Figure 1.1) (ICCAT 2016b). The dominant gear type for yellowfin tuna landings in the Atlantic Ocean is purse-seiners, which account for ~70% of current reported catches (Pecoraro *et al.* 2016b; ICCAT 2016b), and the most productive fishing grounds for yellowfin tuna in this ocean occur in the eastern tropical region, where ~80% of this catch is taken (ICCAT 2016b).

1.2.1.2 Indian Ocean fisheries

A greater proportion of the global yellowfin tuna catch is taken in the Indian Ocean than in the Atlantic Ocean. Industrial longlining for yellowfin tuna in this ocean began in the early 1950s; this remained the dominant gear type until the 1980s when industrial purse-seining began and catches increased from ~70,000 MT in the early 1980s to more than 400,000 MT in 1990 (Figure 1.1) (IOTC-SC19 2016; Pecoraro *et al.* 2016b). The Indian Ocean yellowfin tuna catch peaked in 2004 at more than 525,000 MT, after which it declined until 2009 and then began to increase, with catches of ~415,000 MT in 2014 (Figure 1.1) (IOTC-SC19 2016) and ~408,000 MT in 2015 (IOTC-WPTT18 2016). Yellowfin tuna was the dominant tropical tuna species caught in the Indian Ocean from 2012 to 2015, making up ~45% of Indian Ocean tropical tuna catches (IOTC-WPTT18 2016). Purse-seiners remain the dominant industrial gear type for the stock, followed by longliners, accounting for ~34% and ~19% of the average annual catch from 2012 to 2015 respectively (IOTC-WPTT18 2016). Unlike the tuna fisheries of the Atlantic and Pacific Oceans, a large proportion of the total catch is taken by artisanal fisheries using gears such as pole-and-line, trolling, handlines and gillnets. Until the early 2000s, artisanal catches accounted for ~30% of the total Indian Ocean yellowfin tuna catch, and have since increased to almost 50% of the catch, as this species serves as an important source of protein for many developing coastal nations (IOTC-SC17 2014; IOTC-WPTT18 2016). The greatest proportion of the Indian Ocean yellowfin tuna catch is taken in the western region (IOTC-WPTT18 2016; Pecoraro *et al.* 2016b).

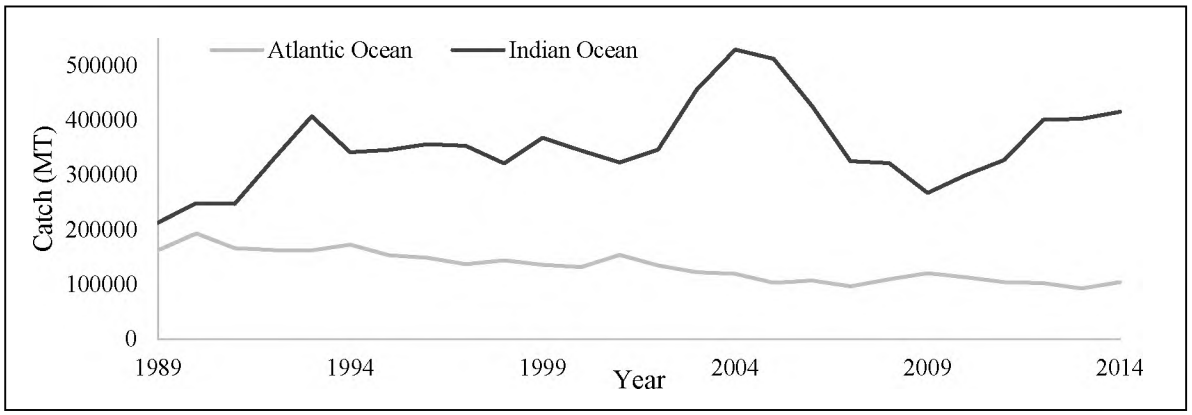

Figure 1.1: Time series of reported catches (metric tons) of the Atlantic Ocean and Indian Ocean yellowfin tuna stocks from 1989 to 2014 (Adapted from: ICCAT 2015b, 2016b; IOTC 2006; IOTC-SC17 2014; IOTC-SC18 2015).

1.2.2 Management of stocks by tuna Regional Fisheries Management Organisations

Yellowfin tuna is a highly migratory species with stocks and fisheries that extend across the boundaries of national Exclusive Economic Zones (EEZs) and into the high seas. Effective

management and conservation of such species therefore requires co-operation among nations that exploit the same stocks, such as through international management bodies, as was recognised by the 1982 United Nations Convention of the Law of the Sea (UNCLOS) (Allen 2010; Cullis-Suzuki & Pauly 2010). The stocks of tuna and tuna-like species are managed by tuna Regional Fisheries Management Organisations (tRFMOs), which are legally mandated to manage these stocks according to the 1995 United Nations Straddling Fish Stocks Agreement, and whose regulations the nations exploiting their stocks must adhere to (Cullis-Suzuki & Pauly 2010). Yellowfin tuna falls under the mandates of four tRFMOs, each of which assesses and manages the yellowfin tuna occurring within its area of competence as a single stock (FAO 2017), based on the currently assumed global population structure of the species (see section 1.3): the International Commission for the Conservation of Atlantic Tunas (ICCAT) is responsible for management of the Atlantic Ocean stock, the Indian Ocean Tuna Commission (IOTC) manages the Indian Ocean stock, the Inter-American Tropical Tuna Commission (IATTC) manages the eastern Pacific Ocean stock and the Western and Central Pacific Fisheries Commission (WCPFC) manages the western-central Pacific Ocean stock (Miyake *et al.* 2010; Cullis-Suzuki & Pauly 2010; FAO 2017). As this study is concerned with yellowfin tuna stocks of the Atlantic and Indian Oceans, these stocks and their relevant tRFMOs, ICCAT and the IOTC respectively, will be discussed further.

ICCAT is the oldest tRFMO and was established in 1969, while the IOTC was established in 1993 and became effective in 1996 (Allen 2010). These tRFMOs assess the status of their stocks against maximum sustainable yield- (MSY) based reference points, based on the recommendations of two international instruments, the FAO Code of Conduct for Responsible Fisheries and the United Nations Fish Stocks Agreement (UNFSA) (Allen 2010), whereby a stock is considered to be in an overfished state if its biomass is less than that required to achieve MSY ($B<B_{MSY}$), and subject to overfishing if its fishing mortality exceeds that required for MSY ($F>F_{MSY}$). The overall objective of both ICCAT and the IOTC is to maintain the stocks under their mandates at sustainable levels, defined as those which allow for MSY to be achieved (Allen 2010; ICCAT 2016c; Anon. 2009). While the Pacific Ocean tRFMOs undertake stock assessment research, ICCAT and the IOTC rely on reports from member states and co-operating non-contracting states that exploit the stocks under their mandates (Kolody *et al.* 2016). These tRFMOs have set data reporting guidelines, which require these states to report annually on various aspects of their fishing activities, including annual catch and effort data (IOTC Secretariat 2014; ICCAT 2016a; Pons *et al.* 2016). The

scientific bodies of the tRFMOs use these data to estimate MSY-based reference points, using population dynamics modelling approaches, and to assess the status of the stocks against these reference points, and results of stock assessments and associated management advice are presented to the relevant Commissions, which use this information to set management recommendations and measures (Allen 2010; Anon. 2009; Miyake *et al.* 2010; ICCAT 2016c).

1.2.2.1 Management of Atlantic Ocean yellowfin tuna by ICCAT

Current management measures in place for the Atlantic Ocean stock, under ICCAT's mandate, are described in ICCAT Recommendations 14-01 and 16-01 (Recommendation by ICCAT on a Multi-Annual Conservation and Management Program for Tropical Tunas). An annual total allowable catch (TAC) of the stock has been set at 110,000 t since 2012, however no catch quotas have been allocated and if this TAC is exceeded ICCAT will review the conservation and management measures in place for the stock. An annual time-area closure was introduced in 2013, with the objective of reducing the catch of juvenile tropical tunas, whereby surface fishing on fish aggregating devices (FADs) is not permitted in the Gulf of Guinea region during January and February; this measure has however been found to be ineffective in achieving this objective (ICCAT 2016b, c). A number of further controls are in place regarding fishing on FADs (Recommendation 16-01). Fishing capacity is also regulated, with vessels >20m requiring authorization to fish for tropical tunas and limitations on capacity existing for some longline and purse seine fleets (ICCAT 2016b).

1.2.2.2 Management of Indian Ocean yellowfin tuna by the IOTC

Prior to the 2015 assessment of the Indian Ocean stock, management measures for this stock were described in IOTC Resolutions 14/02 (For the conservation and management of tropical tunas stocks in the IOTC area of competence), 15/11 (On the implementation of a limitation of fishing capacity of Contracting Parties and Cooperating Non-Contracting Parties) and 15/08 (Procedures on a fish aggregating devices (FADs) management plan, including a limitation on the number of FADs, more detailed specifications of catch reporting from FAD sets, and the development of improved FAD designs to reduce the incidence of entanglement of non-target species). Under these Resolutions, measures were in place to control the fishing capacity of certain fleets and to regulate fishing on FADs. No TAC has been defined for this stock, however it was stated in Resolution 14/02 that a TAC and catch quotas should be set to sustainably manage the stock. Following the finding of this stock being overfished and subject to overfishing in 2015 (see section 1.2.3), the IOTC adopted

Resolution 16/01 (On interim plan for rebuilding the Indian Ocean yellowfin tuna stock in the IOTC area of competence) in 2016, a Conservation and Management Measure (CMM) binding to member states from January 2017. Under this CMM gear-specific catch limits have been set for certain flag states whose fisheries exceeded threshold catch levels in 2014 (>2000 MT for gillnets, >5000 MT for other gears). This is an interim measure, the effectiveness of which will be evaluated and reviewed by the IOTC by 2019 (IOTC-WPTT18 2016).

1.2.3 Status of the stocks

Yellowfin tuna management units, "stocks", refer to those individuals occurring within the areas of competence of each of the four tRFMOs mandated to manage them, and are assessed by the scientific bodies of their relevant tRFMOs against MSY reference points, based on catch and effort data reported by the member and co-operating states that harvest the stocks (Miyake *et al.* 2010; Kolody *et al* 2016).

1.2.3.1 Atlantic Ocean stock

The most recent stock assessment of the Atlantic Ocean stock by ICCAT's Standing Committee on Research and Statistics (SCRS) was carried out in 2011 using catch and effort data from 2010. Based on a production model and an age-structured model, the annual MSY of this stock was estimated to be ~144,600 MT. The stock was found to be in an overfished state in 2010, with the ratio of estimated stock biomass (B_{2010}) to that required to maintain the stock's MSY levels (B_{MSY}) being equivalent to <1 (B_{2010}/B_{MSY}=0.86). The stock was not found to be experiencing overfishing, with the observed level of fishing mortality, F_{2010}, being less than F_{MSY}, the limit reference point for MSY; the level of fishing mortality was however close to reaching F_{MSY} (F_{2010}/F_{MSY}=0.96; ICCAT 2016b). The most recent catch statistics for this stock presented by the SCRS found the total reported catch of the Atlantic Ocean stock for 2014 to be 104,000 MT, with an average annual catch between 2010 and 2014 of 106,000 MT, indicating that this stock has not been fished at levels above its MSY in recent years (ICCAT 2016b).

Although the status of this stock has improved with decreased catches, the continued harvesting of large numbers of juvenile yellowfin tuna (related to surface fishing on FAD-associated schools) suggests that the current MSY estimate may be biased downwards, and that the current levels of fishing which are close to MSY may be resulting in overfishing (Miyake *et al.* 2010; ICCAT 2016c). It has therefore been recommended that catch quotas

should be allocated among member states to ensure this stock is harvested below its TAC, as has been implemented by ICCAT for the Atlantic Ocean stock of bigeye tuna *Thunnus obesus*. Additionally, further measures should be implemented to control fishing on FADs to reduce the catch of juveniles (ICCAT 2016c).

1.2.3.2 Indian Ocean stock

The most recent assessment of the Indian Ocean stock was carried out by the Working Party on Tropical Tunas (WPTT) of the IOTC's Scientific Committee (SC) in 2014. Three models were applied, and the assessment found the MSY of this stock to be ~422,000 t (IOTC-WPTT18 2016). A Stock Synthesis III (SS3) model assessment found a 94% chance of this stock being overfished, with a spawner biomass (SB) ratio relative to MSY levels of SB_{2014}/SB_{MSY}=0.66, and subject to overfishing (F_{2014}/F_{MSY}=1.34). The total catch of this stock in 2014 of ~430,300 t exceeded the stock's estimated MSY. The catch in 2015 declined to ~407,575 t, however un update of the stock assessment found a 68% chance that the stock remains overfished (SB_{2015}/SB_{MSY}=0.89) and subject to overfishing (F_{2015}/F_{MSY}=1.11) (IOTC-WPTT18 2016; IOTC-SC19 2016).

The stock was found not to be overfished or subject to overfishing in 2014, based on the previous stock assessment carried out in 2012 and data available in 2014 (IOTC-SC 2014); it has been suggested that the stock's status has changed as a result of catches in recent years exceeding MSY due to increased industrial and artisanal fishing effort, combined with low recruitment as estimated by the assessment model (IOTC-WPTT18 2016). The SC recommended that the catch of the stock should be reduced to 80% of that in 2014 (~344,200 t), as is has been estimated that this will allow for a 50% probability of the IOTC objectives of $SB>SB_{MSY}$ and $F<F_{MSY}$ being met in 2024 (IOTC-WPTT18 2016). The IOTC therefore adopted Resolution 16/01 (On interim plan for rebuilding the Indian Ocean yellowfin tuna stock in the IOTC area of competence) in 2016 (see section 1.2.2.2) to end overfishing of the stock and rebuild its biomass, which set catch limits for certain flag states and fisheries to reduce the catches of this stock below that of 2014 (IOTC-SC19 2016; IOTC-WPTT18 2016).

1.3 *South African yellowfin tuna fisheries*

Yellowfin tuna is the second most important species in the catches of tuna and tuna-like species by South African fisheries, following albacore tuna *Thunnus alalunga*. The management boundary between the ICCAT and IOTC areas of competence occurs off South

Africa at 20°E (Figure 2.1) (see section 1.5); South Africa is therefore a member state of both of these tRFMOs and presents annual reports on fishing activities of yellowfin tuna (and other tuna and tuna-like species under the tRFMOs' mandates) west of this boundary to ICCAT, and east of this boundary to the IOTC, in accordance with the tRFMOs' data reporting requirements and guidelines. The total yellowfin tuna catch by South African fisheries has shown an increasing trend over time, with a catch of 1,985 t in 2014 (Figure 1.2) (Ndudane 2015; West & Kerwath 2015), with the greatest portion of the annual catch in most years being taken off South Africa's east coast in the IOTC area of competence (Figure 1.2). No catch quota for yellowfin tuna has been allocated to South Africa by either ICCAT or the IOTC, and thus fisheries for this species are managed by total allowable effort (TAE) input controls, rather than catch-related output controls (Ndudane *et al.* 2015; West & Kerwath 2015). The catch limits set by Resolution 16/01 of the IOTC for rebuilding of the Indian Ocean yellowfin tuna stock do not apply to South Africa, as 2014 catches by South African fleets fell below the thresholds set by this Resolution.

Yellowfin tuna is exploited by two main South African commercial fisheries subsectors, the large pelagic longline fishery and the tuna pole and line fishery (West *et al.* 2014; Ndudane 2015; West & Kerwath 2015), as well as by the recreational linefish subsector, and in small numbers by the commercial linefish subsector (West *et al.* 2014; West & Kerwath 2015).

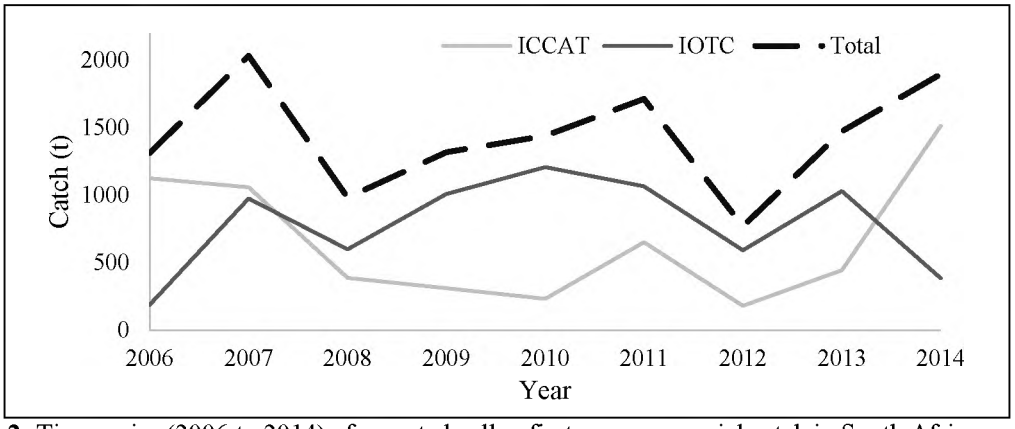

Figure 1.2: Time series (2006 to 2014) of reported yellowfin tuna commercial catch in South African waters, by domestic- and foreign-flagged vessels, in the ICCAT and IOTC areas of competence and in total (Adapted from: West *et al.* 2014; Ndudane 2015; West & Kerwath 2015).

1.3.1 The pelagic longline subsector

The large pelagic longline subsector operates along the extent of South Africa's coastline, within the country's EEZ and extending into the high seas, and has accounted for the greatest proportion of South African yellowfin tuna catch in recent years (Figure 1.3) (West *et al.* 2014; West & Kerwath 2015). Pelagic longlining by South African fishers began in 1995 and

the subsector became commercialised in 2005, following an experimental fishing phase from 1997 to 2005. Prior to 1995 rights were issued to foreign Japanese and Taiwanese fishers, but since its commercialisation fishing rights for this subsector are issued only to South African rights holders (for a detailed history of this subsector, see Penney and Griffiths (1999) and Sauer *et al.* (2003)). South Africa's large pelagic longline subsector currently consists of domestic South African-flagged vessels, and foreign Japanese-flagged vessels under joint venture with South African rights holders, and catches of yellowfin tuna by these fleets are taken mainly within the IOTC area of competence.

The South African-flagged pelagic longline fleet operates in Atlantic and Indian Ocean waters off South Africa and employs mainly swordfish *Xiphius gladius* targeting methods, however 40 to 50% of its catch consists of tropical tunas and sharks (West *et al.* 2014; West & Kerwath 2015), and targeting has recently shifted to include tropical tuna (West & Kerwath 2015). The Japanese-flagged fleet operates mainly in Indian Ocean waters off South Africa, and targets mainly tropical tunas further offshore than the South African fleet. Foreign vessel owners fishing in joint venture with South African rights holders have been encouraged to reflag their vessels, transferring skills to South Africans (West *et al.* 2014). The total yellowfin tuna catch by the large pelagic longline subsector in 2014 showed a 50% decline from 2013 (and a 63% decline in the IOTC area of competence) (Figures 1.2 and 1.3a); this decline is due to fewer Japanese-flagged vessels fishing under joint venture with South African rights holders in 2014 resulting in seven fewer fishing vessels being active in 2014 and a 52% decline in fishing effort in terms of number of hooks set (West & Kerwath 2015).

This subsector is managed by TAE control, whereby 50 permits are issued to rights holders with one permit required per fishing vessel. In 2005, 50 long-term (10 year) permits were issued, of which 30 were for tuna-directed fishing (the remainder being for swordfish targeting) (West *et al.* 2014; West & Kerwath 2015). In 2015 and 2016, 15 year permits were allocated with tuna- and swordfish-directed fishing effort being merged, due to the shift in targeting effort by South African vessels (West & Kerwath 2015).

The South African large pelagic longline fishery is a developing subsector, and catches of tropical tunas (including yellowfin tuna) are expected to continue to increase. South African vessels began to focus fishing effort in Indian Ocean waters off South Africa's east coast in 2001 (prior to which most effort was focused in the Atlantic Ocean), and found catches of

tropical tunas to be greater in these waters. South Africa has submitted a fishing plan to the IOTC outlining intentions to expand and develop this fishery, with catches of bigeye tuna in excess of 1,000 t in future years (West *et al.* 2014). Due to the multispecies nature of tropical tuna-targeting pelagic longlining, catches of yellowfin tuna by the South African pelagic longline fishery are likely to increase as the fishery develops.

1.3.2 The tuna pole and line subsector

The tuna pole and line subsector forms part of South Africa's linefishery, and is the country's primary tuna fishery. The primary target species of this fishery is albacore tuna *Thunnus alalunga*, however yellowfin tuna is caught in smaller quantities with its catch being relatively high in some years (e.g. 2014; Figure 1.3a), when availability of this species off the west coast is higher. The subsector operates from September to May, based on albacore tuna abundance, with effort focused mainly in the Atlantic Ocean in the ICCAT area off South Africa's west coast and into Namibian waters, within the countries' EEZs (Sauer *et al.* 2003; West *et al.* 2014). Yellowfin tuna are caught in the southern portion of the fishery range, on the border of the Agulhas Bank (Sauer *et al.* 2003). Effort by this fleet rarely extends east into Indian Ocean waters; in 2013, for example, no fishing by this fleet occurred in the Indian Ocean (IOTC-SC17 2014; West *et al.* 2014; ICCAT 2015b).

South Africa's tuna pole fishery became commercialised in the 1970s (for a detailed history, see Sauer *et al.* (2003)). All vessels registered in this fishery are South African-flagged, and it is controlled by means of a TAE input control, whereby a maximum of 200 permits are issued to South African rights holders, with one permit being required per vessel (West *et al.* 2014).

Yellowfin tuna is also caught by two additional components of the South African linefishery, the commercial and recreational linefish fisheries. The commercial linefish fishery is a multispecies rod and reel fishery, and occasionally targets yellowfin tuna, mainly in the IOTC area of competence, when other linefish species are not available (West *et al.* 2014). Accurate catch data for each species in this fishery is not available, and tuna are not always identified to species level but rather grouped (ICCAT 2015b); as a result the annual catches of yellowfin tuna by this fishery are uncertain. Yellowfin tuna is not a primary target species of this multispecies nature of this fishery, and its catches are relatively small and do not contribute significantly to the total South African catch (IOTC-SC17 2014; West *et al.* 2014). South Africa's rod and reel recreational fisheries sector is an important component of South

Africa's fisheries sector (Van der Elst and Everett 2015), and targets gamefish, including yellowfin tuna, in Atlantic and Indian Ocean waters of South Africa using rod and reel from ski boats (Sauer *et al.* 2003). All fishers are required to possess fishing permits, and catch is controlled by output controls: a limit of 10 tuna (all species) per permit holder per day is permitted, with a minimum weight of 3.2 kg for yellowfin tuna landed. Catch from this fishery may not be sold (West *et al.* 2014). Accurate catch and effort data for this subsector is not available, however it has been estimated that the combined annual catch of yellowfin tuna and king mackerel (*Scomberomorus cavalla*) for this fishery is greater than 100 t for the entire South African coastline, and that the annual catch of yellowfin tuna in Atlantic Ocean waters is between 20 and 40 t (IOTC-SC17 2014; West *et al.* 2014; ICCAT 2015b).

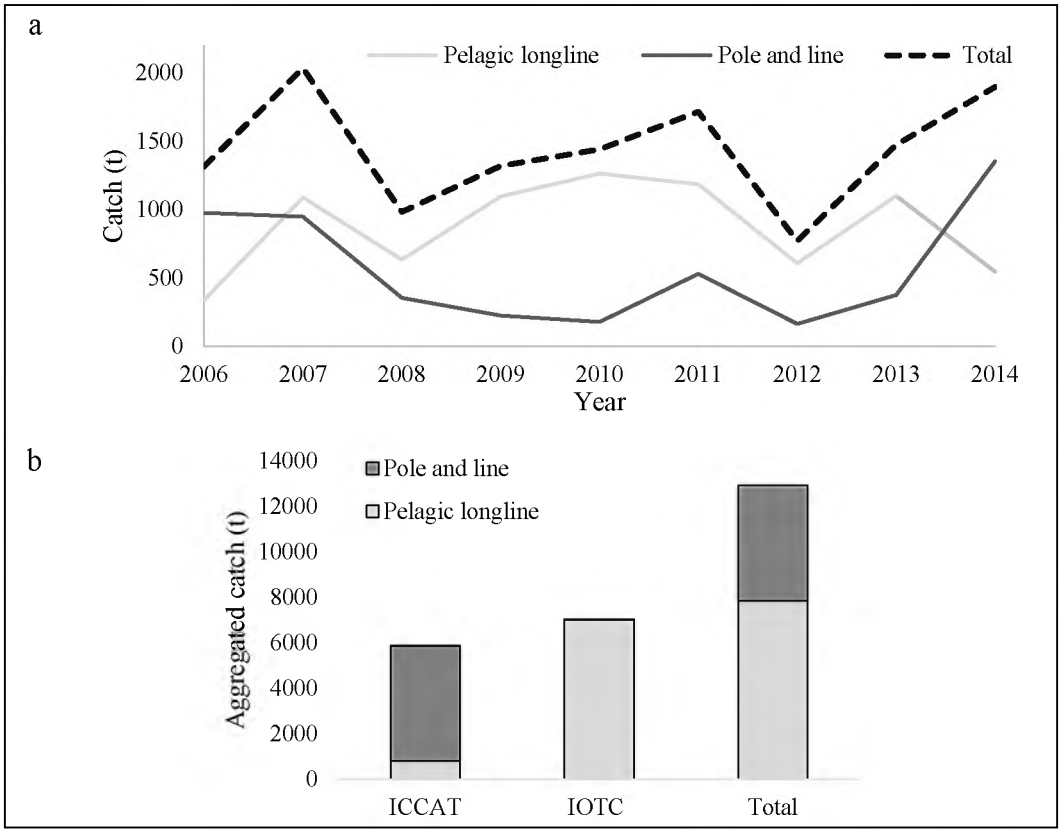

Figure 1.3: Reported yellowfin tuna commercial catch taken in South African waters by domestic- and foreign-flagged vessels, per gear type, a) time-series of catch (2006 to 2014), and b) aggregated catch (2006 to 2014) in the ICCAT and IOTC areas of competence and in total (Adapted from: West *et al.* 2014; Ndudane 2015; West & Kerwath 2015).

1.4 *Population structure in fisheries management*

In developing a management plan for the exploitation of a fisheries resource, it is important that the population structure of the species is incorporated into the definition of its management units, referred to as stocks (Pawson & Jennings 1996; Begg & Waldman 1999; Begg *et al.* 1999; Reiss *et al.* 2009). Populations of a species can be defined as units that are

independent at various scales, depending on the context and objectives of the definition; in the context of fisheries management it is important that management units are demographically independent, and populations may be defined according to the ecological definition of Waples and Gaggiotti (2006), which requires "co-occurrence in space and time" of individuals within a population to enable demographic interaction (Waples & Gaggiotti 2006, p. 1412). According to this definition, it is suggested that units of intraspecific individuals are independent populations if the migration rate among them is lower than 10% per generation (Hastings 1993; Waples & Gaggiotti 2006). At this level of connectivity, populations can be considered demographically independent as the relative contribution of net immigration to their total recruitment, and therefore population dynamics, is low compared to the contribution of net auto-recruitment (Lowe & Allendorf 2010). Accurate life history parameters must be incorporated into the assessment of stocks (Juan-Jordá *et al.* 2013a); these parameters, and responses to harvesting, are independent among discrete populations, and it is therefore necessary that demographically isolated units are separately assessed and managed for long-term sustainability.

Yellowfin tuna stocks are managed based on a level of harvestable surplus, MSY, that can theoretically be removed from the stocks without affecting their long-term sustainability (Carvalho & Hauser 1994; Miyake *et al.* 2010). Stock MSY levels are determined based stock assessments of population parameters (such as recruitment and mortality) that determine stock productivity (Miyake *et al.* 2010). These parameters are assumed to be homogenous across each stock, however if cryptic population structure does occur within management stocks this assumption is violated as the effects of harvesting on population parameters are independent among populations, and their levels of sustainable yield may differ (Pawson & Jennings 1996; Begg & Waldman 1999). Failure to assess and manage discrete populations independently may result in overfishing of less productive units, which may result in population collapses, which may in turn lead to a reduction in species' genetic diversity and adaptive potential (Kenchington *et al.* 2003; Duncan *et al.* 2015). It is also important to manage populations with low demographic connectivity independently as with a low degree of exchange of individuals per generation, such units are not able to replenish each other's biomass through 'rescue effects', and a harvestable surplus of individuals may not be restored by migration following overharvesting (Waples 1998; Hauser & Carvalho 2008). The effect of fishing pressure on discrete populations is independent, and their management units should reflect this. A number of techniques exist to assess the degree of

connectivity among units of a species directly, such as through tagging studies, or indirectly, such as through analysis of genetic or morphometric differences among units, and an ideal population structure study would incorporate a number of these; this is however not always feasible, and population genetic analysis is the single most informative tool for elucidating stock structure (Ward 2000). Genetic markers, which may also be applied to study 'real-time' dispersal (Castric & Bernatchez 2004), are the only tools that can describe the effective dispersal of individuals across generations (i.e. interbreeding).

1.4.1 Population genetics to assess population structure

Indirect methods of delineating populations within species assess the degree of differentiation among sampled units, which indicates whether the units are likely to be demographically independent or connected (Waples 1998). Population genetic analyses of species may be used to infer population connectivity among sampled units of a species either directly, using parentage and assignment tests, or indirectly, based on the level of genetic differentiation among samples at molecular markers from mitochondrial DNA (mtDNA) or nuclear DNA (nDNA) genomes (Waples 1998; Waples & Gaggiotti 2006). Differentiation in allele frequencies at molecular markers can be used to infer their level of connectivity, and thus infer reproductive isolation, among sampled regions as a result of the neutral evolutionary processes of mutation, genetic drift and gene flow that alter population allele frequencies (Waples 1998; Ward 2000; Shaklee & Currens 2003; Allendorf *et al.* 2010). Genetic drift is a random process that results in the fixation of alleles (in neutral genomic regions) within finite populations over generations, as a result of a finite subsample of individuals within the population contributing to the allelic frequencies of subsequent generations (Charlesworth 2009), and therefore increases allele frequency heterogeneity among isolated populations (Waples 1998; Charlesworth 2009). Gene flow refers to the exchange of genes among populations, as a result of effective migrants that successfully reproduce in, and contribute to the gene pool of, recipient populations, countering the effects of genetic drift among populations (Allendorf *et al.* 2010). Measures of genetic differentiation based on allele frequency differences, such as Wright's F_{ST} (Wright 1951), may therefore indicate whether units of a population are reproductively isolated (discrete) or panmictic.

Early genetic studies relied on indirect assessment of allelic frequencies based on differentiation in expressed proteins, such as allozymes. The development of DNA sequencing technology has allowed DNA fragments to be sequenced following anonymous (e.g. restriction fragment length polymorphism (RFLP) analysis using restriction enzymes) or

targeted isolation (using primers) of genomic regions (Garvin *et al.* 2010), allowing molecular markers to be developed from mitochondrial and nuclear genomes, such as nDNA microsatellites (Carvalho & Hauser 1994; Wright & Bentzen 1995; Sunnucks 2000; Vignal *et al.* 2002). Markers from neutral loci, not subject to selection pressures, have traditionally been applied to studies of population genetics as these can allow for inferences of connectivity and isolation to be made, either through indirect (F_{ST}-based; Hellberg 2009) methods, or directly using parentage assignment based tests (Christie *et al.* 2010) (Morin et al. 2004; Helyar et al. 2011).

There are two fundamental limitations to standard population genetic based analysis for fisheries. Firstly, deriving quantitative estimates of gene flow and dispersal from low levels of genetic differentiation among samples of highly abundant marine fish species can be problematic as the methods may not be sufficiently sensitive to detecting dispersal limitations among large populations (Whitlock & McCauley 1999; Palsbøll *et al.* 2007; Hellberg 2009; Allendorf *et al.* 2010). In such cases it is often unclear to what extent low levels of structuring reflect correlations among populations due to retained signatures of historical connectivity, or actual recurrent connectivity. A second issue is the discrepancy between level of gene flow required to limit genetic differentiation, and that of dispersal required to replenish stocks which is of interest to fisheries management objectives (Hauser & Carvalho 2008). This is particularly challenging in studies of the marine environment due to the relative lack of physical barriers to movement compared to freshwater habitats, allowing for potential exchange of individuals among populations (Ward *et al.* 1994b; Waples 1998; Ward 2000).

The number and type of genetic markers employed, and sample sizes, may lack sufficient resolving power to detect structuring due to restricted gene flow, based on the limitations described above. Population genetics studies have traditionally relied on a small numbers of neutral molecular markers, however recent developments in sequencing technology ("next-generation sequencing technology") have the potential to overcome the limitations of traditional population genetics approaches, by increasing enormously the number of independent loci assayed simultaneously per individual, and so increasing power.

1.4.2 Next-generation sequencing and population genomics

1.4.2.1 Next-generation DNA sequencing

Population genetic studies have traditionally relied on relatively small numbers of molecular markers, based on the limitations of Sanger sequencing technology to produce large numbers of sequence reads (Garvin *et al.* 2010; McCormack *et al.* 2013). Recent advances in sequencing technology have resulted in the development of next-generation sequencing (NGS), performed on massively parallel, high-throughput sequencing platforms which are able to generate millions of sequence reads in a single run (e.g. Mardis 2008; Hudson 2008; McCormack *et al.* 2013). A number of sequencing techniques have thus been developed which allow for thousands of genome-wide molecular markers to be discovered and genotyped for any organism, resulting in the recent advance from population genetics to population genomics whereby population differentiation is assessed at a genome-wide scale, defined by Stinchcombe and Hoekstra (2008) as "population genetic analysis of a large number of loci, distributed throughout the genome". Population genomic analysis allows for many of the limitations associated with population genetic analysis of small numbers of molecular markers to be overcome, and has proven to be of particular use in resolving population structure in marine fish species where traditional population genetics has revealed no, or ambiguous patterns of, population structure, such as yellowfin tuna.

The development of NGS technology has enabled population genomics analyses, as the potential to sequence millions of DNA fragments in single reactions on NGS platforms allows for novel discovery of single nucleotide polymorphism (SNP) markers among multiple intraspecific individuals, at homologous loci across their genomes (Hudson 2008; Mardis 2008; Garvin *et al.* 2010; Narum *et al.* 2013). The low-throughput of traditional Sanger sequencing limits the number of high-quality sequence reads that can be produced and therefore the number of markers that may be discovered and genotyped (Hudson 2008; Garvin *et al.* 2010; McCormack *et al.* 2013). Sequencing of SNP markers depends on the targeting of known marker loci, which requires prior genomic knowledge and primer development, or discovery of novel makers, which requires homologous loci of large numbers of individuals to be sequenced to avoid ascertainment bias in the identification of SNPs (Morin *et al.* 2004; Garvin *et al.* 2010; Seeb *et al.* 2011; McCormack *et al.* 2013). The ability to discover and genotype anonymous markers among individuals overcomes much of the ascertainment bias associated with targeting of known loci; it is important however to have samples representative of species' distributions to avoid ascertainment bias in the

discovery of SNP makers (Garvin *et al.* 2010). High-throughput NGS platforms enable sequence reads from across the genomes of multiple individuals to be sequenced in parallel, allowing for the discovery of novel SNPs among individuals.

The sequence read output of NGS platforms results in a trade-off between the number of individuals that can be sequenced, and the proportion of individuals' genomes that can be represented, in a single run (Narum *et al.* 2013). To address this trade-off, reduced representation (RR) genome sequencing, enabled by NGS, reduced effective genome complexity (size), reducing and so controlling the available number of genome-wide DNA fragments to be sequenced, and so allowing for multiple individuals to be screened in parallel for genome-wide variation to a suitable read depth (Van Tassell *et al.* 2008; Garvin *et al.* 2010; Ogden 2011; Seeb *et al.* 2011; Andrews *et al.* 2016). SNP discovery through RR genome sequencing is appropriate for population genomic applications, which do not need to capture the entire genome but instead analyse subsets of molecular markers from across the organism's genomes to reflect genome-wide evolutionary processes and variability (McCormack *et al.* 2013).

1.4.2.1.1 Genotyping-by-sequencing

Genotyping-by-sequencing (GBS; Elshire *et al.* 2011) approaches have been the most widely applied NGS technique for population genomics studies (da Fonseca *et al.* 2016); these refer to a group of RR genome sequencing approaches, many of which are based on restriction site-associated DNA (RAD) sequencing (Baird *et al.* 2008), and employ restriction enzymes in their genome-reduction step, as described in Figure 1.4. These reduce genome complexity by cutting DNA at its recognition sites across the genome, resulting in thousands of genome-wide fragments which are PCR amplified and sequenced on NGS platforms, in parallel reactions for multiplexed individuals. Post-sequencing bioinformatic analyses align sequence reads of homologous loci across demultiplexed individuals, allowing SNPs to be identified as polymorphisms among individuals at single nucleotides (Davey *et al.* 2011, 2013; Davey & Blaxter 2011; Elshire *et al.* 2011; McCormack *et al.* 2013; Narum *et al.* 2013). To distinguish SNPs from rare allelic variants and sequencing errors, identified based on an assigned minor allele frequency (MAF; the occurrence of the less common polymorphic allele among individuals) threshold (Davey & Blaxter 2011; McCormack *et al.* 2013). A number of GBS approaches exist (Andrews *et al.* 2016), which differ in the number and types of restriction

enzymes used, and with variations to some steps, but all follow the same basic steps (Figure 1.4).

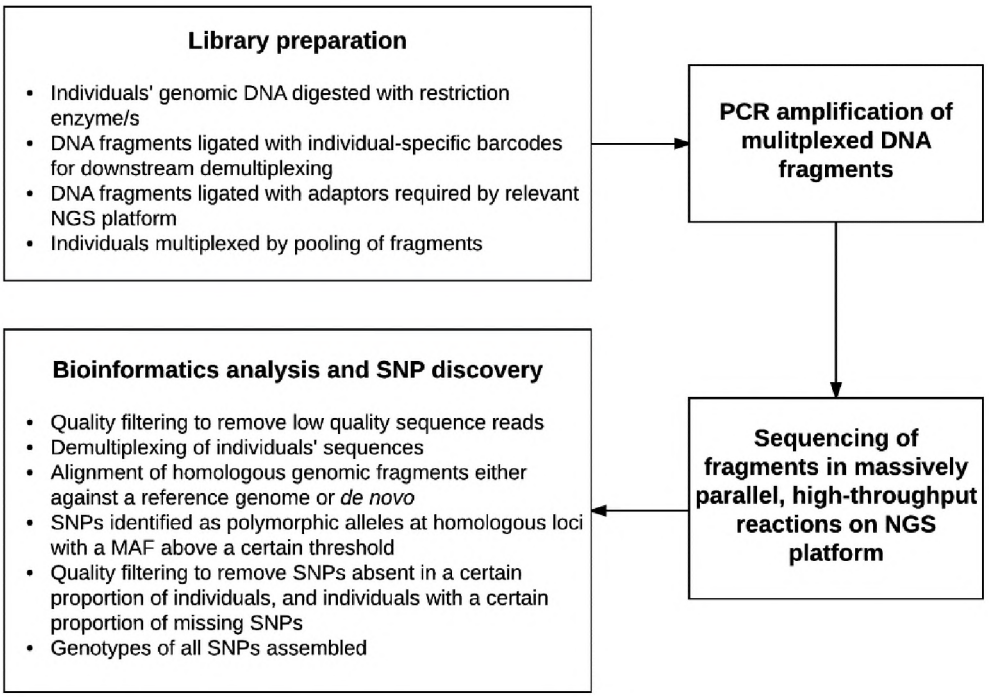

Figure 1.4: Flowchart describing steps common to all genotyping-by-sequencing (GBS) approaches (Adpated from: Baird *et al.* (2008); Elshire *et al.* (2011); McCormack *et al.* (2013); Narum *et al.* (2013); Andrews *et al.* 2016).

The flexibility of GBS approaches make them appropriate for investigations of population structure, as the genome reduction level (GRL), and therefore the number of SNPs potentially discovered, can be modified based on the type (rare- or common-cutting) and number of restriction enzymes used (Garvin *et al.* 2010; Elshire *et al.* 2011), and so a suitable trade-off with number of individuals per population sample made. The trade-off between the number of fragments sequenced and the read depth per fragment on an NGS platform means that with a higher GRL, each fragment may have a higher read depth and therefore a better quality sequence read, reducing sequencing and genotyping error (Davey & Blaxter 2011; Elshire *et al.* 2011; McCormack *et al.* 2013). Quality filtering of sequencing reads removes those with low quality reads that may produce genotyping errors, resulting in homologous reads being absent in some individuals (Davey & Blaxter 2013; McCormack *et al.* 2013). Investigations of intraspecific differentiation assess divergence over relatively short time scales and require relatively small amounts of data (tens to thousands of SNPs) (McCormack *et al.* 2013; Andrews *et al.* 2016; da Fonseca *et al.* 2016). It is therefore more important for population genomics that fewer sites are cut and sequenced, each with a relatively high read depth and

low probability of genotyping error, to allow for a lower rate of missing SNPs across individuals.

Another benefit of GBS is that it allows for genome-wide SNPs to be discovered for any organism, including non-model species without sequenced genomes. The restriction enzymes used are universal and can digest the genome of any organism to isolate cut sites for sequencing, and prior genomic knowledge for the development of primers to target and isolate DNA fragments is not required (Garvin *et al.* 2010; Narum *et al.* 2013; da Fonseca *et al.* 2016). The alignment of homologous fragments in the bioinformatics analysis is more easily done against a reference genome, however they can be aligned *de novo* using algorithms that compare pairwise fragments (Garvin *et al.* 2010; McCormack *et al.* 2013); genome-wide SNP markers can thus be discovered and genotyped for population genomic analysis in any organism.

1.4.2.2 Population genomics analyses

Population genomics analyses of genome-wide molecular makers have a number of advantages over traditional population genetics to resolve population structure, and are thus particularly useful for population structure studies of marine fish species which tend to exhibit low levels of genetic structure among populations (Ward *et al.* 1994b; Ward 2000), as has been demonstrated by a number of studies (e.g. Krück *et al.* 2013; Corander *et al.* 2013; Larson *et al.* 2014).

Population genomics analyses generally rely on genotypes of thousands of SNP markers, discovered using NGS techniques. SNPs are point-source changes in DNA sequences, arising either through transitions or transversions, that become fixed in at least 1% of natural populations, distinguishing them from rare allelic variants (Brookes 1999; Vignal *et al.* 2002; Morin *et al.* 2004). SNPs occur in high densities throughout organisms' genomes, in both coding and non-coding regions (Vignal *et al.* 2002; Baird *et al.* 2008; Helyar *et al.* 2011), have slow mutation rates (Foll & Gaggiotti 2008; Helyar *et al.* 2011) and a simple infinite alleles mutation model (IAM) of evolution (Vignal *et al.* 2002), and exhibit co-dominance (Brookes 1999; Vignal *et al.* 2002; Garvin *et al.* 2010). SNPs may theoretically be multi-allelic, with up to four possible allelic states, however they are generally bi-allelic polymorphisms as their slow mutation rates result in a low probability of two independent nucleotide substitutions occurring at a single locus (Brookes 1999; Vignal *et al.* 2002). Individual SNPs have low genetic diversity levels and low per-locus information content,

however multilocus genotypes of many independent SNPs can be highly informative compared to traditionally used markers, such as nDNA microsatellites and mtDNA sequences (Ryynänen *et al.* 2007). The low individual genetic diversity and slow mutation rate of SNPs relative to multi-allelic markers means that fewer individuals must be assessed as sampling variance is lower (Nei 1978; Kalinowski 2005), and that SNP genotypes are less likely to be subject to homoplasies (Morin *et al.* 2004, 2009). The high density of SNPs throughout genomes, and the usefulness of large genome-wide genotypes of these markers, make them ideal markers for discovery using NGS for population genomics analysis.

Greater precision in the estimation of genetic variability within and between populations may be achieved with increased number of molecular markers (Waples 1998; Allendorf *et al.* 2010). The effects of evolutionary processes may be heterogeneous at different genomic loci, and independent markers will therefore each be associated with a level of error in the estimation of these processes on the genome (Waples 1998; Helyar *et al.* 2011). With increasing numbers of loci, however, this error may be reduced and observed variation may be considered more representative of evolutionary forces acting on individual's genomes, rather than at specific loci (Waples 1998), and genome-wide SNP genotypes may therefore offer increased precision in estimates of population differentiation. Genome-wide SNP markers also take advantage of this genome-wide heterogeneity in nucleotide diversity in species that do not experience the effects of genetic drift strongly. Subsets of more variable and informative loci can be identified from genotypes of thousands of SNPs, and assessed for population structuring (Helyar *et al.* 2011). The high density of SNPs may allow for large numbers of more informative markers to be identified across the genome, allowing greater precision of estimates from these markers.

Finally, population genomics allows for different evolutionary forces acting on organisms' genomes to be incorporated into assessments of population genetic structure. SNPs occur in both coding (selected) and non-coding (neutral) regions, which are subject to different evolutionary forces, and SNP loci potentially subject to these different forces can be identified through outlier tests and separately assessed (Nielsen *et al.* 2009; Gagnaire *et al.* 2015). Population genetic studies have traditionally focused on neutral markers to indicate patterns of population differentiation, however loci linked to regions under selection may reveal patterns of differentiation different to those in neutral regions, and in addition to revealing locally adapted demes may be highly powerful stock markers (Canino *et al.* 2005).

1.4.3 Global population structure of yellowfin tuna

Yellowfin tuna global population structure is currently uncertain, despite the species' commercial and economic value and the importance of incorporating population structure into the definition of fisheries management units. Four discrete populations are currently assumed to occur across the species global distribution: an Atlantic Ocean population, an Indian Ocean population, an eastern Pacific Ocean population and a western-central Pacific Ocean population, each of which is managed as a single stock by its relevant tRFMO. Population genetic studies have revealed inconsistent patterns of structuring within and between regions (Pecoraro *et al.* 2016b), and more recent studies have revealed that the species population structure may be more complex than previously assumed.

The presently accepted four stocks are based on a genetic study using allozyme analysis, which indirectly assesses DNA allelic frequencies based on expressed enzymes. Significant interoceanic allele frequency differentiation was detected by Ward *et al.* (1997) at one of four loci analysed, which grouped individuals sampled from the Atlantic and western-central Pacific Oceans, and those from the Indian and eastern Pacific Ocean. This finding, combined with the geographic locations of samples and evidence of scales of movement available for yellowfin tuna (e.g. Hunter *et al.* 1986, Lewis 1992), suggested that at least four discrete populations of yellowfin tuna occur globally: an Atlantic Ocean population, an Indian Ocean population, an eastern Pacific Ocean population and a western-central Pacific Ocean population. Sharp (1978) and Ward *et al.* (1994) previously detected significant differentiation between eastern and western-central Pacific Ocean individuals at the same allozyme locus (Table 1.1). Early studies of mitochondrial and nuclear DNA sequence data however failed to detect patterns of interoceanic genetic heterogeneity across the same regions, as described in Table 1.1 (Scoles & Graves 1993; Ward *et al.* 1994a, 1997; Appleyard *et al.* 2001); it was therefore suggested that the allozyme differentiation observed may have arisen due to selection rather than reproductive isolation, as allozyme alleles may be subject to differential selection pressures (Appleyard *et al.* 2001).

Interoceanic genetic differentiation has however been observed at various neutral DNA makers by more recent studies (Table 1.1), supporting the structuring hypothesis of at least four populations. Significant (although weak) differentiation was detected through RFLP analysis of the mtDNA ATCO gene region (a coding region flanking the ATPase 6 and cytochrome oxidase subunit II genes; Chow & Inoue 1993) between Atlantic and eastern Pacific Ocean samples (Ely *et al.* 2005), and at the mtDNA cytochrome c oxidase subunit I

(COI) region between Atlantic and Indian Ocean samples (Henriques 2011) and between eastern and western-central Pacific Ocean samples (Li *et al.* 2015). More recently, analyses of single nucleotide polymorphism (SNP) genotypes have found strong and significant evidence for differentiation between eastern and western-central Pacific Ocean samples (Grewe *et al.* 2015) and among Atlantic, Indian and Pacific Ocean samples (Pecoraro *et al.* 2016a). Lack of global panmixia of yellowfin tuna is also supported by tagging studies which suggest the species migrates within, rather than between, ocean basins (e.g. Itano & Holland 2000; Maury *et al.* 2001; Sibert & Hampton 2003; Zagaglia *et al.* 2004; Schaefer *et al.* 2007, 2011; Langley & Million 2012). Inconsistent genetic structuring has however been observed within and between studies, indicating a shallow genetic structure of this species.

A number of studies, employing various markers, have failed to detect significant spatial heterogeneity among regions and samples for which differentiation has been detected by other studies and/or markers. Ely *et al.* (2005) failed to detect significant differentiation among samples at the mtDNA Control Region (CR), while Henriques (2011) and Pecoraro *et al.* (2016b) failed to detect significant differentiation among samples with microsatellite analysis, despite findings of significant differentiation by these authors among the same samples using different markers (Table 1.1). Wu *et al.* (2010) failed to detect significant differentiation between Indian and western-central Pacific Ocean samples at the mtDNA CR, while structure between these regions has since been detected (Pecoraro *et al.* 2016a) (Table 1.1). This indicates that genetic structure among discrete yellowfin tuna populations is shallow, and neutral molecular markers may lack resolving power to detect this structure in some cases.

Table 1.1: Summary of previous population genetic studies investigating the population structure of yellowfin tuna between ocean basins. AO: Atlantic Ocean; EAO: Eastern Atlantic Ocean; WAO: Western Atlantic Ocean; IO: Indian Ocean; WIO: Western Indian Ocean; NIO: Northern Indian Ocean; EPO: Eastern Pacific Ocean; WCPO: Western-central Pacific Ocean. RE: restriction enzymes.

Study	Regions sampled	Markers	Significant spatial heterogeneity
Sharp (1978)	EPO and WCPO	Allozyme analysis	Yes
Scoles & Graves (1993)	AO, EPO & WCPO	MtDNA-RFLP analysis: 12 RE	No
Ward *et al.* (1994a)	EPO & WCPO	Allozyme analysis: 4 loci	Yes
		MtDNA-RFLP analysis: 2 RE	No
Ward *et al.* (1997)	WAO, IO, EPO & WCPO	Allozyme analysis: 4 loci	Yes
		MtDNA-RFLP analysis: 2 RE	No
Appleyard *et al.* (2001)	EPO & WCPO	Microsatellite analysis: 5 loci	No
Ely *et al.* (2005)	AO, IO & EPO	MtDNA CR-I sequence	No
		MtDNA ATCO RFLP analysis	Yes
Wu *et al.* (2010)	WIO & WCPO	MtDNA CR-I sequence	No
Henriques (2011)	WAO & IO	MtDNA COI sequence	Yes
		Microsatellite analysis: 6 loci	No
Li *et al.* (2015)	EPO & WCPO	MtDNA COI sequence	Yes
Grewe *et al.* (2015)	EPO & WCPO	SNP genotypes	Yes
Pecoraro *et al.* (2016a)	AO, WIO, EPO & WCPO	SNP genotypes	Yes
Pecoraro *et al.* (2016b)	AO, WIO, EPO & WCPO	Microsatellite analysis: 12 loci	No

Recent studies have also found evidence for genetic differentiation among samples within ocean basins, suggesting that yellowfin tuna population structure may be more complex than is currently reflected by its four management units. In the Pacific Ocean, significant spatial heterogeneity has been detected within the eastern Pacific Ocean population through microsatellite analysis (Díaz-Jaimes & Uribe-Alcocer 2006), within the western-central Pacific Ocean population using microsatellite (Qiu & Miyamoto 2011; Aguila *et al.* 2015) and SNP genotype analyses (Grewe *et al.* 2015), and within the central Pacific Ocean at the mtDNA COI gene (Li *et al.* 2015). This may indicate independent sub-populations within the eastern and western-central Pacific Ocean yellowfin tuna populations, arising through site fidelity to spawning grounds resulting in genetic differentiation among separate spawning units (Grewe *et al.* 2015; Pecoraro *et al.* 2016b). The occurrence of site fidelity of yellowfin tuna within the Pacific Ocean is supported by results of tagging studies (Sibert & Hampton 2003; Schaefer *et al.* 2007, 2011; Pecoraro *et al.* 2016b).

Spatial genetic heterogeneity has also been detected within the northern Indian Ocean in the waters around Sri Lanka, at mtDNA ATPase 6 and 8 regions and through microsatellite

analysis (Dammannagoda *et al.* 2008), and at the mtDNA CR (Kunal *et al.* 2013). Sub-structuring in the Indian Ocean may also be related to site fidelity to spawning sites in different regions of the Indian Ocean, with feeding migrations to this region resulting in the mixing of individuals from different sub-populations (Dammannagoda *et al.* 2008; Kunal *et al.* 2013). Tag-recapture evidence from the Indian Ocean Tuna Tagging Project (IOTTP) suggests yellowfin tuna make large scale movements throughout the Indian Ocean (Langley & Million 2012; Pecoraro *et al.* 2016b), however it is possible that individuals show a level of site fidelity and return to their natal spawning grounds to spawn. Chow *et al.* (2000a) and Nishida *et al.* (2001) did not detect sub-structuring among samples from throughout the Indian Ocean basin using RFLP analyses of mtDNA and nDNA respectively; this may be related to a lack of sub-structuring, or insufficient resolving power of these markers to detect significant differentiation.

Table 1.2: Summary of previous population genetic studies investigating the population structure of yellowfin tuna within ocean basins. EIO: Indian Ocean; NIO: Northern Indian Ocean; WIO: Western Indian Ocean; EPO: Eastern Pacific Ocean; WCPO: Western-central Pacific Ocean. RE: restriction enzymes.

Study	Regions sampled	Markers	Significant spatial heterogeneity
Díaz-Jaimes & Uribe-Alcocer (2006)	Within EPO	7 polymorphic microsatellites	Yes
Dammannagoda *et al.* (2008)	Within NIO	MtDNA ATPase 6 and 8 regions	Yes
		3 polymorphic microsatellites	Yes (weak)
Chow *et al.* (2000a)	EIO & WIO	MtDNA-RFLP analysis: 2 RE	No
Nishida *et al.* (2001)	EIO & WIO	NDNA-RFLP analysis: 2 RE	No
Qiu & Miyamoto (2011)	Within WCPO	10 polymorphic microsatellites	Yes
Li et al. (2015)	Within CPO	MtDNA CO-I gene segment	Yes
Kunal *et al.* (2013)	Within NIO	MtDNA CR segment	Yes (weak)
Aguila *et al.* (2015)	Within WCPO	9 polymorphic microsatellites	Yes
Grewe *et al.* (2015)	Within WCPO	SNP genotypes	Yes

1.4.3.1 Shallow genetic structure among yellowfin tuna populations

A shallow genetic structure among yellowfin tuna populations can be seen by the inconsistent results across population genetic analyses of neutral molecular markers, with a number of studies failing to detect the presence of significant population differentiation. This is likely a result of the species' large N_e/N_c ratio (where N_e refers to effective population size and N_c to census population size) resulting in ancestral patterns of gene flow among populations being retained, and consequently low levels of genetic drift occurring among populations.

A large N_e of yellowfin tuna populations was inferred by Ely *et al.* (2005). Yellowfin tuna has a high global abundance (large N_c), indicated by high global catches relative to other Scombrid and billfish species (with the exception of skipjack tuna). The species also has high N_e/N_c ratios, characteristic of Scombrids' opportunistic life history strategies, demonstrated by high haplotype diversities (e.g. Ely *et al.* 2005; Wu *et al.* 2010; Kunal *et al.* 2013) which indicate low variance in female reproductive success and high female N_e. Yellowfin tuna additionally has a high potential for dispersal of individuals among discrete units in its pelagic egg and larval life stages, and its highly migratory oceanic adult life stage (Collette *et al.* 2011), between the Atlantic and Indian Oceans around southern Africa, and between the Indian and western-central Pacific Oceans via Indonesian waters. The possibility of a few migrants per generation, combined with large N_e, is therefore the most likely explanation for the low levels of genetic differentiation among demographically isolated population units, rather than a sufficient level of gene flow to maintain demographic connectivity.

The uncertainty of yellowfin tuna population structure, based on its shallow genetic structuring and the weak resolving power of small numbers of neutral markers, suggest that NGS and population genomics approaches may be more effective in detecting spatial heterogeneity of this species among regions that may not be easily detected through traditional population genetics. This has been demonstrated by recent studies that applied large SNP arrays obtained through NGS approaches to investigating yellowfin tuna population structure. Grewe *et al.* (2015) detected significant differentiation between yellowfin tuna samples from the eastern and western-central Pacific Oceans, as well as within the western-central Pacific Ocean (Tables 1.1 and 1.2), for clustering analyses using a subset of SNP loci putatively under selection. Pecoraro *et al.* (2016a) detected significant differentiation among samples from the Atlantic, Indian and Pacific Oceans (Table 1.1), which was consistent across clustering and F_{ST} analyses.

1.5 *Atlantic and Indian Ocean management boundary*

The management boundary between the ICCAT and IOTC areas of competence occurs at 20°E at Cape Agulhas of South Africa, the approximate location of the confluence of the Atlantic and Indian Oceans. The western boundary of the IOTC originally occurred at 30°E, however in 1998 it was extended west to include the area off South Africa between 20°E and 30°E, as fisheries in this area were previously not under the mandate of either tRFMO (IOTC Secretariat 2014). The fine-scale genetic population structure between Atlantic and Indian

Ocean stocks of tuna and tuna-like species managed by ICCAT and the IOTC was therefore not incorporated into defining this management boundary. A recent investigation into the genetic structure of swordfish off South Africa suggests that the management boundary between the ICCAT and IOTC areas of competence does not reflect the biological population structure of the species, with the possibility of a mixed stock occurring off South Africa, and that the boundary dividing ICCAT and IOTC stocks of this species off South Africa should be revised (West 2016).

Delineation of the Atlantic and Indian Ocean yellowfin tuna populations by Ward *et al.* (1997) is based on western Atlantic Ocean and central Indian Ocean samples, and the fine-scale structure of yellowfin tuna between the tRFMOs' areas of competence is not known. It has been suggested that yellowfin tuna caught west of 20°E off the Western Cape province of South Africa, in waters of the south eastern Atlantic Ocean and thus in the ICCAT area of competence, may be migrants from the Indian Ocean rather than the Atlantic Ocean population (Hayasi 1974; Henriques 2011). The lack of spawning grounds of the species in South African waters indicates that these individuals are migrants to the region from the Atlantic or Indian Ocean (or a mixture of individuals from both), while the seasonality of regional catch and abundance suggests that they are of Indian Ocean origin. To date, however, sufficient genetic evidence does not exist to resolve the population origin of these individuals, and investigation into this is considered a research priority for South Africa (West *et al.* 2014).

1.5.1 Non-genetic evidence for Indian Ocean origin of Western Cape yellowfin tuna

Environmental conditions of waters off South Africa, particularly in the southern Benguela region of the south eastern Atlantic Ocean, are not suitable for the spawning of yellowfin tuna which spawns in tropical waters with stable, non-fluctuating conditions and temperatures of generally 24°C or higher, required for optimal growth and survival of the species' early life stages which lack the endothermic adaptations of juveniles and adults (Graham & Dickson 2001; Wexler *et al.* 2007, 2011; Reglero *et al.* 2014; Pecoraro *et al.* 2016b; Muhling *et al.* 2017). Gonadal surveys in the southern Benguela region off South Africa did not detect any ripe yellowfin tuna individuals (Shannon 1987), while larval surveys in this region and off the east coast of South Africa did not detect sufficient larvae to indicate spawning of yellowfin tuna (Shannon & Pillar 1986; Shannon 1987; Beckley 1995; Beckley & Leis 2000), suggesting a lack of spawning of this species off South Africa. This suggests that individuals

in the region are migrants from more tropical areas of the Atlantic or Indian Ocean, or a mixture of individuals from both oceans that return to tropical spawning grounds, and so do not facilitate gene flow between these oceans. Based on the seasonal variation of catches and environmental conditions, it is suggested that yellowfin tuna off South Africa are Indian Ocean individuals whose movement into the region is facilitated by the input warm waters from the Indian Ocean western boundary Agulhas Current.

The influence of Agulhas waters in the southern Benguela region exhibits seasonal and inter-annual variation; as does yellowfin tuna abundance in the region. Oceanographic features of the Agulhas Current that facilitate the input of warm waters into the southern Benguela region extend further west in summer months (Lutjeharms 2006; Dencausse *et al.* 2010), coinciding with increased abundance of yellowfin tuna in the region (Penney & Griffiths 1968). Unusually high yellowfin tuna abundance in this region observed in some years, for example 1979 and 2014 (Penney *et al.* 1992; Ndudane 2015), may be related to favourable environmental conditions for the species associated with a stronger influence of the Agulhas current. It has thus been suggested that individuals occurring in this region are migrants from the Indian Ocean (Hayasi 1974; Penney *et al.* 1992; West *et al.* 2014; Ndudane 2015; West & Kerwath 2015).

It is possible that Indian Ocean yellowfin tuna move into southern Benguela waters as feeding migrations. As pelagic predatory species, yellowfin tuna have high energy requirements, but the distribution of primary productivity in the oceanic environmental is sporadic (Roger 1994; Martin 2003; Potier *et al.* 2004; Tew Kai & Marsac 2010). Yellowfin tuna is therefore characterised by an opportunistic life history strategy, feeding on a wide range of prey organisms, with adults consuming various species of fish, crustaceans and cephalopods (Ménard *et al.* 2006; Juan-Jordá *et al.* 2013b; Zudaire *et al.* 2015; Pecoraro *et al.* 2016b), and shifting of main prey species relative to availability (Potier *et al.* 2004, 2007; Pecoraro *et al.* 2016b). They must locate areas rich in prey organisms for efficient feeding (Sund *et al.* 1981; Bertrand *et al.* 2002; Potier *et al.* 2004), and the distribution and abundance of yellowfin tuna (within their range of favourable environmental conditions) are therefore largely influenced by bottom-up control through prey density (Roger 1994; Bertrand *et al.* 2002; Schaefer *et al.* 2007, 2011). Yellowfin tuna therefore aggregate in regions where oceanographic features result in high productivity and high abundance of prey species, such as include ocean fronts, upwelling and convergence zones, gyres, eddies and seamounts (Sund *et al.* 1981; Blackburn *et al.* 1968, 1969; Kirby *et al.* 2000; Young *et al.*

2001; Tew Kai & Marsac 2010; Schaefer *et al.* 2011); in the Mozambique Channel, for example, yellowfin tuna are associated with eddies as a result of the foraging opportunities they present (Sabarros *et al.* 2009). Yellowfin tuna undertake vast advective migrations to feeding grounds, for example, making trans-Atlantic migrations from spawning grounds in the east to feeding grounds in the west (Maury *et al.* 2001). The southern Benguela region off South Africa is a highly productive area, and it is thus suggested that seasonal migrations and aggregations into the southern Benguela region occur to exploit the high levels of productivity and foraging opportunities available.

In summer months, this region is characterised by ocean fronts and intense upwelling, caused by the mixing of water from the south Atlantic Ocean, the Benguela Upwelling System, and the Agulhas Current and Agulhas Bank (Lutjeharms & Valentine 1984; Hutchings *et al.* 2009), as well as fronts at the edges of Agulhas Rings shed into the southern Atlantic Ocean (Hutchings *et al.* 2009). Productivity in this region is thus increased in summer months (Lutjeharms & Meeuwis 1987), presenting a suitable feeding ground for yellowfin tuna, and has been suggested to serve as a feeding ground for migratory tropical bigeye tuna from the Atlantic and Indian Oceans (Durand *et al.* 2005). The Agulhas Bank is an area of high productivity (Demarqc *et al.* 2003), with an abundance of the prey species of yellowfin tuna such as chokka squid *Loligo reynaudii* (Drapeau *et al.* 2004; Sauer *et al.* 2013), and serves as an important feeding ground for predatory species (Smale *et al.* 1994; Drapeau *et al.* 2004; Götz *et al.* 2014); it is therefore expected to provide an important foraging area for migrant yellowfin tuna to the region. The large size of individuals caught in this region and the compact migrations that have been observed (Talbot & Penrith 1968) support this hypothesis.

Tagging data of yellowfin tuna from the Indian Ocean Tuna Tagging Project (IOTTP), which took place from 2002 to 2009, supports the suggestion that yellowfin tuna migrate to the southern Benguela region off South Africa from the Indian Ocean (Murua *et al.* 2015). No analyses or quantification of yellowfin tuna movement within the Indian Ocean have been published from tag-recapture data of this programme, however data presented show that a small proportion of individuals tagged in the region of Tanzania, in the equatorial western Indian Ocean, were recaptured off South Africa in the region of 20°E (Eveson *et al.* 2015; Murua *et al.* 2015; IOTC-SC17 2014; IOTC-WPTT18 2016), demonstrating the occurrence of movement of yellowfin tuna individuals from the Indian Ocean basin to the southern Benguela region.

1.5.2 Genetic evidence for Indian Ocean origin of Western Cape yellowfin tuna

Henriques (2011), investigating the fine-scale structure between the Atlantic and Indian Ocean yellowfin tuna populations, found weak but significant evidence to support the hypothesis that yellowfin tuna caught off the Western Cape of South Africa, in the ICCAT area of competence, are migrants of Indian Ocean origin. This finding is based on population genetic analyses of South African samples from the ICCAT (Western Cape) and IOTC (Eastern Cape and KwaZulu-Natal) areas of competence, as well as from the western Atlantic Ocean (Brazil) and western central Indian Ocean (Seychelles, La Réunion and Kerala, south west India), at the mtDNA COI region and six polymorphic microsatellite loci. COI analyses revealed weak, but significant and consistent, evidence to suggest the that Western Cape yellowfin tuna form part of the Indian Ocean population, with global and pairwise F_{ST} tests revealing significant differentiation between the sample from Brazil and the remaining samples (with the exception of the Seychelles sample, although this was likely due to its small size), with no significant differentiation detected among remaining samples. This result was weakly supported by reconstruction of phylogeographic relationships among haplotypes. Results of microsatellite analyses provided weak but inconclusive support for the pattern of differentiation revealed by mtDNA analyses; exact tests of differentiation revealed significant differentiation between the Brazil sample and remaining samples, with no differentiation among remaining samples, however F_{ST} tests detected significant differentiation only between the Brazil and KwaZulu-Natal sample and Bayesian clustering and assignment analyses did not detect genetic structuring among samples.

While results of Henriques (2011) suggest that yellowfin tuna caught off the Western Cape of South Africa are an extension if the Indian Ocean population, limitations of the markers employed did not allow for potential fine-scale structure between these populations to be assessed. The confluence of the Atlantic and Indian Oceans occurs off the Western Cape of South Africa, and it is thus possible that fine-scale genetic structure of individuals from the populations of these oceans occurs in this region. The low level of variability of the mtDNA markers, and the small number of nDNA gene loci screened, result in limited power to detect small but potentially significant genetic structuring between the Atlantic and Indian Ocean populations in this region, such as locally adapted gene complexes or admixture of individuals. Population genomic analysis using highly variable, genome-wide molecular

makers may allow for the fine-scale structuring of yellowfin tuna populations within this region, and therefore for the boundary between these populations, to be resolved.

1.6 *Aims and objectives*

Yellowfin tuna is a very important global fisheries species, however knowledge regarding the species' population structure is incomplete due to shallow genetic divergence and weak structuring, and as a result fisheries management units cannot be defined according to demographic connectivity among regions. The population boundary between Atlantic and Indian Ocean yellowfin tuna is thus a research priority, with both fisheries management and conservation implications for this heavily exploited species. The increased fishing pressure and the current status of the Atlantic and Indian Ocean stocks make it essential that the boundary between these stocks is defined, in order for stocks to reflect discrete populations for accurate stock assessment and effective management by their tRFMOs.

The overall aim of the present study was therefore to assess the fine-scale genetic population structure of yellowfin tuna caught in the waters off southern Africa, to determine the origin (Atlantic Ocean stock or Indian Ocean stock) of individuals occurring in this region, and to make recommendations regarding their management. Assuming that yellowfin tuna populations in the Atlantic Ocean and Indian Ocean are distinct, the null hypothesis of the study was that individuals caught off the Western Cape province in the ICCAT area of competence (west of 20°E) were of Atlantic Ocean origin, whereas individuals caught east of this point in the IOTC area of competence were of Indian Ocean origin. The alternative hypothesis was that all individuals caught in South African waters are of Indian Ocean origin, or that they represent a mixed stock of individuals from both oceans, or an admixed stock of hybrid individuals from both Atlantic and Indian Ocean populations.

This research employed a population genomics approach, specifically genome-wide SNP analysis using a GBS approach, to overcome the resolution limitations in previous genetic studies and provide a powerful test of the alternative hypothesis. Population genetic analyses of the same samples were also conducted using a sequence of the mtDNA Control Region (CR), a neutral molecular marker that has been used in investigations of yellowfin tuna population structure by previous studies, in order to compare the resolving power of genome-wide SNP genotypes to this traditionally used marker. This marker was additionally required to verify that all tissue samples were from yellowfin tuna.

A sample of yellowfin tuna from the eastern Atlantic Ocean (Ghana) was included in analyses to address the possibility of sub-structuring within the Atlantic Ocean.

Chapter 2 outlines the methods employed to test population structure, including the sampling process, the sequencing and genotyping of molecular makers and the population genetic and genomic analyses applied to markers.

Chapter 3 presents the results of genetic and genomic analyses of population structure of yellowfin tuna, using both indirect and direct approaches that allow for genetic differentiation, mixture or admixture within and among samples to be detected.

Chapter 4 provides a discussion of the results obtained, explores possible explanations for patters observed, makes recommendations regarding the stock structure of yellowfin tuna off South Africa based on these results, and discusses the use of NGS and population genomic approaches in investigations of the population structure of this species.

CHAPTER 2
MATERIALS AND METHODS

2.1 *Sampling*

2.1.1 Sampling sites

Tissue samples of yellowfin tuna were taken from sampling sites around South Africa, as well as from the species' wider Atlantic and Indian Ocean distributions. A sample was taken from the Western Cape province of South Africa (WC), representing yellowfin tuna caught in the Atlantic Ocean waters of South Africa and assessed and managed as part of the Atlantic Ocean operational stock by the International Commission for the Conservation of Atlantic Tunas (ICCAT) (Figure 2.1, Table 2.1). Samples were taken from the Eastern Cape (EC) and KwaZulu-Natal (KZN) provinces of South Africa to represent individuals caught in the Indian Ocean and managed by the Indian Ocean Tuna Commission (IOTC) as part of the Indian Ocean stock (Figure 2.1, Table 2.1). Additional samples were collected from the Ghana in the Gulf of Guinea (GG) and from Ponta do Ouro in Mozambique (MOZ) to represent the Atlantic and Indian Ocean spawning populations, respectively, as the Gulf of Guinea is recognised as the main spawning ground for yellowfin tuna in the Atlantic Ocean while the Mozambique Channel is a known secondary spawning ground for the species in the Indian Ocean (ICCAT 2016b; IOTC 2014).

Tissue samples were collected in the form of fin clippings, fixed in 95% ethanol immediately after capture or as soon after capture as possible, as high quality DNA that has not been subject to degradation at large portions of the genome is required for restriction-site associated DNA (RAD) sequencing, to ensure a low level of missing data (Graham *et al.* 2015). The EC and KZN samples were collected by the author and collaborators, respectively, at fishing competitions during the course of the study, the GG and MOZ samples were collected by collaborators during research trips during the course of the study, and the WC sample was collected by collaborators during research trips prior to the study commencing. Yellowfin tuna individuals of multiple size classes were present among samples. A summary of sampling information is presented in Table 2.1, with more detailed information in Table 5.1 of Appendix 1.

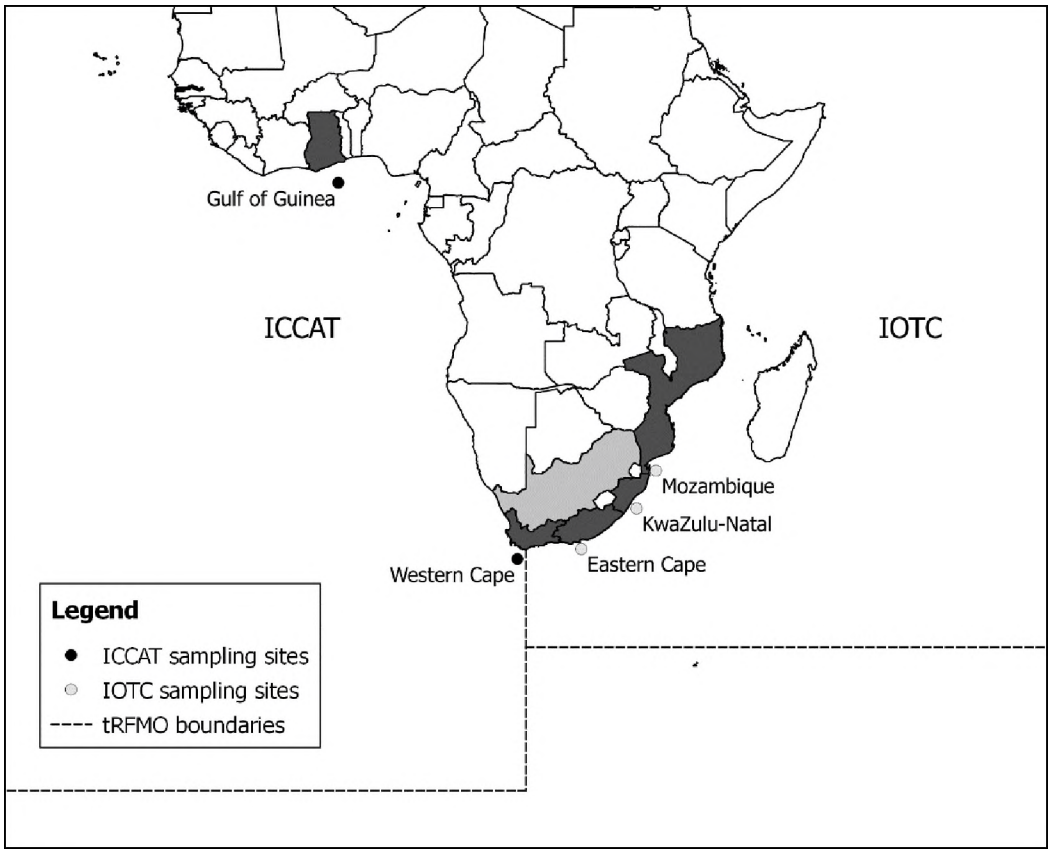

Figure 2.1: Locations of samples obtained from the Atlantic and Indian Oceans, within the ICCAT and IOTC areas of competence respectively, with management boundaries of these tRFMOs shown; boundary between the tRFMOs occurs at 20°E. Map created in QGIS v2.18.2 (QGIS Development Team 2016).

2.1.2 DNA extraction

DNA was extracted from the tissue samples of 96 yellowfin tuna individuals: 20 each from the GG, KZN and MOZ samples, 19 from the WC sample and 17 from the EC sample. It is important for RAD sequencing that no contamination of tissue samples occurs, as the restriction enzyme which cuts the genome at recognition sites for sequencing is not species-specific. DNA extraction protocols commonly used for analyses of ancient DNA were thus employed, as described in Nyström *et al.* (2010), including extraction in a dedicated laboratory, and the decontamination of all surfaces and equipment used.

To obtain DNA of a high quality, a pilot study was conducted whereby DNA of a number of tissue samples was extracted following two different protocols: using a QIAGEN DNAeasy^{TM} kit, and by a more traditional Phenol/chloroform/isoamyl alcohol (PCIA) method, according to the methods of Winnepenninckx *et al.* (1993). The quality and amount of DNA obtained using each method was tested by (1) visualisation by agarose gel electrophoresis and (2) quantification of RNA:DNA using a nanodrop. This process found that the PCIA extraction method resulted in DNA of a higher concentration and molecular

weight, and this method was therefore used for the extraction of DNA for the remaining tissue samples for use in subsequent sequencing reactions.

Table 2.1: Sampling strategy for yellowfin tuna from the Atlantic and Indian Oceans. **N**: sample size; **CR**: mitochondrial DNA Control Region; **SNPs**: nuclear DNA single nucleotide polymorphisms; **FL**: fork length.

Ocean	Region	Sample code	Approximate site co-ordinates	Dates of collection	N		FL range (cm)
					CR	SNPs	
Atlantic Ocean	Gulf of Guinea	GG	Ghana: 2.500°N 0.308°W	August 2015	20	20	35 – 45*
	Western Cape, South Africa	WC	Hermanus: 35.617°S 18.700°E	October 2014	16	13	111 – 153
Indian Ocean	Eastern Cape, South Africa	EC	Port Elizabeth: 34.726°S 25.180°E Port St Francis: -34.738°S 24.902°E	June – July 2015	17	16	93 – 100
	KwaZulu-Natal, South Africa	KZN	Protea Banks: 30. 833°S 30.483°E Richards Bay: -28.829°S 32.173°E	May – July 2015	20	20	58 – 125
	Mozambique	MOZ	Ponto do Ouro: 26.537°S 34.500°E	August 2015	20	20	58 – 68*

*FL not known for all individuals sampled

2.2 *Mitochondrial DNA Control Region*

2.2.1 Control region sequencing

A fragment of the mitochondrial DNA (mtDNA) Control Region (CR) was amplified by polymerase chain reaction (PCR) for 96 yellowfin tuna individuals from five sampling locations (Table 2.1) using species-specific primers developed from sequences on GenBank: forward primer: 5'-TCCTACCCCTAACTCCCAAAG-3'; and reverse primer: 5'-AAACTGTGGGGATTCTCAC-3'. PCR was performed in 10µL reaction solution volumes, each containing 5µL 5x BioMix™ (Bioline, comprised of a reaction buffer, $MgCl_2$, dNTPs, *Taq* DNA polymerase and additives), 0.5 µL each of forward and reverse primer (10% dilution), 2µL DNA (20% dilution) and 2µL double distilled water.

PCR amplification was conducted on a BioRad C1000 Touch™ thermal cycler under the following cycling conditions: an initial denaturation step at 95°C for 60 seconds, followed by 35 cycles of denaturation at 95°C for 90 seconds, annealing at 60°C for 60 seconds and extension at 72°C for 60 seconds, and a final extension step at 72°C for three minutes, after which the product was held at °4C. The success of PCR amplification was tested using gel electrophoresis to visualise the length of fragments: 1µL 5x DNA Loading Buffer BlueTM (Bioline) was added to 2µL of each PCR product, which was loaded into wells of 2% agarose gel alongside a 50bp HyperLadder™ (Bioline). The expected PCR product was approximately 900bp in length, therefore a strong band of DNA against this point of the ladder indicated that the PCR was successful in amplifying the desired fragment.

PCR products were purified to remove excess primers and unused nucleotides. The details of the purification process are described in Appendix 2. Amplicons were sequenced using the forward primer with 6µL reaction mixes, each comprised of 1µL purified PCR product, 1µL forward primer and 4µL double distilled water following AB BigDye Technology on an ABI 3500 sequencer.

2.2.2 Control Region genetic analysis

Control Region sequences were aligned using ClustalW (Thompson *et al.* 1994) in BioEdit Sequence Alignment Editor v7.2.5 (Hall 1999) and manually edited in BioEdit by inspection of sequence chromatograms in Chromas v2.5.0 (Technelysium Pty Ltd). Sequences were entered into the Basic Local Alignment Search Tool (BLAST) available online at https://blast.ncbi.nlm.nih.gov/Blast.cgi, to confirm that all sequences were from yellowfin tuna individuals, as yellowfin tuna closely resemble bigeye tuna *Thunnus obesus*, especially as juveniles, and are therefore often misidentified. Thereafter sequences were analysed for levels of genetic diversity, potential genetic differentiation among samples within the dataset and demographic history of the species.

2.2.2.1 Genetic diversity

Genetic variation was quantified in Arlequin v3.4.2.2 (Excoffier & Lischer 2010) using estimates of standard genetic diversity indices. The number of haplotypes present in the dataset was determined, and the number of these that were unique (present only in one individual in the dataset), private (present only in one sample but in more than one individual within the sample) and shared (present in individuals from different samples). Haplotype diversity (*h*; Nei 1987), the probability (from 0.0 to 1.0) that two randomly drawn haplotypes

are not identical, and nucleotide diversity (π; Tajima 1983), the probability (from 0.0 to 1.0) of nucleotides differing between homologous sites of two randomly drawn haplotypes, and therefore the level of variation present, were estimated.

2.2.2.2 Population structure

The haplotype dataset was assessed for evidence of genetic population structure among individuals, to test the alternative hypothesis of the WC sample being part of the Indian Ocean population, by means of hierarchical analyses of molecular variance (AMOVAs), measures of global and pairwise tests of genetic differentiation among samples, and reconstruction of relationships between haplotypes. The details of these analyses are described below.

Analyses of molecular variance

Hierarchical AMOVAs were performed among haplotypes, in Arlequin, to test two alternative hypotheses of population structure within the dataset. Hypothesis 1 grouped samples into one group comprising of the Atlantic Ocean samples (GG and WC) and another comprising of the Indian Ocean samples (EC, KZN and MOZ), while Hypothesis 2 grouped the WC sample with the Indian Ocean samples, as described in Table 2.2.

Table 2.2: Groupings of samples under two hypotheses of population structure within the dataset, each tested for structure by three-level AMOVAs.

	Atlantic Ocean grouping	**Indian Ocean grouping**
Hypothesis 1	GG and WC	EC, KZN and MOZ
Hypothesis 2	GG	EC, KZN, MOZ and WC

For each grouping hypothesis, the total variance within the dataset was partitioned into covariance components at three hierarchical levels: variance between groups defined by each hypothesis (between-group variance), variance among samples within defined groups (within-group variance) and variance among individuals within samples (within-sample variance). Each covariance component, and its percentage of the total variance, was estimated based on its sum of squares. The fixation index (F-statistic; Wright 1951, 1965) associated with each covariance component was estimated based on the ratios of these components, and the significance of each covariance component and its associated F-statistic was tested by 10 000 non-parametric permutations of the relevant haplotypes and populations (Excoffier *et al.* 1992).

F_{ST} tests of genetic differentiation

Global genetic differentiation within the dataset and pairwise genetic differentiation between samples were computed in Arlequin, as measured by Φ_{ST} values, using pairwise genetic distances among haplotypes.

Significance (at α=0.05) of Φ_{ST} values was assessed after 10 000 permutations of haplotypes to obtain a null distribution of F_{ST} values expected under the hypothesis of no differentiation, with the resulting P-value the proportion of permuted Φ_{ST} values greater than or equal to that observed. The critical significance level for pairwise Φ_{ST} value was corrected for multiple comparisons using sequential Bonferroni correction following the methods of Rice (1989).

Exact tests of genetic differentiation

Pairwise exact tests of differentiation were conducted in Arlequin to test the null hypothesis of a random distribution of haplotypes between each pair of samples by constructing contingency tables of haplotypes x samples for each pair of samples (Raymond & Rousset 1995). A random walk between all states of a Markov chain (MC), with a length of 100 000 steps (following a dememorization of 1 000 steps), explored all potential states of each table (Excoffier & Lischer 2010), with the resulting P-value being the probability of observing a table configuration less than or equally to that expected under the null hypothesis of panmixia (critical α=0.05).

Reconstruction of haplotype phylogenies

A median-joining (MJ) maximum parsimony haplotype network was reconstructed in Network v5.0.0.0 (Bandelt *et al.* 1999) to describe the evolutionary relationships among haplotypes within the dataset, and identify geographic patterns among haplotypes which may provide evidence of potential geographic population structure (Posada & Crandall 2001).

2.2.2.3 Demographic history

Demographic histories of putative yellowfin tuna populations (defined based on genetic structure results for the CR) were inferred based on genetic diversity indices, neutrality test statistics, and haplotype mismatch distributions.

The genetic diversity indices h and π (as described in section 2.2.2.1) may be influenced by past changes in population size, and as a result their values, and the relationships between them, can provide indications of populations' demographic histories. Grant and Bowen

(1998) described four categories of demographic histories for marine fish species' populations based on these two parameters, as described in Table 2.3.

Table 2.3: Demographic histories of fish population associated with different levels of *h* and π, as described by Grant and Bowen (1998). Adapted from Henriques (2011).

	Low π (<0.5%)	**High π (>0.5%)**
Low *h* (<0.5)	Suggests recent bottleneck or founder event	Suggests recent bottleneck event acting on population with a previously large N_e, or secondary contact
High *h* (>0.5)	Suggests historical rapid population expansion event from small N_e (possibly following a bottleneck)	Suggests population at demographic equilibrium with large, stable N_e, or secondary contact

Significance of Tajima's *D* and Fu's F_S neutrality statistics may occur as a result of selection acting on a gene region, or as a result of past demographic changes in population sizes. Tajima's *D* statistic (Tajima 1989) tests for selective neutrality based on segregating differences among haplotypes, while Fu's F_S tests for neutrality based on the number of alleles observed (Fu 1997). These indices were estimated, and their significance (at α=0.05) assessed after 10 000 permutations.

Negative values of these test statistics indicate an excess of recent mutations, and are also compatible with scenarios of past population expansions.

Mismatch distribution parameters were calculated in Arlequin to test the null hypothesis of populations having undergone recent demographic expansions. Populations that have been at demographic equilibrium for a long period of time are expected to exhibit stable bi- or multimodal distributions of pairwise haplotypic differences, while populations that have undergone demographic expansions are expected to exhibit unimodal distributions (Slatkin & Hudson 1991; Rogers & Harpending 1992). The expected (null) distribution of pairwise haplotypic differences under the sudden expansion model was estimated for each putative population and plotted against the observed distribution of haplotypic differences. For each putative population, the sum of squares deviation (SSD) between the observed and expected distributions was calculated to test the validity of the expansion model, and Harpending's raggedness index (*Hri*; Harpending 1994) was calculated to test the goodness of fit of the observed distribution to the null distribution. Significance (at α=0.05) of the SSD and *Hri*

statistics was assessed after 10 000 bootstrap replicates using a parametric bootstrapping approach (Excoffier 2004).

If SSD and *Hri* test statistics found the mismatch distributions of putative populations to be consistent with those expected following historical expansion, demographic expansion parameters were estimated based on the means and variances of mismatch distributions using a generalised non-linear least-square approach (Rogers 1995). These parameters are time since expansion in mutational units (τ, Li 1977) and the mutation parameters before (θ_0) and after (θ_1) the expansion event. Time, in generations, since population expansion was estimated using the equation $\tau=2\mu t$ (Rogers 1995), where t is the time since population expansion in generational units, and 2μ is the mutation rate over the entire haplotype per generation, estimated as $2\mu=um_T$, where u is the mutation rate per nucleotide per generation and m_T is the number of base pairs in the haplotype. This was estimated using the mismatch calculator spreadsheet developed by Schenekar and Weiss (2011) to reduce the rate of errors in expansion time estimated based on mismatch distribution parameters. The generation time was set at 2.5 years, as estimates of the length at which 50% of individuals have reached maturity (L_{50}) has been estimated to correspond to ages of approximately 2 to 3 years in Atlantic and Indian Oceans (Zudaire *et al.* 2013; Diaha *et al.* 2015; Pecoraro *et al.* 2016b).The mutation rate was set at 3.69% per million years, as estimated by Donaldson and Wilson (1999) for the CR of a number of teleost fish species.

2.3 *Nuclear DNA single nucleotide polymorphisms*

2.3.1 SNP genotyping

Extracted and purified DNA (see 2.1.2) of 96 yellowfin tuna individuals was subjected to Tunable Genotyping-by-Sequencing (tGBS®) technology (Ott *et al.* 2017), carried out by Data2Bio (http://www.data2bio.com/) to develop genome-wide single nucleotide polymorphism (SNP) genotypes. tGBS is a form of GBS, a highly multiplexed NGS approach with a high GRL (see section 1.4.2.1.1, resulting in fewer cut sites being obtained and sequenced, allowing for a greater read depth per cut site and therefore more robust genotyping with a reduced rate of genotyping error, and a lower rate of missing data across individuals (Ott *et al.* 2017).

Sequencing library preparation for tGBS involved digestion of genomic DNA with two restriction enzymes (*NspI* and *BfuCI*) resulting in cut fragments to which an individual-specific barcoded oligonucleotide (for downstream demultiplexing) and a universal

oligonucleotide (for sequencing) were ligated at the 3' and 5' overhangs respectively. Cut fragments were pooled and two rounds of PCR amplification were performed (see Ott *et al.* (2017) for a detailed description) resulting in a sequencing library. Sequencing of cut sites was then performed using two runs on an Ion Proton™ System, resulted in raw sequence reads on which bioinformatics analyses were performed at Data2Bio.

Sequence reads for each individual were separated by barcode sorting, and adaptor sequences were removed. Fragments were subjected to quality filtering and those containing bases of low sequence quality were removed. Homologous fragments across individuals were aligned to a reference genome of Pacific bluefin tuna *Thunnus orientalis* (Nakamura *et al.* 2013). The depth of SNP stacks (the number of homologous sequence fragments per individual) was set to a minimum of 25, and SNPs were called by identifying base pairs with alleles that differed among individuals above a minor allele frequency (MAF) threshold of 1%. Identified SNPs were assembled to form genome-wide SNP genotypes from homologous genome regions for all individuals.

2.3.2 SNP genetic analysis

2.3.2.1 Genetic diversity

Indices of genetic diversity were calculated for each sample as an average of the indices for each SNP locus, as mean number of alleles (MNA), mean allelic richness (A_R), observed and expected heterozygosities (H_O and H_E respectively) using the *diveRsity* package (Keenan *et al.* 2013) in R v3.2.2 (R Core Team 2015). Wright's inbreeding coefficient (F_{IS}) values were estimated in per locus for each population in Arlequin, with significance assessed after 10 000 permutations and corrected with Bonferroni correction.

Conformation to Hardy-Weinberg Equilibrium (HWE) expectations within each sample was assessed in Arlequin following Guo and Thomson (1992). Loci were assessed for significant departure from HWE expectations after sequential Bonferroni correction of the critical significance level (Rice 1989).

Tests of linkage disequilibrium (LD) were conducted in Arlequin, to test for significant non-random association of alleles at different loci, for each pair of loci in each sample for smaller datasets. For each pair of loci, the probability of LD being absent was estimated using exact tests to obtain a P-value for each loci pair. Loci pairs were assessed for significance of LD occurring after sequential Bonferroni correction of the critical significance level (Rice 1989).

2.3.2.2 Population structure

In order to detect possible signals of population structure, the dataset of SNP genotypes was analysed for population genetic structure using analyses of molecular variance (AMOVAs), pairwise tests of population differentiation, Bayesian clustering and assignment analyses and multivariate analyses.

Analyses of molecular variance

Hierarchical AMOVAs were performed in Arlequin to test two alternative hypotheses of population structure as described in Table 2.2. Methods for estimating covariance components at three levels (between-group variance, within-group variance and within-sample variance) and each covariance component's percentage of the total variation and associated F-statistic, and for assessing significance (at α=0.05) were the same as those described for the mtDNA CR dataset (section 2.2.2.2). Loci with more than 5% missing data were excluded from analyses.

F_{ST} tests of genetic differentiation

Global and pairwise genetic differentiation, as measured by Wright's F_{ST}, were estimated in Arlequin, based on the methods described by Weir and Cockerham (1984) to test the null hypothesis of samples being drawn from a single, randomly breeding population. F_{ST} measures the proportion of the total variance in a dataset that can be attributed to variance among samples, based on allele frequencies and observed heterozygosities (H_O) in a dataset, under the assumption that an isolated population at equilibrium will have a characteristic set of alleles as a result of allele fixation by genetic drift. As heterozygosity within a dataset increases, F_{ST} decreases. F_{ST} values range from 0.0 (indicating no genetic differentiation among samples) to 1.0.

Significance of F_{ST} values, at α=0.05 for global estimates and using sequential Bonferroni correction (Rice 1989) for pairwise estimates, was assessed according to the methods described for the mtDNA CR dataset (section 2.2.2.2). Loci with more than 5% missing data were excluded from analyses.

Exact tests of genetic differentiation

Global and exact tests of population differentiation were conducted in Genepop v4.5.1 (Rousset 2008). Global tests of genic differentiation tested the null hypothesis of all alleles in the dataset being drawn from the same distribution, by construction of a contingency table of all allelic states for each locus. The overall differentiation at each locus was computed using a

Markov chain (MC) algorithm (of 100 batches of 5 000 iterations each, following a dememorization of 10 000 steps), to assess all possible states and compute an unbiased P-value representing the probability of alleles being drawn from the same distribution (Raymond & Rousset 1995). Overall significance (at α=0.05) was obtained using Fisher's method (Fisher 1935) to combine P-values across all loci. The level of differentiation was assessed as being highly significant if any tests being combined had a P-value of zero.

Pairwise exact tests of genic differentiation to test for sources of potential differentiation in the dataset obtained P-values per sample pair following the same methods as for the global test, with the MC algorithm computing an unbiased P-value for each pairwise comparison at each locus' contingency table rather than an overall P-value.

The same methods were followed for global and pairwise tests of genotypic differentiation, however these tests were concerned with diploid genotypic differentiation (the combination of alleles) per locus, and tested the null hypothesis of genotypes being drawn from the same distribution using contingency tables for genotypes at each locus.

Bayesian clustering and assignment analysis

The model-based Bayesian clustering method implemented in the program Structure v2.3.4 (Pritchard *et al.* 2000) was used to investigate genetic structure among samples. The analysis aims to identify the most probable number of genetic clusters (populations), K, within a dataset and assign each individual a membership probability to each of the identified clusters. Bayesian clustering in Structure does not require sampling location information and thus allows for the analysis of population differentiation without any potential bias associated with *a priori* partitioning of individuals into samples.

The clustering and inherent assignment are based on an algorithm that minimises Hardy-Weinberg and linkage disequilibria within clusters (Pritchard *et al.* 2000). The programme simultaneously estimates the values of the model parameters – the allele frequencies of each population and the population of origin of each individual – through a number of repeats of a Markov chain Monte Carlo (MCMC) algorithm, assuming for each parameter that the other parameter is known. This is done for a number of hypotheses of the number of clusters present (K-values), and for each K-value each individual in the dataset is assigned a membership probability (q-value) to each cluster, representing the posterior probability that its genotype originated in each of the clusters based on clusters' allele frequencies (Pritchard *et al.* 2000).

A posterior probability (lnPr) is estimated for each K-value. The most likely K-value for the dataset was identified as that with the highest associated Evanno's ΔK statistic (Evanno *et al.* 2005), which was calculated in Structure Harvester web v0.6.94 (Earl & vonHoldt 2012) based on the rate of change in lnPr values with each successive K-value tested (averaged across all runs). The ΔK statistic cannot identify the most likely K-value in the absence of structure (where true K=1), therefore in cases where K=2 presented the highest ΔK value, the most likely K-value was determined based on the assignment probabilities of individuals for K=2; if no individuals were assigned to either cluster with q>0.8, K=1 was identified as being most likely (Grewe *et al.* 2015). A hierarchical approach to determine K was implemented: if structure was detected (K>1), each cluster identified was analysed to test for further structuring which may be weaker than the major structure and therefore not detectable through analysis of the entire dataset (e.g. Vähä *et al.* 2007).

Parameters of analysis:

The SNP dataset was tested for structure for K=1 to K=3, with three runs of the analysis per K-value. The "No admixture" ancestry model was used, which assumes that each individual originated in one of K discrete populations and is recommended for datasets with weak structure (Pritchard *et al.* 2000), as was expected due to the large N_e of yellowfin tuna population sizes and associated weak genetic drift effects on the change in allele frequencies (Ely *et al.* 2005). Sampling location information was included in the model using the "Use LOCPRIOR" option to assist, without introducing bias, in detecting clustering where the level of genetic differentiation is weak (Hubisz *et al.* 2009). The "Correlated allele frequencies" model was used, which assumes that allele frequencies among each of the K populations will be similar due to independent drift from allele frequencies of an ancestor population (Falush *et al.* 2003).

Pilot runs were conducted to select appropriate burn-in and run lengths of the MCMC algorithm. The algorithm begins with a random set of parameters and with each repeat of the algorithm the resulting parameters and lnPr depend only on those of the previous repeat. Each MCMC run consisted of a burn-in of 200 000 steps, to allow parameters and lnPr to reach equilibrium and minimise the effects of their starting configuration, followed by 300 000 steps, which resulted in similar lnPr values among runs for each K-value, indicating it was sufficiently long for accurate model parameter estimates to be obtained (Pritchard *et al.* 2000).

Assignment tests

Individual assignment tests were implemented in GeneClass v2.0 (Piry *et al.* 2004) to assign yellowfin tuna individual to their most likely population of origin, with the GG and MOZ samples used as Atlantic and Indian Ocean reference populations respectively.

An initial self-assignment test was performed for individuals of the GG and MOZ samples, to ascertain the level of differentiation between the Atlantic Ocean sample and the most easterly Indian Ocean sample. Following the high levels of self-classification to each of these two samples (see Results section 3.2.2.2), a second batch of assignment tests was performed wherein individuals from the remaining samples (WC, EC and KZN) were treated as unknown fish and assigned to either the GG or MOZ sample. In all cases assignment probabilities were estimated according to the Rannala and Mountain (1997) computation approach and the Paetkau *et al.* (2004) MC resampling algorithm with 10 000 simulations.

A self-classification test of individuals excluding the GG sample was performed, to assess the level of mixing of individuals among these samples. The WC, EC, KZN and MOZ samples were used as reference populations, and assignment probabilities of individuals to each population were calculated using the same parameters described above.

Multivariate analyses

Multivariate analyses were implemented in the *adegenet* package v2.0.1 (Jombart 2008; Jombart & Ahmed 2011) in R (R Core Team 2015), to identify genetic clusters and describe the relationships among them, based on allele frequencies, in each dataset. An advantage of multivariate analyses over Bayesian clustering algorithms are not based on the assumptions of conformance to HWE and absence of LD in clusters/populations (Jombart *et al.* 2009; McKeown *et al.* 2015), and can therefore identify weak structuring which may exist within a dataset.

A principal component analysis (PCA) was performed by transforming the allele frequency data into principal components (PCs), linear combinations of alleles that maximise the total variation in the data (Jombart *et al.* 2009). Individuals were plotted against these PCs to visualise clustering among samples.

A Discriminant Analysis of Principal Components (DAPC; Jombart *et al.* 2010) was then performed to describe the variation between genetic clusters, by identifying clusters and maximising the variation among, and minimising the variation between, these clusters. Clusters were defined using a k-means clustering algorithm (with the "find.clusters" function

in *adegenet*); this algorithm maximises the allele frequency variation among, and minimises the variation within, clusters, for successive hypotheses of the number of clusters (k-values) within a dataset. The clustering solution with the best fit to the data is defined as that with the lowest corresponding Bayesian Information Criterion (BIC) value. The DAPC was then performed by transforming a number of PCs into discriminant functions (DFs), linear synthetic variables that maximise the allele frequency variation among, and minimise this variation within, identified clusters. The number of PCs retained in the DAPC was selected using a cross-validation tests which identifies the optimal number of PCs required to maximise the predictive success of the PCA while minimising the associated error (Jombart 2008; Jombart & Ahmed 2011). Individuals were plotted against DFs to visualise the relationships among genetic clusters within the dataset.

Identification of most informative loci

The coefficients of the variables (alleles) used to define linear DFs of DAPCs, referred to as "loadings", can be used to assess the contribution of alleles to the DFs. Alleles with the greatest loadings account for the greatest proportion of among-cluster variance described by the DFs, and these can therefore be used to identify the most informative loci in the dataset. The "loadingplot" function in *adegenet* was used to identify loci whose alleles contributed the greatest proportion of variation to the genetic structure observed (Jombart 2008; Jombart *et al.* 2010).

2.3.2.3 Comparison of neutral and outlier SNPs

Outlier tests were performed to identify SNP loci putatively under selection, whose allele frequencies in populations may be influenced by the non-neutral process of selection, and putatively neutral loci, whose allele frequencies may be influenced by the neutral processes of gene flow and genetic drift. Results of these tests were used to define a number of datasets, comprising of different subsets of SNP loci, to compare the differences in resolving power between putatively neutral and putatively adaptive genome regions, as well as to identify the most variable loci in the dataset.

Outlier tests

F_{ST} outlier tests were used to identify SNP loci with levels of genetic differentiation (measured by F_{ST}) substantially greater or smaller than the predicted distribution under neutrality, and whose allele frequencies may be under the effects of diversifying or balancing selection (Beaumont & Nichols 1996; Narum & Hess 2011; Russello *et al.* 2012; Lotterhos &

Whitlock 2014). It is expected that at neutral genomic regions, allele frequencies (and resulting F_{ST} values) are influenced by neutral processes and as such all neutral loci are expected to exhibit similar F_{ST} values. At loci subject to selection pressures, and those in linkage with such loci, allele frequencies are expected to be affected by locus-specific effects different from those acting on the rest of the genome, and their F_{ST} value are expected to differ significantly from the genome-wide average (Lewontin & Krakauer 1973; Beaumont & Nichols 1996; Lotterhos and Whitlock 2014). Two F_{ST} outlier tests were used to identify loci with F_{ST} values that differ significantly from the expected range of neutral values.

FDIST2 outlier detection

The FDIST2 outlier detection method (Beaumont & Nichols 1996) was implemented in Lositan (Antao *et al.* 2008). A null distribution of the level of genetic differentiation expected under neutral processes was estimated. The genome-wide mean (putatively neutral) F_{ST} value was estimated using the "Neutral mean F_{ST}" option, following Russello *et al.* (2012). A null model distribution of the mean F_{ST} as a function of expected heterozygosity (H_E) under an island model of population subdivision (Wright 1931) was estimated using 50 000 simulations and plotted with 99% confidence intervals. Each locus' F_{ST} as a function of heterozygosity was plotted against this null distribution, with loci falling above the upper 0.99 quantile identified as being putatively under diversifying selection, and those below the 0.01 quantile identified as putatively under balancing selection.

BayeScan outlier detection

A second outlier test was implemented in BayeScan (Foll & Gaggiotti 2008). This method detects loci potentially under selection based on their F_{ST} values however it does not assume a specific demographic history, but rather assumes independent divergence of all individuals from a common ancestor (Lotterhos & Whitlock 2014), and uses a Bayesian approach to estimate the posterior probability that each locus is under selection (Foll & Gaggiotti 2008). Two evolutionary models were tested: a null model including only population-specific (neutral) effects, and an alternative model including population-specific and locus-specific (selective) effects. The posterior probability of each alternative model was estimated per locus by a reversible-jump Markov chain Monte Carlo (MCMC) approach of 20 000 pilot runs of 20 000 iterations each, and 750 000 MCMC iterations (sample size of 50 000 and thinning interval of 15), as well as a burn-in of 50 000 iterations. Loci with posterior probabilities of $P \geq 0.95$ for the alternative evolutionary model were identified as being subject to selection. The direction of selection for each locus was determined by its posterior

distribution mode (alpha value), with $\alpha<0$ indicating balancing selection and $\alpha>0$ indicating diversifying selection (Foll & Gaggiotti 2008).

Dataset definition

Datasets were defined based on the results of these tests, taking into account the limitations associated with each. FDIST2 has been associated with a relatively high Type I error rate in detection of outliers, while BayeScan is more conservative and detects only stronger outliers (Narum & Hess 2011; Russello *et al.* 2012; Lotterhos & Whitlock 2014). Population structure analyses were applied to each dataset according to the methods described in section 2.3.2.2 for the dataset containing genotypes of all SNPs.

An exception to this was for the Bayesian clustering and assignment analysis in Structure, which differed among datasets. Datasets of less than 1 000 SNP loci were analysed for K=1 to K=5, with five runs per K-value, due to the reduced computational time required per run of K. In datasets found to have a higher level of structure (based on F_{ST}), a different ancestry model, the "With admixture" model (without "Use LOCPRIOR") was applied; this assumes that individuals may be of mixed ancestry and that genotypes may be inherited from different populations with different allele frequencies. This allows individuals to be assigned to more than one cluster, with an admixture coefficient (q-value) indicating the proportion of each genotype to each cluster (Pritchard *et al.* 2000). Datasets that violated the assumptions of loci being in HWE across samples and absence of LD among loci in samples were not analysed in Structure.

CHAPTER 3

RESULTS

3.1 *Mitochondrial DNA Control Region*

3.1.1 Control Region sequencing

Following trimming and editing of Control Region (CR) sequences, a 633 base pair (bp) homologous fragment could be aligned across 93 individuals (Table 2.1). No indels were present in these sequences. The Basic Local Alignment Search Tool (BLAST) revealed all sequences to be from yellowfin tuna individuals.

3.1.2 Control Region genetic analyses

3.1.2.1 Genetic diversity

From 633 bp sequences of the mtDNA CR of 93 yellowfin tuna individuals from five sampling locations (Table 2.1), 91 haplotypes were distinguished by 137 variable sites, of which, 95 were parsimony informative. Of the 91 haplotypes, 89 were unique haplotypes each observed for only one individual in the dataset, one was a private haplotype observed for two individuals within the MOZ sample, and one was a shared haplotype observed for one individual from the KZN sample and one individual from the WC sample. The haplotype diversity (h) observed was therefore high, with an overall value of h=0.999, and ranging from 0.995 in the MOZ sample to 1.000 in the remaining samples (Table 3.1). The observed nucleotide diversity (π) was high for a marine fish species, with an overall value of $\pi = 0.024$ and similar values among all samples, ranging from 0.022 (KZN) to 0.026 (WC; Table 3.1).

The results of Tajima's D test of neutrality exhibited a relatively low negative value overall (D= –1.001) and for all samples, ranging from –1.229 (GG) to –0.71 (EC; Table 3.1); none of the observed Tajima's D values were found to be significant (P>0.05; Table 3.1). Results of Fu's F_S test of neutrality however exhibited highly significant (P<0.01) negative values overall (F_S= –7.552) and for all samples, ranging from –9.702 (KZN) to –5.625 (WC; Table 3.1).

Table 3.1: Genetic diversity indices and neutrality test statistics (and associated P-values) for each sample and overall for a 633 bp fragment of mtDNA Control Region, where **n**: number of individuals; **H**: number of haplotypes; ***h***: haplotype diversity with standard deviation; **π**: nucleotide diversity with standard deviation; **D**: Tajima's neutrality test statistic; and $\mathbf{F}_s$: Fu's neutrality test statistic. Test values significant at α=0.05 displayed in bold, significant at α=0.01 in bold and denoted by *.

Sample	n	H	*h*	π	Tajima's		Fu's	
					D	P	F_S	P
EC	17	17	1.000	0.024	-0.710	0.256	**-6.866***	0.005
GG	20	20	1.000	0.025	-1.299	0.087	**-8.966***	0.002
KZN	20	20	1.000	0.022	-1.044	0.142	**-9.702***	0.001
MOZ	20	19	0.995	0.024	-1.218	0.103	**-6.602***	0.008
WC	16	16	1.000	0.026	-0.734	0.244	**-5.625**	0.010
Overall	93	91	0.999	0.024	-1.001	0.167	**-7.552***	0.005

3.1.2.2 Population structure

Exact tests of population differentiation

Pairwise exact tests of differentiation among haplotypes were conducted for each pair of samples to explore the partitioning of potential differentiation within the dataset. All pairwise comparisons were found to be non-significant, with P-values ranging from 0.484 to 1.000 (Table 3.2).

Table 3.2: P-values for pairwise exact tests of genetic differentiation between samples based on 633 bp mtDNA sequences. No significant differences (at α=0.05) observed.

	EC	GG	KZN	MOZ
GG	1.000			
KZN	1.000	1.000		
MOZ	0.484	0.493	0.505	
WC	1.000	1.000	1.000	0.492

Analyses of molecular variance

A three-level analysis of molecular variance (AMOVA) was implemented to test two alternative hypotheses of population structure, as described in Table 2.2.

For both hypotheses of population structure, the majority of variance was partitioned among individuals within samples, at the within-sample level (>99.5%), with the between-group and within-group covariance components combined accounting for less than 0.5% of the total variance. The among-group covariance component was greater under Hypothesis 1, with the WC sample grouped with the GG sample rather than the Indian Ocean samples, however this difference was small (0.26%). None of the covariance components or associated fixation indices, at any level of structure for either of the hypotheses, were found to be significant

after 10 000 permutations (P>0.05; Table 3.3). Therefore, no significant genetic population structure was detected by AMOVAs for the mtDNA haplotype dataset.

Table 3.3: Results of AMOVAs: covariance components, percent variation, fixation indices and associated P-values for three hierarchical levels of structure, for two alternative hypotheses of population structure, for the mtDNA CR haplotype dataset. No significant variance components or fixation indices (at α=0.05) observed.

	Source of variation	d.f.	S.S.	Covariance component	Percent variation	Fixation index	P-value
H1 GG+WC vs EC+KZN+MOZ	Between groups	1	8.81	0.0236	0.31	F_{CT}=0.003	0.312
	Within groups	3	23.30	0.0085	0.11	F_{SC}=0.004	0.384
	Within samples	88	669.48	7.6077	99.58	F_{ST}=0.001	0.303
	Total	*92*	*701.58*	*7.6398*			
H2 **GG** vs WC+ EC+KZN+MOZ	Between groups	1	8.14	0.0040	0.05	F_{CT}=0.001	0.601
	Within groups	3	23.96	0.0208	0.27	F_{SC}=0.003	0.326
	Within samples	88	669.48	7.6077	99.67	F_{ST}=0.003	0.323
	Total	*92*	*701.58*	*7.6325*			

F_{ST} tests of genetic differentiation

The global F_{ST} value across all samples for the mtDNA CR haplotype dataset was low (F_{ST}=0.00295) and non-significant (P=0.309). Pairwise comparisons to investigate the partitioning of this variation within the dataset were low, ranging from 0.000 for six comparisons to 0.0097 for the comparison between the GG and MOZ samples. One comparison (KZN and WC) showed significant genetic differentiation (P<0.05; Table 3.4), however after sequential Bonferroni correction of the critical P-value, no pairwise comparisons were found to be significant.

Table 3.4: mtDNA CR-based Φ_{ST} values (below diagonal) and associated P-values (above diagonal) among yellowfin tuna samples. Comparisons significant at α=0.05 displayed in bold; comparisons significant after sequential Bonferroni correction in bold and denoted by *.

	EC	GG	KZN	MOZ	WC
EC	-	0.610	0.494	0.429	0.719
GG	-0.0058	-	0.519	0.182	0.231
KZN	-0.0024	-0.0022	-	0.621	0.244
MOZ	0.0003	0.0097	-0.0048	-	0.049
WC	-0.0112	0.0086	0.0087	**0.0263**	-

Reconstruction of haplotype network

Reconstruction of the evolutionary history of yellowfin tuna by means of a median-joining haplotype network did not reveal any discernible patterns of geographic structuring of haplotypes, with haplotypes from various regions dispersed throughout the network (Figure 3.1). The maximum number of site differences between adjacent haplotypes being 24 (between two individuals from the WC sample; Figure 3.1).

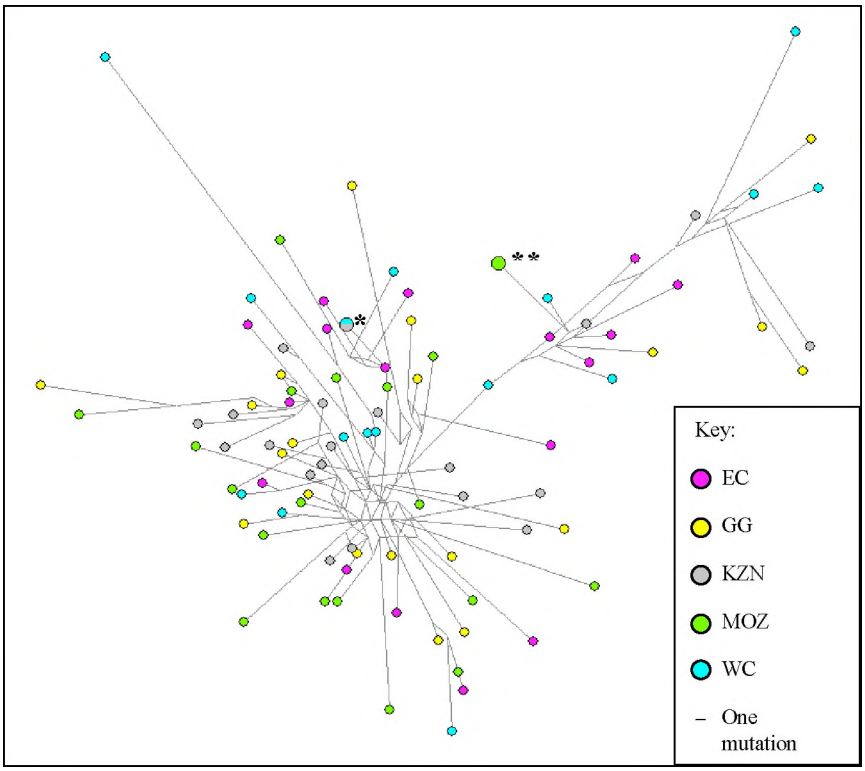

Figure 3.1: Median-joining haplotype network of 91 haplotypes identified from 93 yellowfin tuna mtDNA CR sequences from five sampling locations. Each node represents an observed haplotype; node size is proportional to the frequency of the haplotype in the dataset and node colours denote the sample origin of the haplotypes. Branch length is proportional to the number of mutational steps between haplotypes, as indicated in the key. Shared haplotypes denoted by * and private haplotypes denoted by **. Intermediate nodes not represented.

Demographic history

As no significant genetic differentiation was detected among regional samples for the mtDNA CR, all samples were pooled and the demographic history of yellowfin tuna was reconstructed for one putative population.

The observed haplotype diversity for the putative population of pooled samples was high (h>0.05), while the nucleotide diversity observed was low (π<0.05; Table 3.5), based on the categories described by Grant and Bowen (1998). Both Tajima's D test and Fu's F_S test resulted in negative values with significant departures from neutrality for the putative population of pooled samples (P<0.05; Table 3.5).

Table 3.5: Genetic diversity indices and neutrality test statistics for putative population of pooled samples to reconstruct demographic history. Significant results at α=0.05 displayed in bold.

	Putative population
h	0.999
π	0.024
Tajima's *D*	**-1.4455**
P-value	0.044
Fu's F_S	**-24.1239**
P-value	0.0001

A mismatch distribution for the putative population consisting of all samples pooled was observed to be unimodal (Figure 3.2), characteristic of a population that has undergone a past demographic expansion. The associated sum of squared deviations (SSD) was non-significant (P>0.05; Table 3.6), indicating that the pattern of observed pairwise haplotypic differences distribution did not deviate significantly from that expected under the model of sudden population expansion. The Harpending's raggedness index (H*ri*) values was low (H*ri*<0.01), characteristic of relatively smooth, unimodal mismatch distributions, and non-significant (P>0.05; Table 3.6), suggesting a good fit of the observed distribution to the expansion model distribution.

These results provide support for the occurrence of a past demographic expansion of yellowfin tuna populations, and as a result the time since population expansion of the putative population was estimated from the mismatch parameters obtained from the mismatch distribution's observed mean and variance (Table 3.6). The time of expansion events was estimated to have occurred at approximately 318 thousand years before the present (KYA) for the putative population assessed (Table 3.6).

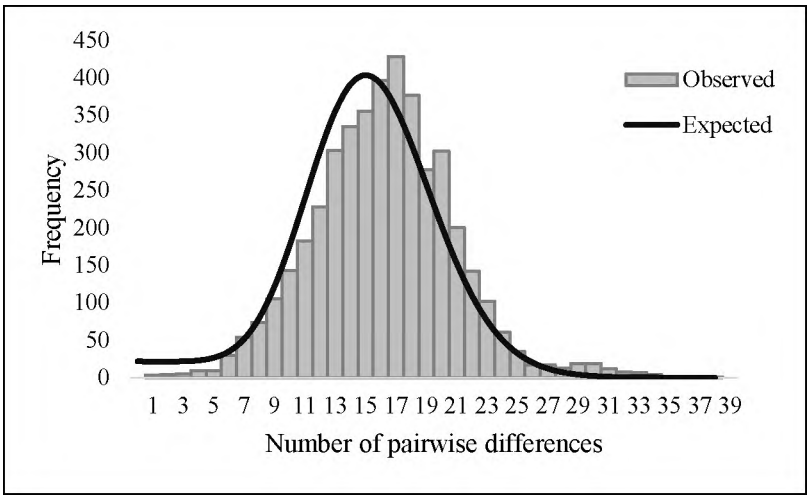

Figure 3.2: Mismatch distribution of observed and expected pairwise nucleotide differences between haplotypes, under the null model of population expansion for putative population of pooled samples.

Table 3.6: Mismatch distribution parameters for putative population of pooled samples. **SSD**: sum of squares deviation; **H*ri***: Harpending's raggedness index; **T_{exp} (KY)**: time of expansion in thousands of years. 95% CI values displayed in parentheses. No significant values observed at α=0.05.

	Putative population
Mean	15.252
Variance	21.620
SSD	0.0006
P-value	0.640
H*ri*	0.0027
P-value	0.739
τ	14.873 (12.252 – 17.350)
θ_0	0.786 (0.000 – 2.714)
θ_1	197.422 (128.203 – 99999.000)
T_{exp} (KY)	318 (262 – 371)

3.2 *Nuclear DNA single nucleotide polymorphisms*

3.2.1 SNP genotyping

Two sequencing runs of the tunable genotyping by sequencing (tGBS®) protocol resulted in a total of 179,521,855 raw sequenced reads across 96 yellowfin tuna individuals, ranging from 56,871 to 10,442,486 raw reads per individual, with an average of 1,870,018 raw reads per individual. Following sequencing, low quality bases were removed and reads were aligned based on a reference genome of Pacific bluefin tuna (*T. orientalis*; Nakamura *et al.* 2013). SNPs were then identified with a minor allele frequency (MAF). Seven individuals from the WC sample for which more than 60% of identified SNPs were missing were removed from the dataset (Figure 3.3), resulting in a dataset of genotypes of approximately 30 000 SNP loci for 89 individuals. The size of genotypes was further reduced by removing SNP loci from the dataset that were missing in more than 10% of individuals; this resulted in a final SNP genotype dataset of 11 862 SNP loci for 89 yellowfin tuna individuals.

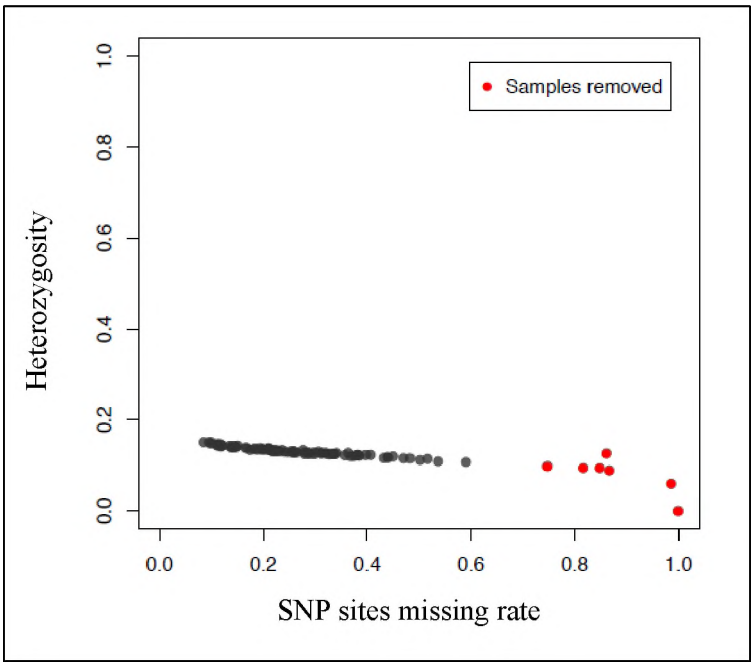

Figure 3.3: Heterozygosity as function of missing SNP sites for all individuals, with individuals removed from the SNP genotype dataset due to missing data displayed in red.

3.2.2 SNP genetic analyses

3.2.2.1 Genetic diversity

All SNPs in the dataset were bi-allelic, therefore the mean number of alleles across all loci and the allelic richness across all loci was observed to be less than two for all samples. Mean number of alleles (MNA) per sample ranged from 1.829 (WC) to 1.908 (KZN), while mean allelic richness (A_R) per sample ranged from 1.772 (EC) to 1.784 (KZN). Observed heterozygosity (H_O) per sample, across all loci, ranged from 0.200 (WC) to 0.207 (GG), and expected heterozygosity (H_E) per sample ranged from 0.220 (WC) to 0.227 (GG; Table 3.7).

Wright's inbreeding coefficient (F_{IS}) values per locus, averaged across all samples, revealed 119 loci to have significant positive F_{IS} values following Bonferroni correction, ranging from 0.3741 to 0.8813, with non-significant values for the remaining 11 743 loci, ranging from -0.4536 to 0.8509. F_{IS} values averaged across all loci were positive and significant ($P<0.05$) for all populations, ranging from 0.71 (MOZ) to 0.88 (KZN) (Table 3.7), however no population F_{IS} values remained significant following Bonferroni correction.

All loci were however found to be in Hardy-Weinberg Equilibrium (HWE), with no loci differing significantly from HWE expectations following sequential Bonferroni correction. No pairs of loci were found to be in linkage after sequential Bonferroni correction.

Table 3.7: Genetic diversity indices per sample for the SNP genotype dataset. **N**: sample size; **MNA**: mean number of alleles per locus, across all loci; A_R: allelic richness, averaged across all loci; H_O: observed heterozygosity averaged across all loci; H_E: expected heterozygosity, averaged across all loci F_{IS}: Wright's inbreeding coefficient (F_{IS} P-value in parenthesis and significant values at α=0.05 in bold).

	N	MNA	A_R	H_O	H_E	F_{IS}
EC	16	1.877	1.772	0.206	0.224	**0.082** (0.009)
GG	20	1.903	1.783	0.207	0.227	**0.083** (0.005)
KZN	20	1.908	1.784	0.204	0.226	**0.088** (0.002)
MOZ	20	1.907	1.781	0.206	0.226	**0.071** (0.011)
WC	13	1.829	1.741	0.200	0.220	**0.080** (0.019)

3.2.2.2 Population structure

F_{ST} tests of genetic differentiation

A global test of genetic differentiation, measured by Wright's F_{ST}, found low but significant structure within the SNPs dataset (F_{ST}=0.0043; P<0.0001).

Pairwise F_{ST} tests, to identify sources of this structure within the dataset, found that all comparisons including the GG sample were higher than comparisons among the remaining samples; GG comparisons ranged from 0.0068 (GG–KZN) to 0.0072 (GG–MOZ), while remaining comparisons ranged from 0.0008 (EC–WC) to 0.0043 (MOZ–WC; Table 3.8). All GG comparisons were observed to be significant after 10 000 permutations (P<0.0001), and remained significant after sequential Bonferroni correction, while none of the comparisons between remaining samples were observed to be significant (P>0.05; Table 3.8).

Table 3.8: SNP-based F_{ST} values (below diagonal) and associated P-values (above diagonal) among yellowfin tuna samples. Comparisons significant at α=0.05 displayed in bold; comparisons significant after sequential Bonferroni correction in bold and denoted by *.

	EC	GG	KZN	MOZ	WC
EC	-	0.000	0.618	0.014	0.997
GG	**0.0071***	-	0.000	0.000	0.000
KZN	0.0023	**0.0068***	-	0.628	0.974
MOZ	0.0037	**0.0072***	0.0018	-	0.555
WC	0.0008	**0.0070***	0.0012	0.0043	-

Exact tests of genetic differentiation

Exact tests of differentiation found global genic and genotypic differentiation within the SNPs genotype dataset to be highly significant (P<0.0001).

Pairwise tests to investigate sources of observed genic differentiation in the dataset found all pairwise comparisons including the GG sample to be significant, whilst all pairwise comparisons among the remaining samples were non-significant (P=1.000; Table 3.9).

Pairwise tests of genotypic differentiation, however, found only the comparison between the GG and WC samples to be significantly different, with comparisons among all remaining samples being non-significant (P=1.000; Table 3.9).

Table 3.9: P-values for pairwise exact tests of genic differentiation (below diagonal) and genotypic differentiation (above diagonal) within SNP genotype dataset. Highly significant comparisons displayed in bold.

	EC	GG	KZN	MOZ	WC
EC	-	1.000	1.000	1.000	1.000
GG	**<0.0001**	-	1.000	1.000	**<0.0001**
KZN	1.000	**<0.0001**	-	1.000	1.000
MOZ	1.000	**<0.0001**	1.000	-	1.000
WC	1.000	**<0.0001**	1.000	1.000	-

Analyses of molecular variance

Three-level analyses of molecular variance (AMOVAs) were performed for the SNP genotype dataset to test two alternative hypotheses of population structure, as described previously for the mtDNA CR dataset (see section 3.1.2.2). AMOVAs included 8 450 of 11 862 loci, for which less than 5% of data was missing among individual genotypes.

These tests found the majority of variance to be partitioned among individual genotypes at the within-sample level (>99%), with the between-group and within-group covariance components accounting for less than 1% of the total variance, under both hypotheses of population structure (Table 3.10). The within-sample covariance component was significant (P<0.01) under both hypotheses after 10 000 permutations, but decreased under Hypothesis 2 (however this difference was small = 0.26%). The between-group covariance component was greater under Hypothesis 2, with the WC sample grouped with the Indian Ocean sample rather than the GG (Atlantic Ocean) sample, however this difference was small (0.46%) and values of between-group differentiation were non-significant for both hypotheses (P<0.0001; Table 3.10). Under Hypothesis 1, with the WC sample grouped with the Atlantic Ocean GG sample, The proportion of variance attributed to variation among samples within groups was significant (P<0.0001); under Hypothesis 2 this covariance component decreased and was non-significant (P>0.05), indicating a lack of significant differentiation among samples within groupings (Table 3.10).

Table 3.10: AMOVA tests of SNP variation in yellowfin tuna samples: covariance components, percent variation, fixation indices and associated P-values for three hierarchical levels of structure, for two alternative hypotheses of population structure. Significant variance components and fixation indices at α=0.05 displayed in bold; significant values at α=0.0001 in bold and denoted by *.

	Source of variation	d.f.	S.S.	Covariance component	Percent variation	Fixation index	P-value
H1 GG+WC vs EC+KZN+MOZ	Between groups	1	1122.09	0.6686	0.07	F_{CT}=0.0007	0.303
	Within groups	3	3195.16	3.6362	**0.39***	**F_{SC}=0.0039***	0.000
	Within samples	173	162069.17	936.8160	**99.54***	**F_{ST}=0.0046***	0.000
	Total	*177*	*166386.42*	*941.1208*			
H2 GG vs WC+ EC+KZN+MOZ	Between groups	1	1318.682	5.0030	0.53	F_{CT}=0.0053	0.196
	Within groups	3	2998.57	1.8355	0.19	F_{SC}=0.0020	0.765
	Within samples	173	162069.17	936.8160	**99.28***	**F_{ST}=0.0073***	0.000
	Total	*177*	*166386.42*	*943.6545*			

Bayesian clustering and assignment analysis

The Bayesian clustering analysis implemented in Structure (Pritchard *et al.* 2000) detected no genetic structure in the SNPs genotype dataset, with the most likely number of clusters being identified as K=1. Evanno's method (Evanno 2005) could not be used to assess the most likely K-value as the calculations of this statistic do not allow for the ΔK statistic of the first or last K-value to be computed. Identification of the most likely K-value was therefore based on the log-likelihood (LnP(K)) value averaged across all runs for each K-value. The LnP(K) was found to be greatest for K=1, and the standard deviation among LnP(K) values across all for each K-value was notably lower for K=1 (2.23) than for K=2 and K=3 (greater than 80 000 and 17 000 respectively; Table 3.11), indicating that all runs returned similar LnP(K) values. Additionally, the assignment plot of individual genotypes generated for K=2, averaged across all runs, revealed that all genotypes from the EC, GG and KZN clusters were assigned to the first cluster with a membership coefficient of q=1.0; no genotypes were assigned to the second cluster. Two genotypes from the MOZ sample and one from the KZN sample were assigned to the second cluster with low membership coefficients of 0.065, 0.005 and 0.020 respectively (Figure 3.4a). When K=3 was tested, no individuals were assigned to the third cluster, with one individual from the MOZ sample being assigned to the second cluster with a low membership coefficient of q=0.012 (Figure 3.4b).

These results indicate a lack of any detectable structure by means of Bayesian clustering within the SNPs genotypes dataset. As a result, only one round of the analysis (including all samples) was performed.

Table 3.11: Mean log-likelihood value associated with each K-value tested for the SNP genotypes dataset, averaged across all runs for each K, and associated standard deviation. Delta K value obtained using Evanno's method.

K	No. runs	Mean LnP(K)	Stdev LnP(K)	Delta K
1	3	-771 881.91	2.36	-
2	3	-819 725.93	82502.33	1.04
3	3	-791 919.57	17203.11	-

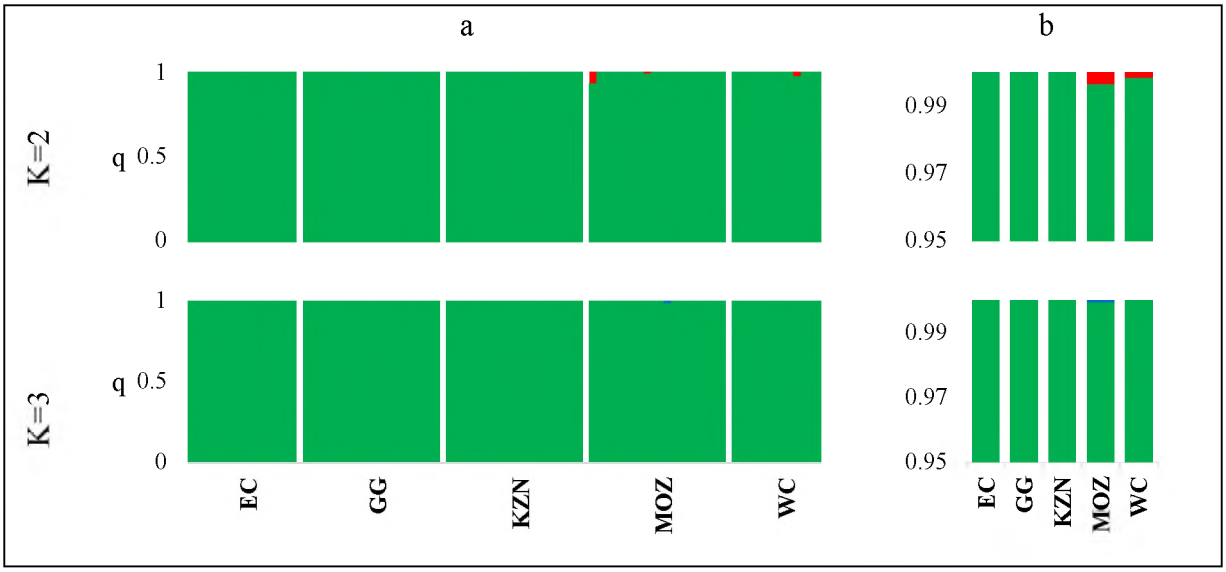

Figure 3.4: Membership coefficients for two alternative hypotheses of K, K=2 and K=3, for a) each individual genotype, where each vertical bar represents one individual's genotype; and b) genotypes grouped by samples, for the Bayesian clustering analysis of the SNP genotypes dataset. For each hypothesis of K, different colours represent each of the K populations defined within the dataset.

Assignment tests

The self-classification step of the GeneClass2 assignment test between the GG and MOZ samples revealed one GG individual assigned to the MOZ sample, and one MOZ individual assigned to the GG sample. All individuals that correctly self-assigned did so with a probability of >0.95. One MOZ individual had an equal probability of assignment to both baseline populations, perhaps representing either a hybrid or an uninformative genotype.

For the assignment of 'unknown' fish, all individuals assigned to either GG or MOZ with a probability of >0.95. All 13 WC individuals were assigned to the MOZ sample. For both the EC and KZN samples, one individual assigned to GG while the remaining individuals all assigned to MOZ.

Self-classification assignment tests excluding the GG sample revealed 22% correct self-assignment (14 fish), with many misassigned fish assigning to more than one sample. .

Multivariate analyses

A principal component analysis (PCA) performed on the SNPs dataset revealed genetic clusters separated along the first PC axis, with one cluster consisting of all individuals from the GG sample and the other consisting of individuals from the remaining four samples, among which there was no separation observed along the first PC axis (Figure 3.5). No further clustering of individuals was observed along the second or third PCs. The eigenvalues of PCs show that all PCs account for a similar level of variation (Figure 3.5).

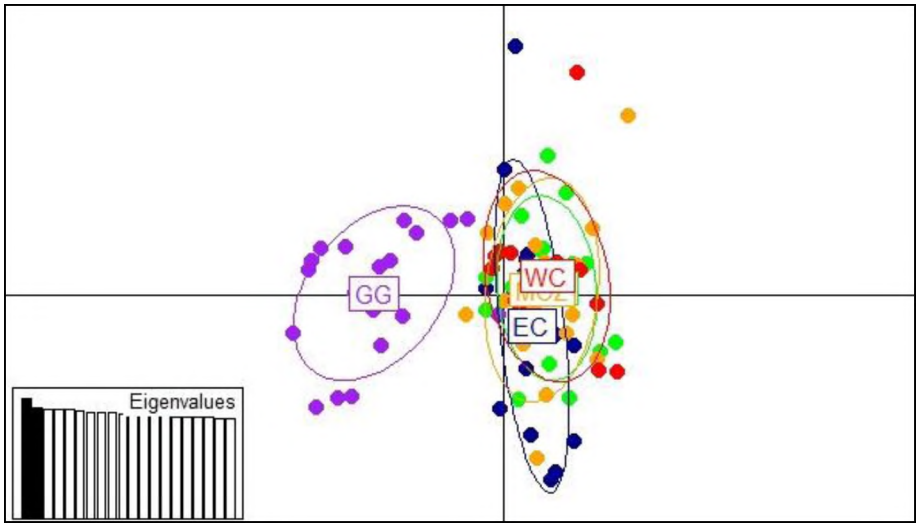

Figure 3.5: Scatterplot of results of PCA for the All SNPs dataset, against the first two PC axes, with eigenvalues of PCs displayed.

The k-means clustering algorithm found the most likely number of genetic clusters in the All SNPs dataset to be one, as the k=1 clustering hypothesis was associated with the lowest BIC value (BIC=653.54; Appendix 3: Figure 5.1). A cross-validation test with 100 replicates indicated that retaining the first 60 PCs in the dimension-reducing step of a DAPC would allow for the optimum level of correct assignment of individuals to their original sample, with the lowest associated mean squared error (Appendix 3: Figure 5.2; Table 5.1).

A discriminant analysis of principal components (DAPC) was performed by transforming the first 60 PCs into discriminant functions (DFs), thereby retaining 72.9% of the total variation described by the PCA. The resulting DAPC scatterplot of individuals against the first two DF axes shows two genetic clusters, with a clear separation of the GG sample from the remaining samples against the first DF axis, and no separation of remaining samples against this axis; this DF which describes the majority of the variation, as shown by the eigenvalues of the DFs (Figure 3.6). No clear genetic clustering was observed along the second DF, with overlapping of the EC, GG, KZN and WC samples, and of the GG, KZN and MOZ samples, but no overlap between the MOZ and EC samples and slight overlap of the MOZ and WC samples;

this DF however does not account for a large proportion of the total variation observed, as seen by its eigenvalue (Figure 3.6).

Results of the PCA and DAPC of the SNPs genotype dataset provide support for the grouping of the WC sample with the Indian Ocean samples, rather than with the Atlantic Ocean GG sample, and the differentiation of GG individuals and those of the remaining samples.

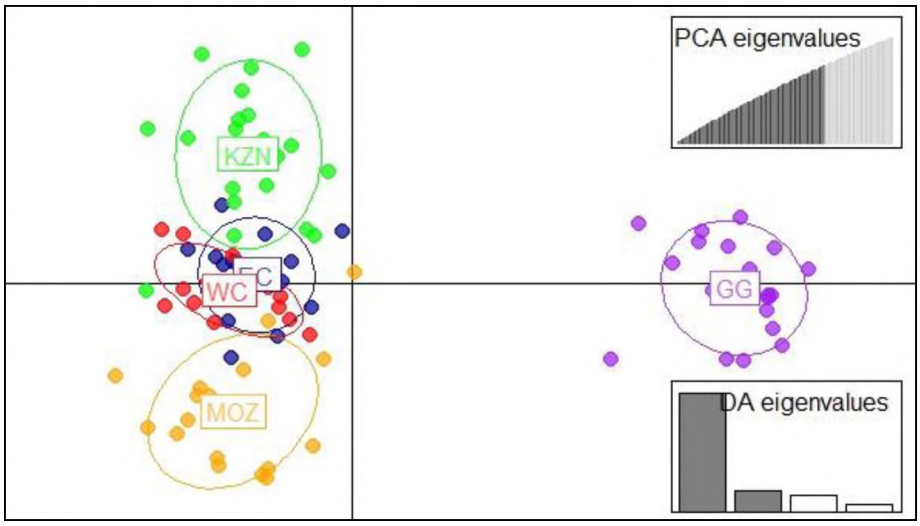

Figure 3.6: Scatterplot of DAPC for the All SNPs dataset, retaining the first 60 PCs of the PCA, against the first two DF axes with eigenvalues of DFs displayed.

3.2.3 Comparison of neutral and outlier SNPs

3.2.3.1 Outlier tests

The FDIST2 F_{ST} outlier detection test implemented in Lositan (Antao *et al.* 2008) with 99% confidence intervals identified 869 SNP loci putatively under positive (diversifying) selection, falling above the upper confidence interval of the null distribution expected under neutral evolution (P≥0.99), and 512 loci putatively under balancing selection, falling below the lower confidence interval (P≤0.01) (Figure 3.7). The mean (putatively neutral) F_{ST} value obtained for this test was 0.003, and was found by defining two putative populations, one comprising only the GG sample and the other comprising the remaining four samples, based on the results of pairwise exact tests and pairwise F_{ST} tests of genetic differentiation between SNP genotypes (Tables 3.8 and 3.10). The 'Mean neutral F_{ST}' option in Lositan, to estimate the null distribution of differentiation, identified a subset of 1 340 outliers putatively under diversifying selection (P≥0.99) with a genetic differentiation over all genotypes of F_{ST}=0.0455 which was significant after 10 000 permutations (P=0.000), whilst the 'Force mean F_{ST}' option resulted in the subset of 869 putatively diversifying outliers with a similar level of significant genetic differentiation, F_{ST}=0.0408 (P=0.000).

The outlier test implemented in BayeScan (Foll and Gaggioti 2008) identified 11 SNP loci whose posterior probability indicated they were likely to be under the influence of selection at a FDR of 0.05 (P≥0.95) (Figure 3.8); of these 11 loci, all were identified as being putatively under positive (diversifying) selection, rather than balancing selection (α>0.0). All 11 BayeScan outlier SNPs were also represented in the final FDIST2 outlier SNP dataset (Figures 3.7 and 3.8).

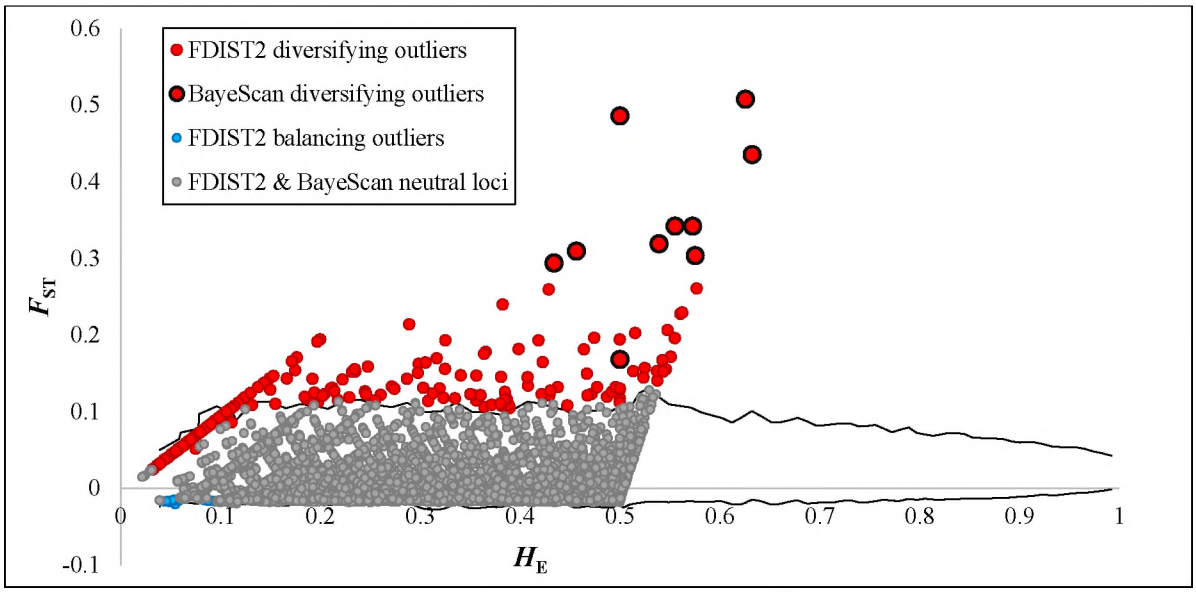

Figure 3.7: Genetic differentiation of SNP loci, measured by F_{ST} as a function of expected heterozygosity (H_E), against the 99% confidence intervals of the null distribution of genetic differentiation expected under neutral conditions, as estimated using the FDIST2 approach in Lositan after 50 000 simulations. SNP loci falling above the upper CI identified by Lositan as being putatively under diversifying selection, those falling below the lower CI identified as being putatively under balancing selection. Outliers also identified by BayeScan displayed.

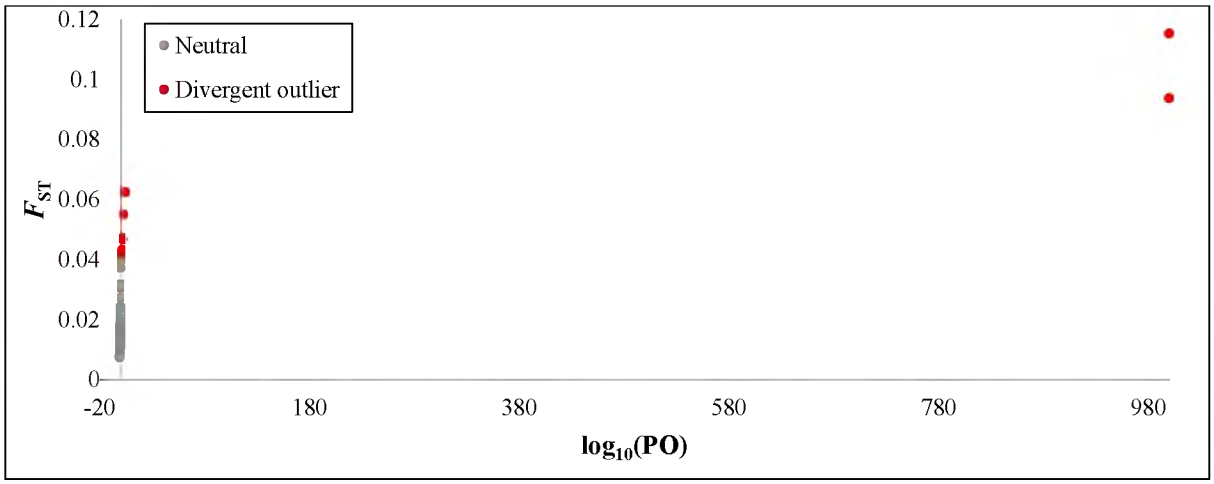

Figure 3.8: Genetic differentiation of SNP loci, measured by F_{ST} as a function of the logarithm of Posterior odds for the model including selection, estimated in BayeScan.

The two outlier detection methods implemented have been associated with different errors in outlier detection: FDIST2 has a relatively high Type I error rate, and BayeScan is more

conservative, identifying only stronger outliers and often failing to detect outliers under balancing selection (Narum & Hess 2011; Russello *et al.* 2012; Lotterhos & Whitlock 2014). Results of these two outlier detection methods were therefore combined to define a number of datasets, as described in Table 3.12, which were used in subsequent analyses of population structure to test the difference in resolving power between putatively neutral and adaptive genome regions.

Table 3.12: SNP datasets defined based on results of FDIST2 and BayeScan outlier tests

Name of dataset	Number of SNP loci	Description
FDIST2 Neutral SNPs	10 482	SNPs identified by FDIST2 as not being under diversifying or balancing selection
BayeScan Neutral SNPs	11 851	SNPs identified by BayeScan as not being under diversifying selection
Combined Neutral SNPs	11 339	All SNPs excluding those identified by BayeScan as being under diversifying selection, and those identified by FDIST2 as being under balancing selection
FDIST2 Outlier SNPs	869	SNPs identified by FDIST2 as being under diversifying selection
BayeScan Outlier SNPs	11	SNPs identified by BayeScan as being under diversifying selection
Variable Neutral SNPs	858	SNPs identified by FDIST2 as being under diversifying selection, excluding those identified by BayeScan as being under diversifying selection

The "Combined Neutral SNPs" and "BayeScan Outlier SNPs" were identified as being most likely to represent putative neutral and putative adaptively diversifying SNPs in the yellowfin tuna SNP genotype data respectively, while the "Variable Neutral SNPs" dataset was identified as being likely to represent a subset of highly variable SNPs that may or may not be under diversifying selection (see section 4.1.3.1 of the Discussion). Results of population structure analyses performed on these datasets are thus presented below.

3.2.3.2 Combined Neutral SNPs

Tests of Hardy-Weinberg Equilibrium and linkage disequilibrium

All SNP loci in the Combined Neutral SNPs dataset were found to be in HWE, with no loci deviating significantly from HWE expectations following sequential Bonferroni correction. No pairs of loci in the dataset were found to be in linkage after sequential Bonferroni correction.

F_{ST} tests of genetic differentiation

The global F_{ST} value for the Combined Neutral SNPs dataset found low but significant structure within this dataset (F_{ST}=0.0040; P=0.000).

Pairwise F_{ST} tests found all comparisons including the GG sample to be higher than comparisons among remaining samples; GG comparisons ranged from 0.0064 (GG–KZN) to 0.0068 (GG–MOZ), while remaining comparisons ranged from 0.0006 (EC–WC) to 0.0037 (EC–MOZ; Table 3.13). All GG samples were found to be significant after 10 000 permutations (P<0.0001), and remained significant after sequential Bonferroni correction. No comparisons among remaining samples were found to be significant (P>0.05; Table 3.13).

Table 3.13: Combined Neutral SNP-based F_{ST} values (below diagonal) and associated P-values (above diagonal) among yellowfin tuna samples. Comparisons significant at α=0.05 displayed in bold; comparisons significant after sequential Bonferroni correction in bold and denoted by *.

	EC	GG	KZN	MOZ	WC
EC	-	0.0000	0.6682	0.0139	0.9975
GG	**0.0066***	-	0.0000	0.0000	0.0000
KZN	0.0021	**0.0064***	-	0.6053	0.9809
MOZ	0.0037	**0.0068***	0.0018	-	0.5500
WC	0.0006	**0.0065***	0.0010	0.0022	-

A Bayesian clustering and assignment analysis was not performed in Structure for the Combined Neutral SNPs dataset based on the low level of differentiation observed within the dataset (F_{ST}<0.005).

Exact tests of genetic differentiation

Exact tests of differentiation for the Combined Neutral SNPs dataset found both the global genic and genotypic differentiation to be highly significant (P<0.0001). Pairwise tests to investigate sources of this differentiation however found all pairwise comparisons of genic and genotypic differentiation to be non-significant (P=1.000; Table 3.14).

Table 3.14: P-values for pairwise exact tests of genic differentiation (below diagonal) and genotypic differentiation (above diagonal) within Combined Neutral SNP genotype dataset. Significant comparisons at α=0.01 displayed in bold.

	EC	GG	KZN	MOZ	WC
EC	-	1.000	1.000	1.000	1.000
GG	1.000	-	1.000	1.000	1.000
KZN	1.000	1.000	-	1.000	1.000
MOZ	1.000	1.000	1.000	-	1.000
WC	1.000	1.000	1.000	1.000	-

Analyses of molecular variance

Three-level AMOVAs, using 8 066 of 11 339 loci with less than 5% missing data, found the greatest proportion of total variation within the dataset (>99%) to be partitioned among individuals at the within-sample level under both hypotheses. This covariance component and its associated fixation indices were found to be significant under both hypotheses (P<0.05), and decreased under Hypothesis 2 (with the WC sample grouped with the Indian Ocean samples), however this increase was small (0.25%; Table 3.15). The within-group covariance component was significant (P<0.0001) under Hypothesis 1, and decreased (although this decrease was small = 0.19%) and was non-significant (P>0.05) under Hypothesis 2 (Table 3.15). The between-group covariance component increased under Hypothesis 2, however this increase was small (0.44%) and the component was not significant under either hypothesis (P>0.05; Table 3.15).

Table 3.15: AMOVA tests of Combined Neutral SNP variation in yellowfin tuna samples: covariance components, percent variation, fixation indices and associated P-values for three hierarchical level of structure, for two alternative hypotheses of population structure. Significant covariance components and fixation indices at α=0.05 displayed in bold; significant values at α=0.0001 in bold and denoted by *.

	Source of variation	d.f.	S.S.	Covariance component	Percent variation	Fixation index	P-value
H1 GG+WC vs EC+KZN+MOZ	Between groups	1	1092.45	0.563	0.06	F_{CT}=0.001	0.309
	Within groups	3	3132.93	3.401	0.37	F_{SC}=**0.004***	0.000
	Within samples	173	159913.76	924.357	99.57	F_{ST}=**0.004***	0.000
	Total	*177*	*164139.14*	*928.321*			
H2 GG vs WC+ EC+KZN+MOZ	Between groups	1	1277.63	4.625	0.50	F_{CT}=0.005	0.200
	Within groups	3	2947.75	1.704	0.18	F_{SC}=0.002	0.780
	Within samples	173	159913.76	924.357	99.32	F_{ST}=**0.007***	0.000
	Total	*177*	*164139.14*	*930.686*			

Multivariate analyses

The PCA of the Combined Neutral SNPs dataset revealed genetic clusters separated along the first PC axis, with one cluster consisting of all individuals from the GG sample and the other consisting of individuals from the remaining four samples, among which there was no separation observed along the first principal component (PC) axis (Figure 3.9). No further clustering of individuals was observed along the second or third PCs. The eigenvalues of PCs show that all PCs account for a similar level of variation (Figure 3.9).

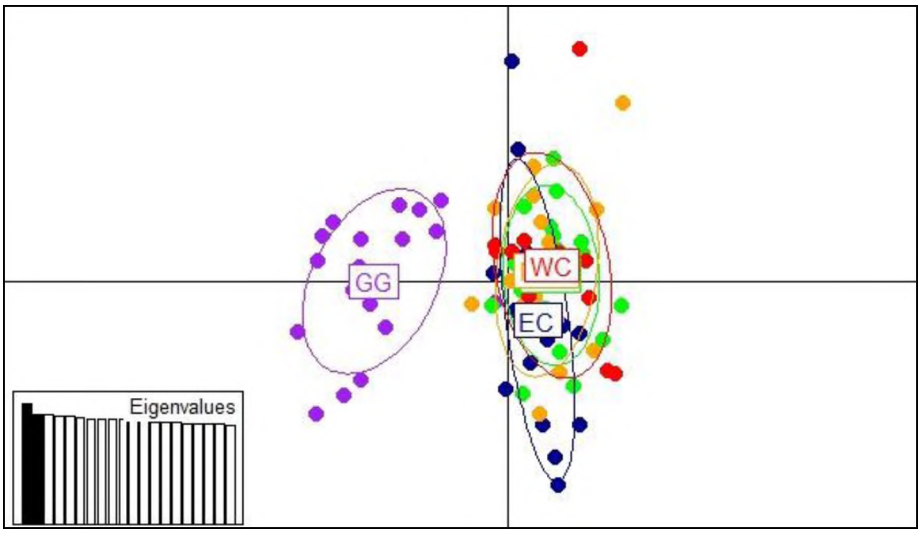

Figure 3.9: Scatterplot of results of PCA for the Combined Neutral SNPs dataset, against the first two PC axes, with eigenvalues of PCs displayed.

The k-means clustering algorithm found the most likely number of clusters within the Combined Neutral SNPs dataset to be one, as the k=1 clustering hypothesis was associated with the lowest Bayesian Information Criterion value (BIC=897.08; Appendix 3: Figure 5.3). A cross-validation test with 100 replicates indicated that retaining the first 40 PCs in the dimension-reducing step of a discriminant analysis of principal components (DAPC) would allow for the optimum level of correct assignment of individuals to their original sample, with the lowest associated mean squared error (Appendix 3: Figure 5.4; Table 5.2).

A DAPC was performed by transforming the first 40 PCs into DFs, retaining 50.2% of the total variance described by the PCA. The resulting DAPC scatterplot of individuals revealed a clear separation of the GG sample from the remaining samples, and no separation among remaining samples, against the first DF axis which describes the majority of variation as shown by the DF eigenvalues (Figure 3.10). No clear genetic clustering was observed along the second DF, with all samples overlapping along this axis except for the EC and MOZ samples for which there was no overlap; this DF however did not count for a large proportion of the total variation observed as seen by its eigenvalue (Figure 3.10).

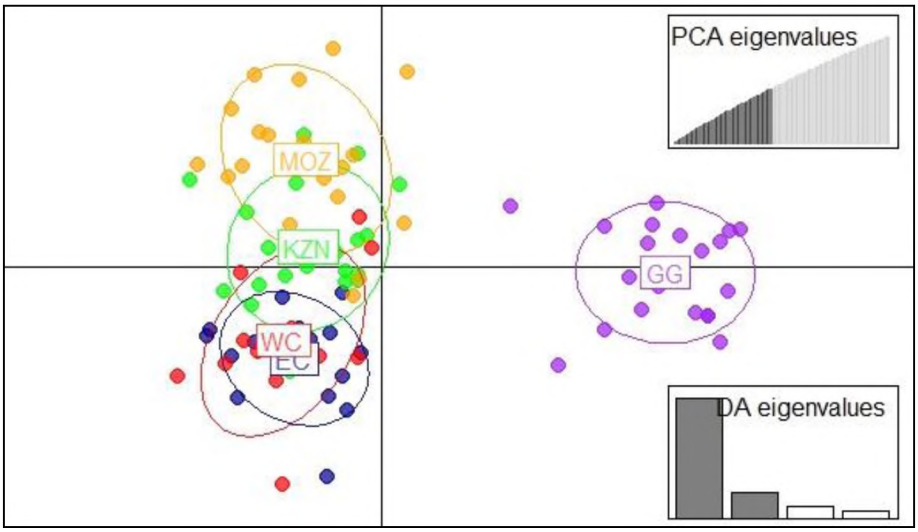

Figure 3.10: Scatterplot of DAPC for the Combined Neutral SNPs dataset, retaining the first 40 PCs of the PCA, against the first two DF axes with eigenvalues of DFs displayed.

3.2.3.3 Variable Neutral SNPs

Tests of Hardy-Weinberg Equilibrium and linkage disequilibrium

All SNP loci in the Variable Neutral SNPs dataset were found to be in HWE, with no loci deviating significantly from HWE expectations following sequential Bonferroni correction. No pairs of loci in the dataset were found to be in linkage after sequential Bonferroni correction.

F_{ST} tests of genetic differentiation

The global F_{ST} value for the Variable Neutral SNPs dataset indicated a relatively high level of significant variation within the dataset (F_{ST}=0.0409; P=0.000).

Pairwise F_{ST} tests revealed the same pattern of structure within the dataset as that observed for the Combined Neutral SNPs dataset, with a greater level of differentiation. All GG comparisons revealed high levels of differentiation, with F_{ST} values ranging from 0.1019 (GG–KZN) to 0.1251 (GG–WC), with all comparisons being highly significant (P<0.0001) and remaining significant after sequential Bonferroni correction (Table 3.16). All remaining comparisons revealed low F_{ST} values, ranging from 0.000 (MOZ–WC) to 0.0071 (KZN–WC), and were non-significant (P>0.05; Table 3.16).

Table 3.16: Variable Neutral SNP-based F_{ST} values (below diagonal) and associated P-values (above diagonal) among yellowfin tuna samples. Comparisons significant at α=0.05 displayed in bold; comparisons significant after sequential Bonferroni correction in bold and denoted by *.

	EC	GG	KZN	MOZ	WC
EC	-	0.0000	0.9295	0.9669	0.8249
GG	**0.1107***	-	0.0000	0.0000	0.0000
KZN	0.0021	**0.1019***	-	0.7495	0.5014
MOZ	0.0008	**0.1033***	0.0037	-	0.9924
WC	0.0039	**0.1251***	0.0071	-0.0005	-

Exact tests of genetic differentiation

Exact tests of differentiation for the Variable Neutral SNPs dataset found both the genic and genotypic differentiation within the dataset to be highly significant (P<0.0001).

Pairwise tests found this variation to be partitioned between the GG sample and the remaining samples. All pairwise comparisons including the GG sample were found to be highly significant (P<0.0001), whilst all pairwise comparisons among remaining samples were found to be non-significant (P=1.000) for both genic and genotypic tests of differentiation (Table 3.17).

Table 3.17: P-values for pairwise exact tests of genic differentiation (below diagonal) and genotypic differentiation (above diagonal) within Variable Neutral SNPs dataset. Significant comparisons at α=0.01 displayed in bold.

	EC	GG	KZN	MOZ	WC
EC	-	**<0.0001**	1.000	1.000	1.000
GG	**<0.0001**	-	**<0.0001**	**<0.0001**	**<0.0001**
KZN	1.000	**<0.0001**	-	1.000	1.000
MOZ	1.000	**<0.0001**	1.000	-	1.000
WC	1.000	**<0.0001**	1.000	1.000	-

Analyses of molecular variance

Three-level AMOVAs performed on the Variable Neutral SNPs dataset, using 575 of 858 loci for which less than 5% of data was missing, found the majority of variance to be partitioned among individuals at the within-sample level under both hypotheses of population structure (described in section 3.1.2.2), with this component being smaller than for the Combined Neutral SNPs dataset (<96%; Table 3.18). This component decreased by approximately 3% under Hypothesis 2, and was significant (P<0.05) under both hypotheses. The within-group covariance component also decreased under Hypothesis 2 (by ~2%), and was significant only under Hypothesis 1 (P<0.05). The between-group covariance component increased by approximately 5% under Hypothesis 2, however this component was non-significant under both hypotheses (P>0.05; Table 3.18).

Table 3.18: AMOVA tests of Variable Neutral SNP variation in yellowfin tuna samples: covariance components, percent variation, fixation indices and associated P-values for three hierarchical level of structure, for two alternative hypotheses of population structure. Significant covariance components and fixation indices at α=0.05 displayed in bold; significant values at α=0.0001 in bold and denoted by *.

	Source of variation	d.f.	S.S.	Covariance component	Percent variation	Fixation index	P-value
H1 GG+WC vs EC+KZN+MOZ	Between groups	1	110.501	0.5186	1.59	F_{CT}=0.0159	0.104
	Within groups	3	201.013	1.0192	**3.13**	F_{SC}=**0.0318***	0.000
	Within samples	173	5373.143	31.0563	**95.28**	F_{ST}=**0.0472***	0.000
	Total	*177*	*568.657*	*32.5965*			
H2 GG vs WC+ EC+KZN+MOZ	Between groups	1	191.849	2.4300	7.20	F_{CT}=0.0720	0.202
	Within groups	3	119.665	0.2585	0.77	F_{SC}=0.0083	0.983
	Within samples	173	5373.143	31.0586	**92.03**	F_{ST}=**0.0797***	0.000
	Total	*177*	*5684.657*	*33.7471*			

Bayesian clustering and assignment analysis

The Bayesian clustering analysis performed on the Variable Neutral SNPs dataset detected two distinct genetic clusters for the first round of the analysis (in which all samples were included), with K=2 being identified as the most likely number of clusters. Evanno's method (Evanno 2005) identified K=2 as being the most likely K-value, with the greatest Delta-K statistic (Table 3.19); Evanno's method is however unable to detect the most likely clustering hypothesis when this value is K=1. A steep increase in mean LnP(K) was observed from K=1 to K=2, where after increases were smaller (Table 3.19), supporting the hypothesis of K=2. The standard deviation of LnP(K) averaged across all runs of K=2 is relatively small (1.53; Table 3.19), indicating similar values of LnP(K) for each run of the analysis and further supporting K=2 as the most likely value of K in the dataset.

Table 3.19: Mean log-likelihood value associated with each K-value tested for Round 1 of the Bayesian clustering analysis of the Variable Neutral SNP genotypes dataset, averaged across all runs for each K with standard deviation of this average. Delta K value obtained using Evanno's method.

K	No. runs	Mean LnP(K)	Stdev LnP(K)	Delta K
1	3	-32 791.90	0.89	-
2	3	-30 750.17	1.53	1 274.98
3	3	-30 600.17	35.49	2.89
4	3	-30 582.57	32.05	1.47
5	3	-30 518.00	152.91	-

Assignment plots of individuals to genetic clusters were not averaged across all runs of each K-value, but based on the first run for each K-value, as the observed standard deviations among LnP(K) values for each K-value were relatively low (ranging from 0.89 to 152.91; Table 3.19), in comparison to those observed for the All SNPs dataset (ranging from 2.4 to >82 000; Table 3.10).

Bayesian assignment of individuals to each cluster for K=2 showed that all individuals from the GG sample were assigned to one of the two genetic clusters with a high admixture coefficients (q>0.98 for all individuals; Figure 3.11a). All individuals from the four remaining samples were assigned to the second genetic cluster with high admixture coefficients; q>0.82 for EC individuals, q>0.79 for KZN individuals, q>0.85 for MOZ individuals and q>0.84 for WC individuals (Figure 3.11a). This result provides support for the hypothesis of genetic structure between the GG samples and the remaining samples, and the grouping of the WC sample with the Indian Ocean samples, and indicates the genetic differentiation can be detected in the dataset when the most variable (putatively adaptive) loci are analysed.

Assignment analyses for K=3 to K=5 displayed the same patterns of clustering between the GG sample and remaining samples, with all GG individuals being assigned to one cluster with a high admixture coefficient and all individuals from remaining samples having a low admixture coefficient for this cluster. For K=3, individuals from the each of the remaining four samples were assigned to the remaining two clusters with relatively high admixture coefficients, with a similar proportion of individuals from each sample being assigned to each of these two clusters (Figure 3.11). A second round of Bayesian clustering and assignment analysis was therefore conducted, excluding the GG sample, to test for possible structure in the Variable Neutral SNPs dataset among individuals from the EC, KZN, MOZ and WC samples that may be too weak to be detected with the inclusion of the strongly differentiated GG genotypes.

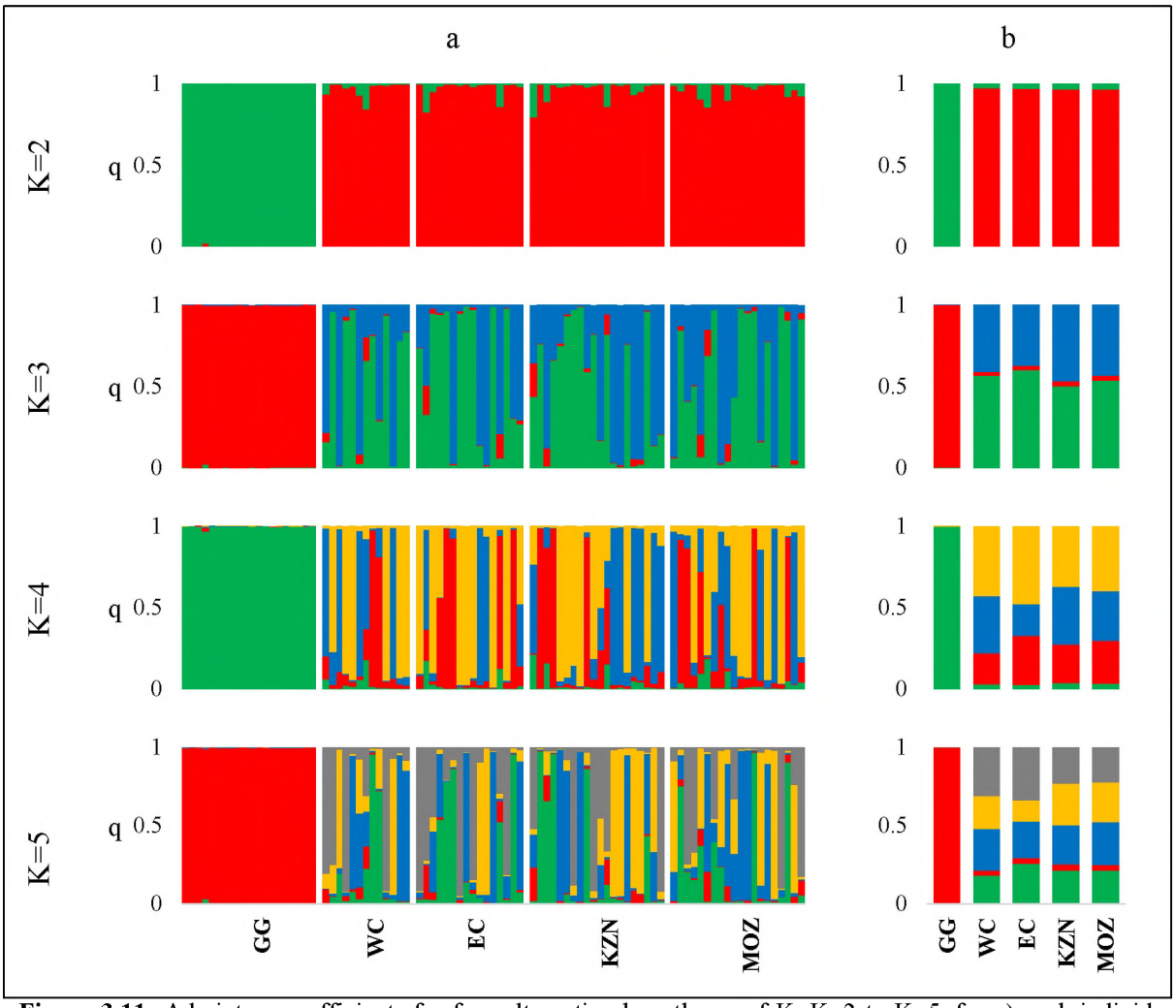

Figure 3.11: Admixture coefficients for four alternative hypotheses of K, K=2 to K=5, for a) each individual genotype, where each vertical bar represents one individual's genotype; and b) genotypes grouped by samples for Round 1 of the Bayesian clustering analysis of the Outlier SNPs dataset. For each hypothesis of K, different colours represent each of the K populations defined within the dataset.

Evanno's method identified K=3 to be the most likely K-value for Round 2 of the Variable Neutral SNPs dataset, with the greatest ΔK value (Table 3.20).

Table 3.20: Mean log-likelihood value associated with each K-value tested for Round 2 of the Bayesian clustering analysis of the Outlier SNP genotypes dataset, averaged across all runs for each K with standard deviation of this average. Delta K value obtained using Evanno's method.

K	No. runs	Mean LnP(K)	Stdev LnP(K)	Delta K
1	6	-28 008.90	1.78	-
2	6	-27 915.37	26.54	0.04
3	6	-27 822.95	15.31	1.74
4	3	-27 757.23	23.57	0.05
5	3	-27 690.37	45.88	-

Membership assignment plots for K=2 revealed that individuals from each sample were assigned to each of the two identified clusters with high admixture coefficients (Figure

3.12a). The proportion of membership assignment of individuals to each cluster was however similar for all samples, for both K=2 and K=3 (Figure 3.12b), indicating a lack of geographic structure among the EC, KZN, MOZ and WC samples.

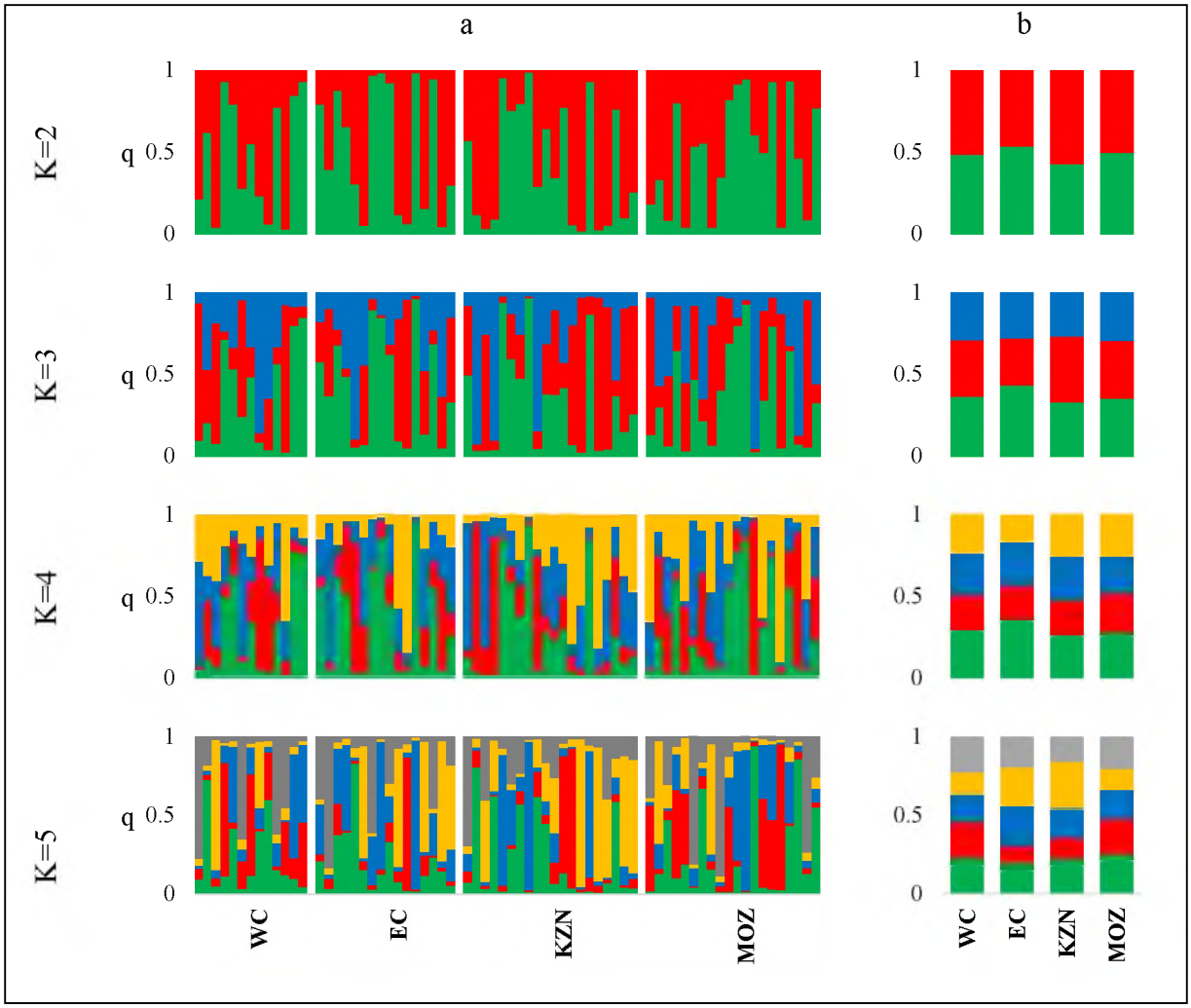

Figure 3.12: Admixture coefficients for two alternative hypotheses of K, K=2 and K=3, for a) each individual genotype, where each vertical bar represents one individual's genotype; and b) genotypes grouped by samples for Round 2 of the Bayesian clustering analysis of the Variable Neutral SNPs dataset. For each hypothesis of K, different colours represent each of the K populations defined within the dataset.

Multivariate analyses

The PCA of the Variable Neutral SNPs dataset revealed clear clustering of samples along the first PC axis, accounting for the greatest proportion of total variance. The GG sample was clearly separated from the remaining samples along this axis, with the remaining samples overlapping. No clustering was observed along the second PC, with individuals from all samples overlapping (Figure 3.13).

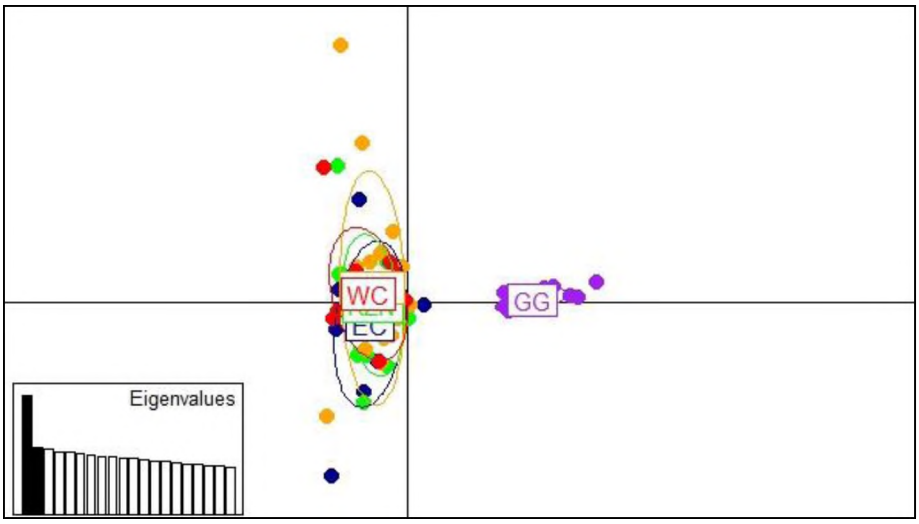

Figure 3.13: Scatterplot of results of PCA for the Variable Neutral SNPs dataset, against the first two PC axes, with eigenvalues of PCs displayed.

The k-means clustering algorithm found the most likely number of clusters to be either one or two; the clustering solution with the lowest associated BIC value was k=1 (BIC=667.34), with a small increase in BIC for k=2 (BIC=668.59), after which BIC values increased at a greater rate for subsequent clustering hypotheses (Appendix 3: Figure 5.5). Under the clustering solution of k=2, all individuals from the GG sample were assigned to one cluster, while individuals of all remaining samples were assigned to the other cluster. A cross-validation test with 100 replicated indicated that retaining the first 50 PCs in the dimension-reducing step of the DAPC allowed an optimum level of correct assignment of individuals to their original sample, with the lowest associated mean squared error (Appendix 3: Figure 5.6; Table 5.3).

By retaining the first 50 PCs, 77.2% of the total variance described by the PCA was conserved in the DAPC. The resulting DAPC scatterplot of individuals revealed a clear separation of individuals from the GG sample from individuals from the remaining samples, with overlap among remaining samples, along the first DF axis, which accounted for the largest proportion of variation observed (Figure 3.14). Separation of the KZN sample from the EC, MOZ and WC samples was observed along the second DF axis, however this axis did not account for a high level of the variation described (Figure 3.14).

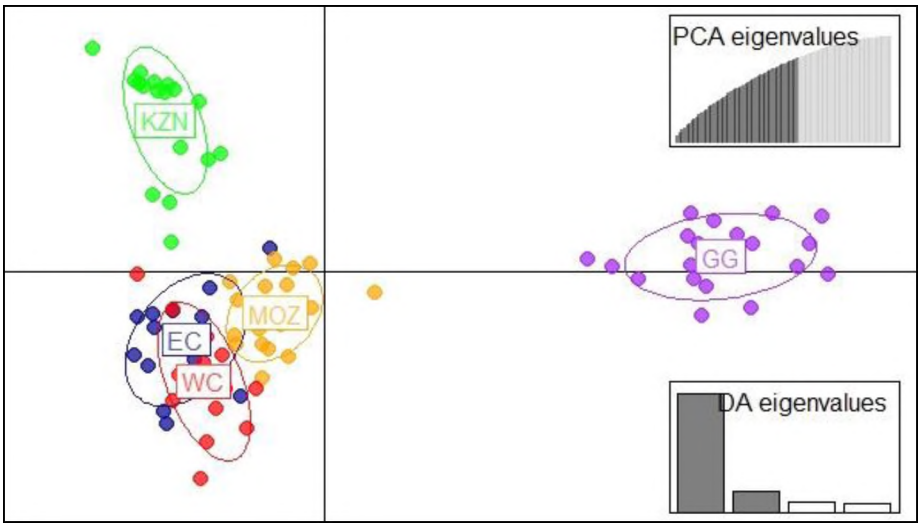

Figure 3.14: Scatterplot of DAPC for the Variable Neutral SNPs dataset, retaining the first 50 PCs of the PCA, against the first two DF axes with eigenvalues of DFs displayed.

3.2.3.4 BayeScan Outlier SNPs

Tests of Hardy-Weinberg Equilibrium and linkage disequilibrium

Four of eleven SNP loci in the BayeScan Outlier SNPs dataset (Loci 7, 8, 9 and 11) were found to deviate significantly from HWE expectations, following sequential Bonferroni correction, in the GG sample. All SNP loci from the remaining four samples were found to conform to HWE expectations.

Tests for linkage disequilibrium found pairs of loci to be in linkage, following sequential Bonferroni correction, of which three pairs were in linkage in all five samples, three were in linkage in all samples except for the MOZ sample (although these pairs were found to be in linkage in this sample prior to correction), and three pairs of loci were each found to be in linkage in one sample (Figure 3.15). The four SNP loci (SNPs 9902, 9903, 9904 and 10582) found to deviate from HWE in the GG sample appeared to be in linkage; these were SNPs 9902 (BADN01095305.1-742), 9903 (BADN01095305.1-772), 9904 (BADN01095305.1-780) and 10582 (BADN01133019.1-952). In order to remove this effect, three of these loci (SNPs 9902, 9903 and 10582) were removed from the dataset for subsequent analyses.

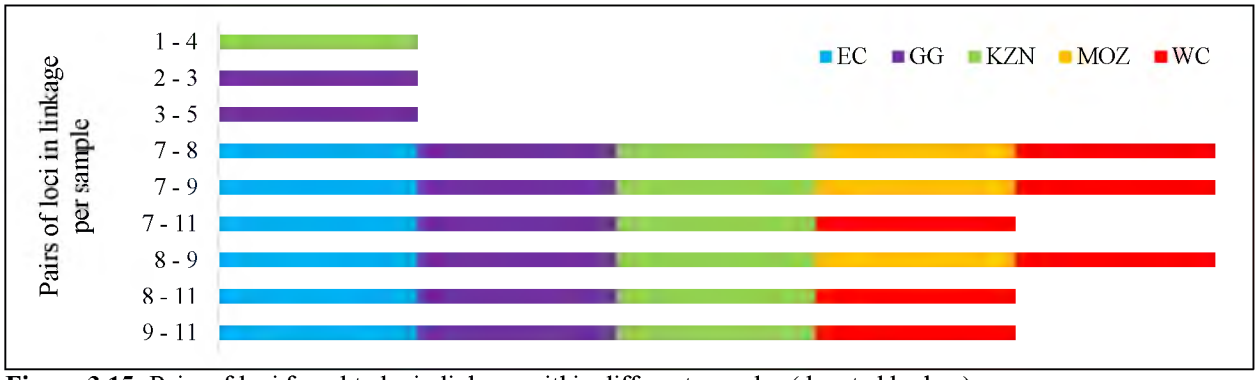

Figure 3.15: Pairs of loci found to be in linkage within different samples (denoted by key).

A Bayesian clustering and assignment analysis was not performed on the BayeScan Outlier SNPs dataset in Structure, as a result of this dataset violating Structure assumptions of all loci being in HWE and the small number of loci following removal of SNPs forming linkage groups (Pritchard *et al.* 2000).

Exact tests of genetic differentiation

Exact tests of differentiation for the BayeScan Outlier SNPs dataset found both the genic and genotypic differentiation within the dataset to be highly significant ($P<0.0001$).

As with the Variable Neutral SNPs dataset, pairwise tests found the variation within the BayeScan Outliers dataset to be portioned between the GG sample and the remaining samples. All pairwise comparisons including the GG sample were found to be highly significant ($P<0.0001$), whilst all pairwise comparisons among remaining samples were found to be non-significant ($P>0.05$) for both genic and genotypic tests of differentiation (Table 3.21).

Table 3.21: P-values for pairwise exact tests of genic differentiation (below diagonal) and genotypic differentiation (above diagonal) within BayeScan Outlier SNP dataset. Significant comparisons at $\alpha=0.01$ displayed in bold.

	EC	**GG**	**KZN**	**MOZ**	**WC**
EC	-	**<0.0001**	0.0719	0.9432	0.8854
GG	**<0.0001**	-	**<0.0001**	**<0.0001**	**<0.0001**
KZN	0.1596	**0.0000**	-	0.5769	0.3219
MOZ	0.9307	**0.0000**	0.7305	-	0.8430
WC	0.8770	**0.0000**	0.5159	0.8651	-

Analyses of molecular variance

Three-level AMOVAs of the BayeScan Outlier SNPs dataset using 6 of the 8 loci, for which less than 5% of data was missing, found the majority of variance in the dataset to be partitioned among individuals at the within-sample level under both hypotheses of population structure (described in section 3.1.2.2), with this component being smaller than for the

Combined Neutral and Variable Neutral SNPs datasets (<74%; Table 3.22). This component was significant under both hypotheses (P<0.05) and decreased by approximately 13% under Hypothesis 2. The within-group covariance component also decreased under Hypothesis 2 (by approximately 14%), and was significant under both hypotheses (P<0.05), while the between-group covariance component increased by approximately 27% under Hypothesis 2, however this component was non-significant under both hypotheses (P>0.05; Table 3.22).

Table 3.22: AMOVA tests of BayeScan Outlier SNP variation in yellowfin tuna samples: covariance components, percent variation, fixation indices and associated P-values for three hierarchical level of structure, for two alternative hypotheses of population structure. Significant covariance components and fixation indices at α=0.05 displayed in bold; significant values at α=0.0001 in bold and denoted by *.

	Source of variation	d.f.	S.S.	Covariance component	Percent variation	Fixation index	P-value
H1 GG+WC vs EC+KZN+MOZ	Between groups	1	17.960	0.130	11.35	F_{CT}=0.114	0.101
	Within groups	3	21.339	0.178	15.58	F_{SC}=**0.176***	0.000
	Within samples	173	144.381	0.835	73.06	F_{ST}=**0.269***	0.000
	Total	*177*	*183.680*	*1.142*			
H2 GG vs WC+ EC+KZN+MOZ	Between groups	1	34.824	0.536	38.56	F_{CT}=0.386	0.199
	Within groups	3	4.475	0.019	1.38	F_{SC}=**0.023**	0.025
	Within samples	173	144.381	0.835	60.05	F_{ST}=**0.399***	0.000
	Total	*177*	*183.680*	*1.390*			

F_{ST} tests of genetic differentiation

The global F_{ST} value for the BayeScan Outlier SNPs dataset indicated a high level of significant differentiation within the dataset (F_{ST}=0.2335; P=0.000).

Pairwise F_{ST} tests revealed a greater level of differentiation than those of previous datasets, with significant F_{ST} values ranging from 0.0956 (EC–KZN) to 0.3613 (GG–MOZ). As observed for the Combined Neutral and Variable Neutral SNP datasets, all comparisons including the GG sample were significant following sequential Bonferroni correction (Table 3.21). Unlike the Combined Neutral and Variable Neutral SNP datasets, this dataset also detected significant differentiation, following sequential Bonferroni correction, between the EC and KZN samples. The level of differentiation between these samples was relatively low (F_{ST} =0.0956) in comparison to the GG comparisons (F_{ST} >0.33; Table 3.23).

Table 3.23: BayeScan Outlier SNP-based F_{ST} values (below diagonal) and associated P-values (above diagonal) among yellowfin tuna samples. Comparisons significant at α=0.05 displayed in bold; comparisons significant after sequential Bonferroni correction in bold and denoted by *.

	EC	GG	KZN	MOZ	WC
EC	-	0.0000	0.0045	0.3170	0.5979
GG	**0.3343***	-	0.0000	0.0000	0.0000
KZN	**0.0956***	**0.3460***	-	0.1650	0.0578
MOZ	0.00382	**0.3613***	0.0221	-	0.4609
WC	-0.00929	**0.3306***	0.0626	-0.0002	-

Multivariate analyses

The PCA of the BayeScan Outlier SNPs dataset revealed slight clustering of samples. Along the first PC axis, accounting for the greatest proportion of total variance, the GG sample was largely separated from the remaining samples, with overlap of only a few individuals of the remaining samples. There was no clear separation among remaining samples along the first PC axis. Along the second PC axis, accounting for less variation the first PC, all samples were observed to overlap, however the GG sample overlapped with other samples to a lesser extent with some individuals of the GG sample not overlapping with individuals of other samples along this axis (Figure 3.16).

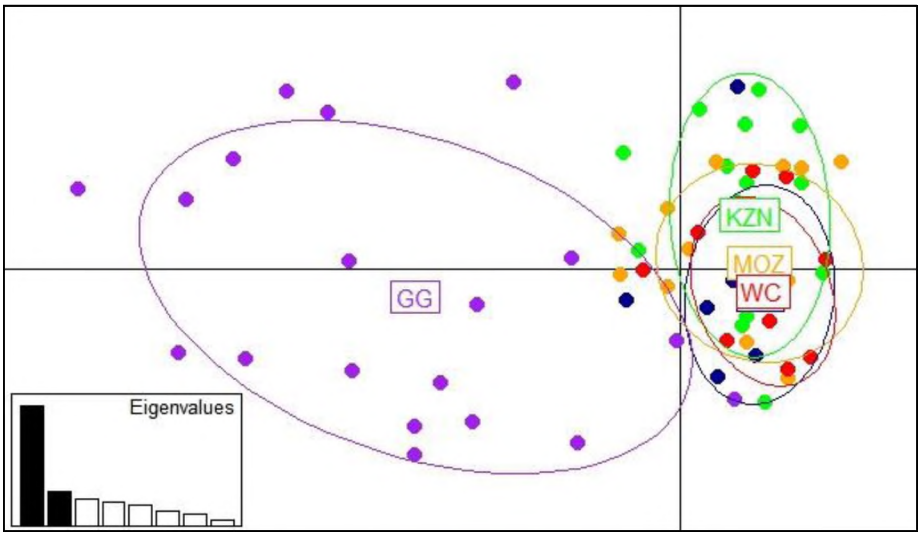

Figure 3.16: Scatterplot of results of PCA for the BayeScan Outlier SNPs dataset, against the first two PC axes, with eigenvalues of PCs displayed.

The k-means clustering algorithm found the BIC value to show a trend of continuing decrease with clustering solutions of increasing k-values, however the largest decrease in BIC values between subsequent hypotheses, from k=1 to k=20, occurred between k=1 and k=2 (Appendix 3: Figure 5.7), therefore k=2 (BIC=213.95) was selected as the most likely clustering solution. Under this clustering solution, 15 individuals from the GG sample were assigned to one cluster, while the remaining five GG individuals and all individuals from the remaining samples were assigned to the other cluster. A cross-validation test with 100 replicates indicated that retaining the first six PCs in the dimension-reducing step of DAPC would allow for the optimum level of correct assignment of individuals to their original sample, with the lowest associated mean squared error (Appendix 3: Figure 5.8; Table 5.4).

By retaining the first 6 PCs, 93.3% of the total variance described by the PCA was conserved in the DAPC. The resulting DAPC scatterplot of individuals shows that individuals from the

GG sample are largely separated from the remaining samples along the first DF, with only a few individuals of this sample overlapping with remaining samples along the first DF, accounting for the largest proportion of variance observed (Figure 3.17). No genetic clustering was observed along the second DF, which did not account for a large portion of the total variation, with individuals from all samples overlapping along this axis except for the EC and MOZ samples for which there was no overlap (Figure 3.17).

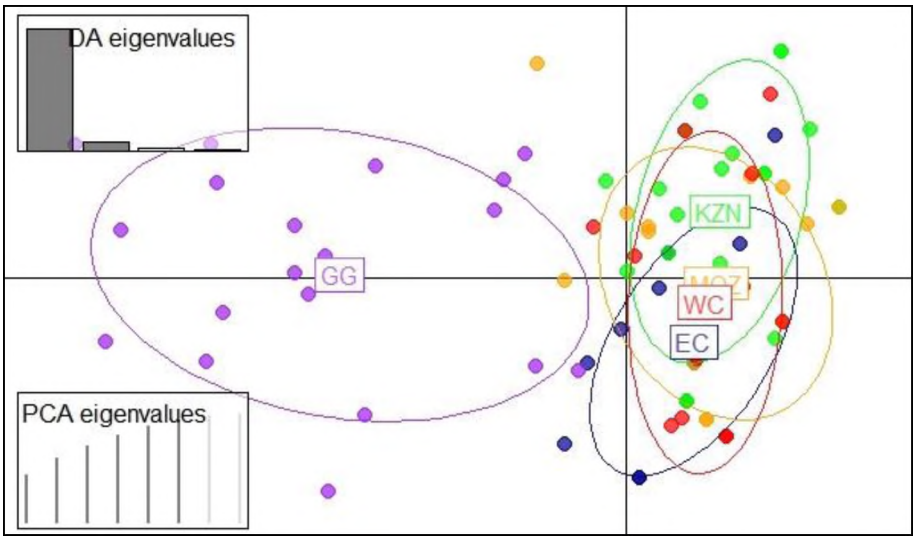

Figure 3.17: Scatterplot of DAPC for the BayeScan Outlier SNPs dataset, retaining the first 5 PCs of the PCA, against the first two DF axes with eigenvalues of DFs displayed.

3.2.3.5 Most informative SNP loci

Loadings of SNP loci to the first DF of the Combined Neutral SNPs DAPC were used to identify the most informative loci, accounting for the greatest differentiation between clusters, in this dataset. The 99.75% quantile of loadings to this DF identified 34 loci falling above this threshold (, of which 27 were also present in the Variable Neutral SNPs dataset, while 13 loci fell above the 99.9% quantile, all of which were also in the Variable Neutral SNPs dataset (Figure 18; Table 3.24).

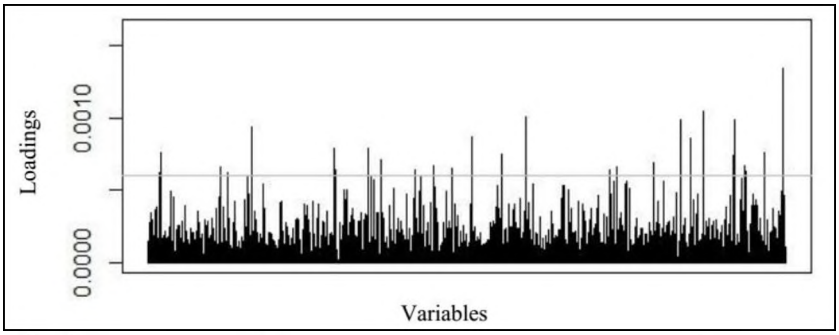

Figure 3.18: Contributions of the Combined Neutral SNPs dataset alleles to loadings of the DAPC. **Variables**: individual alleles within the dataset; **Loadings**: contributions of alleles to variation described by the first DF.

Table 3.24: Neutral SNP loci with alleles with the greatest contributions (>97.75) to the variation described by the first DF of the DAPC. Bold text denotes loci with contributions >99.90%.

SNP number	SNP name	SNP number	SNP name
SNP 199	BADN01001554.1-2225	**SNP 7051**	**BADN01062909.1-1369**
SNP 233	**BADN01001873.1-2496**	SNP 8600	BADN01079393.1-868
SNP 235	BADN01001873.1-2595	SNP 8733	BADN01080603.1-1163
SNP 1330	BADN01010292.1-2018	SNP 9422	BADN01088510.1-1714
SNP 1471	BADN01011166.1-12104	**SNP 9905**	**BADN01095305.1-848**
SNP 1840	BADN01014065.1-485	**SNP 10104**	**BADN01098010.1-4795**
SNP 1922	**BADN01014737.1-293**	SNP 10339	BADN01101482.1-447
SNP 3449	**BADN01027942.1-10841**	**SNP 10340**	**BADN01101482.1-471**
SNP 3490	BADN01028389.1-10316	SNP 10878	BADN01109115.1-3224
SNP 4088	**BADN01034232.1-859**	SNP 10920	BADN01109523.1-2085
SNP 4128	BADN01034457.1-11954	**SNP 10922**	**BADN01109523.1-2103**
SNP 4317	BADN01036303.1-5684	SNP 11102	BADN01113373.1-565
SNP 4960	BADN01042533.1-6406	SNP 11116	BADN01113648.1-2507
SNP 5306	BADN01045751.1-7003	SNP 11117	BADN01113648.1-2508
SNP 5668	BADN01049434.1-3369	**SNP 11134**	**BADN01113977.1-265**
SNP 6020	**BADN01052320.1-3254**	**SNP 11810**	**BADN01131998.1-547**
SNP 6597	BADN01058307.1-10972	**SNP 11812**	**BADN01131998.1-577**

3.2.3.6 Remaining SNP datasets

Figures and tables referred to below are displayed in Appendix 4.

FDIST2 Neutral SNPs dataset

When excluding all outlier loci detected by the FDIST2 outlier method, both balancing and diversifying, the power to detect structure within the dataset was reduced. This can be observed by the dataset's low and non-significant global differentiation (F_{ST}=0.0027; P=0.068), as well as by the non-significance (P>0.05) of pairwise exact tests of differentiation and three-level AMOVAs, and the absence of clear clustering for PCA and DAPC multivariate tests (Tables 6.1 and 6.2, and Figures 6.1 and 6.2, respectively).

Weak structure was detected for this dataset by a highly significant level of global genic differentiation (P<0.0001). Pairwise F_{ST} tests of genetic differentiation found two comparisons to be low but significantly different from zero after sequential Bonferroni correction, the GG–KZN (F_{ST}=0.0037) and GG–MOZ (F_{ST}=0.0039) comparisons (Table 6.3).

BayeScan Neutral SNPs dataset

When excluding loci detected by BayeScan as being under diversifying selection but not excluding those identified by FDIST2 as being under balancing selection, the same pattern,

and a similar but slightly lower level, of structure was observed as for the Combined Neutral SNPs datasets.

Exact tests of global genic and genotypic differentiation were both highly significant (P<0.0001), however pairwise tests revealed no significant comparisons (Table 6.4).

Three-level AMOVAs revealed an increase in between-group variance, and a decrease in within-group and within-sample variance under Hypothesis 2 of structure, however this variance was not significant at all levels of structure (α=0.05; Table 6.5).

The global F_{ST} value of this dataset was low (and similar to the of the Combined Neutral SNPs dataset) but significant (F_{ST}=0.0040; P=0.000). Pairwise F_{ST} tests found all comparisons including the GG sample to be significant after Bonferroni, with remaining comparisons being non-significant (P>0.05); the level of structure for this dataset was similar to observed for the Combined Neutral SNPs dataset, with significant F_{ST} values ranging from 0.0062 (GG–KZN) to 0.0065 (GG–MOZ; Appendix 4: Table 6.6).

Multivariate PCA and DAPC analysis revealed clustering between the GG sample and the remaining samples (Figures 6.3 and 6.4) as observed for other datasets, excluding the FDIST2 Neutral SNPs.

FDIST2 Outlier SNPs dataset

The FDIST2 Outlier SNPs dataset, which included diversifying outlier loci detected by FDIST2 as well as those detected by BayeScan, revealed the same pattern of structure as the Variable Neutral SNPs dataset, which excluded the BayeScan outliers, with a similar but slightly higher level of structure. Exact tests of genetic differentiation found the global genic and genotypic differentiation within the dataset to be highly significant (P<0.0001), and all pairwise exact tests including the GG sample to be highly significant (P<0.0001), with comparisons among remaining samples non-significant (P=1.000; Table 6.7). Three-level AMOVAs found the between-group variation to increase and the within-group and within-sample variation to decrease under Hypothesis 2 of population structure, with the same levels of structure showing significance as for the Variable Neutral SNPs; the fixation indices for the FDIST2 Outlier SNPs dataset were however slightly stronger (Table 6.8).

The global differentiation of this dataset was high and significant, with a similar F_{ST} value to that of the Variable Neutral SNPs (F_{ST}=0.0408; P=0.000). Pairwise F_{ST} tests revealed the same pattern of structure as the Variable Neutral SNPs dataset, with all GG comparisons

being significant after sequential Bonferroni correction and remaining comparisons being non-significant (P>0.05); the significant F_{ST} levels were however slightly higher for this dataset, ranging from 0.1176 (GG–KZN) to 0.1393 (GG–WC; Table 6.9).

The Bayesian clustering and assignment analysis (implemented in Structure) detected K=2 to be the most likely number of clusters within the dataset (Table 6.10), with all GG individuals being assigned to one cluster and individuals from remaining samples being assigned to the other with high admixture coefficients (Figure 6.5), as was observed for the Variable Neutral SNPs dataset. Multivariate PCA and DAPC analyses also revealed clustering between the GG sample and the remaining samples (Figures 6.6 and 6.7), as observed for other datasets excluding the FDIST2 Neutral dataset.

3.2.4 Summary of SNP results

Population structure analyses of SNPs identified from sites across the genome of yellowfin tuna revealed a pattern of structure between individuals from the Gulf of Guinea (GG) sample, and individuals from the Eastern Cape (EC), KwaZulu-Natal (KZN), Mozambique (MOZ) and Western Cape (WC) samples. This pattern was present in all datasets except for that which excluded SNPs identified by FDIST2 as being outliers (the "FDIST2 Neutral SNPs" dataset), and the level of this structure differed between datasets including different subsets of loci. The strongest level of differentiation was observed for the "BayeScan Outlier SNPs" dataset, including only loci identified by BayeScan as being putatively under diversifying selection (see section 3.2.3.4). The "Variable Neutral SNPs" dataset revealed a relatively high level of structure (section 3.2.3.3), while the "Combined Neutral SNPs" (section 3.2.3.2) and "FDIST2 Neutral SNPs" datasets, and the dataset including all SNP loci (section 3.2.2.2), revealed a lower level of structure.

CHAPTER 4
DISCUSSION

4.1 *Overview*

The salient features of the nuclear DNA data were (i) the differentiation of the Gulf of Guinea sample from all other samples, for all datasets except for that wherein all FDIST2 outliers were removed, and (ii) the non-significant differentiation of the Western Cape sample from all Indian Ocean samples. These patterns can be readily applied to stock management and assessment, and permit robust inferences of fisheries-relevant stock dynamics: specifically (i) restricted connectivity across the Benguela Upwelling System, and (ii) cohesion among the Western Cape and Indian Ocean individuals. In contrast to the nuclear data, mitochondrial DNA revealed no population structure, but revealed signals of historical population expansion and connectivity. This research therefore additionally supports the enhanced resolution of NGS methods for studies of stock structure and traceability of yellowfin tuna.

4.2 *Results of genetic analysis*

4.2.1 Mitochondrial DNA Control Region

4.2.1.1 Genetic diversity

High levels of genetic diversity were observed for the mitochondrial DNA Control Region (CR), which were similar across all sampling sites providing no evidence of spatial variation in mtDNA genetic diversity. The high genetic diversity indices observed here are consistent with those of previous studies employing the mtDNA CR to assess the population structure of yellowfin tuna across various geographic localities. Levels of haplotype diversity (overall h=0.999) and nucleotide diversity (π=0.024) were similar to those observed by Ely *et al.* (2005), Wu *et al.* (2010) and Kunal *et al.* (2013), with overall h values among these studies ranging from 0.996 (Wu *et al.* 2010) to 0.999 (Kunal *et al.* 2013), and overall π ranging from 0.025 (Wu *et al.* 2010) to 0.101 (Kunal *et al.* 2013). These studies found the majority of observed haplotypes to be singletons (present in only one individual), as was observed in this study. This is characteristic of fishes of the Scombridae family which tend to exhibit high levels of mtDNA CR variability, with similar levels of h and π having been observed for mackerel species *Scomber scombrus* and *S. japonicus* (Zardoya *et al.* 2004), Atlantic bluefin tuna *Thunnus thynnus* (Ely *et al.* 2002; Alvarado Bremer *et al.* 2005) and swordfish *Xiphius*

gladias (Alvarado Bremer *et al.* 1996, 2005; Chow *et al.* 1997). Other regions of the mtDNA genome, with slower rates of mutation accumulation, have been observed to be less variable than the CR in yellowfin tuna with lower levels of *h* and π observed among samples at the ATPase 6 and 8 regions (Dammannagoda *et al.* 2008) and the cytochrome c oxidase subunit I (COI) region (Henriques 2011; Li *et al.* 2015). The high level of genetic diversity at the yellowfin tuna CR may be due to the hypervariable nature of this non-coding region, which accumulates mutations at a high rate relative to other mtDNA regions resulting in high levels of π (Meyer 1993), combined with large female effective population sizes (N_e) of the species (Ely *et al.* 2005), resulting in a high diversity of maternally inherited mtDNA genomes among individuals, and therefore high levels of *h* (Alvarado Bremer *et al.* 1997; Avise 1998; Ely *et al.* 2005).

4.2.1.2 Population structure

No discernible genetic structure was detected at the mtDNA CR among yellowfin tuna from different sampling locations, as can be observed by the weak and non-significant overall and pairwise F_{ST} values, exact tests of differentiation and hierarchical analyses of molecular variance (AMOVAs), and the lack of a geographic pattern for the haplotype network. A lack of structure at the CR is consistent with the results of Wu *et al.* (2010), who did not detect significant geographic differentiation among yellowfin tuna individuals from the western Indian and western Pacific Oceans, despite structure between these oceans having been detected at one allozyme locus (Ward *et al.* 1997), at the mtDNA ATCO gene (Ely *et al.* 2015) and with genome-wide SNP genotypes (Pecoraro *et al.* 2016a). Kunal *et al.* (2013) additionally detected significant but weak differentiation among yellowfin tuna samples from northern Indian Ocean waters at the CR, while stronger structure was detected within this region at mtDNA ATPase genes and two microsatellite loci (Dammannagoda *et al.* 2008).

The lack of, or weak, resolving power of the yellowfin tuna CR may be a result of the high level of genetic variation present at this mtDNA region. A high level of genetic diversity is desirable for detecting significant differentiation among samples, however too high a level of diversity may make regions uninformative as molecular markers (Duncan *et al.* 2015). The high *h* observed suggests that the sample sizes employed by this study (N=16 to 20) were not sufficiently large to detect geographic structure using haplotype frequencies; CR analyses of larger samples by Wu *et al.* (2010) (N=51 to 73) and Kunal *et al.* (2013) (N>48) also failed to detect conclusive structuring, indicating that substantially larger samples may be required for this region to become informative. The hypervariable nature of the CR may also result in

homoplasies, whereby the same mutations arise independently among individuals rather than through inheritance and common ancestry, resulting in convergence that complicates the reconstruction of geographic structure among haplotypes (Ely *et al.* 2005). Other mtDNA regions with slower rates of mutation, i.e. the ATCO, ATPase and COI genes (Ely *et al.* 2005; Dammannagoda *et al.* 2008; Henriques 2011; Li *et al.* 2015), have been more useful in detecting significant genetic structure among samples of yellowfin tuna from various region; thus, based on its high genetic diversity, it appears that the mtDNA CR is not able to detect fine scale, recurrent genetic structure within yellowfin tuna and is therefore not a suitable molecular marker for resolving population/stock structuring of the species of interest to fisheries management. This marker was however useful in this study for the confirmation that all sampled individuals were yellowfin tuna.

4.2.1.3 Demographic history

The mtDNA CR was useful in inferring the past demographic processes shaping current yellowfin tuna populations, and suggests that the species has undergone a past population expansion. Genetic diversity indices for the samples examined here (h>0.05 and π<0.05) are characteristic of populations that have undergone past rapid expansion from a relatively small population size, possibly following a bottleneck or founder event (Grant & Bowen 1998). Signatures of demographic expansions were also evident from the negative and significant values of Tajima's D and Fu's F_S (Excoffier 2004), and the unimodal distribution of nucleotide differences among haplotypes (Harpending 1994). A past population expansion of yellowfin tuna populations has also been suggested by Ely *et al.* (2005), Wu *et al.* (2010), Henriques (2011), Kunal *et al.* (2013) and Li *et al.* (2015).

Mismatch distribution parameters observed here indicate that a population expansion occurred approximately 318 thousand years before the present (KYA). This time since expansion may however be inflated as molecular mutation rates have been demonstrated to be time-dependent, rather than constant over time, whereby mutation rates within populations are higher over short time scales and decrease with longer periods of time (Ho *et al.* 2011). The mutation rate of the CR over the short time scale of relevance here (the last glacial event) may thus be higher than that used in calculation of time since expansion (3.69% per million years) based on longer time scales of between species divergence following closure of the Isthmus of Panama ~3.5 million YA (Donaldson & Wilson 1999); it is therefore possible that expansion may have occurred more recently than the 318 KYA calculated here. Additionally,

a general teleost CR mutation rate was used for calculations and may not reflect the rates of mutation of the yellowfin tuna CR.

Previous studies have found a broad range of population expansion time estimates using different mtDNA genome regions. At the mtDNA CR, Ely *et al.* (2005) estimated an expansion time of 522 KYA for pooled Atlantic, Indian and Pacific Ocean samples, using a less conservative mutation rate (4.9% per million years) than that employed here, while Kunal *et al.* (2013) estimated an expansion at 360 KYA for Indian Ocean samples using the same mutation rate (3.6% per million years) as used here. At the mtDNA COI region, Henriques (2011) estimated more recent expansion times of 34.2 and 19.8 KYA for putative Atlantic and Indian Ocean populations, respectively, while Li *et al.* (2015) estimated an expansion time of 820 KYA for eastern and western Pacific Ocean samples; the time-dependent effect on molecular mutation rates may however differ across mtDNA regions (Ho *et al.* 2011) and as such it is not meaningful to compare estimates across regions. A broad range of estimates of time since expansion of yellowfin tuna populations has thus been found across studies and mtDNA regions (from 19.8 KYA; (Henriques 2011) to 820 KYA (Li *et al.* 2015)), and while it cannot be conclusively stated when expansion of yellowfin tuna populations occurred, the combined evidence suggests that these population expansions occurred during or following the period of Pleistocene glacial maxima (Hewitt 1996).

4.2.2 Nuclear DNA single nucleotide polymorphisms

4.2.2.1 Sequencing success

Following quality filtering of sequence reads, seven individuals from the Western Cape sample were removed from the dataset as their rate of missing SNPs was above the 60% threshold as a result of lower sequencing success and a resulting high rate of missing data. Sampling of Western Cape yellowfin tuna was conducted prior to sampling in other regions, and tissue samples from the Western Cape were preserved in ethanol for six to ten months longer than the remaining samples. Degradation of genomic DNA over time can lead to the missing loci or restriction enzyme recognition sites, and the genome being cut at fewer sites (Graham *et al.* 2015). It is therefore possible the high rate of missing data within this sample is due to degradation of DNA, resulting in fewer cut sites and missing homologous sequence reads and SNPs.

4.2.2.2 Genetic diversity

The SNPs analysed were bi-allelic and as such, had relatively low levels of genetic diversity compared to highly polymorphic, multi-allelic molecular markers such as microsatellites (Brookes 1999; Vignal *et al.* 2002; Helyar *et al.* 2011). The mean number of alleles (MNA) across bi-allelic SNP loci cannot exceed MNA=2, while the expected heterozygosity (H_E), based on the observed allele frequency per locus, cannot exceed H_E=0.50 (Ryynänen *et al.* 2007; Helyar *et al.* 2011; DeFaveri *et al.* 2013), although this value depends on the estimator of H_E employed. The genetic diversity observed here for 11 869 SNP markers was thus relatively high and very similar among sampling locations, showing little evidence of spatial genetic variability. Genetic diversity indices were slightly lower than, but similar to, those observed by Grewe *et al.* (2015) across 6 217 SNP loci for three Pacific Ocean samples of similar sizes (N=22 to 24) to those analysed here (Table 4.1); results between this study and that of Grewe *et al.* (2015) however cannot be directly compared as genetic diversity of SNP genotypes however depends on the bioinformatics filtering of raw sequence data to obtain them, as the stringency of this affects the number and genetic diversity of SNPs discovered.

Although positive and significant (P<0.05) inbreeding coefficient (F_{IS}) values were observed for each sample, suggesting a deficiency of heterozygotes (Beebee & Rowe 2005), these values did not remain significant following Bonferroni correction. Only 1% of loci had significant F_{IS} values overall after Bonferroni correction, and no loci were found to deviate significantly from Hardy-Weinberg Equilibrium (HWE) expectations; it is thus most likely that the no mixing of genetic cohorts is present within samples.

Genetic diversity indices of yellowfin tuna across SNP loci were low compared to those found by previous studies using microsatellite markers (Appleyard *et al.* 2001; Díaz-Jaimes & Uribe-Alcocer 2006; Henriques 2011; Qiu and Miyamoto 2011; Aguila *et al.* 2015), due to the bi-allelic nature of SNPs analysed, summarised in Table 4.1. Levels of genetic diversity between SNPs and microsatellites within species' genomes are likely to be largely correlated (Ryynänen *et al.* 2007; DeFaveri *et al.* 2013), however genome-wide SNPs genotypes are more likely to reflect genome-wide levels of genetic diversity within samples than small numbers of microsatellite markers from few genomic regions (Helyar *et al.* 2011; Gagnaire *et al.* 2015), and these results are therefore likely to provide a genome-wide estimate of yellowfin tuna genetic diversity.

Table 4.1: Genetic diversity indices obtained for this and previous studies by analysis of SNP and microsatellite nDNA molecular markers, as average across all loci, for a range of samples.

Study	Marker number and type	MNA range	H_E range	H_O range
Present study	11 869 SNPs	1.83 – 1.91	0.220 – 0.227	0.200 – 0.207
Grewe *et al.* 2015	6 712 SNPs	1.98 – 1.98	0.234 – 0.259	0.225 – 0.293
Appleyard *et al.* 2001	5 microsatellites	11.2 – 15.8	0.602 – 0.644	0.562 – 0.630
Díaz-Jaimes & Uribe-Alcocer 2006	7 microsatellites	5.14 – 7.86	0.523 – 0.604	0.478 – 0.535
Henriques 2011	6 microsatellites	11 – 12	0.741 – 0.773	0.478 – 0.535
Qiu & Miyamoto 2011	10 microsatellites	3 – 11	0.353 – 0.897	0.160 – 0.844
Aguila *et al.* 2015	7 microsatellites	6.9 – 9.1	0.634 – 0.715	0.566 – 0.674

The high levels of genetic diversity observed for yellowfin tuna, by this and previous studies, suggest that the species' fitness and ability to adapt to changing environmental conditions may be relatively resilient to fishing pressure. Genetic diversity, measured by A_R, is an indication of a population's potential to adapt to changes in environmental conditions, as diversity allows for the evolution of species based on selection pressures (Allendorf 1986; Caballero & Garcia-Dorado 2013; Greenbaum *et al.* 2014). Intense fishing pressure and resultant reduced population sizes may however result in a loss of genetic diversity and adaptive potential of species (Kenchington *et al.* 2003), which may lead to population collapses with changed selection pressures. The large N_e of yellowfin tuna populations may allow the species to maintain high levels of genetic diversity due to lower levels of reproductive variance (Charlesworth 2009), as observed for a European anchovy *Engraulis encrasicolus* population that suffered a collapse, but maintained high levels of genetic diversity as a result of its large N_e (Montes *et al.* 2016). This is as opposed to species with large N_c but low N_e/N_c ratios, which may lose genetic diversity and reproductive potential with high levels of fishing pressure, despite maintaining large census population sizes as observed for New Zealand snapper *Pagrus auratus* by Hauser *et al.* (2002). Despite the large N_e of yellowfin tuna populations, and its resultant evolutionary potential, its numbers may still suffer from collapses from which it may not be able to recover if its populations are harvested above MSY levels for sustained periods (Pecoraro *et al.* 2016b); it is therefore important that populations of this heavily fished species are accurately defined and assessed, in order to be fished at sustainable levels.

4.2.2.3 Population structure

Analyses of genotypes of all 11 869 identified and filtered SNPs (hereafter referred to as "All SNPs" dataset) identified two clear patterns of genetic differentiation: firstly, significant differentiation was observed between the Atlantic Ocean and Indian Ocean samples of yellowfin tuna and secondly, the sample of Western Cape individuals was observed to group with the Indian Ocean samples (Eastern Cape, KwaZulu-Natal and Mozambique) rather than the Atlantic Ocean (Gulf of Guinea) sample.

The relatively weak but highly significant global genetic differentiation (F_{ST}=0.0043; P<0.0001) within the dataset resulted from a consistent pattern between the Atlantic Ocean (Gulf of Guinea) sample and the remaining Indian Ocean samples (including the Western Cape sample), with highly significant differentiation of the Gulf of Guinea sample in both F_{ST} (>0.006; P<0.0001) and exact tests but no significant differentiation among any of the remaining samples (P>0.05). This pattern was also supported by AMOVA and multivariate analyses (PCA and DAPC). Bayesian clustering analysis in Structure (Pritchard *et al.* 2000) failed to detect genetic structure within the dataset; this may however be due to this method's lack of power to identify genetic clusters when structure within a dataset is low (F_{ST}<0.03) (Latch *et al.* 2006).

Results of GeneClass2 assignment tests have provided complementary information to the results of traditional equilibrium-based analyses, such as F_{ST}. Assignment patterns were congruent with other tests in supporting differentiation of GG individuals from those of remaining samples through high levels of self-assignment and low levels of misassignment between these groups (GG vs remaining samples), while the low levels of self-classification among samples excluding the GG sample are consistent with a shared genepool and support the genetic similarity among Indian Ocean and WC samples.

The assignment tests did identify first generation migrants between the GG sample and the Indian Ocean grouping (within the EC and KZN samples). While these numbers were low, they should be sufficient to erode genetic differentiation and create genetic homogeneity between Atlantic and Indian Ocean populations. The occurrence of genetic differentiation between Atlantic and Indian Ocean groupings detected here therefore suggests that such migrants do not interbreed, or their respective hybrids have low survival rates. Collectively, this implies that factors other than physical restriction to gene flow, such as behaviour and local adaptation, may contribute to restricted gene flow despite some level of straying, as has

been observed for other species such as Atlantic herring *Clupea harengus* (Gaggiotti *et al.* 2009).

The structure presented here between Atlantic and Indian Ocean samples of yellowfin tuna supports the findings of previous studies and was therefore not unexpected. Ward *et al.* (1997) identified significant differentiation between yellowfin tuna samples from the western central Atlantic Ocean and the northern central Indian Ocean at one allozyme locus (GPI-A*), and it is upon these findings that the currently assumed global population and stock structure of the species is based. This allozyme locus may be under selection, however the results presented here for neutral loci (excluding the BayeScan diversifying outliers) support this structure, and suggest a break down in gene flow and connectivity between Atlantic and Indian Ocean yellowfin tuna. Demographic isolation between these regions has also been inferred from neutral markers by previous studies; Henriques (2011) found a western central Atlantic Ocean sample to be weakly but significantly differentiated from two south western and two central Indian Ocean samples at the mtDNA COI region, and from a south western Indian Ocean sample at six microsatellite loci, and Pecoraro *et al.* (2016a) found further evidence for this structure with strong and significant differentiation between four Atlantic Ocean samples (two eastern central and two western central) and two western central Indian Ocean samples with genotypes of 6 772 SNP loci. Based on this structuring between Atlantic and Indian Ocean yellowfin tuna, the ultimate aim of this study was to determine the geographic boundary between these two populations by assessing the population origin of individuals occurring off the Western Cape province of South Africa.

The results presented here provide strong and consistent evidence to suggest that the yellowfin tuna occurring off the Western Cape, within south eastern Atlantic Ocean waters, are migrants to the region from the Indian Ocean population rather than the Atlantic Ocean population. This structure supports that previously observed by Henriques (2011); a sample of yellowfin tuna from the Western Cape exhibited a relatively strong and significant level of mtDNA COI haplotype differentiation from a western Atlantic Ocean sample from Brazil (F_{ST}=0.098; P<0.05), and no significant differentiation (P>0.05) from five Indian Ocean samples, including three central Indian Ocean samples. This was weakly supported by nuclear data; at six microsatellite loci, significant genic and genotypic differentiation (P<0.005) was observed between the Brazil and Western Cape samples with no significant differentiation between the Western Cape and Indian Ocean samples, however F_{ST} tests did not detect this structure. The weak structure revealed by these markers, which was not

consistent across all analyses, was not sufficient to include Western Cape yellowfin tuna individuals in the Indian Ocean population. Sub-structuring within the Atlantic Ocean basin due to different spawning populations could not be rejected as the cause of differentiation between eastern and western Atlantic Ocean samples, and the weak and inconsistent results did not assess fine-scale structuring off South Africa, such as admixture or mixture of Atlantic and Indian Ocean individuals (Henriques 2011). Tagging data for yellowfin tuna in the Atlantic Ocean suggests a single population in this ocean basin (Maury *et al.* 2001; Zagaglia *et al.* 2004). The inclusion of an eastern Atlantic Ocean sample, from the main spawning area of Atlantic Ocean yellowfin tuna (Pecoraro *et al.* 2016b), in the present study addresses this issue, and the clear differentiation observed between Western Cape and Gulf of Guinea samples supports the hypothesis that the results of Henriques (2011) were due to the Western Cape individuals being of Indian Ocean origin.

4.2.3 Next-generation sequencing and population genomics in the study of yellowfin tuna populations

These results demonstrate that population genomic analyses of datasets of genotypes of thousands of SNP markers, obtained using next-generation sequencing (NGS) technology, are efficient in resolving population structure in yellowfin tuna, as previously observed by Grewe *et al.* (2015) and Pecoraro *et al.* (2016a). Increased statistical power to detect genetic differentiation among populations was observed for this study, compared to that of small numbers of traditionally used markers. Datasets of thousands of SNP loci are expected to increase the power and precision of population genetic analyses to identify genetic differentiation among weakly structured populations (Waples 1998; Willette *et al.* 2014; Gagnaire *et al.* 2015). The low genetic diversity levels of SNP loci, as discussed in section 4.2.2.2, mean that individual SNPs have relatively low information content when compared to multi-allelic markers such as microsatellites (Brookes 1999; Vignal *et al.* 2002; Rosenberg *et al.* 2003; Helyar *et al.* 2011). The statistical power of genetic marker datasets is however dependent on the total number of independent alleles assessed rather than the number alleles per marker (Kalinowski 2002; Helyar *et al.* 2011). Kalinowski (2002) equated the statistical power to discriminate among populations of 100 SNP loci to that of between 10 and 20 microsatellites. Analyses of traditional molecular markers (such as mtDNA sequences and nDNA microsatellites) have had varied and limited success in detecting structure among the Atlantic and Indian Oceans, as seen by previous studies (Ward *et al.* 1997; Ely *et al.* 2005; Henriques 2011) and in the present study for the mtDNA CR. Building on the findings of

Grewe *et al.* (2015) and Pecoraro *et al.* (2016b), the results presented here show that population genomics analyses of thousands of SNP markers provide increased statistical power and precision to resolve population structure in yellowfin tuna.

4.2.3.1 Definition of Neutral and Outlier datasets

The definition of neutral and outlier datasets was based on the shortcomings of each of the two outlier detection tests employed. The FDIST2 F_{ST} method (Beaumont & Nichols 1996) is associated with a high Type I error rate (false-positive detection of outliers) (Narum & Hess 2011; Russello *et al.* 2012; Lotterhos & Whitlock 2014). This method defines outlier loci as those with genetic differentiation levels among putative populations that fall outside of a null distribution of F_{ST} values expected under neutral processes. The null distribution is however based on an island model of population subdivision (Wright 1931), while in reality, demographic processes resulting in population differentiation may be more complex than this model. The range of neutral F_{ST} values defined by the island model may therefore be narrower than a species' actual range, resulting in some neutral loci by chance falling outside the null distribution (Lotterhos & Whitlock 2014). To account for the relatively high Type I error rate of FDIST2 the more conservative dataset obtained by means of the 'Force mean F_{ST}' approach was considered to be a more reliable set of outliers putatively under diversifying selection. Bayesian methods of outlier detection, such as that implemented in BayeScan (Foll & Gaggiotti 2006) have a lower Type I error rate and have more power to detect loci under diversifying selection (Narum & Hess 2011; De Mita *et al.* 2013). In cases where population structure is low, BayeScan has a high power to detect outliers under diversifying selection that differ from the genome-wide average, however in these cases it lacks power to discriminate between loci under balancing selection and neutral loci, and is associated with a relatively high Type II error rate in the detection of balancing outliers (Helyar *et al.* 2011). As expected, and consistent with previous studies (e.g. Freamo *et al.* 2011; Russello *et al.* 2012), the FDIST2 method detected a large number of diversifying and balancing outlier SNPs (869 and 512 loci respectively), while the BayeScan test detected only 11 diversifying outliers (all of which were identified by diversifying outliers by FDIST2), and no balancing outliers. Subsets of neutral and outlier SNP loci were defined based on these properties of each method.

The genome-wide analysis of SNP markers identified through tGBS® permitted increased power to detect markers under selection and thus to disentangle the roles of selection and neutral processes in shaping genetic structure. Three datasets identified through outlier tests

were considered to represent genomic regions under the influence of neutral processes and not subject to selection pressures: the "FDIST2 Neutral SNPs", "Combined Neutral SNPs" and "Variable Neutral SNPs" datasets. The "BayeScan Neutral SNPs" dataset was considered to represent regions under the influence of neutral processes and regions potentially under balancing selection. The "BayeScan Outlier SNPs" dataset was considered to consist of regions potentially under the influence of selection pressures, while the "FDIST2 Outlier SNPs" dataset is considered to consist of a combination of highly differentiated neutral and possibly selected regions.

Of the original 11 862 SNPs, BayeScan identified 11 positive outliers potentially under diversifying selection. The 858 loci identified as being under selection by the FDIST2 method but not by BayeScan were considered to be neutral regions with a relatively high level of differentiation among putative populations, which therefore fell outside of the FDIST2 null distribution, as the levels of genetic differentiation may differ among genomic loci related to the different nucleotide diversity levels (Helyar *et al.* 2011). The probability of detecting loci in, or linked to, selected regions is relatively low due to the large size of genomes and the rapid decay of linkage disequilibrium (LD) in natural populations (Bonin 2008), so it was therefore considered unlikely that >3% of SNPs would be under the influence of diversifying selection. The 512 loci identified by FDIST2 as being under balancing selection were considered to consist of a combination of regions potentially under balancing selection and neutral regions with low levels of differentiation between putative populations, which would be uninformative in population genetics analyses. The remaining 11 339 loci were considered to be neutral regions with varying levels of differentiation among samples.

4.2.3.2 Neutral SNP loci

The three neutral SNPs datasets are considered to represent different subsets of neutral loci with different levels of genetic differentiation among populations. The "FDIST2 Neutral SNPs" dataset of 10 482 loci is the most weakly differentiated, as it excludes the loci identified by FDIST2 as being outliers assumed to be the most variable neutral SNPs. This dataset of >10 000 SNPs did not reveal either significant overall structure (F_{ST}<0.005; P>0.05), or a clear pattern of geographic structure. This demonstrates that in species with large N_e that prevents the accumulation of genetic drift effects in populations, thousands of neutral loci may not improve the statistical power or precision of datasets to detect structure, as the signature of genetic drift present remains unchanged, even with increased numbers of

markers, as suggested by Gagnaire *et al.* (2015); however it is likely that the lack of structure observed is a result of the high Type I error rate of the FDIST2 method, resulting in exclusion of more variable neutral SNPs in this dataset and a downward bias of the pattern of differentiation.

The "Variable Neutral SNPs" dataset, a subset of 858 of the most variable neutral SNPs, however demonstrated the benefits of reduced-representation genome sequencing for molecular marker discovery; with thousands of genome-wide loci, neutral markers that are more variable among populations than the genome-wide average can be identified and analysed to detect signatures of genetic drift and isolation not present at the majority of neutral loci (Helyar *et al.* 2011). This dataset revealed relatively strong and highly significant population structure (F_{ST}>0.04; P=0.000), with strong (F_{ST}>0.10) differentiation partitioned between Atlantic and Indian Oceans samples and the Western Cape sample grouping with the Indian Ocean samples. This pattern was consistent across all genetic analyses, including multivariate analyses. The greater level of genetic differentiation in this dataset allowed for a Bayesian clustering and assignment analysis to be performed in Structure (Pritchard *et al.* 2000); this analysis found all Western Cape individuals to cluster with the Indian Ocean population with no mixing or admixture, suggesting that all individuals occurring in this region are migrants from the Indian Ocean. Individual assignment for K-values greater than two indicated a pattern of sub-structure among Indian Ocean (including Western Cape) individuals, however the conformance of all samples to HWE expectations indicates a lack of mixed cohorts and sub-structuring.

Individual-based analyses of population structure, such as Bayesian clustering and multivariate analyses, may provide less biased pictures of population structure than approaches based on differentiation between pre-defined (and possibly biased) groupings (such as F_{ST} analyses) and they do not rely on assumptions of underlying models of differentiation (Wilson & Rannala 2003; Sham *et al.* 2009; Allendorf *et al.* 2010). Structure analyses can additionally reveal the origin of individuals without including sampling information in the analyses, and can reveal mixing or admixture (hybridisation) of individuals within samples (Pritchard *et al.* 2000); this is essential for the assessment of fine-scale genetic structure at the boundary between populations.

The pattern of population structure detected for the Combined Neutral SNPs dataset of 11 339 SNPs remained the same as that of the Variable Neutral SNPs dataset. The differentiation

within this dataset was however weaker (F_{ST}=0.004; P=0.000), as expected with inclusion of the less variable FDIST2 neutral SNPs. This indicates that the 858 highly variable SNPs identified as outliers by FDIST2 but not by BayeScan account for the majority of differentiation within the dataset.

The same pattern of structure as that detected for the genotypes of all SNPs (see section 4.2.2.3) was revealed by analyses of subsets of neutral SNPs. This suggests that the pattern of population structure inferred by this study reflects limited gene flow between yellowfin tuna from the Atlantic Ocean and those from the Indian Ocean, including the Western Cape, and that individuals from these two regions are effectively reproductively isolated from one another and represent two discrete populations.

4.2.3.3 Outlier SNP loci

The BayeScan Outlier SNPs dataset of 11 loci that may be under strong diversifying selection showed a similar pattern of population structure to that of the total SNPs dataset. The overall structure within the BayeScan Outlier SNPs dataset was higher than that of the neutral datasets (F_{ST}>0.23; P=0.000) however the pattern of structure was less stable and consistent across analyses, which is likely due to the small number of loci (eight following removal of loci in LD) and the high level of variation within the dataset. This can be seen by pairwise F_{ST} tests which found the Eastern Cape and KwaZulu-Natal samples to differ significantly, which is likely to be an artefact of the sampling based on the geographic positions of these sampling sites relative to other sites from which they did not differ significantly. It is also possible that this differentiation may arise due to some level of sub-structuring within the Indian Ocean population of yellowfin tuna, or may be due to the effect of sampling different size classes (representing different cohorts), as a number of smaller individuals (six fish of ≤60 cm FL) were sampled from KZN, compared to the other South African sites (all fish from EC >93 cm FL; WC fish >111 cm FL). The k-means clustering algorithm of the DAPC also revealed high levels of variation among individuals within the dataset, with the likelihood of clustering hypotheses rising with increasing k-values, suggesting more than 20 clusters within the dataset. The multivariate clustering analyses also showed a higher degree of overlap of the Gulf of Guinea individuals with the remaining individuals, indicating the high level of variation in this dataset is not as clearly partitioned among populations.

Population genetics has traditionally relied on neutral markers to assess the influence of evolutionary genetic drift effects. However, loci under the influence of selection can provide

evidence of contemporary geographically-based patterns of interbreeding (Hauser & Carvalho 2008). In species with large N_e, such as yellowfin tuna, the efficiency of selection in allele fixation and in countering the homogenising effect of gene flow is greater than that of genetic drift, the effects of which are retarded by very large population sizes (Allendorf *et al.* 2010; Gagnaire *et al.* 2015). Yellowfin tuna have an opportunistic life history strategy, and therefore have high levels of fecundity (Arocha *et al.* 2001; Juan-Jordá *et al.* 2013b; Zudaire *et al.* 2013; Diaha *et al.* 2015; Pecoraro *et al.* 2016b) and juvenile mortality (Hampton 2000; Pecoraro *et al.* 2016b); this means that survival of individuals may be related to environmental conditions, that allowing favourable adaptations to environmentsto become fixed rapidly within populations (Allendorf *et al.* 2010). Differentiation at adaptive loci may therefore represent allele fixation through inheritance, or difference in survival based on short-term environmental conditions. The differentiation observed at outlier loci may thus be related to variable environmental conditions across habitats.

Despite some variation in the overall pattern of population differentiation, the analyses of SNPs potentially under adaptive selection suggest that individuals from the Indian Ocean and Western Cape samples originated in the same population and were spawned under similar environmental conditions and selection pressures, different to those experienced by the Gulf of Guinea sample. Differentiation of the Gulf of Guinea sample at neutral loci indicates a level of reproductive isolation, while differentiation at potentially diversifying (outlier) loci suggests that these populations are also locally adapted. These results therefore suggest that both neutral evolutionary forces (breakdown of gene flow), and contemporary selective forces (different environments) are driving the same pattern of population isolation between Atlantic and Indian Ocean populations, and that the pattern of genetic differentiation observed for neutral SNPs is being maintained on ecological time scales that are of relevance to fisheries management objectives.

It is not known whether these outlier markers are in fact located in (or linked to) coding genomic regions, and therefore their functional significance; the high linkage disequilibrium among four outlier SNP loci (SNPs 9902, 9903, 9904 and 10582) within populations however suggests that these may represent true genomic loci, located in relatively small genomic regions potentially under selection, as has been observed for linkage groups of outlier SNP loci associated with temperature in Atlantic cod (Bradbury *et al.* 2010, 2013). To assess conclusively whether they are in genic regions, these loci will need to be aligned to the

reference genome for Pacific bluefin tuna *Thunnus orientalis*, sequenced by Nakamura *et al.* (2013).

4.3 *Evolutionary history of yellowfin tuna*

4.3.1 Historical population isolation

It is suggested that cold sea surface water temperatures during Pleistocene glacial maxima drove vicariance between Atlantic and Indian Ocean yellowfin tuna populations. During these periods, colder water at higher latitudes would have contracted the habitats of tropical and subtropical marine fish species towards the equator, and colder water off southern Africa is likely to have restricted the inter-oceanic movement and gene flow of intraspecific individuals between the Atlantic and Indian Oceans leading to population isolation (Alvarado Bremer *et al.* 1998, 2005; Graves & McDowell *et al.* 2003). This evolutionary history has been proposed to account for the structure observed between the Atlantic and Indo-Pacific Ocean populations of other tropical and subtropical large pelagic fish species; separate mtDNA clades have been identified between these regions for bigeye tuna (Alvarado Bremer *et al.* 1998; Chow *et al.* 2000b; Durand *et al.* 2005; Chiang *et al.* 2008), blue marlin *Makaira nigricans* (Buonaccorsi *et al.* 2001; Graves & McDowell 2003, 2015), swordfish (Alvarado Bremer *et al.* 1996, 2005; Smith *et al.* 2015) and sailfish *Istiophorus platypterus* (Graves & McDowell 2003, 2015), and it is suggested that yellowfin tuna shares this evolutionary history of Pleistocene vicariance.

Yellowfin tuna preferentially inhabit waters with temperatures of 18 to 31°C (Pecoraro *et al.* 2016b), and it is expected that cooling of sea temperatures around southern Africa during glacial maxima would have resulted in contraction of the species' range toward the equator, leading to a break in gene flow between Atlantic and Indian Ocean populations. Genetic diversity indices observed here from CR haplotypes, and by previous studies (see section 4.2.1.3), suggest the rapid expansion of both Atlantic and Indian Ocean populations of yellowfin tuna from small population sizes, possibly following a bottleneck or founder event (Grant & Bowen 1998). Based on estimated times since expansion of yellowfin tuna populations (see section 4.2.1.3), it is possible that these populations expanded during warm interglacial period of the Pleistocene, during the series of intense glacial events which began approximately 700 KYA (Hewitt 1996).

It is proposed that contemporary gene flow of yellowfin tuna individuals between the Atlantic and Indian Oceans is absent or limited, and not at a level ecologically relevant to fisheries

management applications. Based on the absent or weak signal of genetic differentiation observed between these regions by early population genetics studies (e.g. Scoles & Graves 1993; Ward *et al.* 1994; Ely *et al.* 2005), Díaz-Jaimes *et al.* (2010) suggested that contemporary gene flow exists between the Atlantic and Indian Oceans for this species. Significant heterogeneity presented here however rejects the null hypothesis of panmixia between these regions, supporting the suggestion by Ely *et al.* (2005) that the weak genetic structure detected at neutral molecular markers in yellowfin tuna compared to other Scombrids, reflects the species' relatively large N_e rather than the presence of substantial recurrent gene flow.

4.3.2 Present population isolation

Yellowfin tuna are physiologically restricted to warmer waters as their heart temperatures cannot be maintained by their counter-current heat exchangers, and their cardiac function is restricted at ambient water temperatures of 15°C or lower (Brill & Bushnell 2001; Graham & Dickson 2001; Galli *et al.* 2009); this species therefore does not remain in cold waters for long periods of time (Brill *et al.* 1999; Brill & Lutcavage 2001; Galli *et al.* 2009). It has been proposed that the presence of cold water of the Benguela Upwelling System, along the west coast of southern Africa from approximately 19°S to 34°S (Hutchings *et al.* 2009), prevents the movement of yellowfin tuna from the Atlantic Ocean basin north of the Benguela Current system into the southern Benguela region in the south eastern Atlantic Ocean off the west coast of South Africa (Shannon *et al.* 1987). The presence of this species off South Africa is therefore restricted to the warm waters of the Agulhas Current, and various authors have suggested that the movement of yellowfin tuna individuals into the southern Benguela region is facilitated by the influence of this current (Hayasi 1974; Shannon *et al.* 1987; Penney *et al.* 1992), a hypothesis supported by the results of this study which suggest that Western Cape yellowfin tuna originate in the Indian Ocean, and are reproductively isolated from Atlantic Ocean individuals (see sections 4.2.2.3 and 4.2.3.2). The lack of spawning grounds of yellowfin tuna off South Africa, and large size classes, seasonality and inter-annual variability of individuals caught in the southern Benguela region (see section 1.5.1 of Chapter 1) suggest that the individuals occurring in this region are migrants from spawning populations elsewhere in the Atlantic or Indian Ocean basins. The increase in abundance of yellowfin tuna in the southern Benguela region in summer months coincides with the influence of the warm Agulhas waters extending further west into this region, and tagging data has indicated the movement of Indian Ocean individuals to the southern Benguela region

(Eveson *et al.* 2015; Murua *et al.* 2015) (see section 1.5.1), supporting the suggestion that the migrants to this region are of Indian Ocean origin.

This hypothesis is supported by the genetic data presented here; 13 yellowfin tuna individuals of size class 103 – 159 cm FL caught in the Western Cape off South Africa during October 2015 were all found to be genetically similar to Indian Ocean samples (see sections 4.2.2.3 and 4.2.3.2), suggesting a migration of these individuals to this region from Indian Ocean waters. The sample of individuals from the Western Cape was small, and thus does not rule out the possibility that migrants from the Atlantic Ocean also occur in this region and were not sampled by chance, or are present in the region during different months; the genetic difference however indicates that even if dispersal does occur, gene flow is limited. It is however considered that the pattern of structure detected is stable, and that there is a lack of mixing of individuals from the two oceans in this region. The high levels of *h* observed at the CR within each sample and overall confirm that individuals within each sample were not spawned from the same, or closely related, females for all but two individuals. This suggests that the genetic differentiation observed has arisen due to restricted gene flow, rather than sweepstake events of recruitment which may bias estimations of differentiation among samples (Waples 1998; Pecoraro *et al.* 2016b), and that the pattern of differentiation among samples is temporally stable. The inclusion of adults in the analyses presented here support this temporal stability, as yellowfin tuna juveniles form more compact schools and samples of adults are thus more likely to represent a mixture of different cohorts (Hallier & Gaertner 2008). Additionally, Henriques (2011) found evidence for yellowfin tuna individuals caught off the Western Cape of South African between 2008 and 2009 to be of Indian Ocean origin (see section 1.5.2). It is suggested that the cold waters of the Benguela current, and particularly the permanent cold water upwelling cell off southern Namibia, represent the contemporary and historical barrier to gene flow between Atlantic and Indian Ocean populations of yellowfin tuna, as reported for other fish species in this region (Henriques *et al.* 2012, 2014, 2015, 2016), and represents a biogeographical barrier within this global species.

4.3.2.1 Migration of Indian Ocean individuals to the south eastern Atlantic Ocean

The movement of yellowfin tuna aggregations to the southern Benguela region may be for feeding purposes (see section 1.5.1). This movement, into the southern extent of the species' distribution in this region, is similar to the migration of yellowfin tuna off south eastern Australia, at the species' western-central Pacific Ocean southern extent, as described by

Young *et al.* (2001). Yellowfin tuna move into the southern region to feed, as this area is characterised by a high prey density related to a warm-core eddy and the continental shelf (Young *et al.* 2001). Inter-annual variation in the presence and abundance of the species in this region occurs, and has been observed to coincide with favourable environmental conditions (such as water temperature), generally related to La Niña events, which result in warm waters of the East Australian Current extending further south (Young *et al.* 2001). It is therefore suggested that prey density is not the primary factor driving yellowfin tuna migrations to this region, with restriction of the species to favourable conditions being more important in determining abundance (Kirby *et al.* 2000; Young *et al.* 2001). It is thus suggested that yellowfin tuna migrations to feeding grounds in the southern Benguela area may be largely determined by environmental conditions, resulting in the observed inter-annual fluctuations in abundance in the region.

4.4 *Implications of findings*

4.4.1 Management boundary

At present, all yellowfin tuna catches taken in South African waters west of 20°E are reported to the International Commission for the Conservation of Atlantic Tunas (ICCAT) and are included in the stock assessment of the Atlantic Ocean management unit (stock), while catches east of this point are reported to the Indian Ocean Tuna Commission (IOTC) for inclusion in stock assessments of the Indian Ocean stock. In managing fisheries species, it is critical that stocks are defined based on species' biological population structure, as reproductively isolated units (populations) have independent responses to fishing pressure, have unique levels of genetic diversity and adaptation to their specific environmental conditions, and are unable to replenish each other's biomass (see section 1.4). The evidence presented here strongly suggests that the yellowfin tuna individuals caught off the Western Cape of South Africa, within the ICCAT area of competence, are migrants from the Indian Ocean population and these catches should thus be included in stock assessments of the Indian Ocean stock, rather than the Atlantic Ocean stock.

The most recent stock assessment of the Indian Ocean yellowfin tuna stock, carried out by the Scientific Committee (SC) of the IOTC, found this stock to be both overfished and subject to overfishing in 2014 (IOTC–SC18 2015), with a reported catch of >430 000 metric tons (MT), exceeding the MSY of ~422,000 MT estimated in 2014. Based on this, the Indian Ocean yellowfin tuna stock is currently in a rebuilding phase. Catch limits have been allocated for a

number of states whose 2014 catches exceeded gear-specific thresholds, under Resolution 16/01 of the IOTC. For effective rebuilding of this stock to biomass levels which will permit MSY to be achieved in the long-term, and prevent a collapse of the stock, it is important that accurate catch data be incorporated into stock assessments to allow fishing mortality to return to a level allowing $F<F_{MSY}$ to be achieved. Additionally, the Atlantic Ocean stock, which was found in 2011 to be overfished but not subject to overfishing (ICCAT 2016b), should not include catches taken off South Africa in its stock assessments, which results in overestimation of the catch of this stock relative to its MSY level of 145 MT (ICCAT 2016b).

Evidence presented here strongly suggests that, while there may be a low level of intermingling of Atlantic and Indian Ocean stocks off South Africa, the majority of South Africa's yellowfin tuna catches are from the Indian Ocean biological stock. However only a portion of these catches are reported to the IOTC for inclusion in its assessments. The proportion of this catch taken west of 20°E and thus not reported to the IOTC fluctuates annually, and in some years makes up a large proportion. In 2014, for example, 80% of the 1 985 t yellowfin tuna catch off South African was taken west of 20°E (Ndudane 2015; West & Kerwath 2015) and thus not reported to the IOTC, but instead included in the Atlantic Ocean stock assessment (Ndudane 2015). The pelagic longline catch of yellowfin tuna (of which most is taken off the east coast in the IOTC area of competence) decreased from 2013 to 2014 due to a reduced number of Japanese permit holders, however South Africa has presented plans to the IOTC to increase the catch of tropical tunas taken by local fleets (West *et al.* 2014). This fishery is likely to extend into ICCAT waters in years of favourable environmental conditions and increased yellowfin tuna availability, as has been seen in previous years (e.g. 2014); it is thus expected that South African catches of yellowfin tuna in both IOTC and ICCAT waters will increase in future.

It is therefore of critical importance for the sustainability of the Indian Ocean population that all catches of Indian Ocean yellowfin tuna taken in South African waters, including those taken west of 20°E, be reported to the IOTC for inclusion in the Indian Ocean stock. It is therefore recommended that the management boundary of this species between the two tRFMOs be moved west to the western extent of South Africa's EEZ at 13.34°E (Figure 4.1), to incorporate the southern Benguela Sub-system, the proposed biogeographical barrier to gene flow between Atlantic and Indian Ocean yellowfin tuna populations.

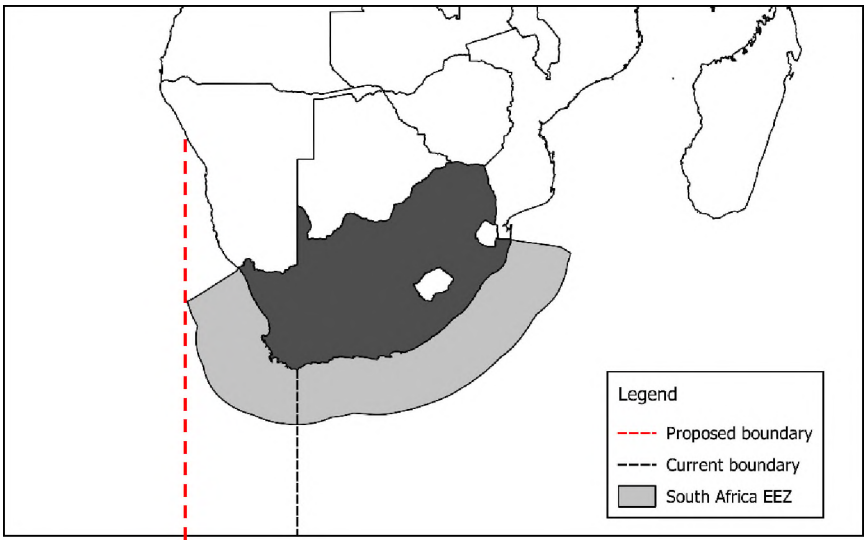

Figure 4.1: Current (20°E) and proposed (13.34°E) longitudes of the boundary between the Atlantic and Indian Ocean tRFMOs, ICCAT and the IOTC, based on population genomic data presented here. Map created in QGIS v2.18.2 (QGIS Development Team 2016).

4.4.2 Management implications for South Africa and the tRFMOs

It has been previously suggested that the yellowfin tuna caught in the ICCAT area of competence by South African fisheries are migrants from the Indian Ocean rather than the Atlantic Ocean, based on catch data such as the large size classes and seasonality of catch taken in the region (e.g. Hayasi 1974; Penney *et al.* 1992; Ndudane 2015; West & Kerwath 2015). This was suggested in a scientific report for ICCAT, shortly after the tRFMO was established in 1969, however genetic evidence was not available to support this hypothesis (Hayasi 1974). Prior to the present study, no conclusive genetic evidence has become available to assess the origin of yellowfin tuna individuals caught in the ICCAT area of competence by South African fisheries and address this concern, which has thus remained a research priority for South African tuna fisheries management (West *et al.* 2014). Results presented here provide strong evidence to reject the null hypothesis of yellowfin tuna caught west of 20°E in South African waters being of Atlantic Ocean origin. It is suggested that based on this, the most recent and conclusive scientific evidence available, the management boundary of yellowfin tuna between the two ocean basins be moved further west, to approximately 13.34°E.

South Africa's Department of Agriculture, Forestry and Fisheries (DAFF) prepares annual reports on catch and effort data for yellowfin tuna (as well as for other species under the mandates of the tRFMOs) for both ICCAT and the IOTC, for fisheries operating west and east of the 20°E management boundary, respectively. The implication of a changed boundary for South African fisheries management would thus require all South African yellowfin tuna

fisheries data to be included in annual reports to the IOTC, with no yellowfin tuna data reported to ICCAT. The implications for the tRFMOs would be the inclusion of all South African fisheries data in Indian Ocean yellowfin tuna stock assessments by the SC of the IOTC, and exclusion of South African data in the Atlantic Ocean yellowfin tuna stock assessments of the Standing Committee on Research and Statistics (SCRS) of ICCAT.

The Atlantic and Indian Ocean tRFMOs set MSY based limits (F_{MSY} and SB_{MSY}) for their entire yellowfin tuna stocks, based on their most recent stock assessments; ICCAT has not allocated per-country catch quotas while the IOTC has allocated quotas only to states with the greatest catch. The current rebuilding plan for the Indian Ocean stock is in an interim phase and will be reviewed in 2019, and if the stock remains overfished and subject to overfishing it is possible that South Africa may be allocated a quota. If the Atlantic Ocean stock remains overfished, or becomes subject to overfishing, quotas may be allocated for this stock in the future too. It is therefore important that South Africa's Western Cape yellowfin tuna catches are included in the total Indian Ocean catch of the species, to ensure that possible yellowfin tuna catch quotas are allocated to South Africa by the correct tRFMO.

4.5 *Conclusion and recommendations*

At present, catches of yellowfin tuna taken in Atlantic Ocean waters off South Africa, west of 20°E, are reported to and managed by ICCAT as part of the Atlantic Ocean population of this species. The results presented here however provide strong evidence to reject the hypothesis that yellowfin tuna occurring in the Atlantic Ocean waters of South Africa are part of the wider Atlantic Ocean population, whereas the results support the alternative hypothesis that these south eastern Atlantic individual are migrants from the Indian Ocean population. The strong and consistent results presented here should provide sufficient evidence for an immediate change in the management boundary of this species between the Atlantic and Indian Oceans. It is therefore recommended that at the next meetings of ICCAT and the IOTC, changing this boundary to the western extent of South Africa's EEZ at 13.34°E should be considered. If implemented, this will require all yellowfin tuna catches in South African waters to be reported to, and assessed managed by, the IOTC as part of the Indian Ocean yellowfin tuna stock.

4.6 *Future work*

The results of the Bayesian clustering and assignment analysis revealed the possibility of sub-structuring among yellowfin tuna samples from within the Indian Ocean, although this was

not supported by tests for conformance to HWE expectations. Dammannagoda *et al.* (2008) and Kunal *et al.* (2013) detected fine-scale structuring in northern Indian Ocean waters around India, which suggests the possibility of different spawning populations within this ocean basin. Fine-scale genetic structure has also been observed for the species within the Pacific Ocean (Díaz-Jaimes & Uribe-Alcocer 2006; Aguila *et al.* 2015; Grewe *et al.* 2015; Li *et al.* 2015), suggesting that yellowfin tuna may display fidelity to spawning sites leading to genetic sub-structuring within oceans (Pecoraro *et al.* 2016b). This is supported by evidence of yellowfin tuna within the Pacific Ocean from tagging studies (Schaefer *et al.* 2007, 2011; Pecoraro *et al.* 2016b). If more than one isolated spawning population exists within the Indian Ocean, it is important to analyse this substructure to determine admixture or mixture proportions of fisheries, so that sustainable harvesting levels can be established to avoid the collapse of more vulnerable subpopulations (Ward 2000; Reiss *et al.* 2009). Future genetic work would also benefit from cohort-specific analysis to disentangle recruitment variability from actual population sub-structure (Burford *et al.* 2011).

The mechanical mixing detected by GeneClass2 assignment tests needs to be explored further, as it may indicate mixed stock dynamics occurring within the region with both populations being exploited within South African fisheries. The proportions of these populations within catches, with potential spatial and temporal variability, should be established to avoid indiscriminate harvesting of smaller population units or inaccurate reporting of stock specific fishery data for inclusion in stock assessments by the tRFMOs.

Non-genetic methods should also be explored to further resolve the population structure of yellowfin tuna, as these may provide complementary information to genetic evidence (Pecoraro *et al.* 2016b). Otolith microchemistry is an especially useful technique; this allows for direct assessment of an individual's origins and ontogenetic dispersal patterns (McKeown *et al.* 2015).

The finer-scale structure of the yellowfin tuna population boundary between the Atlantic and Indian Oceans should be further tested, and samples from further offshore, west of 13°E, should therefore be analysed. Finally, further sampling of yellowfin tuna individuals off southern Africa should be conducted, to establish whether the pattern of structure detected here is temporally stable and consistent throughout the seasonal abundance of the species in the southern Benguela region. Those loci identified as likely to be under diversifying selection (SNPs 9902, 9903, 9904 and 10852), as well as the most informative neutral SNPs

described in section 3.2.3.5, may be useful regions to target for future work. Based on the results presented here, however, it is suggested that sufficient evidence exists to manage southern Benguela catches of yellowfin tuna as part of the Indian Ocean population.

REFERENCES

Aguila, R.D., Perez, S.K.L., Catacutan, B.J.N., Lopez, G.V., Barut, N.C. & Santos, M.D. (2015) Distinct Yellowfin Tuna (*Thunnus albacares*) Stocks Detected in Western and Central Pacific Ocean (WCPO) Using DNA Microsatellites. *PLoS ONE* **10**(9), e0138292. doi: 10.1371/journal.pone.0138292.

Allen, R. (2010) International management of tuna fisheries: Arrangements, challenges and a way forward. *FAO Fisheries and Aquaculture Technical Paper*. No. 536. FAO, Rome.

Allendorf, F.W., Hohenlohe, P.A. & Luikart, G. (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11**, 697–709.

Allendorf, F.W. (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* **5**, 181–190.

Alvarado Bremer, J. R., Mejuto, J., Greig, T. W. & Ely, B. (1996) Global population structure of the swordfish (*Xiphias gladius* L.) as revealed by analysis of the mitochondrial DNA control region. *Journal of Experimental Marine Biology and Ecology* **197**, 295–310.

Alvarado Bremer, J.R., Naseri, I. & Ely, B. (1997) Orthodox and unorthodox phylogenetic relationships among tunas revealed by the nucleotide sequence analysis of the mitochondrial DNA control region. *Journal of Fish Biology* **50**, 540–554.

Alvarado Bremer, J.R., Stequert, B., Robertson, N.W. & Ely, B. (1998) Genetic evidence for inter-oceanic subdivision of bigeye tuna (*Thunnus obesus*) populations. *Marine Biology* **132**, 547–557.

Alvarado Bremer, J.R., Viñas, J., Mejuto, J., Ely, B. & Pla, C. (2005) Comparative phylogeography of Atlantic bluefin tuna and swordfish: the combined effects of vicariance, secondary contact, introgression, and population expansion on the regional phylogenies of two highly migratory pelagic fishes. *Molecular Phylogenetics and Evolution* **36**, 169–187.

Andrews, K.R. & Luikart, G. (2014) Recent novel approaches for population genomics data analysis. *Molecular Ecology* **23**, 1661–1667.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**, 81–92. doi: 10.1038/nrg.2015.28.

Anonymous (2009) *Report of the IOTC Performance Review Panel: January 2009*. Indian Ocean Tuna Commission.

Antao, T., Lopes, A., Lopes, R.J., Beja-Pereira, A. & Luikart, G. (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{st}-outlier method. *BMC Bioinformatics* **9**(1), 323. doi: 10.1186/1471-2105-9–323.

Appleyard, S.A., Grewe, P.M., Innes, R.H. & Ward, R.D. (2001) Population structure of yellowfin tuna (*Thunnus albacares*) in the western Pacific Ocean, inferred from microsatellite loci. *Marine Biology* **139**, 383–393.

Arocha, F., Lee, D. W., Marcano, L. A., & Marcano, J. S. (2001) Update information on the spawning of yellowfin tuna, *Thunnus albacares*, in the Western Central Atlantic. *Collective Volume of Scientific Papers, ICCAT* **52**, 167–176.

Avise, J. (1998) *Phylogeography*. Havard University Press, Cambridge, USA.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **3**(10), e3376. doi: 10.1371/journal.pone.0003376.

Bandelt, H.J., Forster, P. & Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37–48.

Beaumont, M.A. & Nichols, R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B* **263**, 1619–1926.

Beckley, L.E. (1995) The Agulhas Current Ecosystem with Particular Reference to Dispersal of Fish Larvae. In: Okemwa, E., Ntiba, M.J. & Sherman, K. (eds) *Status and Future of Large Marine Ecosystems of the Indian Ocean: A Report of the International Symposium and Workshop*. Gland, Switzerland, IUCN.

Beckley, L.E. & Leis, J.M (2000) Occurrence of tuna and mackerel larvae (Family: Scombridae) off the east coast of South Africa. *Marine and Freshwater Research* **51**, 777–82.

Beebee, T. & Rowe, G. (2005) *An Introduction to Molecular Ecology*. Oxford University Press, Oxford, UK.

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

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

Bertrand, A., Bard, F.-X. & Josse, E. (2002) Tuna food habitats related to the micronekton distribution in French Polynesia. *Marine Biology* **140**, 1023–1037.

Blackburn, M. (1968) Oceanography and the ecology of tunas. *Oceanography and Marine Biology* **3**, 299–322.

Blackburn, M. (1969) Conditions related to upwelling which determine distribution of tropical tunas off western Baja California. *United States Fish and Wildlife Service Fishery Bulletin* **68**, 147–176.

Block, B.A. & Finnerty, J.R. (1994) Endothermy in fishes: a phylogenetic analysis of constraints, predispositions, and selection pressures. *Environmental Biology of Fishes* **40**, 283–302.

Bonin, A. (2008) Population genomics: a new generation of genome scans to bridge the gap with functional genomics. *Molecular Ecology* **17**, 3583–3584.

Bradbury, I.R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I.G., Snelgrove, P.V., Morris, C.J., Gregory, R.S., Hardie, D.C. & Hutchings, J.A. (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society of London B: Biological Sciences* **277**, 3725–3734.

Bradbury, I.R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I.G., Snelgrove, P.V., Morris, C.J., Gregory, R.S., Hardie, D. & Hutchings, J.A. (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications* **63**, 450–461.Brill, R.W. & Bushnell, P.G. (2001) The cardiovascular system of tunas. In: Block, B.A. & Stevens, E.D. (eds) *Tunas: Physiology, Ecology and Evolution. Fish Physiology* **19**, 79–120. Academic Press, San Diego.

Brill, R.W. & Lutcavage, M.E. (2001) Understanding Environmental Influences on Movements and Depth Distributions of Tunas and Billfishes Can Significantly Improve Population Assessments. *American Fisheries Society Symposium* **25**, 179–198.

Brill, R.W., Block, B.A., Boggs, C.H., Bigelow, K.A., Freund E.V. & Marcinek., D.J. (1999) Horizontal movements and depth distribution of large adult yellowfin tuna (*Thunnus albacares*) near the Hawaiian Islands, recorded using ultrasonic telemetry: implications for the physiological ecology of pelagic fishes. *Marine Biology* **133**, 395–408.

Brookes, A.J. (1999) The essence of SNPs. *Gene* **234**, 177–186.

Buonaccorsi, V.P., McDowell, R. & Graves, J.E. (2001) Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). *Molecular Ecology* **10**, 1179–1196.

Burford, M.O., Carr, M.H., Bernardi, G. (2011) Age-structured genetic analysis reveals temporal and geographic variation within and between two cryptic rockfish species. *Marine Ecology Progress Series* **442**, 201–215.

Caballero, A. & García-Dorado, A. (2013) Allelic diversity and its implications for the rate of adaptation. *Genetics* **195**, 1373–1384.

Canino, M.F., O'Reilly, P.T., Hauser, L. & Bentzen, P. (2005) Genetic differentiation in walleye Pollock (Theragra chalcogramma) in response to selection at the pantophysin (PanI) locus. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 2519–2529.

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

Castric, V. & Bernatchez, L. (2004) Individual assignment test reveals differential restriction to dispersal between two salmonids despite no increase of genetic differences with distance. *Molecular Ecology* **13**, 1299–1312.

Charlesworth, B. (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195–205. doi: 10.1038/nrg2526.

Chiang, H.-C., Hsu, C.-C., Wu, G.C.-C., Chang, S.-K. & Yang, H. (2008) Population structure of bigeye tuna (*Thunnus obesus*) in the Indian Ocean inferred from mitochondrial DNA. *Fisheries Research* **90**, 305–312. doi: 10.1016/j.fishres.2007.11.006.

Chow, S. & Inoue, S. (1993) Intra-and interspecific restriction fragment length polymorphism in mitochondrial genes of *Thunnus* tuna species. *Bulletin of the National Research Institute of Far Seas Fisheries* **30**, 207–225.

Chow, S., Okamoto, H., Uozumiá, Takeuchi, Y. & Takeyama, H. (1999) Genetic stock structure of the swordfish (*Xiphias gladius*) inferred by PCR-RFLP analysis of the mitochondrial DNA control region. *Marine Biology* **127**, 359–367.

Chow, S., Hazama, K., Nishida, T., Ikame, S. & Kurihara, S. (2000a) A preliminary genetic analysis on yellowfin tuna stock structure in the Indian Ocean using mitochondrial DNA variation. *WPTT00–11, IOTC Proceedings* **3**, 312–316.

Chow, S., Okamoto, H., Miyabe, N., Hiramatsu, K. & Barut, N. (2000b) Genetic divergence between Atlantic and Indo-Pacific stocks of bigeye tuna (*Thunnus obesus*) and admixture around South Africa. *Molecular Ecology* **9**, 221–227.

Christie, M.R., Johnson, D.W., Stallings, C.D. & Hixon, M.A. (2010) Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. *Molecular Ecology* **19**, 1042–1057.

Collette, B.B. & Nauen, C.E. (1983) *FAO Species Catalogue. Scombrids of the world: an annotated and illustrated catalogue of tunas and mackerels, bonitos and related species known to date*. FAO Fisheries Synopsis 125. FAO, Rome.

Collette, B.B., Reeb, C. & Block, B.A. (2001) Systematics of the tunas and mackerels (Scombridae). In: Block, B.A. & Stevens, E.D. (eds) *Tuna: Physiology, Ecology, and Evolution. Fish Physiology* **19**, 1–33. Academic Press, San Diego.

Collette, B., Acero, A., Amorim, A.F., Boustany, A., Canales Ramirez, C., Cardenas, G., Carpenter, K.E., Chang, S.-K., de Oliveira Leite Jr., N., Di Natale, A., Die, D., Fox, W., Fredou, F.L., Graves, J., Guzman-Mora, A., Viera Hazin, F.H., Hinton, M., Juan Jorda, M., Minte Vera, C., Miyabe, N., Montano Cruz, R., Masuti, E., Nelson, R., Oxenford, H., Restrepo, V., Salas, E., Schaefer, K., Schratwieser, J., Serra, R., Sun, C., Teixeira Lessa, R.P., Pires Ferreira Travassos, P.E., Uozumi, Y. & Yáñez, E. (2011) *Thunnus albacares. The IUCN Red List of Threatened Species 2011*, e.T21857A9327139. doi: 10.2305/IUCN.UK.2011-2.RLTS.T21857A9327139.en.

Corander, J., Majander, K. K., Cheng, L. & Merilä, J. (2013) High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. *Molecular Ecology* **22**, 2931–2940.

Cullis-Suzuki, S. & Pauly, D. (2010) Failing the high seas: A global evaluation of regional fisheries management organizations. *Marine Policy* **34**, 1036–1042.

da Fonseca, R.R., Albrechtsen, A., Themudo, G.C., Ramos-Madrigal, J., Sibbesen, J.A., Maretty, L., Zepeda-Mendoza, M.L., Campos, P.F., Heller, R. & Pereira, R.J. (2016) Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics* **30**, 3–13. doi: 10.1016/j.margen.2016.04.012.

Dagorn, L., Holland, K.N., Hallier, J., Taquet, M., Moreno, G., Sancho, G., Itano, D.G., Aumeeruddy, R., Girard, C., Million, J. & Fonteneau, A. (2006) Deep diving behavior observed in yellowfin tuna (*Thunnus albacares*). *Aquatic Living Resources* **19**, 85–88.

Dammannagoda, S.T., Hurwood, D.A. & Peter, B.M. (2008) Evidence for fine geographical scale heterogeneity in gene frequencies in yellowfin tuna (*Thunnus albacares*) from the north Indian Ocean around Sri Lanka. *Fisheries Research* **90**, 147–157.

Davey, J.W. & Blaxter, M.L. (2011) RADSeq: next generation population genetics. *Briefings in Functional Genomics* **9**, 416–423.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499–510.

Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K. & Blaxter, M.L. (2013) Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology* **22**, 3151–3164.

De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J. & Vigouroux, Y. (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* **22**, 1383–1399.

DeFaveri, J., Viitaniemi, H., Leder, E & Merilä, J. (2013) Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. *Molecular Ecology Resources* **13**, 377–392.

Demarcq, H., Barlow, R. G. & Shillington, F. A. (2003) Climatology and variability of sea surface temperature and surface chlorophyll in the Benguela and Agulhas ecosystems as observed by satellite imagery. *African Journal of Marine Science* **25**, 363–372.

Dencausse, G., Arhan, M. & Speich, S. (2010) Spatio-temporal characteristics of the Agulhas Current retroflection. *Deep-Sea Research I* **57**, 1392–1405.

Diaha, N.C., Zudaire, I., Chassot, E., Pecoraro, C., Bodin, N., Amandè, M.J., Dewals, P., Roméo, M.U., Irié, Y.D., Barryga, B.D., Gbeazere, D.A. & Kouadio, D. (2015) Present and future of reproductive studies of yellowfin tuna (*Thunnus albacares*) in the eastern Atlantic Ocean. *Collective Volume of Scientific Papers, ICCAT* **71**, 489–509.

Díaz-Jaimes, P. & Uribe-Alcocer, M. (2006) Spatial differentiation in the eastern Pacific yellowfin tuna revealed by microsatellite variation. *Fisheries Science* **72**, 590–596.

Díaz-Jaimes, P., Uribe-Alcocer, M., Rocha-Olivares, A., García-de-León, F.J., Nortmoon, P. & Durand, J.D. (2010) Global phylogeography of the dolphinfish (*Coryphaena hippurus*): The influence of large effective population size and recent dispersal on the divergence of a marine pelagic cosmopolitan species. *Molecular Phylogenetics and Evolution* **57**, 1209–1218.

Donaldson, K.A., Wilson, R.R. (1999) Amphi-panamaic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rates in the mitochondrial DNA control region of fishes. *Molecular Phylogenetics and Evolution* **13**, 208–213.

Drapeau, L., Pecquerie, L., Fréon, P. & Shannon, L.J. (2004) Quantification and representation of potential spatial interactions in the southern Benguela ecosystem. *African Journal of Marine Science* **26**, 141–159. doi: 10.2989/18142320409504054.

Duncan, M., James, N., Fennessy, S.T., Mutombene, R.J. & Mwale, M. (2015) Genetic structure and consequences of stock exploitation of Chrysoblephus puniceus, a commercially important sparid in the South West Indian Ocean. *Fisheries Research* **164**, 64–72.

Durand, J.-D., Collet, A., Chow, S., Guinand, B. & Borsa, P. (2005) Nuclear and mitochondrial DNA markers indicate unidirectional gene flow of Indo-Pacific to Atlantic bigeye tuna (*Thunnus obesus*) populations, and their admixture off southern Africa. *Marine Biology* **147**, 313–322.

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

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. & Mitchell, S.E. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **6**(5), e19379. doi: 10.1371/journal.pone.0019379.

Ely, B., Stoner, D.S., Alvarado Bremer, J.R., Dean, J.M., Addis, P., Cau, A., Thelen, E.J., Jones, W.J., Black, D.E., Smith, L. & Scott, K. (2002) Analyses of nuclear ldhA gene and mtDNA control region sequences of Atlantic northern bluefin tuna populations. *Marine Biotechnology* **4**, 583–588.

Ely, B., Viñas, J., Alvarado Bremer, J.R., Black, D., Lucas, L., Covello, K., Labrie, A.V. & Thelen, E. (2005) Consequences of the historical demography on the global population structure of two highly migratory cosmopolitan marine fishes: the yellowfin tuna (*Thunnus albacares*) and the skipjack tuna (*Katsuwonus pelamis*). *BMC Evolutionary Biology* **5**(1), 19. doi: 10.1186/1471-2148-5-19.

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

Eveson, J.P., Million, J., Sardenne, F. & Le Croizier, G. (2015) Estimating growth of tropical tunas in the Indian Ocean using tag-recapture data and otolith-based age estimates. *Fisheries Research* **163**, 58–68.

Excoffier, L., Smouse, P. & Quattro, J. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Excoffier, L. (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology* **13**, 853–864.

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

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

FAO (2017) Fishery Statistical Collections: Global Tuna Catches by Stock. Available at: http://www.fao.org/fishery/statistics/tuna-catches/en [Accessed 23/01/2017].

Fisher, R. A. (1935) The logic of inductive inference (with discussion). *Journal of the Royal Statistical Society* **98**, 39–82.

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

Fonteneau, A. (1997) *Atlas of tropical tuna fisheries: World catches and environment*. ORSTOM, Paris, France.

Freamo, H., O'Reilly, P., Berg, P., Lien, S. & Boulding, E. (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Molecular Ecology Resources* **11**, 243–256.

Fromentin, J.M. & Fonteneau, A. (2001) Fishing effects and life history traits: a case study comparing tropical versus temperate tunas. *Fisheries Research* **53**, 133–150.

Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

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

Gagnaire, P-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S. & Bierne, N. (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications* **8**, 769–786.

Galli, G.L., Shiels, H.A. & Brill, R.W. (2009) Temperature Sensitivity of Cardiac Function in Pelagic Fishes with Different Vertical Mobilities: Yellowfin Tuna (*Thunnus albacares*), Bigeye Tuna (*Thunnus obesus*), Mahimahi (*Coryphaena hippurus*), and Swordfish (*Xiphias gladius*). *Physiological and Biochemical Zoology: Ecological and Evolutionary Approaches* **82**, 280–290.

Garvin, M.R., Saitoh, K. & Gharrett, A.J. (2010) Application of single nucleotide polymorphisms to non-model species: a technical review. *Molecular Ecology Resources* **10**, 915–934.

Götz, A., Kerwath, S.E., Samaai, T., da Silva, C. & Wilke, C.G. (2014) An exploratory investigation of the fish communities associated with reefs on the central Agulhas Bank, South Africa. *African Zoology*, **49**, 253–264.

Graham, J.B. & Dickson, K.A. (2001) Anatomical and physiological specializations for endothermy. In: Block, B.A. & Stevens, E.D. (eds) *Fish Physiology: Tuna, Physiology, Ecology, and Evolution. Fish Physiology* **19**, 121-165. Academic Press, San Diego.

Graham, C.F., Glenn, T.C., McArthur, A.G., Boreham, D.R., Kieran, T., Lance, S., Manzon, R.G., Martino, J.A., Pierson, T., Rogers, S.M., Wilson, J.Y. & Somers, C.M. (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources* **15**, 1304–1315.

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

Graves, J.E. & McDowell, J.R. (2003) Stock structure of the world's istiophorid billfishes: a genetic perspective. *Marine and Freshwater Research* **54**, 287–289.

Graves, J.E. & McDowell, J.R. (2015) Population structure of istiophorid billfishes. *Fisheries Research* **166**, 21–28.

Greenbaum, G., Templeton, A.R., Zarmi, Y. & Bar-David, S. (2014) Allelic Richness following Population Founding Events – A Stochastic Modeling Framework Incorporating Gene Flow and Genetic Drift. *PLoS ONE* **9**(12), e115203. doi: 10.1371/journal.pone.0115203.

Grewe, P.M., Feutry, P., Hill, P.L., Gunasekera, R.M., Schaefer, K.M., Itano, D.G., Fuller, D.W., Foster, S.D. & Davies, C.R. (2015) Evidence of discrete yellowfin tuna (*Thunnus albacares*) populations demands rethink of management for this globally important resource. *Scientific Reports* **5**, 16916. doi: 10.1038/srep16916.

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

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Hallier, J.-P. & Gaertner, D. (2008) Drifting fish aggregation devices could act as an ecological trap for tropical tuna species. *Marine Ecology Progress Series* **353**, 255–264.

Hampton, J. (2000). Natural mortality rates in tropical tunas: size really does matter. *Canadian Journal of Fisheries and Aquatic Sciences* **57**, 1002–1010.

Harpending, R. C. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66**, 591–600.

Hastings, A. (1993) Complex interactions between dispersal and dynamics: lessons from coupled logistic equations. *Ecology* **74**, 1362–1372.

Hauser, L., Adcock, G.J., Smith, P.J., Ramirez, J.H.B. & Carvalho, G.R. (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences* **99**, 11742–11747. doi: 10.1073/pnas.1772242899.

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

Hayasi, S. (1974) A hypothesis on population structure of yellowfin tuna in the Atlantic Ocean mainly based on longline data. *Collective Volume of Scientific Papers, ICCAT* **2**, 40–48.

Hellberg, M.E. (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution, and Systematics* **40**, 291–310.

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T., Cariani, A., Maes, G.E., Diopere, E., Carvalho, G.R. & Nielsen, E.E. (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**, 123–136.

Henriques, R. (2011) *Influence of the Benguela Current in genetic sub-structuring of commercially exploited fish species* (Unpublished doctoral thesis). Royal Holloway University of London, London, United Kingdom.

Henriques, R., Potts, W.M., Sauer, W.H.H. & Shaw, P.W. (2012) Evidence of deepgenetic divergence between populations of an important recreational fishery species, *Lichia amia* L. 1758, around southern Africa. *African Journal of Marine Science* **34**, 585–591.

Henriques, R., Potts, W.M., Santos, C.V., Sauer, W.H.H. & Shaw, P.W. (2014) Population Connectivity and Phylogeography of a Coastal Fish, *Atractoscion aequidens* (Sciaenidae), across the Benguela Current Region: Evidence of an Ancient Vicariant Event. *PLoS ONE* **9**(2), e87907. doi: 10.1371/journal.pone.0087907.

Henriques, R., Potts, W.M., Sauer, W.H.H. & Shaw, P.W. (2015) Incipient genetic isolation of a temperate migratory coastal sciaenid fish (*Argyrosomus inodorus*) within the Benguela Cold Current system. *Marine Biology Research* **11**, 423-429. doi: 10.1080/17451000.2014.952309.

Henriques, R., Potts, W.M., Sauer, W.H.H., Santos, C.V., Kruger, J., Thomas, J.A. & Shaw, P.W. (2016) Molecular genetic, life-history and morphological variation in a coastal warm-temperate sciaenid fish: evidence for an upwelling-driven speciation event. *Journal of Biogeography* **43**, 1820–1831. doi: 10.1111/jbi.12829.

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

Ho, S.Y.W., Lanflear, R., Bromham, L., Phillips, M.J. Soubrier, J., Rodrigo, A.G. & Cooper, A. (2011) Time-dependent rates of molecular evolution. *Molecular Ecology* **20**, 3087–3101.Holsinger, K.E. & Weir, B.S. (2009) Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics.* **10**, 639–650.

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

Hudson, M.E. (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources* **8**, 3–17.

Hunter, J.R., Argue, A.W., Bayliff, W.H., Dizon, A.E., Fonteneau, A., Doodman, D. & Seckel, G.R. (1986) The dynamics of tuna movements: an evaluation of past and future research. *FAO Fisheries Technical Paper.* No. 277. FAO, Rome.

Hutchings, L., van der Lingen, C.D., Shannon, L.J., Crawford, R.J.M., Verheye, H.M.S., Bartholomae, C.H., van der Plas, A.K., Louw, D., Kreiner, A., Ostrowski, M., Fidel, Q., Barlow, R.G., Lamont, T., Coetzee, J., Shillington, F., Veitch, J., Currie, J.C. & Monteiro,

P.M.S. (2009) The Benguela Current: An ecosystem of four components. *Progress in Oceanography* **83**, 15–32.

ICCAT (2016a) *Report for biennial period, 2014-15. Part II (2015). Vol. 1, COM*. Madrid, Spain.

ICCAT (2016b) *Report for biennial period, 2014-15. Part II (2015). Vol. 2, SCRS*. Madrid, Spain.

ICCAT (2016c) *Report of the Independent Performance Review of ICCAT*. Madrid, Spain.

IOTC (2006) *Report of the Ninth Session of the Scientific Committee*. Victoria, Seychelles, 6-10 November 2006. IOTC-2006-SC-R[EN].

IOTC (2014) *Report of the Seventeenth Session of the Scientific Committee*. Victoria, Seychelles, 8–12 December 2014.

IOTC Secretariat (2014) *Guidelines for the reporting of Fisheries Statistics to the IOTC*. IOTC Secretariat, Mahé, Seychelles, January 2014.

IOTC–SC17 (2014) *Report of the Seventeenth Session of the IOTC Scientific Committee*. Seychelles, 8–12 December 2014. IOTC–2014–SC17–R[E].

IOTC–SC18 (2015) *Report of the 18th Session of the IOTC Scientific Committee*. Bali, Indonesia, 23–27 November 2015. IOTC–2015–SC18–R[E]..

IOTC–SC19 (2016) *Report of the 19th Session of the IOTC Scientific Committee*. Seychelles, 1–5 December 2016. IOTC–2016–SC19–R[E].

IOTC–WPTT18 2016. R*eport of the 18th Session of the IOTC Working Party on Tropical Tunas*. Seychelles, 5–10 November 2016. IOTC–2016–WPTT18–R[E].

Itano, D. & Holland, N. (2000) Movement and vulnerability of bigeye (*Thunnus obesus*) and yellowfin tuna (*Thunnus albacares*) in relation to FADs and natural aggregation points. *Aquatic Living Resources* **13**, 213–223.

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

Jombart, T. & Ahmed, I. (2011) *adegenet* 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 10.1093/bioinformatics/btr521.

Jombart, T., Pontier, D. & Dufour, A-B. (2009) Genetic markers in the playground of multivariate analysis. *Heredity* **102**, 330–341.

Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94; 10.1186/1471-2156-11-94.

Juan-Jordá, M.J., Mosqueira, I., Freire, J. & Dulvy, N.K. (2013a) The Conservation and Management of Tunas and Their Relatives: Setting Life History Research Priorities. *PLoS ONE* **8**(8), e70405. doi: 10.1371/journal.pone.0070405.

Juan-Jordá, M. J., Mosqueira, I., Freire, J. & Dulvy, N. K. (2013b) Life in 3-D: life history strategies in tunas, mackerels and bonitos. *Reviews in Fish Biology and Fisheries* **23**, 135–155.

Kalinowski, S.T. (2002) How many alleles per locus should be used to estimate genetic distances? *Heredity* **88**, 62–65.

Kalinowski, S.T. (2005) Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity* **94**, 33–36.

Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W. & Prodohl, P.A. (2013) diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* **4**, 782–88.

Kenchington, E., Heino, M. & Nielsen, E.E. (2003) Managing marine genetic diversity: time for action? *ICES Journal of Marine Science* **60**, 1172–1176.

Kirby, D.S., Fiksen, O. & Hart, P.J.B. (2000) A dynamic optimisation model for the behaviour of tunas at ocean fronts. *Fisheries Oceanography* **9**, 328–342.

Kolody, D. S., Eveson, J. P. & Hillary, R. M. (2016) Modelling growth in tuna RFMO stock assessments: Current approaches and challenges. *Fisheries Research* **180**, 177–193.

Krück, N. C., Innes, D. I. & Ovenden, J. R. (2013) New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). *Molecular Ecology Resources* **13**, 715–725.

Kumar, S., Stecher, G. & Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**, 1870–874.

Kunal, S.P., Kumar, G., Menezes, M.R. & Meena, R.M. (2013) Mitochondrial DNA analysis reveals three stocks of yellowfin tuna *Thunnus albacares* (Bonnaterre, 1788) in Indian waters. *Conservation Genetics* **14**, 205–213.

Langley, A. & Million, J. (2012) Determining an appropriate tag mixing period for the Indian Ocean yellowfin tuna stock assessment. Indian Ocean Tuna Commission 14th Working Party on Tropical Tuna Working Paper, IOTC-2012-WPTT14-31.

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D. & Seeb, J. E. (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications* **7**, 355–369.

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

Lewis, A.D. (1992) *Stock structure of Pacific yellowfin tuna – a review*. South Pacific Commissin, Nourmea, New Caledonia.

Lewontin, R. C. & Krakauer, J. (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74, 175–195.

Li, W.H. (1977) Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics* **85**, 331–337.

Li, W., Xinjun, C., Qianghua, X., Jiangfeng, Z., Xiaojie, D. & Liuxiong, X. (2015) Genetic Population Structure of *Thunnus albacares* in the Central Pacific Ocean Based on mtDNA COI Gene Sequences. *Biochemical Genetics* **53**, 8–22.

Lotterhos, K.E. & Whitlock, M.C. (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* **23**, 2178–2192.

Lowe, W. H. & Allendorf, F. W. (2010) What can genetics tell us about population connectivity? *Molecular Ecology* **19**, 3038–3051. doi: 10.1111/j.1365-294X.2010.04688.x.

Lutjeharms, J.R.E. (2006) *The Agulhas Current*. Springer, Berlin.

Lutjeharms, J.R.E. & Meeuwis, J.M. (1987) The extent and variability of South-East Atlantic upwelling. *South African Journal of Marine Science* **5**, 51–62.

Lutjeharms, J.R.E. & Valentine, H.R. (1984) Southern Ocean thermal fronts south of Africa. *Deep Sea Research* **31**, 1461–1475.

Majkowski, J. (2007mmcc) Global fishery resources of tuna and tuna-like species. *FAO Fisheries Technical Paper*. No. 483. FAO, Rome.

Mardis, E.R. (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133–141.

Martin, A.P. (2003) Phytoplankton patchiness: the role of lateral stirring and mixing. *Progress in Oceanography* **57**, 125–174.

Maury, O., Gascuel, D., Marsac, F., Fonteneau, A. & De Rosa, A. (2001) Hierarchical interpretation of nonlinear relationships linking yellowfin tuna (*Thunnus albacares*) distribution to the environment in the Atlantic Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 458–469.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C. & Brumfield, R.T. (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* **66**, 526–538.

McKeown, N.J., Arkhipkin, A.I. & Shaw, P.S. (2015) Integrating genetic and otolith microchemistry data to understand population structure in the Patagonian Hoki (*Macruronus magellanicus*). *Fisheries Research* **164**, 1–7.

Ménard, F., Labrune, C., Shin, Y.J., Asine, A.S. & Bard, F.X. (2006) Opportunistic predation in tuna: a size-based approach. *Marine Ecology Progress Series* **323**, 223–231.

Meyer, A. (1993) Evolution of mitochondrial DNA in fish. In: Hochachka, P.W. & Mommsen, T.P. (eds) *Biochemistry and molecular biology of fish*. Elsevier, Amsterdam.

Miyake, M., Guillotreau, P., Sun, C. & Ishimura, G. (2010) Recent developments in the tuna industry: stocks, fisheries, management, processing, trade and markets. *FAO Fisheries and Aquaculture Technical Paper*. No. 543. FAO, Rome.

Montes, I., Iriondo, M., Manzano, C., Santos, M., Conklin, D., Carvalho, G.R., Irigoien, X. & Estonba, A. (2016) No loss of genetic diversity in the exploited and recently collapsed

population of Bay of Biscay anchovy (*Engraulis encrasicolus*, L.). *Marine Biology* **163**(98), 10.1007/s00227-016-2866-2.

Morin, P.A., Luikart, G., Wayne, R.K. & the SNP Workshop Group. (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution* **19**, 208–216.

Morin, P.A., Martien, K.K. & Taylor, B.L. (2009) Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources* **9**, 66–73.

Muhling, B.A., Lamkin, J.T., Alemany, F., García, A., Farley, J., Ingram, G.W., Berastegui, D.A., Reglero, P. & Carrion, R.L. (2017) Reproduction and larval biology in tunas, and the importance of restricted area spawning grounds. *Reviews in Fish Biology and Fisheries*. doi: 10.1007/s11160-017-9471-4.Murua, H., Eveson, J.P. & Marsac, F. (2015) The Indian Ocean Tuna Tagging Program: Building better science for more sustainability. *Fisheries Research* **163**, 1–6.

Nakamura, Y., Mori, K., Saitoh, K., Oshima, K., Mekuchi, M., Sugaya, T., Shigenobu, Y., Ojima, N., Muta, S., Fujiwara, A., Yasuike, M., Oohara, I., Hirakawa, H., Chowdhury, V.A., Kobayashi, T., Nakajima, K., Sano, M., Wada, T., Tashiro, K., Ikeo, K., Hattori, M., Kuhara, S., Gojobori, K. & Inouye, K. (2013). Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. *Proceedings of the National Academy of Sciences* **110**, 11061–11066.

Narum, S.R. & Hess, J.E. (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11**, 184–194.

Narum, S.R., Buerklee, C.A., Davey, J.W., Miller, M.R. & Hohenlohe, P.A. (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**, 2841–2847.

Ndudane, S. (2015) Annual Report of South Africa. In: ICCAT (2016) *Report for biennial period, 2014-2015*, Part II (2015) – Vol. 3: Annual Reports.

Nei, M, (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583–590.

Nei, M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.

Nielsen, E.E., Hemmer-Hansen, J., Larsen, P.F. & Bekkevold, D. (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* **18**, 3128–3150.

Nishida, T., Chow, S., Ikame, S. & Kurihara, S. (2001) RFLP analysis on single copy nuclear gene loci in yellowfin tuna (*Thunnus albacares*) to examine the genetic differentiation between the western and eastern samples from the Indian Ocean. *IOTC Proceedings* **4**, 437–441.

Nyström, V., Dalén, L., Vartanyan, S., Lidén, K., Ryman, N. & Angerbjörn, A. (2010) Temporal genetic change in the last remaining population of woolly mammoth. *Proceedings of the Royal Society of London B: Biological Sciences*. doi: 10.1098/rspb.2010.0301.

Nyström, V., Humphrey, J., Skoglund, P., McKeown, N.J., Vartanyan, S., Shaw, P.W., Lidén, K., Jakobsson, M., Barnes, I., Angerbjörn, A., Lister, A. & Dalén, L. (2012) Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation. *Molecular Ecology* **21**, 3391–3402.

Ogden, R. (2011) Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources* **11**, 109–116.

Ott, A., Liu, S., Schnable, J.C., Yeh, C.T., Wang, C. & Schnable, P.S. (2017) Tunable Genotyping-By-Sequencing (tGBS®) Enables Reliable Genotyping of Heterozygous Loci. *bioRxiv*. doi: 10.1101/100461.

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

Palsbøll, P. J., Berube, M. & Allendorf, F. W. (2007) Identification of management units using population genetic data. *Trends in Ecology and Evolution* **22**, 11–16.

Pawson, M.G. & Jennings, S. (1996) A critique of methods for stock identification in marine capture fisheries. *Fisheries Research* **25**, 203–217.

Pecoraro, C., Babbucci, M., Villamor, A., Franch, R., Papetti, C., Leroy, B., Ortega-Garcia, S., Muir, J., Rooker, J., Arocha, F., Murua, H., Zudaire, I., Chassot, E., Bodin, N., Tinti, F., Bargelloni, L. & Cariani, A (2016a) Methodological assessment of 2b-RAD

genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*). *Marine Genomics* **25**, 43–48.

Pecoraro, C., Zudaire, I., Bodin, N., Murua, H., Taconet, P., Díaz-Jaimes, P., Cariani, A., Tinti, F. & Chassot, E. (2016b) Putting all the pieces together: integrating current knowledge of the biology, ecology, fisheries status, stock structure and management of yellowfin tuna (*Thunnus albacares*). *Reviews in Fish Biology and Fisheries*. doi: 10.1007/s11160-016-9460-z.

Penney, A.J. & Griffiths, M.H. (1999) A first description of the developing South African pelagic longline fishery. *Collective Volume of Scientific Papers, ICCAT* **49**, 162–173.

Penney, A.J., Krohn, R.G. & Wilke, C.G. (1992) A description of the South African tuna fishery in the southern Atlantic Ocean. *Collective Volume of Scientific Papers, ICCAT* **39**, 247–257.

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L. & Estoup, A. (2004) GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. *Journal of Heredity* **95**, 536–539.

Pons, M., Branch, T.A., Melnychuk, M.C., Jensen, O.P., Brodziak, J., Fromentin, J.M., Harley, S.J., Haynie, A.C., Kell, L.T., Maunder, M.N., Parma, A.M., Restrepo, V.R., Sharma, R., Ahrens, R. & Hilborn, R. (2017) Effects of biological, economic and management factors on tuna and billfish stock status. *Fish and Fisheries* **18**, 1–21. doi: 10.1111/faf.12163.

Posada, D. & Crandall, K.A. (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution* **16**, 37–45.

Potier, M., Marsac, F., Lucas, V., Sabatié, R., Hallier, J.-P. & Ménard, F. (2004) Feeding Partitioning among Tuna Taken in Surface and Mid-water Layers: The Case of Yellowfin (*Thunnus albacares*) and Bigeye (*T. obesus*) in the Western Tropical Indian Ocean. *Western Indian Ocean Journal of Marine Science* **3**, 51–62.

Potier, M., Marsac, F., Cherel, Y., Lucas, V., Sabatié, R., Maury, O. & Ménard, F. (2007) Forage fauna in the diet of three large pelagic fish (lancetfish, swordfish and yellowfin tuna) in the western equatorial Indian Ocean. *Fisheries Research* **83**, 60–72.

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

QGIS Development Team (2016) QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available at: http://www.qgis.org/.

Qiu, F. & Miyamoto, M.M. (2011) Use of Nuclear DNA Data to Estimate Genetic Diversity and Population Size in Pacific Bluefin and Yellowfin Tuna (*Thunnus orientalis* and *T. albacares*). *Copeia* **2**, 264–269.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

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

Raymond, M. & Rousset, F. (1995) An exact test for population differentiation. *Evolution* **49**, 1280–1283.

Reglero, P., Tittensor, D.P., Álvarez-Berastegui, D., Aparicio-González, A. & Worm, B. (2014) Worldwide distributions of tuna larvae: revisiting hypotheses on environmental requirements for spawning habitats. *Marine Ecology Progress Series* **501**, 207–224. doi: 10.3354/meps10666.

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

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

Roger, C. (1994). The plankton of the tropical western Indian Ocean as a biomass indirectly supporting surface tunas (yellowfin, *Thunnus albacares* and skipjack, *Katsuwomus pelamis*). *Environmental Biology of Fishes* **39**, 161–72.

Rogers, A. (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* **49**, 608–615.

Rogers, A. R. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552–569.

Rosenberg, N., Li, L., Ward, R. & Pritchard, J. (2003) Informativeness of genetic markers for inference of ancestry. *American Journal of Human Genetics* **73**, 1402–1422.

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

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

Ryynänen, H.J., Tonteri, A., Vasemägi, A. & Primmer, C.R. (2007) A Comparison of Biallelic Markers and Microsatellites for the Estimation of Population and Conservation Genetic Parameters in Atlantic Salmon (*Salmo salar*). *Journal of Heredity* **98**, 692–704.

Sabarros, P.S., Ménard, F., Lévénez, J.-J., Tew-Kai, E. & Ternon, J.-F. (2009) Mesoscale eddies influence distribution and aggregation patterns of micronekton in the Mozambique Channel. *Marine Ecology Progress Series* **395**, 101–107.

Sauer, W.H.H., Hecht, T., Britz, P.J. & Mather, D. (2003) *An Economic and Sectoral Study of the South African Fishing Industry*. Volume 2: Fishery Profiles. Report prepared for Marine and Coastal Management by Rhodes University.

Sauer, W.H.H., Downey, N.J., Lipinski, M., Roberts, M.J., Smale, M.J., Shaw, P., Glazer, J. & Melo, Y. (2013) *Loligo reynaudii*, Chokka Squid. In: Rosa, R., O'Dor, R. & Pierce, G. (Eds) *Advances in Squid Biology, Ecology and Fisheries, Part I – Myopsid Squids*. Nova Science Publishers, Inc., New York.

Schaefer, K.M., Fuller, D.W. & Block, B.A. (2007) Movements, behavior, and habitat utilization of yellowfin tuna (*Thunnus albacares*) in the northeastern Pacific Ocean, ascertained through archival tag data. *Marine Biology* **152**, 503–525.

Schaefer, K.M., Fuller, D.W. & Block, B.A. (2011) Movements, behavior, and habitat utilization of yellowfin tuna (*Thunnus albacares*) in the Pacific Ocean off Baja California, Mexico, determined from archival tag data analyses, including unscented Kalman filtering. *Fisheries Research* **112**, 22–37.

Schenekar, T. & Weiss, S. (2011) High rate of calculation errors in mismatch distribution analysis results in numerous false inferences of biological importance. *Heredity* **107**, 511–512.

Scoles, D.R. & Graves, J.E. (1993) Genetic analysis of the population structure of yellowfin tuna, *Thunnus albacares*, from the Pacific Ocean. *Fishery Bulletin* **91**, 690–698.

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S. & Seeb, L.W. (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources* **11**, 1–8.

Shaklee, J.B. & Currens, K.P. (2003) Genetic stock identification and risk assessment. In: Hallerman, E.M. (ed.) *Population Genetics: Principles and Applications for Fisheries Scientists*. American Fisheries Society, Bethesda, Maryland.

Sibert, J. & Hampton, J. (2003) Mobility of tropical tunas and the implications for fisheries management. *Marine Policy* **27**, 87–95.

Sham, P., Cherny, S. & Purcell, S. (2009) Application of genome-wide SNP data for uncovering pairwise relationships and quantitative trait loci. *Genetica* **136**, 237–243.

Shannon, L.V. (1987) The tunas of the Benguela region off southern Africa. *Collective Volume of Scientific Papers, ICCAT* **26**, 566–581.

Shannon, L.V. & Pillar, S.C. (1986) The Benguela Ecosystem Part III, Plankton. *Oceanography and Marine Biology: An Annual Review*. **24**, 65–170.

Sharp, G.D. (1978) Behavioral and physiological properties of tunas and their effects on vulnerability to fishing gear. In: Sharp, G.D. & Dizon, A.E. (eds) *The physiological ecology of tunas.* Academic Press, New York.

Slatkin, M. & Hudson, R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562.

Smale, M.J., Klages, N.T., David, J.H.M. & Cockcroft, V.G. (1994) Predators of the Agulhas Bank. *Suid-Afrikaanse Tydskrif vir Wetenskap* **19**, 135-142.

Smith, B. L., Lu, C.-P., García-Cortés, B., Viñas, J., Yeh, S.-Y. & Alvarado Bremer, J.R. (2015) Multilocus Bayesian Estimates of Intra-Oceanic Genetic Differentiation, Connectivity, and Admixture in Atlantic Swordfish (*Xiphias gladius* L.). *PLoS ONE* **10**(6), e0127979. doi: 10.1371/journal.pone.0127979.

Stinchcombe, J.R. & Hoekstra, H.E. (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**, 158–170.

Sund, P.N., Blackburn, M. & Williams, F. (1981) Tunas and their environment in the Pacific Ocean: a review. *Oceanography and Marine Biology: An Annual Review*. **19**, 443–512.

Sunnucks, P. (2000) Efficient genetic markers for population biology. *Trends in Ecology and Evolution* **15**, 199–203.

Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.

Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Talbot F.H. & Penrith M.J. (1968) The tunas of the genus Thunnus in South African waters. *Annals of the South African Museum* **52**, 1–41.

Tew Kai, E. & Marsac, F. (2010) Influence of mesoscale eddies on spatial structuring of top predators' communities in the Mozambique Channel. *Progress in Oceanography* **86**, 214–223.

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

Vähä, J-P., Erkinaro, J., Niemela, E. & Primmer, C.R. (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology* **16**, 2638–2654.

Van der Elst, R.P. & Everett, B.I. (2015) *Offshore fisheries of the Southwest Indian Ocean: their status and the impact on vulnerable species*. Oceanographic Research Institute, Special Publication 10. 448pp.

Van Tassell, C.P., Smith, T.P., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C. & Sonstegard, T.S. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods* **5**, 247–252.

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

Waldman, J.R. (1999) The importance of comparative studies in stock analysis. *Fisheries Research* **43**, 237–246.

Waples, R.S. (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity* **89**, 191–201.

Waples, R.S. & Gaggiotti, O. (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419–1439.

Ward, R.D. (2000) Genetics in fisheries management. *Hydrobiologia* **420**, 191–201.

Ward, R.D. (2002) Genetics of fish populations. In: Hart, P.J.B & Reynolds, J.D. (eds) *Handbook of Fish Biology and Fisheries*, Vol. 1. Blackwell Science, Malden, USA.

Ward, R.D., Elliott, N.G., Grewe, P.M. & Smolenski, A.J. (1994a) Allozyme and mitochondrial DNA variation in yellowfin tuna (*Thunnus albacares*) from the Pacific Ocean. *Marine Biology* **118**, 531–539.

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

Ward, R.D., Elliott, N.G., Innes, B.H., Smolenski, A.J. & Grewe, P.M. (1997) Global population structure of yellowfin tuna, Thunnus albacares, inferred from allozyme and mitochondrial DNA variation. *Fisheries Bulletin* **95**, 566–575.

Weir, B.S. & Cockerham, C. (1984) Estimating *F*-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358–1370.

West, W. (2016) *Genetic stock structure and estimation of abundance of swordfish (*Xiphias gladius*) in South Africa.* (Unpublished doctoral thesis). University of Cape Town, Cape Town, South Africa.

West, W. & Kerwath, S. (2015) South African National Report to the Scientific Committee of the Indian Ocean Tuna Commission, 2015. *Report to the IOTC Scientific Committee*, IOTC–2015–SC18–NR37. IOTC, Victoria, Seychelles.

West, W., Da Silva, C. & Smith, C. (2014) South African National Report to the Scientific Committee of the Indian Ocean Tuna Commission, 2014. *Report to the IOTC Scientific Committee*, IOTC–2014–SC17–NR35. IOTC, Victoria, Seychelles.

Wexler, J.B., Chow, S., Wakabayashi, T., Nohara, K., Margulies, D., 2007. Temporal variation in growth of yellowfin tuna (*Thunnus albacares*) larvae in the Panama Bight, 1990–97. *Fisheries Bulletin* **105**, 1–18.

Wexler, J.B., Margulies, D. & Scholey, V.P. (2011) Temperature and dissolved oxygen requirements for survival of yellowfin tuna, *Thunnus albacares*, larvae. *Journal of Experimental Marine Biology and Ecology* **404**, 63–72.

Whitlock, M.C. & McCauley, D.E. (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/Nm+1$). *Heredity* **82**, 117–125.

Willette, D.A., Allendorf, F.W., Barber, P.H., Barshis, D.J., Carpenter, K.E., Crandall, E.D., Cresko, W.A., Fernandez-Silva, I., Matz, M.V., Meyer, E., Santos, M.D., Seeb, L.W. & Seeb, J.E. (2014) So, you want to use next-generation sequencing in marine systems? Insight form the Pan-Pacific Advanced Studies Institute. *Bulletin of Marine Science* **90**, 79–122.

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

Winnepenninckx, B., Backeljau, T. & De Wachter, R. (1993) Extraction of high molecular weight DNA from molluscs. *Trends in Genetics* **9**, 407.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics* **16**, 97–159.

Wright, S. (1951) The genetical structure of population. *Annals of Eugenics* **15**, 323–354.

Wright, S. (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**, 395–420.

Wright, J. M. & Bentzen, P. (1995) Microsatellites: genetic markers for the future. In: Carvalho, G.R. & Pitcher, T.J. (eds) *Molecular genetics in fisheries*. Springer, Netherlands.

Wu, G.C., Chiang, H., Chou, Y., Wong, Z, Hsu, C., Chen, C. & Yang, H. (2010) Phylogeography of yellowfin tuna (*Thunnus albacares*) in the Western Pacific and the

Western Indian Oceans inferred from mitochondrial DNA. *Fisheries Research* **105**, 248–253.

Young, J. W., Bradford, R., Lamb, T. D., Clementson, L. A., Kloser, R. & Galea, H. (2001) Yellowfin tuna (*Thunnus albacares*) aggregations along the shelf break off south-eastern Australia: links between inshore and offshore processes. *Marine and Freshwater Research* **52**, 463–474.

Zagaglia, C.R., Lorenzzetti, J.A. & Stech, J.L. (2004) Remote sensing data and longline catches of yellowfin tuna (*Thunnus albacares*) in the equatorial Atlantic. *Remote Sensing of Environment* **93**, 267–281.

Zardoya, R., Castilho, R., Grande, C., Favre-Krey, L., Caetano, S., Marcato, S., Krey, G. & Patarnello, T. (2004) Differential population structuring of two closely related fish species, the mackerel (*Scomber scombrus*) and the chub mackerel (*Scomber japonicus*), in the Mediterranean Sea. *Molecular Ecology* **13**, 1785–1798.

Zudaire, I., Murua, H., Grande, M. & Bodin, N. (2013) Reproductive potential of yellowfin tuna (*Thunnus albacares*) in the western Indian Ocean. *Fisheries Bulletin* **111**, 252–264.

Zudaire, I., Murua, H., Grande, M., Goñi, N., Potier, M., Ménard, F., Chassot, E. & Bodin, N. (2015) Variations in the diet and stable isotope ratios during the ovarian development of female yellowfin tuna (*Thunnus albacares*) in the Western Indian Ocean. *Marine Biology* **162**, 2363–2377.

APPENDIX 1: Sampling strategy

Table 5: Detailed sampling strategy of individuals analysed using SNP genotypes. **FL**: fork length; dash (-) denotes unknown FL.

Sample	Fish ID	Vial	Date	Location	FL (cm)
EC	EC1	476	Jun-15	Port Elizabeth	100
	EC2	480	Jul-15	Port St Francis	98
	EC3	494	Jul-15	Port St Francis	95
	EC4	497	Jul-15	Port St Francis	96
	EC5	501	Jul-15	Port St Francis	99
	EC6	506	Jul-15	Port St Francis	100
	EC7	508	Jun-15	Port Elizabeth	95
	EC8	512	Jul-15	Port St Francis	98
	EC9	515	Jun-15	Port Elizabeth	94
	EC10	517	Jun-15	Port Elizabeth	97
	EC11	519	Jun-15	Port Elizabeth	93
	EC12	520	Jun-15	Port Elizabeth	100
	EC13	530	Jul-15	Port St Francis	93
	EC14	532	Jun-15	Port Elizabeth	97
	EC15	537	Jul-15	Port St Francis	93
	EC16	538	Jul-15	Port St Francis	97
GG	GG1	9	Aug-15	Ghana	-
	GG2	18	Aug-15	Ghana	-
	GG3	25	Aug-15	Ghana	-
	GG4	29	Aug-15	Ghana	-
	GG5	32	Aug-15	Ghana	-
	GG6	37	Aug-15	Ghana	-
	GG7	38	Aug-15	Ghana	-
	GG8	44	Aug-15	Ghana	-
	GG9	47	Aug-15	Ghana	-
	GG10	48	Aug-15	Ghana	-
	GG11	49	Aug-15	Ghana	-
	GG12	59	Aug-15	Ghana	-
	GG13	60	Aug-15	Ghana	-
	GG14	63	Aug-15	Ghana	-
	GG15	67	Aug-15	Ghana	-
	GG16	71	Aug-15	Ghana	-
	GG17	76	Aug-15	Ghana	-
	GG18	78	Aug-15	Ghana	-
	GG19	88	Aug-15	Ghana	-
	GG20	91	Aug-15	Ghana	-
KZN	KZN1	110	Jun-15	Protea Banks	99
	KZN2	125	Jul-15	Protea Banks	94
	KZN3	133	Jul-15	Protea Banks	59
	KZN4	136	Jul-15	Protea Banks	59
	KZN5	137	Jul-15	Protea Banks	60
	KZN6	139	Jul-15	Protea Banks	99
	KZN7	140	Jul-15	Protea Banks	100

	KZN8	201	May-15	Richards Bay	125
	KZN9	202	May-15	Richards Bay	58
	KZN10	211	Jun-15	Richards Bay	60
	KZN11	401	Jul-15	Protea Banks	105
	KZN12	403	Jul-15	Protea Banks	104
	KZN13	405	Jul-15	Protea Banks	60
	KZN14	412	Jul-15	Protea Banks	108
	KZN15	413	Jul-15	Protea Banks	98
	KZN16	415	Jul-15	Protea Banks	97
	KZN17	421	Jul-15	Protea Banks	102.3
	KZN18	424	Jul-15	Protea Banks	101.2
	KZN19	430	Jul-15	Protea Banks	103
	KZN20	433	Jul-15	Protea Banks	100
MOZ	MOZ1	P10	Aug-15	Ponto do Ouro	62
	MOZ2	P11	Aug-15	Ponto do Ouro	58
	MOZ3	P12	Aug-15	Ponto do Ouro	58
	MOZ4	P13	Aug-15	Ponto do Ouro	63
	MOZ5	P14	Aug-15	Ponto do Ouro	62
	MOZ6	P17	Aug-15	Ponto do Ouro	58
	MOZ7	P18	Aug-15	Ponto do Ouro	58
	MOZ8	P20	Aug-15	Ponto do Ouro	63
	MOZ9	P21	Aug-15	Ponto do Ouro	60
	MOZ10	P22	Aug-15	Ponto do Ouro	68
	MOZ11	P23	Aug-15	Ponto do Ouro	66
	MOZ12	P24	Aug-15	Ponto do Ouro	-
	MOZ13	P25	Aug-15	Ponto do Ouro	-
	MOZ14	P26	Aug-15	Ponto do Ouro	-
	MOZ15	P5	Aug-15	Ponto do Ouro	-
	MOZ16	P6	Aug-15	Ponto do Ouro	-
	MOZ17	P7	Aug-15	Ponto do Ouro	-
	MOZ18	P8	Aug-15	Ponto do Ouro	-
	MOZ19	P9A	Aug-15	Ponto do Ouro	-
	MOZ20	P9B	Aug-15	Ponto do Ouro	-
WC	WC1	100	Oct-14	Hermanus	152
	WC2	111	Oct-14	Hermanus	131
	WC3	122	Oct-14	Hermanus	111
	WC4	30	Oct-14	Hermanus	153
	WC5	45	Oct-14	Hermanus	136
	WC6	51	Oct-14	Hermanus	136
	WC7	55	Oct-14	Hermanus	137
	WC8	57	Oct-14	Hermanus	115
	WC9	59	Oct-14	Hermanus	131
	WC10	69	Oct-14	Hermanus	133
	WC11	85	Oct-14	Hermanus	142
	WC12	88	Oct-14	Hermanus	143
	WC13	98	Oct-14	Hermanus	151

APPENDIX 2: PCR product cleaning protocol

The PCR product of each individual was combined with SureClean Plus™ (Bioline) in a 1.5mL tube, at volumes of 5µL each, which were left standing at room temperature for 10 minutes to allow for the precipitation of shorter fragments of nucleic acids, facilitated by SureClean™. Tubes were then centrifuged at 13,300 rpm for 15 minutes, in a Sorvall™ Legend™ Micro 17 Centrifuge (Thermo Scientific), to allow the larger (desired) fragments to be bound to the bottom of the tubes, leaving the shorter fragments suspended in the liquid. Liquid was then pipetted out of the tubes, and 10µL 70% ethanol was added; tubes were then vortexed at high speed for 10 seconds each, and centrifuged at 13,300 rpm for a further 15 minutes. Liquid was pipetted from tubes, which were then incubated, with their lids open, at 35°C for 10 minutes to allow for all liquid to be evaporated, leaving behind the cleaned PCR product, to which 5µL double distilled water was added. Cleaned, diluted PCR products were then frozen overnight, after which they were thawed and used in the preparation of sequencing reactions.

APPENDIX 3: Multivariate analysis results

All SNPs dataset

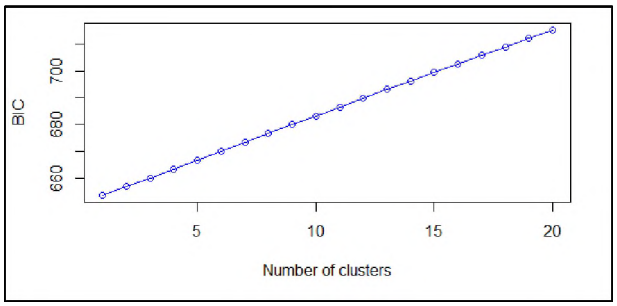

Figure 5.1: BIC values for 20 clustering solutions (k=1 to k=20) of the k-means clustering algorithm for the SNPs dataset

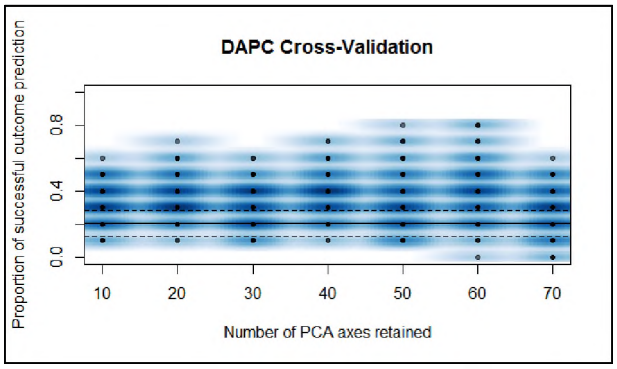

Figure 5.2: Proportion of successful assignment of individuals to original sample with increasing numbers of principal components (PCs) retained in discriminant analyses of principal components (DAPCs), after 100 replicates, for the SNPs dataset.

Table 5.1: Mean successful assignment (MSA) and root mean squared error (MSE) associated with increasing numbers of PCs retained in DAPCs of the SNPs dataset, after 100 replicates. Number of PCs retained for final DAPC denoted by *.

	10	**20**	**30**	**40**	**50**	**60***	**70**
MSA	0.336	0.360	0.317	0.372	0.334	0.393	0.289
MSE	0.673	0.649	0.692	0.639	0.683	0.626	0.722

Combined Neutral SNPs dataset

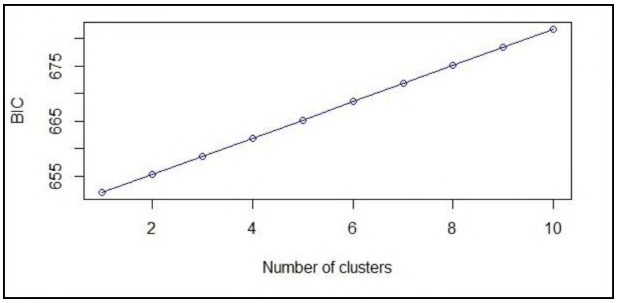

Figure 5.3: BIC values for 20 clustering solutions (k=1 to k=20) of the k-means clustering algorithm for the Combined Neutral SNPs dataset

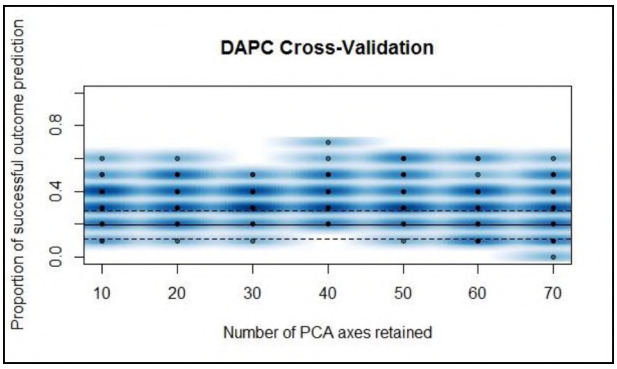

Figure 5.4: Proportion of successful assignment of individuals to original sample with increasing numbers of PCs retained in DAPCs, after 100 replicates, for the Combined Neutral SNPs dataset.

Table 5.2: Mean successful assignment (MSA) and root mean squared error (MSE) associated with increasing numbers of PCs retained in DAPCs of the Combined Neutral SNPs dataset, after 100 replicates. Number of PCs retained for final DAPC denoted by *.

	10	**20**	**30**	**40***	**50**	**60**	**70**
MSA	0.330	0.347	0.330	0.370	0.363	0.303	0.297
MSE	0.679	0.664	0.676	0.641	0.650	0.710	0.718

Variable Neutral SNPs dataset

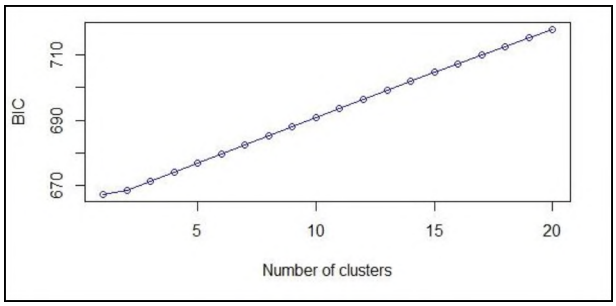

Figure 5.5: BIC values for 20 clustering solutions (k=1 to k=20) of the k-means clustering algorithm for the Variable Neutral SNPs dataset

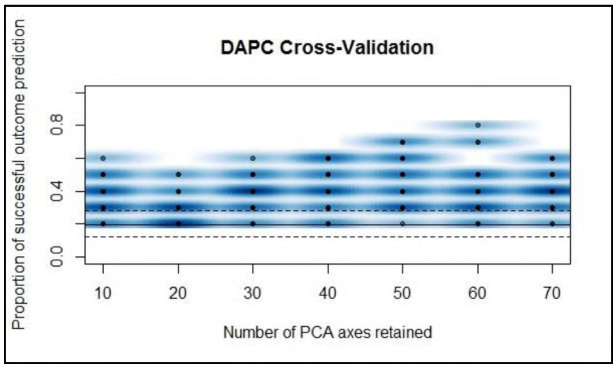

Figure 5.6: Proportion of successful assignment of individuals to original sample with increasing numbers of PCs retained in DAPCs, after 100 replicates, for the Variable Neutral SNPs dataset.

Table 5.3: Mean successful assignment (MSA) and root mean squared error (MSE) associated with increasing numbers of PCs retained in DAPCs of the Variable Neutral SNPs dataset, after 100 replicates. Number of PCs retained for final DAPC denoted by *.

	10	**20**	**30**	**40**	**50***	**60**	**70**
MSA	0.363	0.277	0.383	0.423	0.447	0.400	0.403
MSE	0.645	0.729	0.624	0.591	0.571	0.617	0.605

BayeScan Outlier SNPs dataset

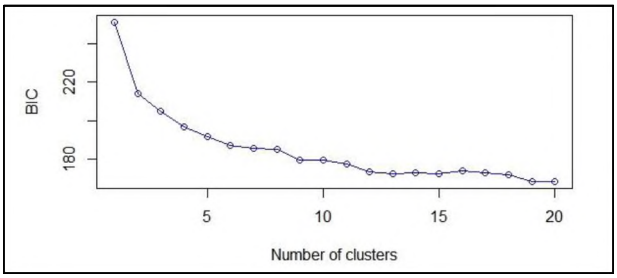

Figure 5.7: BIC values for 20 clustering solutions (k=1 to k=20) of the k-means clustering algorithm for the BayeScan Outlier SNPs dataset

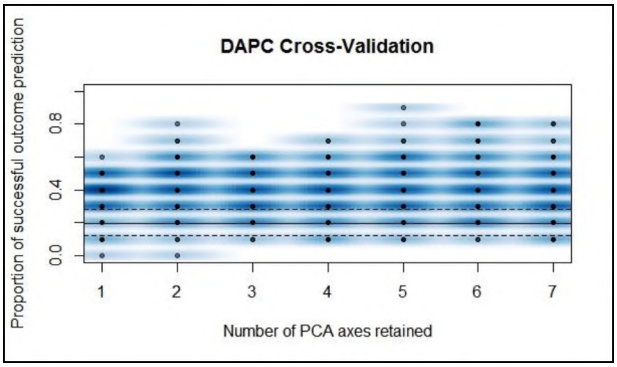

Figure 5.8: Proportion of successful assignment of individuals to original sample with increasing numbers of PCs retained in DAPCs, after 100 replicates, for the BayeScan Outlier SNPs dataset.

Table 5.4: Mean successful assignment (MSA) and root mean squared error (MSE) associated with increasing numbers of PCs retained in DAPCs of the BayeScan Outlier SNPs dataset, after 100 replicates. Number of PCs retained for final DAPC denoted by *.

	1	**2**	**3**	**4**	**5**	**6***	**7**
MSA	0.378	0.416	0.382	0.382	0.417	0.423	0.392
MSE	0.631	0.600	0.630	0.632	0.602	0.599	0.625

APPENDIX 4: Population structure results

FDIST2 Neutral SNPs dataset

Table 6.1: P-values for pairwise exact tests of genic differentiation (below diagonal) and genotypic differentiation (above diagonal) within FDIST2 Neutral SNP genotype dataset. No significant comparisons at α=0.05 observed.

	EC	GG	KZN	MOZ	WC
EC	-	1.000	1.000	1.000	1.000
GG	1.000	-	1.000	1.000	1.000
KZN	1.000	1.000	-	1.000	1.000
MOZ	1.000	1.000	1.000	-	1.000
WC	1.000	1.000	1.000	1.000	-

Table 6.2: AMOVA tests of FDIST2 Neutral SNP variation in yellowfin tuna samples: covariance components, percent variation, fixation indices and associated P-values for three hierarchical levels of structure, for two alternative hypotheses of population structure. No significant variance components or fixation indices observed at α=0.05.

	Source of variation	d.f.	S.S.	Covariance component	Percent variation	Fixation index	P-value
H1	Between groups	1	981.948	0.0429	0.00	F_{CT}=0.0001	0.498
	Within groups	3	2932.192	2.3862	0.27	F_{SC}=0.0027	0.039
	Within samples	173	154531.354	893.2448	99.73	F_{ST}=0.0027	0.064
	Total	*177*	*158445.494*	*985.6738*			
H2	Between groups	1	1085.840	2.1940	0.24	F_{CT}=0.0041	0.800
	Within groups	3	2828.300	1.4495	0.16	F_{SC}=0.0016	0.667
	Within samples	173	154531.354	893.2448	99.59	F_{ST}=0.0041	0.070
	Total	*177*	*158445.494*	*896.8884*			

Table 6.3: FDIST2 Neutral SNP-based F_{ST} values (below diagonal) and associated P-values (above diagonal) among yellowfin tuna samples. Comparisons significant at α=0.05 displayed in bold; comparisons significant after sequential Bonferroni correction in bold and denoted by *.

	EC	GG	KZN	MOZ	WC
EC	-	0.018	0.589	0.008	0.996
GG	**0.0037**	-	0.006	0.000	0.183
KZN	0.0021	**0.0037***	-	0.560	0.982
MOZ	**0.0038**	**0.0039***	0.0017	-	0.400
WC	0.0005	0.0032	0.0007	0.0024	-

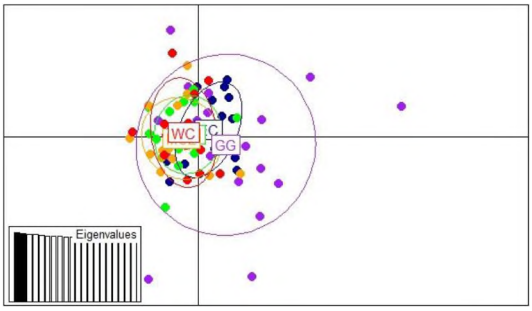

Figure 6.1: Scatterplot of results of PCA for the FDIST2 Neutral SNPs dataset, against the first two PC axes, with eigenvalues of PCs displayed.

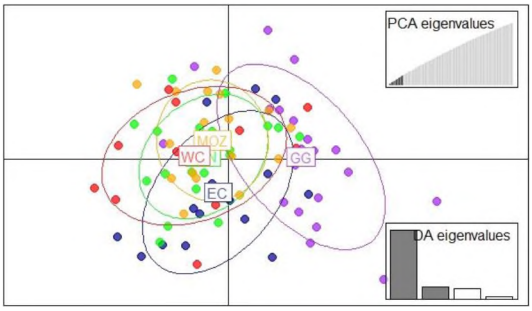

Figure 6.2: Scatterplot of DAPC for the FDIST2 Neutral SNPs dataset, retaining the first 40 PCs of the PCA, against the first two DF axes with eigenvalues of DFs displayed.

BayeScan Neutral SNPs dataset

Table 6.4: P-values for pairwise exact tests of genic differentiation (below diagonal) and genotypic differentiation (above diagonal) within BayeScan Neutral SNP genotype dataset. Highly significant comparisons displayed in bold.

	EC	GG	KZN	MOZ	WC
EC	-	1.000	1.000	1.000	1.000
GG	1.000	-	1.000	1.000	1.000
KZN	1.000	1.000	-	1.000	1.000
MOZ	1.000	1.000	1.000	-	1.000
WC	1.000	1.000	1.000	1.000	-

Table 6.5: AMOVA tests of BayeScan Neutral SNP variation in yellowfin tuna samples: covariance components, percent variation, fixation indices and associated P-values for three hierarchical levels of structure, for two alternative hypotheses of population structure. Significant variance components and fixation indices at α=0.05 displayed in bold; significant values at α=0.0001 in bold and denoted by *.

	Source of variation	d.f.	S.S.	Covariance component	Percent variation	Fixation index	P-value
H1	Between groups	1	1101.285	0.5177	0.06	F_{CT}=0.0006	0.295
	Within groups	3	3170.581	3.4318	0.37	F_{SC}=**0.0037**	0.001
	Within samples	173	161898.741	935.8309	99.58	F_{ST}=**0.0042***	0.000
	Total	*177*	*166170.607*	*939.7804*			
H2	Between groups	1	1277.848	4.3703	0.46	F_{CT}=0.0046	0.199
	Within groups	3	2994.018	1.8199	0.19	F_{SC}=0.0019	0.783
	Within samples	173	161898.741	935.8309	99.34	F_{ST}=**0.0066***	0.000
	Total	*177*	*166170.607*	*942.0211*			

Table 6.6: BayeScan Neutral SNP-based F_{ST} values (below diagonal) and associated P-values (above diagonal) among yellowfin tuna samples. Comparisons significant at α=0.05 displayed in bold; comparisons significant after sequential Bonferroni correction in bold and denoted by *.

	EC	GG	KZN	MOZ	WC
EC		0.0000	0.6560	0.0144	0.9974
GG	**0.0064***		0.0000	0.0000	0.0000
KZN	0.0022	**0.0062***		0.6356	0.9746
MOZ	0.0037	**0.0065***	0.0018		0.5443
WC	0.0008	**0.0063***	0.0012	0.0024	

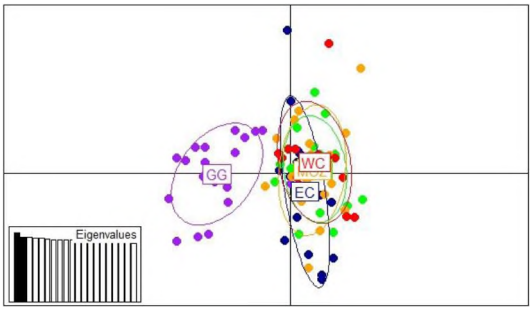

Figure 6.3: Scatterplot of results of PCA for the BayeScan Neutral SNPs dataset, against the first two PC axes, with eigenvalues of PCs displayed.

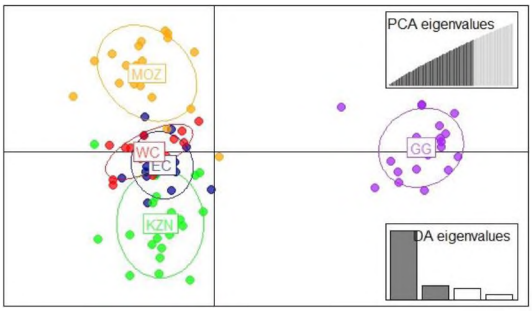

Figure 6.4: Scatterplot of DAPC for the BayeScan Neutral SNPs dataset, retaining the first 40 PCs of the PCA, against the first two DF axes with eigenvalues of DFs displayed.

FDIST2 Outlier SNPs dataset

Table 6.7: P-values for pairwise exact tests of genic differentiation (below diagonal) and genotypic differentiation (above diagonal) within FDIST2 Outlier SNP genotype dataset. Highly significant comparisons displayed in bold.

	EC	GG	KZN	MOZ	WC
EC	-	**<0.0001**	1.000	1.000	1.000
GG	**<0.0001**	-	**<0.0001**	**<0.0001**	**<0.0001**
KZN	1.000	**<0.0001**	-	1.000	1.000
MOZ	1.000	**<0.0001**	1.000	-	1.000
WC	1.000	**<0.0001**	1.000	1.000	-

Table 6.8: AMOVA tests of FDIST2 Outlier SNP variation in yellowfin tuna samples: covariance components, percent variation, fixation indices and associated P-values for three hierarchical levels of structure, for two alternative hypotheses of population structure. Significant variance components and fixation indices at α=0.05 displayed in bold; significant values at α=0.0001 in bold and denoted by *.

	Source of variation	d.f.	S.S.	Covariance component	Percent variation	Fixation index	P-value
H1	Between groups	1	131.307	0.6694	1.97	F_{CT}=0.0197	0.099
	Within groups	3	225.592	1.2236	**3.61***	**F_{SC}=0.0368***	0.000
	Within samples	173	5543.573	32.0148	**94.42***	**F_{ST}=0.0197***	0.000
	Total	*177*	*5900.72*	*33.9348*			
H2	Between groups	1	232.683	3.0627	8.66	F_{CT}=0.0866	0.200
	Within groups	3	124.216	0.27402	0.77	F_{SC}=0.0085	0.956
	Within samples	173	5543.573	32.0438	**90.57***	**F_{ST}=0.0943***	0.000
	Total	*177*	*5900.472*	*35.3805*			

Table 6.9: FDIST2 Outlier SNP-based F_{ST} values (below diagonal) and associated P-values (above diagonal) among yellowfin tuna samples. Comparisons significant at α=0.05 displayed in bold; comparisons significant after sequential Bonferroni correction in bold and denoted by *.

	EC	GG	KZN	MOZ	WC
EC	-	0.000	0.745	0.962	0.849
GG	**0.1264***	-	0.000	0.000	0.000
KZN	0.0040	**0.1176***	-	0.698	0.382
MOZ	0.0008	**0.1189***	0.0040	-	0.992
WC	0.0036	**0.1393***	0.0080	-0.0006	-

Table 6.10: Mean log-likelihood value associated with each K-value tested for the FDIST2 Outlier SNP genotypes dataset, averaged across all runs for each K, and associated standard deviation. Delta K value obtained using Evanno's method.

K	No. runs	Mean LnP(K)	Stdev LnP(K)	Delta K
1	5	-33 788.46	1.19	-
2	5	-31 544.02	4.85	434.15
3	5	-31 404.06	12.70	6.96
4	5	-31 352.48	25.90	0.35
5	5	-31 309.84	46.73	-

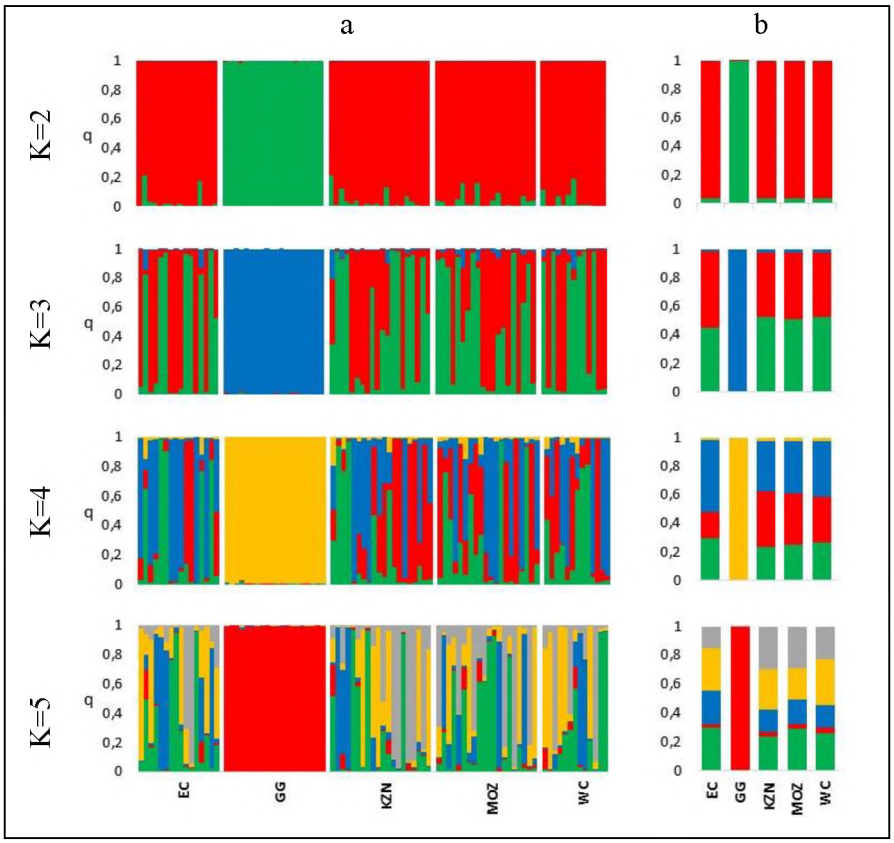

Figure 6.5: Membership coefficients for four alternative hypotheses of (K=2 to K=5) for a) each individual genotype, where each vertical bar represents one individual's genotype; and b) genotypes grouped by samples, for the Bayesian clustering analysis of the FDIST2 Outlier SNP genotypes dataset. For each hypothesis of K, different colours represent each of the K populations defined within the dataset.

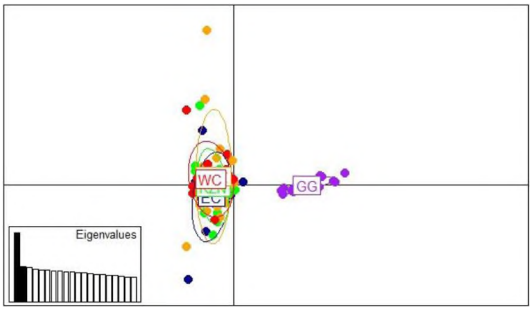

Figure 6.6: Scatterplot of results of PCA for the FDIST2 Outlier SNPs dataset, against the first two PC axes, with eigenvalues of PCs displayed.

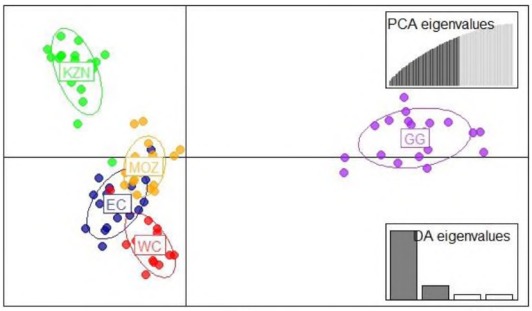

Figure 6.7: Scatterplot of DAPC for the FDIST2 Outlier SNPs dataset, retaining the first 40 PCs of the PCA, against the first two DF axes with eigenvalues of DFs displayed.