Population structure of three commercially important species in the Gulf of Guinea

Alan D Lovell

Thesis submitted for the degree of Doctor of Philosophy

Biological Sciences

University of Warwick

Coventry CV4 7AL, UK

Submission date: August 2000

Contents

List of Figures

List of Tables

Acknowledgements

I have been lucky/unlucky[1] enough to have had three serial official supervisors, and one unofficial supervisor, throughout this PhD project. Thanks must go, in chronological order, to Jacquie McGlade (for funding), E. J. Milner-Gulland (for encouragement), and Graham Medley (for reading the thing) - my three official supervisors. The very unofficial Jack Cohen bullied and cajoled me into believing in the work, and showed a remarkable amount of faith in my ability, despite the obvious shortcomings of his pupil. They've all, official and unofficial, had varying degrees of influence on the final outcome.

Other friends and colleagues have supplied enormous help. Most particularly Paul "Rick" Shaw who allowed me to use his primer sets on my cuttlefish. With his help and guidance I did in two weeks of lab work what would probably have taken me about six months alone. Alex Rogers gave me an invaluable shove in the right direction during my first fumbling months of lab work, Lesley Ward and Susan Davis performed Herculean sequencing feats and Sharon Egan, Joy Watts and Jane Green all helped me to get up and running at Warwick.

A significant number of samples in this project were collected by others, to whom I owe a great deal of thanks. Three names stand out: Ekkehard Klingelhoeffer (of the National Marine Information and Research Centre, Swakopmund, Namibia), Eric Morize (of ORSTOM at Guinea), and Alex Rogers (of the University of Southampton, though then of the Marine Biological Association based at Plymouth). I was also aided by the Gulf of Guinea Large Marine Ecosystem Project based at the Centre for Oceanographic Research in Abidjan, in particular the co-ordinator Prof. Ibe, and the senior scientists

[1] delete as appropriate

Dr. Jacques Abe and Dr. Peter Schern.

The research cruise of the RV Dr. Fridtjof Nansen was of immense benefit as it actually allowed me to get some samples from the Gulf of Guinea! I would like to thank Kwame Koranteng, for alerting me to the possibility of participation on the cruise, and the scientific staff on the vessel, notably: Daniel, Merete, Oddgeir, Paul and Philipe. In particular great thanks go to Joseph Teye and Yaovi Acakpo-Addra, both of whom provided valuable help with the joys of sexing fish. The rest of the crew (even, admittedly, the Man Utd supporters) provided excellent company.

Back home in sunny Warwick, at one time or another, we have Andrew, Ant, Ben, Carlos, Chris, Daz, Del, Dianne, Dom, Harry, Helen, Jeanette, Jo, Joel, John, Kaija, Lisa P, Lisa W, May, Martin, Mick, Nick, Sarah, Seema, Shana, Simon, Young and loads more who I've missed out...(sorry). All of whom have in one way or another made life more entertaining and, at times, joyously farcical. In particular I must thank Ben, Andrew and even Joel for being computer gurus (in particular with regards to UNiX, LiNUX and LaTeX advice), and without whom I would no doubt be trying to write this up with some ugly piece of WYSIWYG software such as MS-Word (fancy). Then of course there is May, who doesn't deserve me. She must have done something *very* bad in a former life[2].

Finally my number one big thank you has to go to my parents and bro Pete. All of whom are ***still*** waiting for me to do something vaguely useful.

[2]this life of mine, however, is benefiting hugely...

Declaration

This thesis is the result of original research conducted by myself, unless otherwise stated in the text or acknowledgements. All sources of information have been specifically acknowledged.

No part of this thesis has been submitted for a degree at any other university.

.....

Alan Lovell

Summary

The Gulf of Guinea Large Marine Ecosystem (LME) extends from the Bissagos Islands to Cape Lopez and takes in the maritime waters of all countries between Guinea Bissau and Gabon. The ecosystem is very productive and the fisheries sector is of great economic importance. This thesis uses molecular markers (mitochondrial DNA and microsatellites) to provide a comparative study of the population structure of three commercially important species in the region: *Trachurus trecae* (Cunene horse mackerel), *Pagellus bellottii* (Red pandora) and *Sepia officinalis* (Common cuttlefish). *T. trecae* showed evidence of population subdivision within the Gulf of Guinea (F_{ST}=0.056) which was explicable by temporal (F_{ST}=0.048), as opposed to spatial (F_{ST}=0.001), structuring. Thus contemporaries from the same length cohort showed genetic similarity, regardless of geographic proximity. A significant correlation (correlation coefficient *D*: r=0.93, p=0.01) was found between the cohort length and Tajima's *D*. *P. bellottii* likewise showed little evidence of spatial subdivision within the Gulf of Guinea (F_{ST}=0.009), however four individuals from a single trawl showed high sequence variation from all other samples (and when included in the analysis F_{ST}=0.095). Both fish species displayed bimodal length frequencies for some trawls and when split according to cohort length there was evidence of within trawl heterogeneity, indicating that shoals are an aggregation of smaller groups. *S. officinalis* revealed no spatial subdivision in the the Gulf of Guinea (F_{ST}=0.00), though four individuals showed highly atypical allele sizes. Possible evidence of selection at one microsatellite locus was found. When compared with outgroups from southwest Africa and Europe *T. trecae* and *S. officinalis* showed great differentiation (F_{ST}=0.642 and F_{ST}=0.301 respectively). Comparative results across species therefore indicate *(i)* that the Gulf of Guinea is a well defined LME and *(ii)* there are no major oceanographic structures within the LME that have caused spatial population subdivision. Given such a lack of spatial subdivision, management needs to operate at a regional level for these species. These results were found for three species with very differing life histories, so they may also be applicable to other marine species in the region.

Abbreviations

AMOVA	Analysis of Molecular Variance
ASD	Average Squared Distance
BDRM	Base de Données Régionale Maritime (Regional Maritime Database)
dNTP	deoxynucleotide
EDTA	ethylenediaminetetraacetic acid
EU	European Union
F	Fixation Index (hierarchical - embellished with subscripts to denote levels of hierarchy compared)
FAO	Food and Agriculture Organisation of the United Nations
FIAS	Fisheries Information and Analysis System
IAM	Infinite Alleles Model
ITCZ	Inter Tropical Convergence Zone
LME	Large Marine Ecosystem
MSA	Mixed Stock Analysis
MST	Minimum Spanning Tree
mtDNA	mitochondrial DNA
Nansen	RV Dr. Fridtjof Nansen

NORAD	Norwegian Agency for Development Cooperation
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
rRNA	ribosomal RNA
RV	Research Vessel
SDS	sodium dodecyl sulfate
SMM	Stepwise Mutation Model
TE	Tris - EDTA buffer
tRNA	transport RNA
Tris	(hydroxymethyl) aminomethane
UK	United Kingdom of Great Britain and Northern Ireland
UNDP	United Nations Development Programme
UPGMA	Unweighted Pair Group Method with Arithmetic Mean

Chapter 1

Introduction

1.1 Preamble

The introductory chapter covers the aims of the project and provides the necessary background to understand the rest of the thesis. The Gulf of Guinea ecosystem is introduced in detail, along with the key features of neighbouring ecosystems. Physical and oceanographical structures are explained, and how they may affect the structure of marine biological populations. The three species to be analysed are introduced, along with justification for their choice. Finally molecular ecology, the analytical tool used in this project, is introduced as a practical study of genetic patterns in biological populations.

The structure of this thesis is designed to allow rapid access to chapters of interest. The first two chapters, the "Introduction" and "Materials and Methods", are self explanatory. The three chapters entitled "Molecular Genetic Analysis of ..." are the heart of the thesis and present the results of genetic analyses of the three species under consideration. The chapter on "Comparative Molecular Ecology and Evolution" ties together the lines of enquiry across the three species chapters, along with introducing some new analyses. Extensive cross referencing allows the reader to begin by reading one of the Results chapters and only to refer back to the "Introduction" or "Materials and Methods" when required (e.g. for explanation of a methodology or model). Each of the Results chapters, with the use of cross referencing, can therefore be read independently of the rest of the thesis. A list of common names for all organisms referred to in this

thesis can be found in the appendices, section A on page 190.

1.2 Aims of the project

There were four main aims of the project. These aims are listed below, while in parentheses the chapters are given in which the respective aims have been addressed.

1. Discover and develop molecular markers for population analyses of three commercially important species in the Gulf of Guinea. Compare the different marker sets and provide information on their strengths and weaknesses. (chapters 2 and 6)

2. Collect samples of the three commercially important species from the study area for genetic analyses. (chapter 2)

3. Use the developed molecular marker sets and collected samples to do a preliminary analysis of the population structure of the three marine species in and around the Gulf of Guinea and the west African coast. (chapters 3, 4 and 5)

4. Describe the direction future work may take, taking into account interesting results from this and other projects and areas where current knowledge is lacking. (chapters 6 and 7)

The work in this thesis was funded by the European Union Project entitled: Impacts of Environmental Forcing on Marine Biodiversity and Sustainable Management in the Gulf of Guinea (Anon, 2000). The two main goals of the EU project are *(i)* To assess the impacts of upwelling and other forms of environmental forcing on marine biodiversity and the dynamics of fisheries; and *(ii)* To develop and implement an information and analysis system for the sustainable management and governance of fisheries resources (Anon, 2000). The "Discussion and Conclusions" chapter (page 158) will tie in how the results of this thesis meet some key objectives of the European Union Project.

1.3 Introduction to the Gulf of Guinea Large Marine Ecosystem

A Large Marine Ecosystem (LME) has been described as a region with distinct physical, oceanographic and biological features that has a distribution of trophically dependent populations (Tilot & King, 1993). The size of an LME is usually about 200,000 km^2 or larger, and boundaries between them are not precisely defined but are regions that may shift position during the seasons (Sherman, 1993). The Gulf of Guinea LME extends from the Bissagos Islands in the North (about 11°N) to Cape Lopez in the South (Binet & Marchal, 1993), thus including the maritime waters of all countries between Guinea Bissau and Gabon, including the island state of Equatorial Guinea. The Gulf of Guinea LME has often been divided into three related but distinct subsystems (see Figure 1.1)[1].

The Sierra Leone and Guinea Plateau subsystem to the west contains many subareas of ecological and economic importance. Within Senegal there are the humid estuarine areas of Sine-Saloum and Casamance that are of particular importance for migratory birds. These and other estuarine regions are characterised by having shallow water regions that are areas of high productivity, and many fish species benefit from a juvenile stage centred in these mangrove-dominated estuarine systems. Both commercially important species, and juvenile stages of the species on which they feed, are often found to have spawning and nursery stages in these estuarine ecosystems (Tilot & King, 1993; Laegdsgaard & Johnson, 1995).

The species for which these mangrove-dominated ecosystems are of tremendous importance include *Pagellus bellottii, Epinephelus aeneus, Plectorhynchus mediterraneus, Brachydeuterus auritus, Sparus caerusostictus*, and *Macroramphosus* spp. as well as the pelagic species *Sardinella aurita, S. maderensis, Pomatomus saltatrix, Argyrosomus regius, Decapterus rhonchus, Trachurus trecae, Scomber japonicus, Sardina pilchardus* and *Scyris alexandrinus*. Inshore migrations occur and many commercially important tropical fish species and their fisheries appear to be dependent on the productivity of

[1]The figure, and all other maps in this thesis, was created with the aid of the online mapping tool at: http://www.aquarius.geomar.de/omc/omc_intro.html

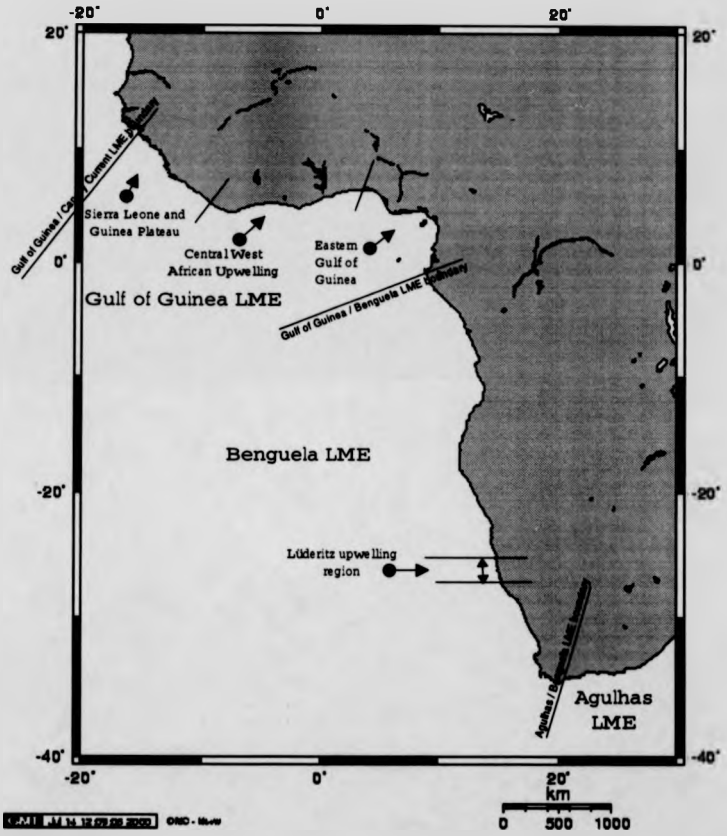

Figure 1.1: An overview of Large Marine Ecosystems (LMEs) of west and south west Africa. The boundaries are not precisely defined but are regions that may shift position during the seasons. The three subsystems of the Gulf of Guinea LME, and the Lüderitz upwelling region in the Benguela LME, are marked (see text).

estuarine and mangrove areas for continued exploitation of finfish stocks (Tilot & King, 1993).

Unfortunately mangrove ecosystems, and thus mangrove associated fisheries, are sensitive to human activities upstream. Because of their nature as interface ecosystems that link marine and terrestrial habitats, activities such as dredging, draining land, aquaculture development (Anon, 1994), or road construction can damage or alter their usual functioning. Deforestation, desertification, urbanisation and hunting all threaten mangrove-dominated estuarine ecosystems.

The Central West African upwelling subsystem also has a variety of subareas that are important as spawning grounds for a variety of finfish species. Many of the Capes along the coast of Ghana and Côte d'Ivoire show a similarity to the physical features of the South-eastern Brazilian Bight and Southern Californian Bight. Such coastal physical features are protected from currents and create a closed geostrophic pattern which may aid in retaining larvae, allowing such larvae to take advantage of a favourable coastal environment. As noted above, many of the favourable spawning grounds are linked with mangrove habitats, and these are often under stress from economic activities. The lagoons along this part of the coast also harbour the juveniles of many species (Tilot & King, 1993).

Two studied species in the region are *Sardinella aurita* and *S. maderensis*. The two species are linked to the flow pattern of the Guinea Current and the Guinea Under Current, and their distribution is most probably defined by the availability of food (mostly phytoplankton and zooplankton) (Chikhi, 1995). There are probably two types of migratory behaviour. The first is the movement inshore to shallower coastal waters during the upwelling period, from the deeper (70 - 100 m) waters that they reside in during warmer periods. The second is migration along the coast, following the displacement of cold waters. Anchovies however are more adaptable than sardinellas and can be found at depth or in upwelling regions along the coast of Togo and Ghana. Such a distribution could be compared to the ecosystems of the southwest Atlantic (south-central Argentina to southern Brazil) (Tilot & King, 1993). Because the fish species of interest

in this project have a trophic relationship similar to sardinellas they may display similar feeding migrations.

The Eastern Gulf of Guinea subsystem is a relatively less well known subsystem and is dominated by variability in the amount of river input because of climatic fluctuations of the monsoon. Since 1970 river inputs and the seasonal equatorial upwelling have decreased (Tilot & King, 1993). Such changes in the physical environment presumably affects the distribution, life cycle and sizes of the fish species found in the subsystem (Tilot & King, 1993), though given the lack of suitable data from the region it is not possible to safely describe or explain such effects.

1.3.1 Biodiversity of Gulf of Guinea LME

Following Koranteng *et al.* (1996), the fisheries resources in the Gulf of Guinea LME can be classified as follows:

- **Small pelagic species**: sardinellas (*Sardinella aurita* and *S. maderensis*), anchovy (*Engraulis encrasicolus*) and chub mackerel (*Scomber japonicus*) are the most abundant and economically important small pelagic species in the ecosystem. In the western Gulf of Guinea the distribution and abundance of these species are affected by the seasonal coastal upwelling. In the southern part of the LME *Ethmalosa fimbriata, S. maderensis* and *Ilisha africana* become more dominate.

- **Coastal demersal species**: demersal species are more variable than the small pelagics and dominance depends on the shelf bottom. In shallow waters and relatively soft bottoms the demersal fish fauna is dominated by croakers (*Pseudotolithus senegalensis, P. typus*); threadfins (*Galeoides decadactylus*); grunts (*Pomadasys incisus* and *P. jubelini*) and soles (*Cynoglossus* and *Solea* spp.). In the Ghana - Côte d'Ivoire sector the most important coastal demersal fish species are of the families Sparidae, Haemulidae, Mullidae, Sciaenidae, Lutjanidae, and Serranidae. A little deeper, seabreams (*Pagellus bellottii, Sparus caeruleostictus*) dominate the fauna.

  Significant faunal shifts in the demersals have occurred, the most striking being

in this project have a trophic relationship similar to sardinellas they may display similar feeding migrations.

The Eastern Gulf of Guinea subsystem is a relatively less well known subsystem and is dominated by variability in the amount of river input because of climatic fluctuations of the monsoon. Since 1970 river inputs and the seasonal equatorial upwelling have decreased (Tilot & King, 1993). Such changes in the physical environment presumably affects the distribution, life cycle and sizes of the fish species found in the subsystem (Tilot & King, 1993), though given the lack of suitable data from the region it is not possible to safely describe or explain such effects.

1.3.1 Biodiversity of Gulf of Guinea LME

Following Koranteng *et al.* (1996), the fisheries resources in the Gulf of Guinea LME can be classified as follows:

- **Small pelagic species**: sardinellas (*Sardinella aurita* and *S. maderensis*), anchovy (*Engraulis encrasicolus*) and chub mackerel (*Scomber japonicus*) are the most abundant and economically important small pelagic species in the ecosystem. In the western Gulf of Guinea the distribution and abundance of these species are affected by the seasonal coastal upwelling. In the southern part of the LME *Ethmalosa fimbriata, S. maderensis* and *Ilisha africana* become more dominate.

- **Coastal demersal species**: demersal species are more variable than the small pelagics and dominance depends on the shelf bottom. In shallow waters and relatively soft bottoms the demersal fish fauna is dominated by croakers (*Pseudotolithus senegalensis, P. typus*); threadfins (*Galeoides decadactylus*); grunts (*Pomadasys incisus* and *P. jubelini*) and soles (*Cynoglossus* and *Solea* spp.). In the Ghana - Côte d'Ivoire sector the most important coastal demersal fish species are of the families Sparidae, Haemulidae, Mullidae, Sciaenidae, Lutjanidae, and Serranidae. A little deeper, seabreams (*Pagellus bellottii, Sparus caeruleostictus*) dominate the fauna.

  Significant faunal shifts in the demersals have occurred, the most striking being

the triggerfish (*Balistes capriscus*) episode of the mid seventies to the late eighties. The species dominated the landings of the trawlers operating in the western Gulf of Guinea, and in 1981 the stock of the triggerfish in Ghana - Côte d'Ivoire waters was estimated by acoustic methods to be over 500 000 tonnes. During the eighties the catch of triggerfish declined and the species had virtually disappeared by 1989. The reason for such faunal shifts is unknown, though physical change in the ecosystem are thought by some to be involved. During and after the triggerfish population decline the cephalopod populations increased in size and economic importance, though the fishery is already considered over-exploited. (Koranteng *et al.*, 1996)

- **Deep water demersal species**: below 200-300 metres depth the demersal species include the blackmouth croaker (*Pentheroscion mbizi*); the silver-rag driftfish (*Paracubiceps ledanoisi*); the deep-sea red crab (*Geryon maritae*); and the deep-water rose shrimp (*Parapenaeus longirostris*). Deep-water demersal species are not yet fully exploited in the region, presumably because of the high cost of deep-water fishing.

Species that migrate extensively throughout the eastern Atlantic include the yellowfin (*Thunnus albacares*), skipjack (*Katsuwonus pelamis*) and bigeye (*Thunnus obesus*). The migration routes of these individuals will often take them outside the EEZs (Exclusive Economic Zones) of coastal states. Other tuna-like species include Atlantic little tunny (*Euthynnus alletteratus*), Atlantic bonito (*Sarda sarda*) and wahoo (*Acanthocybium solandri*).

1.3.2 Species assemblages

Six species assemblages are often attributed to the continental shelf of the eastern Atlantic, from Gambia to the Congo. All but one of them can be equated to a community previously described off the coasts of western Europe (Longhurst, 1969). This situation can arise because of the presence of a permanent, sharp thermocline at depths always shallower than 50m. Because of its shallowness the cool sub-thermocline water occurs at

depths in the tropical region at which a considerable amount of sunlight still penetrates, and thus a number of cool sub-tropical fish are able to exist in the tropical region of the Gulf of Guinea. In fact, the physical conditions of around 100m depth in the Gulf of Guinea is very similar to the conditions off north-western Africa and southern Europe. The region where this thermocline is seasonally disrupted, off the coast from Côte d'Ivoire to Togo, displays assemblages of demersal fish that are not so readily recognised elsewhere.

The assemblages can be listed thus from Longhurst (1969):

Assemblage A: Offshore sciaenid sub-community

Assemblage B: Estuarine sciaenid sub-community

Assemblage C: Super-thermocline sparid community

Assemblage D: Sub-thermocline sparid community

Assemblage E: Deep shelf community

Assemblage F: Slope community

A few species are difficult to allocate to particular assemblages because of their eurybathy (i.e. their ability to change depth), for example *Brachydeuterus auritus*. Assemblage D, the sub-thermocline sparid community, includes the species *Pagellus bellottii*, *Trachurus trachurus* and *T. trecae*, and *Scomber japonicus*.

Trachurus spp. have been associated with two other semi-pelagic species; *Boops boops* and *Scomber japonicus*, so much so that it is suggested that they form combined shoals (Longhurst, 1969). These occur only in a rather narrow depth and temperature range, with their highest abundance being around the bottom of the thermocline at a temperature of about 17°C. In the Canary and Benguela Current LMEs these cold water species form surface shoals. The demersal shoals of the tropical Gulf of Guinea region are also generally less abundant and consist of smaller sized individuals when compared with the surface shoals of the Benguela and Canary Current regions. *Pagellus bellottii*

is more confined to the tropical region than *Trachurus* spp. It is associated particularly with *Dentex angolensis*.

1.3.3 The economic environment

The marine environment off west Africa is one of the richest and most productive in the world. The overriding reason for such productivity is the constant or seasonal upwelling in the area. This productivity has led to a situation where there is both a strong artisanal and industrial sector. The former represents part of the subculture of the region and has strong traditional roots and powerful social and political impacts, while the latter is composed of foreign and local fleets and is a major contributor to the region's economy (Tilot & King, 1993).

The total biological potential of the fisheries off west Africa has been estimated at between 3.2 and 5 million tonnes per year, which represents about 4.5% of the worlds catch (Anon, 1992). The fisheries sector is therefore of great economic importance to the countries of west Africa, providing valuable employment. Furthermore, only south-east Asia is more dependent on fish for its supply of protein (Tilot & King, 1993).

The artisanal fisheries of the area accounts for about 70% of the catch taken by west Africans. Only ten ethnic groups out of about seventy account for 75% of the artisanal catch (Tilot & King, 1993), and thus most of the catch is taken by migrant fishing groups. Though this has occasionally been a source of political argument, the migrant fishing communities land their catch at the local markets and contribute to the regional economy, and often do not compete with the local fishing communities because they target different stocks. The income earned by the migrant fishing communities will invariably be spent on food, accommodation, official and non-official permits and so on, all of which supports the local economy.

The most developed artisanal fisheries are found close to upwelling areas, particularly in Senegal (the Wolof, Lebu, Niominka, and Bijogo), Ghana, Togo and Côte d'Ivoire (fished by the Aladian, Fanti, Ga and Ewe). The variability of the Central African Upwelling system have forced some tribes to develop adaptive fishing strategies, such as

the Ewe and Fanti, while the non-upwelling regions off Sierra Leone are exploited by the Sherbo tribe, and off Casamance and Sine Saloum by the Lebu and Niominka (Tilot & King, 1993).

Domestic industrial fishing accounts for the remaining 30% of catches taken by West African vessels, and is set to rise in the future. Unlike the artisanal fleets there are very few privately owned industrial vessels and most are nationally owned. As expected, the countries with the most industrial vessels are those near productive areas, and they generally target the higher valued demersal species. Some agreements regarding cross national fishing access have been negotiated, and Senegalese vessels can be found off the coasts of Gambia, Guinea Bissau and Guinea, while Nigerian vessels have some access to the stocks off Angola and Gabon. Such agreements are expected to become more common in the future (Tilot & King, 1993).

In addition to west African artisanal and industrial fleets the highly productive waters have also attracted foreign fishing fleets on a large scale since the 1960s. These foreign fleets have predominately come from Europe, Russia, ex-eastern bloc countries, The Korean Republic and Japan (Tilot & King, 1993). A combination of highly efficient fishing practices and virtually a complete lack of governance have led to resources in the area being under great stress.

1.3.4 The Canary Current and Benguela Current LMEs

The two bordering LMEs either side of the Gulf of Guinea are the Canary Current LME to the north and the Benguela Current LME to the south. A review of the important features of the two LMEs follows.

Canary Current LME

The Canary current LME is often split into three further subsystems (Tilot & King, 1993). The southernmost subsystem, that borders the Gulf of Guinea LME is termed the Northwest African upwelling system. It is a permanent upwelling system and is in the passage of the ITCZ (the Inter Tropical Convergence Zone, forming the boundary

between the Canary Current LME and the Gulf of Guinea LME). The boundary is influenced by the Cape Verde peninsula and the Kayer canyon which limits the migration of certain demersal fish species (Tilot & King, 1993).

Further North, i.e. beyond the northernmost limit of the ITCZ, is the Canary Islands shelf subsystem. It has seasonal upwelling and the fauna has multi biogeographical origins (Tilot & King, 1993). Further north again is the Moroccan shelf subsystem which is also a seasonal upwelling area and is under Mediterranean influence. It is transversed by canyons and the northern boundary is represented by the northern limit of the Canary current.

Benguela Current LME

The Benguela upwelling system runs along the southwestern coast of Africa from Angola in the North to Cape Agulhas in the South (see Figure 1.1). It is split into northern and southern subregions by a zone of intense perennial upwelling activity in the Lüderitz region (26°- 27.5° S) (Cole, 1999). The high levels of offshore transport and turbulence in this region acts as a barrier to south-north transport of pelagic eggs and larvae. This, combined with little migration across the upwelling system, often results in populations north and south of the upwelling region being reproductively isolated (Cole, 1999). The upwelling region around Lüderitz has the lowest species diversities of larvae in the system. Species spawn to the north and to the south of the region, but not within it (Pilar Olivar & Shelton, 1993).

1.4 Oceanography and population structuring

The Gulf of Guinea is a unique ecosystem, biologically and oceanographically, because of its east-west coastline combined with seasonal upwelling systems. Here are described some of the oceanographic features found in the region, with a brief discussion on how they may influence the distribution of marine populations in the region.

1.4.1 Tides and fronts

The whole of the west coast of Africa has a tidal range of less than 2m, being highest in the bight of Benin, putting it in the microtidal range (<2m) (Longhurst & Pauly, 1987). It is a common observation globally that there is a narrow strip of ocean above or slightly seaward of the shelf edge that is in a state of permanent upwelling, though possibly quite weakly so (Mann & Lazier, 1991). This feature is observable in the Gulf of Guinea, particularly during the dry season of November through to March, outside the main upwelling event. Palaeoarctic terns, yellowfin tuna and porpoises can be found congregating around the edge of the shelf at this time to take advantage of the large number of small pelagic fish and micronekton. Internal waves of up to 50m amplitude increase the amount of mixing between the photic zone and the relatively nutrient rich zone underneath, and acoustic methods have allowed researchers to follow the plankton populations "riding" these internal wave trains as they break upon the shelf (Longhurst & Pauly, 1987).

As with coastlines in higher latitudes, wind-forcing from moderate winds blowing for two days or more parallel to the coastline can cause further dynamics, including the initiation of upwelling events. Together, these dynamic processes can create fronts between the waters of the shelf and the slope and open ocean, having consequences for the distribution of plankton and pelagic fish (Sabates & Pilar Olivar, 1996).

Though tides are oscillatory, often reflected as diurnal or semi-diurnal tides, there is often a net movement of water in one direction: a residual current. These are stronger in shallow water, around headlands and in bays and estuaries, and water is passed through an area in the direction of the residual current. Where the turbulence induced by bottom friction is strong enough, the stratification of the continental shelf water can be broken down, leading to eddy production. Whether or not this occurs is dependent primarily on the tidal current and the depth of the water. Freshwater run off and the buoyancy of solar heating and evaporation also has some effect. A front often develops between the tidally mixed and stratified areas and nutrients are supplied to the euphotic zone, enhancing biological production (Kingsford, 1990).

An important shelf process with regards to larval transport is the existence of permanent or seasonally recurring coastal gyres in which eggs, larvae and other planktonic organisms may be retained in a favourable environment. Eddies are most likely to be produced in bays or behind headlands (as suggested above), and it is here that nursery grounds could be found, either seasonally or permanently, with water circulation running counter current to the main stream offshore. Permanently tidal mixed areas like Georges Bank, near the Gulf of Maine, and the Dogger Bank in the North Sea, have levels of primary production considerably higher than adjacent stratified areas of shelf. These sites are often selected as breeding grounds by commercially important fish stocks and it is reasonable to assume that similar areas exist throughout the Gulf of Guinea and elsewhere.

Although it is often noted that adult individuals of a species like herring are found in large aggregations, it is often hypothesised that they divide into breeding populations with a characteristic time and place of breeding, returning to the place where they were hatched (Mann & Lazier, 1991; Cury, 1994). For some stocks such breeding grounds may be readily identified as bays or estuaries, for others there is no clear physical boundaries to the nursery area or spawning ground. Tidal fronts may act as defining boundaries for some of these nursery or spawning areas.

The concept that the turbulence generated by tidal action could keep some shallow waters mixed while deeper water nearby becomes stratified was advanced over 60 years ago. The Simpson & Hunter (1974) method of predicting such fronts uses an energy argument, which postulates that a front would be found where the intensity of turbulent mixing was just enough to continuously overcome the barrier to mixing presented by stratification. Tidal mixed areas can be predicted from physical features, and has been done for areas on Georges Bank, Nantucket Shoals, in the mouth of the Bay of Fundy and off the south west coast of Nova Scotia (e.g. see Simpson & Hunter, 1974). Iles & Sinclair (1982) showed that there was a remarkable similarity between the distribution of herring larvae and the occurrence of tidal mixing, on both the east and west coast of the North Atlantic. This trend has been found by other researchers, and has been an

observation central to the creation of the member-vagrant hypothesis (see section 1.4.4).

1.4.2 Thermoclines and upwelling

Some of the sharpest gradients in temperature are found in tropical water masses. The thin skin of the warm tropical water is separated from the cold water below by a sharp thermal discontinuity (thermal stratification) called the thermocline. This barrier can be broken through mixing, found for example during upwelling. Salt concentration is determined by the relative amounts of evaporation to precipitation, and humid tropical regions have a somewhat lower salinity level than that of arid regions.

The "estuarization" of the continental shelf can happen when river discharges and monsoonal rains dilute the surface water to such an extent that salinity changes substantially. Surface salinity maps show that the Gulf of Guinea is one of the regions where discrete plumes of river discharge water are clearly visible (Longhurst & Pauly, 1987), and along with lowering the surface salinity they also tend to be of high turbidity and contain muddy deposits. Such effects are generally seasonal and can be observed particularly in the Bight of Biafra, in the extreme eastern corner of the Gulf of Guinea, where waters appear to move northward into the Bight from south of the equator, thereby crossing the conjunction of the South Equatorial Current and the Guinea Current. As the salinity rises, after the monsoon season, the brackish water mass is then generally driven back southwards.

The subsurface waters immediately beneath the thermocline are usually more uniform than the warm skin of tropical water sitting upon it. Such waters masses are generally formed by sinking near the subtropical divergence. An important feature of these water masses is that they often contain very little oxygen and, along with an O_2 minimum, the water may contain trace elements of the poisonous hydrogen sulphide (H_2S). This combination of factors serves to displace the demersal fauna and benthic crustacea during periods of upwelling, and entrapment of biota in oxygen-poor upwelled water has sometimes been blamed for observed occurrences of mass mortality (though such mass mortality has not been observed off west Africa itself).

The most important aspect of upwelled water however is that it is nutrient rich, principally in nitrogen, most often in the form of ammonia (NH_3). These nutrients are recycled from previous primary production events, where phytoplankton, zooplankton and faeces have sunk into the deeper waters. The upwelled water stimulates primary production by the phytoplankton, when the water has calmed a little. These have a very rapid response time to new conditions because they have a life cycle of just a few hours or days. The grazing zooplankton respond more slowly because they on average have a life cycle of a few weeks. Fish, squid and other organisms can then prey on the abundance of both phytoplankton and zooplankton. Finally, waste, dead organisms and other detritus that is not immediately digested by predators sink down below the thermocline to waters that will be upwelled again in the future, ensuring continued recycling of nutrients.

It is important to note therefore that upwelled waters can have both a positive and negative effect. The water is often poisonous to fish stocks, because of the lack of O_2 and trace amounts of H_2S. In addition the upwelled water can also sweep eggs and larvae out to sea, away from the safety of their nursery areas, though demersal eggs have been shown to be much more resistant to movement by such offshore currents (Norcross & Shaw, 1984). However the advantage of upwelling is the nutrient input, enjoyed mainly by the phytoplankton during the relatively calm periods following upwelling events when stratification develops and the phytoplankton feed, grow and multiply. In other words there is a miniature "spring bloom" during each calm period (Mann & Lazier, 1991). The fish stocks then harvest either the phytoplankton or zooplankton, or both depending on the species, and it is this abundant supply of food that leads to the productive fisheries found in upwelling areas.

The Gulf of Guinea undergoes seasonal upwelling events. Upwelling occurs weakly in January-February and strongly during the rainy season in July-September. The causes of this upwelling are quite complex with various mechanisms involved. A predominant feature that can at least partly explain the seasonality of the upwelling is the seasonal intensification of wind stress off South America, causing a series of eastward-propagating,

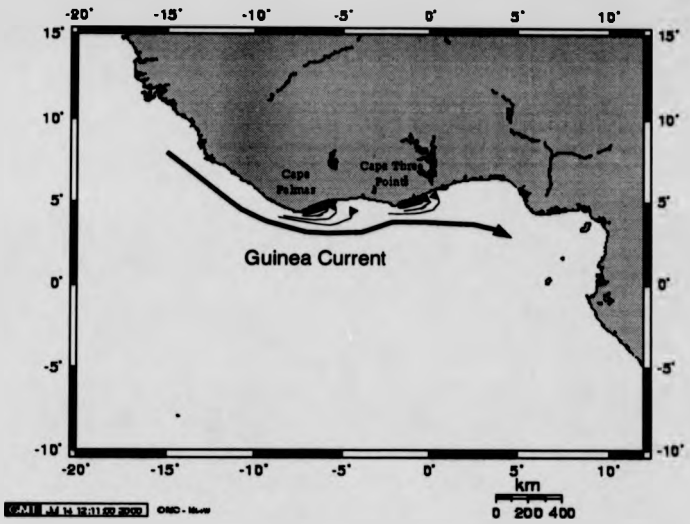

Figure 1.2: Larval retention zones in the Gulf of Guinea. Sea surface temperatures are permanently coldest just east of Cape Palmas and Cape Three Points due to vertical turbulence. The patch to the east of Cape Palmas extends further downstream than that of Cape Three Points.

equatorially trapped Rossby waves which are transformed at the eastern margin of the ocean into poleward-propagating (i.e. north and south) Kelvin waves, a northern series trapped by the coast of the Gulf of Guinea and a southern series on the coast of Angola (Longhurst & Pauly, 1987). Imposed on this seasonal cycle is a series of discreet, rapid drops in sea surface temperature, most strongly during the wet season upwelling when the temperature is already at its coolest. These are imposed by the dissipation of the fortnightly tidal energy (Houghton, 1983). Sea surface temperatures are coldest just east of the two most prominent capes, Cape Palmas and Cape Three Points, because of the eastward flow of the Guinea Current past these capes intensifying the vertical turbulence and ensuring a mixed water column. Such cold water patches at the surface are found all year round, and such a cape effect is most noticeable from July to September (Picaut, 1983). The cold water patch to the east of Cape Palmas extends further downstream than that of Cape Three Points (see Figure 1.2).

The great pair of eastern boundary upwelling regions of Africa, the Canary and

Benguela upwelling systems, are dominated by classical wind driven divergence of the Ekman layer. The morphology of the continental shelf restricts the width of the upwelling strip to 10-20 km. Local differences in the intensity and direction of wind stress will also have an effect on the behaviour of the upwelling system.

Finally, a further region of upwelling occurs because of the creation of cyclonic gyres by the equatorial counter current arriving at the eastern ocean boundary. The waters in the domes are physically and chemically similar to the waters in the equatorial divergence, and a rich development of phytoplankton can often be found downstream of the upwelled water (Mann & Lazier, 1991). A well studied example is the Costa Rica dome, thought to be a permanent feature, whereas the Guinea dome is probably seasonal.

1.4.3 The stock concept

The concept of the unit stock in fisheries arose from the need to identify manageable groups of fish in the sea. Central to the concept is the Harden-Jones (1968) model of the triangle of migration. Fish migrate, often against the prevailing currents, from the feeding ground to their spawning ground. After spawning the spent fish are carried back by the current to their feeding grounds, often in deeper waters. At the same time the larvae drift with the same currents and through behavioural mechanisms, such as vertical migration, find their way to nursery grounds, which are often in shallower waters. After a period of (often) two or more years the immature fish leave the nursery grounds and join the adult stock, either on their feeding or spawning grounds. This closes the triangle of migration.

The two distinct migrations, that of the adults to the spawning grounds and the juveniles to the nursery grounds and then to the adult population, maintain the identity of the stock from generation to generation. The stock theoretically is isolated from other stocks, as far as wind, tide and current permit (Cushing, 1995). It is with regard to such containment that the physical environment is so important, as it is regions of upwelling, currents, fronts and gyres that will act as defining features of stocks. One of the original aims of genetic methods in fisheries was to try and identify these stocks, with the hope

that the unit stock would be a genetically distinct entity because of its isolation. It has not been an entirely successful venture as genetic differences have in general been hard to find except over great distances such as ocean basins. Such problems in finding genetically distinct stocks have lead researchers to propose alternative definitions of stocks (e.g. see Smith *et al.*, 1990; Dizon *et al.*, 1992; Carvalho & Hauser, 1994).

In spite of the disappointments of earlier work on genetics in fisheries there were also some notable successes, particularly in anadromous fishes such as the salmon (e.g. Utter *et al.*, 1980), showing that on occasions there were identifiable genetic differences between stocks. Fish such as anadromous salmon are relatively much easier to work on than purely marine species because they return to easily identifiable spawning grounds (i.e. rivers). It is more difficult to work on marine species because it is hard to follow individuals from spawning ground to nursery ground to feeding ground.

1.4.4 Theories and tests of population structure

The member-vagrant hypothesis is a major theoretical framework for studying the spatial pattern of migrating marine species (Iles & Sinclair, 1982; Sinclair, 1988). It postulates that the population structure of a species evolves primarily as a consequence of *(i)* the selective forces that maximise the probability of encounter among sexually mature individuals and *(ii)* the survival of early life history stages (Bernatchez & Martin, 1996). One of the predictions of the hypothesis is that the number of populations of a species will be defined by the number of environmental settings that allow closure of the triangle of migration. A further implication is that where the spawning sites and larval retention grounds are different, the population structure should reflect the retention zones rather than the spawning sites.

Much of the evidence for the member-vagrant hypothesis comes from the correlation of retention zones and exclusive characteristics of distinct fish groups such as the example of the North Sea herring (Iles & Sinclair, 1982). However the characteristics used (reviewed by Sinclair, 1988) were most often examples of phenotypic variation that could have been caused by local environmental differences, and not necessarily because

of genetic isolation between the populations.

Over the last few years however there have been some attempts to try and test the member-vagrant hypothesis with the use of molecular markers. The advantage of such markers is that they offer the opportunity to distinguish populations on a direct measure of genetic isolation, reflecting the reproductive ecology and behaviour of the populations in question (see section 1.6). Bernatchez & Martin (1996) analysed mitochondrial DNA variation in the anadromous rainbow smelt, *Osmerus mordax*, in the Gulf of St. Lawrence and St Lawrence estuary, Canada, with a view to testing the prediction that the number of populations will reflect the distribution of zones of retention for the larval stages. Significant genetic subdivision was found and results generally supported the member-vagrant hypothesis. However, the authors found that some of the results did not agree with the hypothesis, as there were too many populations in some zones where the physical environment appeared to provide a single retention zone. To explain such anomalies the authors postulated the likely cause to be the effects of the last glaciation, explaining that it seemed unlikely that the two populations sharing the one retention derived from a common ancestor. Instead, the ancestors of the two populations evolved in separate glacial refugia and developed some kind of biological barrier to gene flow so that when recolonisation of the contemporary species range occurred the two populations retained their genetic exclusiveness.

A similar study (Lu *et al.*, 1997) of four carp species in the Yangtze River, China, sampled juvenile fish from three nursery areas located downstream of the spawning sites. Results showed distinctive genetic variance between the nursery areas. This result is of particular interest because a similar study on the same species in the Yangtze River found no significant population subdivision for any of the carp species (Zhao & Li 1995, in Lu *et al.* 1997). It is postulated by the authors of the more recent paper that multiple stocks were undetected because of an insufficient level of resolution of the marker used (allozymes).

Both these genetic studies of the member-vagrant hypothesis came from freshwater examples, where lakes, rivers and estuaries allow relatively well defined larval retention

areas and nursery grounds. No work has addressed directly the member-vagrant hypothesis in marine environments. However, there are some interesting studies suggesting that marine populations could behave according to the member-vagrant hypothesis.

Firstly, studies of the cod population complex with the use of microsatellite markers off Newfoundland indicated that within the complex there are multiple, genetically distinct populations (Bentzen *et al.*, 1996). This despite the fact that adults of the distinct overwintering populations intermingle during the summer feeding periods. Particular cohorts of larvae from distinct spawning events of cod off Newfoundland could be followed in a water mass. The aggregation of larvae, consisting of cohorts of larvae from the distinct spawning events, was not found to be significantly heterogenous as judged by microsatellite data, indicating that although the spawning events were distinct, they originated from subsets of a single population. Slight differences in the time of spawning is hypothesised by the authors to lead to variation in reproductive success of the subsets of the spawning population. This was linked, by the authors, to the idea of match-mismatch proposed by Cushing (1972), where many individuals fail to contribute to recruitment, and which would lead to a smaller effective population size (N_e) and possible genetic drift.

Like the example of the carp above (Lu *et al.*, 1997), these results in favour of genetically distinct populations were found after other studies had failed to find any significance population structure because of the lack of resolution power of the markers used (once again, allozymes). Again it emphasises the importance of *(i)* using a marker with suitable resolution for intra-species population studies, and *(ii)* a sample design that reflects testable hypotheses based on current understanding of spawning sites and nursery areas.

A second case study worth mentioning is Hedgecock *et al.* (1994). The Northern anchovy, *Engraulis mordax*, has always been managed as a single unit stock, with a single spawning area for the entire population. However, allele frequencies revealed significant genetic heterogeneity falsifying the hypothesis that the central stock is a randomly mating population. However the heterogeneity appeared to be geographically chaotic, giving

no indication of spatially distinct panmictic units. The loci that contributed to heterogeneity differed from year to year. Most of the heterogeneity found however was between sample stations, as opposed to within sample stations, which the authors suggested was because of one or more of the following reasons: *(i)* life-long fidelity to schools, *(ii)* assortative movements and grouping, or *(iii)* homing to natal spawning grounds.

Finally, anchovies in both the Atlantic and Pacific show reproductive behaviour that appears to generate genetic heterogeneity. The southwest Atlantic anchovy (*Engraulis anchorita*) spawn in winter and summer. In winter they spawn in the coastal bight, south and downstream of the upwelling off Cabo Frio, depending on the upwelling at the shelf break. Further south, in summer, the anchovies spawn in the region of productive tidal fronts (Cushing, 1995). It would be interesting to test here the hypothesis that different groups of spawners take advantage of the three productive seasons, leading to genetic isolation and therefore population subdivision.

1.4.5 Natal homing

Much of the above discussion on the member-vagrant hypothesis has touched on the question of natal homing, whereby the adults of a species maintain fidelity to the spawning grounds of their parents. Through such behaviour genetic subdivision of populations could occur even if the feeding grounds are shared between offspring from many different spawning sites or nursery grounds.

Cury (1994) introduced the "eternal retour" (perpetual return) hypothesis to reproductive behaviour, whereby natal homing is not peculiar to turtles or salmon, but is a more common mechanism regulating the spatial dynamics of reproduction. The individual is at the heart of the eternal retour hypothesis. Three components are required for successful reproductive behaviour; they are (from Cury, 1994):

1. A conservative mechanism: imprinting or early memorisation of environmental clues allows the fixation within a generation of environmental reproductive possibilities. What has been possible during many generations at an evolutionary scale or what is newly experienced in the environment (by reference to the existing

pattern) persists.

2. Obstinacy at an individual level: the strength of imprinting or other early life memorisation allows the maintenance of the environmental possibilities for many generations. The weakness is it also ensures the ignorance of contemporaneous changes (as imprinting is not altered by any behavioural or environmental information).

3. A small percentage of strays: few individuals within a population stray or a low rate of error at an individual level is necessary to experiment with new, possible solutions that are generated by environmental changes.

It is generally acknowledged that the juvenile salmon imprints on the unique chemical odour of its natal stream and uses such environmental cues to return to the stream during a spawning migration. The fact that juvenile salmon do not leave their river until after spending a few years developing in the nursery regions allows the individual to gain a strong imprint. However, if marine fish return to their (marine or estuarine) native spawning grounds the evidence seems strong that they may have to use another mechanism. Often the nursery ground may not be the same as the spawning ground, and because of the pelagic egg phase the larvae often hatch at some distance from where the adult fish spawn. How would imprinting happen in such a situation? Even if the larvae hatch early, within the spawning site, there is some doubt as to whether the little brain of the fish larvae is competent to acquire an imprint (Cushing, 1995).

However, following work on salmon, the sensitive period appears to be very short, within the life time of a larva. An imprinting period of 36-48 hours has been sufficient to ensure a high degree of homing success in the coho salmon (*Oncorhynchus kisutch*) (Sandercock, 1991) and might occur as early as hatching in the case of the pink salmon (*O. gorbascha*) (Heard, 1991). That imprinting seems to be the obstinate feature of the eternal retour is exemplified in the sardine (*Sardina pilchardus*) off Spain. There is acknowledged to be two spawning populations; one population spawns, as expected, near the coastline, while the second group breeds further out at sea. It was found that the second group is actually spawning at an area where the ancient coastline used to be,

and have apparently retained the location of this spawning ground through successive generations of natal homing behaviour (Chikhi, 1995).

1.4.6 Social facilitation and metapopulations

However natal homing is not the only mechanism whereby newly mature individuals can successfully find a spawning ground. The "social facilitation" hypothesis (Owens *et al.*, 1982) has also been proposed whereby first-time breeders follow experienced breeders to a spawning site and, having had a favourable experience, fix on that site for future spawning. Such a model of migration was put forward with regard to the long travels of marine turtles, but could apply to any migrating species that reproduces more than once in its life. However, at least with regard to the green turtle, studies of the population genetic structure argued against social facilitation, as significant structuring was found (Meylan *et al.*, 1990; Bowen *et al.*, 1992).

The social facilitation hypothesis argues that while adult fish display loyalty to a particular time and space of spawning, such loyalty comes not from their own juvenile experiences (natal homing) but rather from the adult stocks that they are adopted by (repeat homing) (McQuinn, 1997). Not all individual fish develop at the same rate as their contemporaries, either because of genetics, the environment, or chance. Therefore, because maturity is thought to be more dependent on length than on age (Toreson, 1986), individuals will leave nursery areas to join the adult stock at different times, and consequently become "adopted" by a stock that spawns at a particular time and place. The main difference between the two competing hypotheses is that natal homing is seen to be innate and genetically based, while repeat homing is learnt over a lifetime from the elders of the stock.

Evidence for the social facilitation hypothesis comes from the observation that some species show increased homing precision with the number of spawnings (e.g. the walleye, *Stizostedion vitreum* (Olson *et al.*, 1978)) and that fish have the ability to learn spatial patterns such as shapes and levels of brightness. The Atlantic herring (*Clupea harengus*) has been used as evidence both for natal homing and social facilitation. I mentioned

above how Iles & Sinclair (1982) commented on the similarity between environmental characteristics and the morphometric features of the populations of fish found there. However, environmental heterogeneity could explain such features, with migration being due to social facilitation rather than natal homing. Like natal homing, much of the strength of the argument for social facilitation comes from the lack of hard evidence to support its rival.

To try and come to terms with what happens to the vagrants, and whether they should be seen as either peripheral to population dynamics or of central importance in the fate of species and populations of species, the metapopulation concept has been used to create an umbrella over competing theories (McQuinn, 1997). It postulates that the population structure of many species can be considered as an array of local populations that are linked by various degrees of gene flow. The metapopulation concept was first defined by Levins (1968) as:

> " ...a population of local populations which were established by colonists, survive for a while, send out migrants, and eventually disappear. The persistence of a species in a region depends on the rate of colonisation successfully balancing the local extinction rate."

Within such a concept the vagrants play a central role in the dynamics of the species at the metapopulation level, and indeed become increasingly important with increasing environmental and population instability. It is important to note that populations can be defined in dimensions other than purely spatial. With the extension of Levin's original metapopulation model to include more realistic variations, such as variable local population sizes, less frequent local extinctions and immigration into existing local populations, the model allows for the kinds of population pattern and richness described by Iles & Sinclair (1982).

1.4.7 Shoaling behaviour

Shoals are merely any form of social assembly (analogous to flocks in birds), regardless of the behaviour that created the shoal. When fish are assembled and display synchronized

or polarized swimming behaviour the groups are known as schools. Schools are therefore a specialised subset of shoals (Pitcher & Parish, 1993).

It is often postulated that the keys to understanding shoaling behaviour are predator avoidance and food. Synchronized cooperation confuses predators while food gathering in shoals reflects shifting balances between joining, staying in or leaving a shoal (Pitcher & Parish, 1993). While predator avoidance and feeding are of great importance, there is another important function of shoaling behaviour: migration. As with the metapopulation and social facilitation concept above, younger fish need to learn spawning migration routes from elders. The relationships of individuals within a shoal or a school are of key importance, and individual variation within group assemblages has become an area of increasingly intense research (for a review see Magurran, 1993).

Freshwater fish, such as the three-spined sticklebacks, have often been shown to prefer shoaling with kin (Van Havre & Fitzgerald, 1988). Additionally Quinn & Busack (1985) have shown recognition of siblings in shoaling coho salmon, through the use of chemosensory signals. Consequently, any observed altruistic behaviour could have evolved through kin selection.

Traditionally it has been argued that kinship, and related kin-selected altruistic behaviour, can only occur in freshwater or reef fish where mobility is relatively small and kin groups can assemble. Marine species, with potentially highly mobile, pelagic larvae, are assumed not to be able to create such kin groups and thus unable to develop altruistic behaviour (Pitcher & Parish, 1993). However, with growing evidence of natal homing in marine species, some of which has been reviewed above, this assessment is being reviewed. Signs of genetic affinity within shoals, resulting in shoal loyalty, will help indicate whether kinship plays a part in shoaling marine species, perhaps establishing the conditions for altruism to evolve.

1.5 The three species

Pictures of the three species used in the project are shown on plate 1[2]. This section reviews what work has been done on the species of interest, with particular reference to west African populations. Taxonomic issues are also introduced for *Trachurus* spp. and *Sepia* spp. No genetic studies have been done hitherto on the three species in the Gulf of Guinea.

1.5.1 *Trachurus* spp.

Trachurus spp. is used in this project to described both *T. trachurus* (the Atlantic horse mackerel), found in the Benguela LME and Canary Current LME (and elsewhere), and *T. trecae* (the Cunene horse mackerel), found in the Gulf of Guinea (though also extends north to Mauritania and south to southern Angola) (Anon, 1981). Furthermore, *T. trachurus* off South Africa and Namibia are often designated as a further subspecies, *T. trachurus capensis* (the Cape horse mackerel) (Naish *et al.*, 1991). Samples for this project have come from southwest Africa and UK waters (termed *T. trachurus*) and the Gulf of Guinea (termed *T. trecae*). The validity of their taxonomic distinctiveness is questioned in section 3.5.4. The focus of the work will, naturally, be on *T. trecae* from the Gulf of Guinea.

T. trachurus and *T. trecae* are semi-pelagic schooling species, often forming shoals with *Scomber japonicus* and *Boops boops* (see section 1.3.2). In the tropical Gulf of Guinea *T. trecae* is found in deeper waters, is less abundant and of smaller size than *T. trachurus* in the Canary and Benguela Current systems. It has often been associated with sandy bottom types. It is predominately a cold water species and the sharp thermal profile of the Gulf of Guinea forces it to live below the thermocline, moving to shallower waters during periods of upwelling.

The family Carangidae, of which *Trachurus* spp. are members, is an important component of the fisheries of the East Central Atlantic Ocean Area. Three species dominate the catch, *T. trachurus*, *T. trecae* and *Decapterus rhonchus* (the false scad),

[2]The pictures were taken by Oddgeir Alvheim on the RV Dr Fridtjorf Nansen cruise described in section 2.2.3.

Original
In
Colour

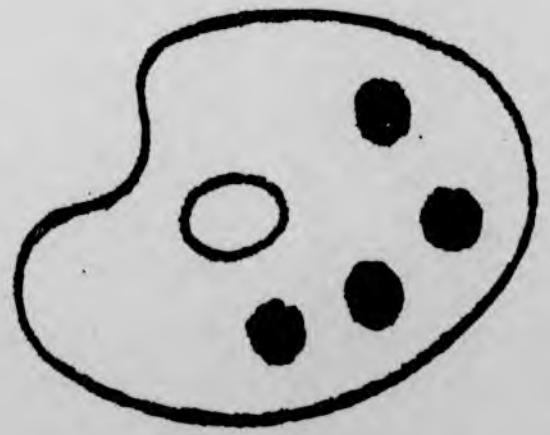

Plate I: The three Gulf of Guinea species studied

I.1 *Trachurus trecae* (Carangidae) - Cunene horse mackerel

I.2 *Pagellus bellottii* (Sparidae) - Red pandora

I.3 *Sepia officinalis* (Sepioidea) - Common cuttlefish

together comprising more than 28% of the total catch from the area (Maxim, 1995). *T. trachurus* catches peaked in the 1970's, and have been falling ever since, though the lack of experimental hauls makes it difficult to make confident predictions of stock health.

Hecht (1990) found that *T. trachurus* attained a maximum age of around 10 years off the southeast coast of South Africa. Fifty percent sexual maturity was attained in the second year, at a total length of around 32 to 33 cm. Horse mackerel first recruited to the fishery in their second year, although fifty percent recruitment only occurs at 5 years of age with a length of around 44 cm. The maximum length of an individual was found to be 49.9 cm. The study furthermore presented evidence for two population-level spawning peaks (i.e. spawning behaviour in the *population* peaks twice; not necessarily twice for individuals within the population) during the reproductive season, extending from June to November. Hecht supported the hypothesis that the adult Cape horse mackerel undertake a spawning migration and that the fish on South Africa's south-east and west coasts, and those off southern Namibia, are a single stock.

Kerstan (1995) found mean annual increments of 15, 7, 5 and 4 cm during the first four years of life of *T. trachurus* from the Agulhas Bank, off southern South Africa, resembling the growth rates of Namibian specimens. Lengths-at-age in most age groups increased with increasing depth and were inversely related to bottom temperature. Bimodal length frequency distributions in several of the age-groups point to two spawning units, or reproductive seasons. Karlouriga & Economidis (1997) found that *T. trachurus* in the Gulf of Saronikos (Greece) had continual growth when juveniles and seasonal growth as adults.

Juveniles and adults of the species feed on a wide variety of pelagic and benthic fishes, crustaceans and squids. Hecht (1990) found that along the southeast coast of South Africa the cape horse mackerel filter feeds on zooplankton. All species of *Trachurus* have a pelagic egg phase (Whitehead *et al.*, 1986).

Little population genetic research has been focused on *Trachurus* spp. worldwide. No work has been done on the west African populations (*T. trecae*), though there have been a few studies on the south and southwest African populations (see Hecht, 1990;

Naish *et al.*, 1991; Kerstan, 1995).

1.5.2 *Pagellus bellottii*

Pagellus bellottii (the red pandora) is distributed from the Straits of Gibraltar to Angola, and around the Canary Islands. It is also found in the southwestern Mediterranean (Whitehead *et al.*, 1986). It is a demersal species inhabiting hard as well as sandy bottoms to depths of about 250 m. It is found in schools, especially in the upper 100 m. During upwelling it is able to move inshore with the cold water, up to a depth of about 20 metres. However during times of thermal stability, with a sharp thermocline, it is restricted to the cool subthermocline waters, 40 metres and below. There is migration towards the coast for spawning behaviour, and this happens intermittently from the second year onwards between May and December according to the latitude (off Tema, Ghana, the peak spawning period was found to be September (Rijavec, 1973)). The age when 50% of *P. bellottii* first come to spawn is about 14 months.

P. bellottii is a protogynic hermaphrodite; i.e. the gender change is from female to male. One possible explanation for this sequential hermaphroditic behaviour is that only the larger fish can successfully be males because of the necessity to protect a harem of females (Levinton, 1995). This model of social behaviour would necessitate the presence of a shoal of females that could be protected by one or a few males, with the change from female to male happening to the strongest female when a male position becomes available.

Rijavec (1973) found the size range of *P. bellottii* off Ghana to be from 6 to 27 cm. The combined length-frequency curves were mostly unimodal or bimodal, and only occasionally polymodal. The length-frequency curves of individual samples were almost always unimodal, which Rijavec suggested might indicate that the shoals were grouped by age. Rijavec's results also indicated that there may be a size distribution by depth, whereby the highest values of mean size were always (except during spawning) around a depth of 40 metres. This depth was also where the sex ratio was found to be constantly in favour of males, possibly arguing against the harem model described above.

There are believed to be three main stocks in west and northwest African waters: Rio de Oro (Western Sahara), Dakar (Senegal) and Takoradi (Ghana) (Koranteng & Pitcher, 1987). Koranteng & Pitcher (1987) showed, by the use of mixture analysis of length frequency data, that two cohorts of *P. bellottii* enter the fishery off Ghana each year. The cohorts could be traced successfully for the first three years of age. Cohorts spawned before the major coastal upwelling in June had faster growth and higher natural mortality than the cohorts spawned in the minor January upwelling.

P. bellottii appear to be eurythermic and euryhaline (i.e. tolerant of a range of temperature and salinity levels) as sexually mature specimens off Dakar (Senegal) and off Takoradi (Ghana) have been found in waters with temperatures from 18-30°C and salinity from 32% to 35.5%. Rijavec (1973) found mature individuals in November in Cameroon and in May in Gabon and Congo. At least moderate sexual activity was shown throughout the year.

The diet of *P. bellottii* is omnivorous, with predominately carnivorous tendencies. Food includes crustaceans, cephalopods, small fish, amphioxus and worms. Rijavec (1973) found that condition factor, C (classically given by $C = W/L^3$ where W and L is weight and length respectively) for *P. bellottii* was constantly highest in the months after the upwelling season, because of the increased food production associated with the upwelling.

No molecular genetic work has been done on *P. bellottii*.

1.5.3 *Sepia* spp.

Sepia spp. is used to describe both *Sepia officinalis officinalis* (the common cuttlefish) and *S. o. hierredda* (the subspecies of the common cuttlefish found from Cape Blanc (21°N) to South Africa). The two subspecies appear to overlap between Cape Verde and Cape Blanc (Anon, 1981). Samples for this project were taken from west Africa, though data sets from samples taken from the Atlantic and Mediterranean coastlines of Spain were also used. The taxonomy of the two subspecies of *S. officinalis* is not settled and as such the term *Sepia* spp. is generally used in this project to refer to all collected

samples.

The common cuttlefish is most abundant in shallow waters of less than 100m depth, though can be found down to a depth of about 200m, and are generally found over sandy areas of the shelf. The larger specimens are generally caught at greater depth. Seasonal migrations (mainly vertical) have been shown to occur in all studied populations. The stock off Senegal is thought to follow a seasonal north-south and an inshore-offshore migratory pattern.

As with most other species of cephalopods they have a short life span with one spawning event followed most often by death. Spawning occurs all year round with a peak around May to August at depths of generally less than 40m and a temperature of between 13 to 15°C (Anon, 1981). After spawning there is massive mortality of the adults, particularly the females.

Males are usually sexually mature at 12 - 14 cm mantle length, though can mature at mantle lengths of 6 - 8 cm (Roddy Williamson, pers. comm.), and carry up to 1,400 spermatophores. Females are normally mature at 11 - 14 cm mantle length. They lay between 150 to 4,000 eggs in grape like clusters (usually coloured black by a covering of ink deposited by the female on laying) on rocks and shells on the sandy bottom (Anon, 1981). The size of the individual eggs is between 8 to 10 mm in diameter and hatching takes place after 30 to 70 days of incubation, depending on the temperature (generally between 16 to 22°C). The hatchlings are 7 to 8 mm in length.

Food consists of small molluscs, crabs, shrimps, other cuttlefishes, and juvenile demersal fishes (Roper *et al.*, 1984). Cannibalism is common and, because of the short life span and consequent rapid growth, daily feeding rates of 10 to 30% of body weight does not seem unlikely. Predators of the cuttlefish include sharks, sparids and other demersal fishes and cuttlefishes.

Although the cuttlefish is often regarded as an opportunist species, becoming more abundant when its predators decline in abundance (Boyle & Boletzky, 1996), evidence from the southern Benguela system would seem to dispute this simple predator prey relationship. It was found that the abundance of ***Sepia australis*** showed no link with

the variation of its most important predator, the shallow-water Cape hake, *Merluccius capensis*. However, the variations in abundance of *S. australis* and one of its prey species, the stomatopod crustacean *Pterygosquilla armata capensis*, showed simultaneous changes, suggesting that both species respond to the same environmental factors (Lipinski *et al.*, 1992). It is unlikely that these changes reflect predator prey cycles as there was no time lag between the two species.

S. officinalis is most often caught during sunlit hours, on sandy mud bottoms from 10 to 110m depth. The peak abundance in the Cape Blanc area is October to March, while in the Spanish Sahara area the peak abundance is around September to January.

Genetic variability measured by allozymes is often found to be low in cephalopods (Boyle & Boletzky, 1996) and the genus *Sepia* is no exception (Sanjuan *et al.*, 1996). However, recent results with microsatellite loci have found far greater variability (Shaw & Perez-Losada, 1999; Shaw *et al.*, 1999; Perez-Losada *et al.*, 2000).

1.6 Molecular ecology and phylogenetics

Molecular ecology describes a field of study that focuses on questions concerning natural and introduced populations and their environments investigated using molecular approaches. Its theoretical basis is in population genetics, the mathematical study of genetic changes that occur in populations. A basic problem in population genetics is how to determine the frequency of a mutant gene in a population, over time, under the influence of various evolutionary forces such as selection and migration. Also addressed is the question of how long-term genetic variability is maintained in a population. Unlike phenotypic variation, most nuclear variation is posited to have relatively little effect on the fitness of an organism and thus chance is likely to play a significant part in determining the frequencies of the various alleles or mutants (see section 1.6.3). Stochastic effects therefore play a key role in many population genetical theories. Two excellent introductions to molecular ecology are Majerus *et al.* (1996) chapter 9, and Avise (1994) chapters 1 and 2, and for a review of molecular ecological techniques in fisheries see Wirgin & Waldman (1994). Two textbooks that cover in detail some of the concepts

and calculations covered in this project are Hartl & Clark (1997) and Li (1997).

1.6.1 Populations and random mating

Problems of a precise definition of population are because of levels of population subdivision within any group of individuals, most typically seen as a "non-random pattern of individuals within a spatial distribution of organisms" (Li, 1997). Examples of non-random distribution in the marine environment include shoaling or schooling behaviour, and the presence of more than one reproductive season in the same geographic region (to which individuals may remain loyal). Thus subdivision can effectively occur in time (temporal subdivision) as well as across space (spatial subdivision), and can be caused by social behaviour as well as geographic division.

Random mating, i.e. the lack of population subdivision, is a key assumption in nearly all models in population genetics. Deviations from it are often sought to illustrate the presence of separate populations or sub-populations, and thus it often acts as a null hypothesis against which to test other hypotheses. The theory of random mating dictates that the chance of an organism mating with another having a particular genotype is equal to the frequency of that particular genotype in the population (Li, 1997).

Although the random mating model appears superficial, the process is not always simple or trivial. Firstly, random mating may be realised for one trait but not for another. For example in human populations random mating may be the norm with respect to blood groups but not to skin colour. A second feature is that a population split into smaller subpopulations may have random mating within all the subpopulations but not satisfy random mating over the entire population.

1.6.2 Hardy-Weinberg principle

Under random mating, and with a number of other assumptions, it is possible to deduce genotypic frequencies from one generation to the next. The assumptions that need to be made are:

- The organism is diploid, reproduction is sexual, and generations are non-overlapping.

- The gene under consideration has two alleles, and the allele frequencies are identical in males and females.
- Mating is random and the population size is very large (in theory, infinite).
- Migration and mutation is negligible (and can effectively be ignored) and natural selection does not affect the alleles under consideration.

The Hardy-Weinberg model (named after the mathematician G. H. Hardy and physiologist W. Weinberg, who independently formulated the model in 1908) states that the mathematical relation between the allele frequencies and the genotype frequencies is given by:

$$AA : p^2 \quad Aa : 2pq \quad aa : q^2$$

in which p^2, $2pq$, and q^2 are the frequencies of genotypes AA, Aa, and aa in zygotes of any generation, p and q are the allele frequencies of A and a in gametes of the previous generation, and $p + q = 1$ (Hartl & Clark, 1997).

The Hardy-Weinberg principle can be used as a reference model in which there are no evolutionary forces at work other than those imposed by the process of reproduction itself. Significant deviations from Hardy-Weinberg expectations (of allele frequencies) indicate that one or more of the assumptions listed above does not apply to the tested sample. However, even if Hardy-Weinberg expectations are met it does not necessarily mean that all assumptions apply to the sample (e.g. see Li, 1988, for an example of how non-random mating can still give rise to Hardy-Weinberg expectations). Random mating is therefore a sufficient condition, not a necessary one, for the attainment of Hardy-Weinberg expectations. Lessios (1992) also highlights the fact that non-significant deviation from Hardy-Weinberg expectations does not necessarily mean that the assumptions are met, given the problems of sampling error. Significant deviations, however, are normally informative.

A related measure to the Hardy-Weinberg model is the Inbreeding co-efficient, denoted F_{IS}. It is the proportionate reduction in heterozygosity, due to inbreeding, relative to random mating (Hartl & Clark, 1997). There is a close relationship between F_{IS} and

the hierarchical F statistics introduced in section 2.8.3 (see Hartl & Clark, 1997, for details).

1.6.3 Neutral mutations

The neutral theory runs to the core of molecular ecology and population genetics and can be a highly divisive issue. Avise (1994), chapter 2, is an excellent introduction to the concept of the neutral theory and its history.

Effectively the neutral theory (or the theory of selective neutrality) maintains that many genetic polymorphisms result from selectively neutral alleles maintained by a balance between the effects of mutation and random genetic drift (Kimura, 1968; King & Jukes, 1969). Mutation introduces new alleles into the population, while drift determines fixation or loss (with loss being the usual outcome because of low initial frequencies). At equilibrium there is a balance between mutation and drift so that, on average, each new allele gained by random mutation is balanced by one being lost (or more rarely, fixed).

The key feature of the theory is the underlying principle that most mutations have such a minor, if any, effect on the phenotype, that they have little or no influence on fitness. The frequencies of the neutral alleles are therefore not determined by natural selection. Such a conclusion lead Kimura (1968) to note that " ... we must recognise the great importance of random genetic drift ... in forming the genetic structure of biological populations." Thus neutral alleles, while unsuitable for the study of genetic adaptation, are ideal for mapping the geographical structure and history of natural populations, and the ancestral relationships of species (for further views on issues such as the "nearly neutral theory" see the debate between Ohta 1996 and Kreitman 1996, while Rand *et al.* 1994, Ballard & Kreitman 1995 and Zouros *et al.* 1995 all urge due caution with the assumption of selective neutrality).

1.6.4 Infinite alleles and stepwise mutation model

A variety of mathematical models for studying genetic variability have been developed. A commonly used model is the infinite alleles model (Kimura & Crow, 1964). The model

assumes that every mutation creates a new and unique allele that does not currently exist in the population. Assuming selective neutrality, the infinite alleles model (IAM) predicts the expected heterozygozity as:

$$h = \frac{4N_e u}{1 + 4N_e u}$$

Where N_e is the effective population size and u the mutation rate per gene per generation. Factors that can lead to deviations from the expected heterozygozity include selection and migration. For example, overdominant selection can increase measured heterozygozity while purifying selection can reduce it. F_{ST} pairwise differences (see section 2.8.2) are based on the infinite alleles model.

Although such a model can be argued to be somewhat representative of nucleotide substitution, questions are raised over its suitability as a model for analysing microsatellite data, where the allele may change in a stepwise manner so that recurrent and backward mutation can occur. In this case, a mutation may not result in a new allele not already present in the population. To make the model more realistic Ohta & Kimura (1973) suggested a stepwise mutation model whereby alternative alleles are viewed as integers on a line. Mutation can cause the allele to move one step either to the right or the left (though the model can be extended to include the possibility of two step changes, such as the two-phase model; DiRienzo *et al.* 1994). Under this stepwise mutation model (SMM) the expected heterozygozity at equilibrium is now given as:

$$h = 1 - 1/\sqrt{1 + 8N_e u}$$

The expected value of heterozygozity under the SMM can be substantially lower than that expected with the infinite allele model. The stepwise mutation model is most often used with microsatellite data. Measurements derived from the SMM model include R_{ST} and $(\delta\mu)^2$ (see sections 2.8.3 and 2.8.4 respectively).

1.6.5 Infinite sites model

The infinite sites model, developed by Kimura (1969, 1971) and Watterson (1975) is an alternative to the infinite alleles model for nucleotide sequence data. Rather than treating each mutation as generating a new allele, the allele is considered as a sequence of nucleotides with mutation altering a site in the sequence. It assumes two features of the data:

1. The mutation rate is sufficiently low such that most sites will be monomorphic and that those sites that are polymorphic will be segregating for just two nucleotides. (Much of the available data on nucleotide variation seems consistent with this.)

2. The sequence is sufficiently long such that the frequency of polymorphic sites is low and that most of the mutations that do occur will be at sites that were previously monomorphic.

By consideration of the patterns of similarities across alleles considerable information about the history and relationships of alleles can be gained. The infinite alleles model, on the other hand, ignores the patterns of similarities across alleles, simply considering them as distinct, and as such much of the available information is lost. The infinite sites model is also appealing because it directly addresses the type of genetic data frequently gathered by molecular population geneticists (i.e. sequence data).

1.6.6 Sequence alignment

Before you can compare two sequences, they first have to be aligned. If the sequences were not aligned it wouldn't be possible to establish which regions of the two sequences are homologous, that is, correspond to the same region of the gene and hence may be compared. In this thesis, however, there are more than two sequences to be aligned. There are various ways of performing such a multiple sequence alignment, such as finding the alignment that minimises the total cost of all the pairwise alignments, or to create a guide tree (Page & Holmes, 1998). Given the high similarity and equal length of all the sequences used herein, alignment of the sequences is not difficult and could easily be

done by hand. However, Clustal *W* was used, with default options, for all sequences and the interested reader should refer to Thompson *et al.* (1994) for details on the multiple sequence alignment algorithm used.

1.6.7 Phylogenetic trees

Population genetic data raise the possibility of reconstructing the path of evolution and *(i)* illustrate the overall genealogical ties between populations within species (i.e. coalescence analysis) and between species themselves, and *(ii)* estimate the time since two species shared a common ancestor.

Such analyses derive phylogenetic trees which are simply graphs composed of nodes and branches, in which one branch connects two adjacent nodes. The nodes represent the taxonomic units, be they species or populations within a species, while the branches represent the relationships between them, in terms of descent and ancestry. We can further distinguish between internal and external nodes. External nodes are those which are found at the edge of the tree (also called terminal nodes), while nodes found nested within the tree are internal nodes (Li, 1997). In phylogenetic studies the external nodes represent the extant taxonomic units under study and are referred to as operational taxonomic units (OTUs).

A phylogenetic tree can be rooted or unrooted. The former defines the evolutionary path, while the latter does not make assumptions or require information about common ancestors. Most methods create unrooted trees and as such it is common to include an outgroup that is known to have branched off earlier than the other taxonomic nodes in the study. The root is then placed between the outgroup and the node connecting it to other groups under study.

The number of trees feasibly generated from just 10 OTUs is more than 2 million unrooted trees and nearly 35 million rooted trees (Li, 1997). Since only one of these trees gives the correct (i.e. historically true) evolutionary relationship between OTUs statistical inference from results remains an unresolved problem.

1.6.8 Gene trees and species trees

The trees discussed hitherto are species trees. They represent the evolutionary pathways of a group of species or populations within a species. However, when the tree is constructed with data from a single gene across all taxonomic units the inferred tree is a gene tree (Nei, 1987). There are two key differences between gene trees and species trees.

1. The divergence of two genes sampled may predate the divergence between the two species, thus overestimating the divergence time between the two taxonomic units. Such a problem is particularly relevant to closely related species.

2. The branching pattern of the gene tree may be different from the pattern of relationships between the species. Such an error is particularly likely when the divergence time between the most closely related species is short.

The only way to avoid the second error is to use more than one gene in the reconstruction of a phylogeny. A large amount of data is needed to be reasonably confident that the gene trees reflect the species tree, particularly when taking into account the stochastic nature of nucleotide mutation that may also lead to erroneous phylogenies being inferred.

1.6.9 Tree reconstruction

Many methods of tree reconstruction have been proposed because not all work well under all conditions. The two tree building methods in this project are unweighted pair group method with arithmetic mean (UPGMA) (Nei, 1975) and Fitch (1977). The UPGMA is an example of a distance matrix method, while Fitch is a maximum parsimony method (see Li 1997 for a review of these and other methods, including algorithms based on maximum likelihood - maximum likelihood was not used in this project because of computational complexity and the need for an explicit model of evolution. The maximum likelihood technique may be considered more useful as a tool for testing models of molecular evolution, than as an all-purpose method of phylogenetic inference; Page &

Holmes 1998). The methodology of each tree reconstruction technique is given in section 2.10.

UPGMA is the simplest method for building tree reconstruction and is based on a method for developing taxonomic phenograms (i.e. trees constructed from differences and similarities between phenotypic forms). However it can be used for phylogenetic trees when the rates of evolution between the various nodes is approximately constant, giving a linear relationship between genetic distance and time of divergence. It employs a simple sequential clustering algorithm and is a commonly used tree reconstruction method. However, the approach is based purely on similarity between current taxa. No attempt is made to recreate evolution that led to present day patterns of similarity.

UPGMA has been chosen to illustrate relationships between samples because *(i)* it follows a simple clustering algorithm that can be easily understood and *(ii)* the aim of the analysis is simply to show the relationship between samples, and not to infer parameters such as time since separation. An important point to consider is that any tree represents an extrapolation from the data (Smouse, 1998). Hence the data, not the tree, is of key importance and as such is given in tables of pairwise comparisons.

The Fitch method uses an algorithm that searches for a tree that requires the smallest number of evolutionary changes to explain differences among the OTUs under study. Such a tree is termed a maximum parsimony tree. Often there are a number of trees that can be created using the minimum number of changes, and thus no unique tree can be found. Fitch (1977) details how choices between competing maximum parsimony trees are made. Note that with the Fitch method, unlike UPGMA, OTUs can be internal as well as external.

1.6.10 Molecular variance parsimony (Minimum spanning trees)

Traditional tree building methods share two common features: *(i)* they concentrate on molecular information, while discarding information on its population frequency and geographic location, and *(ii)* the object of the exercise is to arrive at one tree that is optimal (Excoffier & Smouse, 1994). Excoffier & Smouse (1994) note that not only is

more than one optimal tree found, but that also information on relative frequencies and locations of the haplotypes can be used to aid the construction of a haplotypic tree.

In formulating a methodology for creating trees using molecular variance parsimony Excoffier & Smouse (1994) show a direct relationship between the age of a haplotype and its frequency in a population, with high frequency haplotypes being more likely to have been present in the population for a longer time. Also, as the vast majority of new haplotypes are likely to be derived from the most common haplotypes it is anticipated that rarer variants, generally representing more recent mutations, are most likely to be more closely related to the high frequency, common haplotypes, than to other rarer variants. Using a ranking system of competing trees, the molecular variance parsimony method aims to create optimal trees based on both molecular and population genetic information.

1.7 Mitochondrial DNA and molecular techniques

1.7.1 Arrangement of Mitochondrial DNA

Mitochondrial DNA (mtDNA) is usually maternally transmitted (though see Hoeh *et al.* 1991 for evidence of biparental inheritance of mtDNA in *Mytilus*) and haploid. Absence of recombination (but see Hey, 2000, for a review of evidence of recombination in human mtDNA) means that regions of the mtDNA are very well preserved across taxonomic boundaries and as such "universal" markers are often found that can amplify the same region of mtDNA across a wide range of species (e.g. see Kocher *et al.*, 1989). Figure 1.3 shows a diagram of the mitochondrial genome. Three regions of the mitochondrial genome are analysed in this project, they are: *(i)* the control region, *(ii)* the cytochrome *b* gene and *(iii)* the 12S rRNA gene.

The Control Region

The control region, situated between the tRNA(*Pro*) and tRNA(*Phe*) genes, is the major noncoding region of the mitochondrial genome. The region includes transcriptional promoters for both strands, the heavy strand replication origin and the displacement

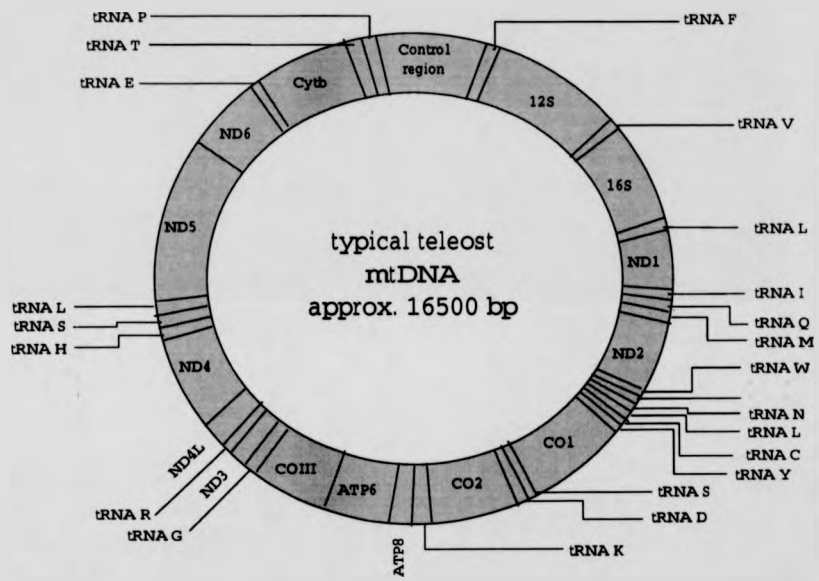

Figure 1.3: The mitochondrial genome is remarkably conserved among a wide range of varied taxa, with only the occasional minor rearrangement found, most often due to translocation of tRNA genes. The tRNA genes are identified here with the single letter amino acid code according to the coding strand. tRNA F, found between the control region and the 12S gene, is referred to throughout the thesis with its fuller name tRNA(*Phe*). A copy of the mitochondrial translation code can be found in the appendices (Table B.1).

or D-loop region. Because of the reduced functional constraints some portions of the control region can evolve at a far higher rate than average for the mitochondrial genome, particularly those regions adjacent to the flanking tRNAs (Palumbi, 1996).

Cytochrome *b*

Cytochrome *b* is a protein in the electron transfer chain, and is the only protein product of the mitochondrial genome that is a fully functional monomer; i.e. it is not a subunit of a larger enzyme complex (Palumbi, 1996). The biochemistry of cytochrome *b* is well known and has been extensively studied in mammals and as such a number of universal primers have been developed for inter and intra-species research (e.g. see Kocher *et al.*, 1989; Cantatore *et al.*, 1994).

12S rRNA

12S rRNA is the gene for the small subunit ribosomal RNA in mitochondria. It is fairly conserved among taxa but there are regions of high sequence substitution. Overall, it mutates at roughly the average rate of the mitochondrial genome (Palumbi, 1996).

1.7.2 The Polymerase Chain Reaction

The polymerase chain reaction (PCR) has become a mainstay of population genetics and molecular ecology. It allows defined segments of isolated DNA to be amplified to microgram quantities from as little as a single template molecule, and for most applications it is a fast, relatively inexpensive and generally simple way to generate ample material for further analyses.

1.7.3 Restriction Fragment Length Polymorphisms

Restriction Fragment Length Polymorphism (RFLP) is a generic technique by which restriction enzymes are used to cut the DNA sequence at enzyme specific sites, leading to different sized fragments that can be viewed on an electrophoretic gel. Variation in the DNA sequence leads to sites being variously present or absent between different

individuals in a population, conferring differing band patterns on the gel. The patterns of the bands therefore allow inference of sequence variation between individuals and populations (Majerus *et al.*, 1996).

1.7.4 Sequencing

Sequencing studies are becoming more common as the cost of sequencing falls. Other methods of analysis, such as RFLPs, can be seen as imperfect attempts to gain access to sequence information, and as such can never provide as much information as direct sequence analysis of the DNA.

1.7.5 Microsatellites

Microsatellites comprise of tandemly repeated strings of short elements, usually 2, 3 or 4 base pairs long, scattered throughout the nuclear and mitochondrial genome of all eukaryotes (Majerus *et al.*, 1996). They are highly polymorphic, showing multiple alleles of various lengths at most loci. This is due to a process termed molecular slippage whereby the two strands of DNA become misaligned during replication and repeat units are added or lost, resulting in alleles of differing lengths (Li, 1997). They are thought to be usually selectively neutral (Majerus *et al.*, 1996), however there are a number of cases where microsatellites are hypothesized to cause autism and schizophrenia in humans (Fischer, 1998) . The mutation mechanism (slippage) is not fully understood, and thus analytical methods are still being developed using a variety of assumptions.

Chapter 2

Materials and Methods

2.1 Preamble

This chapter gives details of sample collection and preservation. Laboratory methods are described that are common to the genetic analyses of all three species. Also included are sections on more specific laboratory procedures such as sequence analysis (of *Trachurus* spp. and *P. bellottii* samples), restriction fragment length polymorphism (RFLP) (of *Trachurus* spp.), and microsatellite analysis (of *Sepia* spp. samples). Finally, numerical methodologies are provided for the various measures of genetic diversity and distance used in the project, such as analysis of molecular variance (AMOVA) and F_{ST}, and methods of phylogenetic tree reconstruction are described.

2.2 Sample collection

Most samples were collected by the author on a research cruise that covered the region from the Benin/Nigerian border to the Liberia/Côte d'Ivoire border, covering the central Gulf of Guinea LME subsystem. The cruise was undertaken by the Research Vessel Dr. Fridtjof Nansen (NORAD - FAO/UNDP Project GLO 92/013), a Norwegian vessel that specialises in tropical marine research, between the dates of 19 April to 6 May 1999 (hereafter often referred to as the Nansen cruise). The tissue samples collected were preserved in high grade pure ethanol on site and sent to the University of Warwick for

analysis.

Samples from the southwest African coast (Angola, Namibia and South Africa) were collected by Ekkehard Klingelhoeffer of the National Marine Information and Research Centre, Swakopmund, Namibia, during two cruises; the first run by the Nansen programme during June 1997 and the second on the Research Vessel Benefit during September and October 1997. Samples from Guinean waters were collected by Eric Morize of ORSTOM (now the Institut de recherche pour le développement), in the Republic of Guinea, on board the Research Vessel Antea during a cruise from the 25 September to 10 October 1998. Guinean and Namibian samples were preserved in high grade pure ethanol. In addition, a sample of *T. trachurus* specimens were available for analysis from the Plymouth coast, taken by vessels of the Marine Biological Association of the United Kingdom, Plymouth, UK. Samples collected off the coast of Plymouth were preserved in liquid nitrogen. Data from *Sepia officinalis* samples collected in Spanish waters were provided by Paul Shaw of the University of Hull in the UK. Figure 2.1 shows an overview of the sample locations.

Those samples collected by collaborators were collected when the researcher had time to do so during a busy research cruise. The locations visited on the Nansen cruise by the author were sampled in greater detail, though sample sizes were limited by *(i)* the catch available, *(ii)* the speed of collection of the tissue samples, i.e. while still fresh, and *(iii)* the logistics of shipping the samples to the UK.

2.2.1 Sample preservation

It was decided early on in the project that the most practical preservative would be pure (~98%) ethanol. There were alternatives that could be used to preserve the DNA of freshly caught animal tissue, most notably the use of liquid nitrogen to freeze the tissue, or the use of the preservative, BLB.

Using liquid nitrogen to freeze collected samples is necessary when allozyme and protein based techniques are the principal tools used to investigate the genetic variation in a population. Freezing in liquid nitrogen is the principle way to retain protein, as

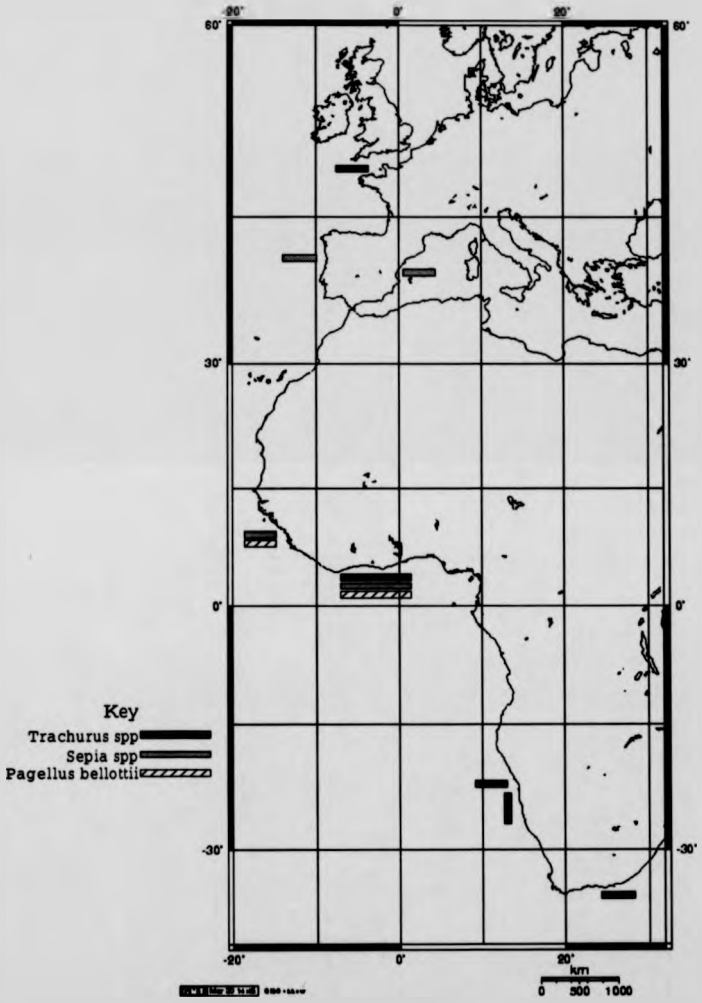

Figure 2.1: An overview of sampling sites for the three species. The key denotes the shading used to indicate the three species. Note that the relative sizes of the markers do not represent the number of individuals collected. More detailed maps of sample sites can be found at the beginning of relevant Results chapters. See text for details of research vessels and dates of sample collection.

well as DNA structure. However the major drawback to the use of liquid nitrogen is the heavy and expensive equipment used to ship a relatively small amount of material, which necessarily greatly increases the transport costs. It was decided that the increased costs and logistical difficulties were not sufficiently compensated by the chance to use allozyme analysis, and consequently a DNA only preservative would be sufficient.

Two DNA preservatives were investigated as to their suitability: *(i)* ~98% ethanol and *(ii)* BLB (which consists of 50mM NaCl, 50mM Tris HCl (pH 8), 100mM EDTA and 1% SDS).

In order to test the preserving characteristics of the two preservatives some samples of *S. officinalis* were preserved in either ethanol or BLB. DNA was then extracted after a period of storage at 4°C of one month and six months using the usual protocol (DNA methodology is described in section 2.3.1) and the amount and quality of DNA measured by mass spectrometry. Both gave good DNA preservation and RNA to DNA ratio. Therefore the choice came down to logistical reasons, and given the general availability of ethanol, and that it can be used neat without earlier preparation and at room temperature, it was decided that it would be a more practical preservative.

2.2.2 Samples collected by collaborators

Samples collected by collaborators were taken from Guinea, Namibia, South Africa, and the UK. It was requested that samples be collected and placed straight into ~98% ethanol, then sent via courier to the University of Warwick. The tissue taken was always a piece of white muscle from along the ventral flank of the fish, or in the case of *Sepia*, the tip of a tentacle was cut off. In the majority of cases this collection of samples resulted in sufficient preservation of DNA for subsequent laboratory extraction.

A general rule of thumb regarding the number of samples taken is that around 50 individuals per site should be enough for molecular genetic analysis of population structure (rather arbitrary and often simply based on what has worked in the past; however with regards to microsatellites see Ruzzante 1998 for a study of sampling variance). For most sites this was possible, and 50 individuals were taken, though for a few sites less

were available. All fieldwork is inherently riddled with unforseen difficulties and it was always emphasised to collaborators that if it were not possible to take 50 individuals from each site then just sample whatever is possible. Therefore numbers of individuals from different collecting stations often differed.

2.2.3 Samples collected on the Dr. Fridtjof Nansen cruise

The cruise of the RV Dr. Fridtjof Nansen surveyed the pelagic and demersal resources of the central subsystem of the Gulf of Guinea LME. The total duration of the cruise lasted from the 19 April to 6 May 1999, starting and finishing in Tema, Ghana, with a brief overnight stay in Abidjan, Côte d'Ivoire. It surveyed the waters of Benin, Togo, Ghana and Côte d'Ivoire with a view to surveying the pelagic and demersal resources.

The survey design consisted of a series of parallel shelf surveys during daylight hours. These were taken 24 nautical miles apart, from approximately 20 metres depth to 100 metres. Occasionally a few bottom trawls were made at depths deeper than 100 metres in areas of suitable trawling grounds. These shelf transects allowed comparisons of samples taken not only from along the coast but also at different depths and distances from the shoreline. A few pelagic trawls with a mid-water trawl were taken during dark hours, when acoustic traces showed a potentially interesting catch of pelagic species.

The demersal trawl, with which all sampled fish in this project were caught, was a "Gisund super bottom trawl". It has a 20mm mesh size in the codend with an inner-net of 10mm meshsize. The estimated opening is about 6m. The SCANMAR system was used on all trawl hauls, providing sensors to relay information on the trawl opening, clearance and bottom contact (Anon, 1999).

All three species under investigation were found and sampled during the Nansen cruise. For reasons highlighted above, the number of individual samples from each collecting station often differed. Which trawling stations were to be used as sample stations for tissue collection was often decided by the appearance of the species in question. Detailed maps and descriptions of sample collection from the Nansen cruise are shown in the relevant Results chapters.

All specimens sampled were measured (fork length for the fish and mantle length for *Sepia*) and sexed. Some of the young specimens were not sufficiently sexually developed to allow discrimination of ovaries or testes, and so could not be sexed visually.

2.3 Basic laboratory methods

Once returned to the laboratory the animal tissue was transferred to fresh ~98% high grade ethanol and kept in a 4°C fridge. The DNA was then extracted and PCR used to amplify the region of interest for either sequencing, RFLP or microsatellite analysis. The protocols used for these various stages of analysis are described below.

2.3.1 DNA extraction

The DNA extraction method used here worked for all samples. Because PCR was used to amplify selected parts of the mtDNA, large quantities of high quality DNA was not needed from the isolation protocol described here. Therefore the protocol could be scaled down to allow rapid isolation from a number of samples.

The protocol described is a widely used method of extracting DNA, particularly from animal tissues. It is based on solubilization of the tissue in a detergent buffer, here containing SDS, followed by protease digestion of cellular proteins with proteinase *K*. The following phenol step effectively separates the DNA from other cellular components, especially proteins. The chloroform step removes any traces of phenol from the DNA solution while precipitation of DNA is overnight in an ethanol/sodium hydroxide solution at -20°C.

The actual protocol used is as follows. Total genomic DNA was prepared from 1-2g of muscle tissue. Each sample was digested in 580μl of proteinase *K* solution (20mg ml^{-1}) for at least 2 hours, with occasional mixing, at 55°C. 600μl of phenol/chloroform (1:1) solution was added to the digested DNA/proteinase *K* solution, mixed thoroughly by hand, and centrifuged at 13,000 rpm (in a microcentrifuge with a rotor diameter of 10cm) for 15 minutes. The aqueous layer was removed and transferred to a new tube, into which 600μl of chloroform/isoanylalcohol (24:1) solution was added. Again the

solution was mixed thoroughly and centrifuged at 13,000 rpm, this time for 10 minutes. Carefully the aqueous layer was removed and added to 100μl 3M sodium hydroxide (NaOH) and an excess (at least 1000μl) of cold ~98% ethanol. The solution was left at -20°C overnight to allow precipitation of the DNA.

To form a DNA pellet the solution was centrifuged at 13,000 rpm for 30 minutes. The ethanol supernatant was carefully discarded and the pellet dried by placing the tubes in a 65°C oven until dry. The pellet was then washed with 70% ethanol and finally the DNA was resuspended in 50μl of TE (1 part 1mM EDTA with 10 parts 10mM Tris, made up to pH 8 with conc. HCl) at 65°C for approximately 90 minutes. Such isolated DNA was used as a template in the PCR analyses described in this chapter.

2.3.2 PCR procedure

The protocol described is the core one used to routinely amplify sequences for sequencing and restriction fragment length polymorphisms (RFLPs). However the PCR reaction used for amplification of microsatellites loci differed slightly and is described in section 2.6.1.

The protocol is as follows. Firstly a PCR MasterMix was created: For each individual PCR the MasterMix contained 2.5μl of MgCl (25mM), 5μl of 10x Taq buffer (supplied by GIBCOBRL™ the manufacturers of the Taq polymerase), 2μl of dNTP mixture (consisting of 1 x [dATP, dCTP, dGTP, dTTP - supplied by GIBCOBRL™] and 6 x ultra pure water), and 1μl of each primer. 1μl of the resuspended DNA as described from the above isolation protocol was used as template, and each sample was bought up to 50μl with ultra pure H_2O (the ultra pure water was treated with Millipore MQ™ treatment, in which it is double distilled and treated with ion exchanges and micro-filters to remove bacteria).

The reactions were overlaid with 45-50μl of mineral oil to prevent evaporation during thermal cycling. PCR conditions consisted of an initial step of 94°C for 5 minutes, followed by a holding stage at 82°C. At this point the 0.25μl of Taq polymerase was added (a "hot start"). Once the Taq was added, the thermocycler continued for 35

cycles of denaturation (94°C, 1 min), annealing (50-55°C, 1min) and extension (72°C, 1-3min). The final stage consisted of 10 minutes at 72°C, to ensure that all annealed template was fully polymerized.

2.4 Sequencing

Following a 50μl PCR of the region of interest, 4μl of the amplified PCR product was run on a 0.8% agarose gel. Following a successful result the remaining PCR product was cleaned using microcon centrifugal cleaners, following manufacturers instructions (Microcon™). An alternative method of cleaning the DNA was by adding 0.5 volume of 7.5M ammonium acetate and 1 volume isopropanol to the PCR product (after removing the mineral oil overlay), and inverting several times and incubate at room temperature for 10 minutes. The mixture was then centrifuged at 13,000 rpm for 10 minutes and washed with 70% ethanol and allowed to dry on the bench. The pellet of cleaned DNA was then resuspended in ultra pure water.

Approximately 2ng of the cleaned DNA was used in a sequencing reaction with 3*p*mol of primer. The sequencing product was loaded onto a Perkin Elmer ABI Prism 377 DNA sequencer (or an Applied Biosystems 373A DNA sequencer) and run according to ABI recommendations overnight.

2.4.1 Sequencing cytochrome *b*

A section of the cytochrome *b* gene was sequenced using universal primers described by Kocher *et al.* (1989). The primers amplify a short section of the cytochrome *b* gene from a wide variety of taxa and have revealed intra-population variation for a variety of fish species.

The primers used were *cytb h* and *cytb l* (Kocher *et al.*, 1989; Palumbi, 1996), giving a sequence of 211bp. Their sequences are shown in Table 2.1. Details of the PCR reaction used are described in section 2.3.2; the annealing temperature was 52°C and the extension time, at a temperature of 72°C, was 1 minute.

Chromatogram files for every nucleotide sequence were viewed using the Chromas

cytb h:	ccc ctc aga atg ata ttt gtc ctc a
cytb l:	cca tcc aac atc tca gca tga tga aa

Table 2.1: Primers used for sequence analysis of a section of the cytochrome *b* gene in *Trachurus* spp.

software, and sequences were aligned using Clustal *W* (Thompson *et al.*, 1994) with default options. DnaSP (DNA Sequence Polymorphism) (Rozas & Rozas, 1999) and Arlequin (Schneider *et al.*, 1997) software was used to produce statistics on the population mutation and polymorphism in the sequences, such as: nucleotide diversity, transition to transversion ratio, synonymous to nonsynonymous mutations and disequilibrium between polymorphic sites. Sequence analysis is described in section 2.4.4.

To test the use of the cytochrome *b* nucleotide sequence as a marker for studies of not only populations within a species but also between species, families and other taxonomic orders, sequences from the order Perciformes were extracted from the Blast database of the National Centre for Biotechnology Information website,[1] and aligned with Clustal *W* (with default options) to the cytochrome *b* sequence in this study. A "Consense" tree, generated with the aid of the PHYLIP (Felsenstein, 1989) software package, was drawn with bootstrapping confidence limits placed on all internal nodes to assess the degree of confidence among different taxonomic levels (further details of the method can be found in the appendices, page 196).

2.4.2 Sequencing 12S rRNA

The primers used for analysis of 12S rRNA in *P. bellottii* were *PBD1* and (the universal primer) *12Sar3* (Palumbi, 1996), giving a sequence of 288 base pairs in length. Their sequences are shown in Table 2.2. Details of the PCR reaction used are described in section 2.3.2; the annealing temperature was 50°C and the extension time, at a temperature of 72°C, was 1 minute.

As with cytochrome *b*, chromatogram files for every nucleotide sequence were viewed using the Chromas software, and aligned using Clustal *W*. Likewise, DnaSP and Arlequin software was used to produce statistics on the population mutation and polymorphism

[1] http://ncbi.nlm.nih.gov

PBD 1: aca aag aaa act ccc cac cc
12Sar3: ata gtg ggg tat cta atc cca gtt

Table 2.2: Primers used for sequence analysis of a section of the 12S rRNA gene in *P. bellottii*.

12Sar3 (CB2-3'): ccc ctc aga atg ata ttt gtc ctc a
Dloop 1 (CB1-5'): cca tcc aac atc tca gca tga tga aa
DNO 1: gcc tac caa ccg gtg ata act
DNO 3: tcc ttg ccc gtc gta aat aa
SNO 3: cca gtt ata gtg ggg tat ct
PBD 1: aca aag aaa act ccc cac cc

Table 2.3: Primers used for sequence analysis of a section of the control region, tRNA (*Phe*) and 12S rRNA genes in *T. trecae* and *P. bellottii*. Rough locations of the primers along the sequence can be found in Figures 3.5 on page 77 for *T. trecae* and Figure 4.4 on page 104 for *P. bellottii*

in the sequences. Genetic distances and pairwise fixation indices were calculated also as for the cytochrome *b* gene (see section 2.4.4).

2.4.3 Sequencing of the control region, tRNA (*Phe*) and 12S rRNA

The control region of *T. trecae* and *P. bellottii* was sequenced with a view to studying its potential use as a region for population analysis. The basic PCR reaction conditions were used as in section 2.3.2; the annealing temperature was 50°C, and the extension time, at 72°C, was 2 minutes. The sequence of the control region, 12S rRNA and tRNA (*Phe*) has no reading frame as none of the sequences codes for any protein.

The primers for *P. bellottii* and *T. trecae* are given in Table 2.3. The Primers *12Sar3* (Palumbi, 1996) and *Dloop 1* (Kocher *et al.*, 1989) are universal primers, while all other primers were designed, with the aid of "Primer3" software on the internet[2], from the heavy and light strand until the complete region of interest was sequenced. *DNO 1*, *DNO 3*, and *SNO 3* were primers designed from *T. trecae*, while *PBD 1* was designed from *P. bellottii*.

[2]Steve Rozen, Helen. J. Skaletsky (1998) Primer3.
Code available at: http://www-genome.wi.mit.edu/genome_software/other/primer3.html

2.4.4 Sequence analysis

Chromatogram files for every nucleotide sequence were viewed using the Chromas software[3], and sequences were aligned using Clustal *W* (Thompson *et al.*, 1994) with default options. DnaSP (DNA Sequence Polymorphism) (Rozas & Rozas, 1999) and Arlequin software (Schneider *et al.*, 1997) were used to produce statistics on the population mutation and polymorphism in the sequences, such as: nucleotide diversity, transition to transversion ratio and estimates of θ (the product of the effective population size and mutation rate). Genetic distances between the sampled stations were estimated by the analysis of molecular variance (AMOVA) as described by Excoffier *et al.* (1992). The analysis of molecular variance approach is essentially similar to other approaches based on analysis of variance of gene frequencies, but it takes into account the number of mutations between molecular haplotypes. Pairwise fixation indices (F_{ST}) were also calculated among samples, with a slight transformation to linearize the distance with population divergence time (Reynolds *et al.*, 1983). Both analyses were performed on the Arlequin software package.

2.5 Restriction Fragment Length Polymorphism

The protocol used was as follows. 10μl of PCR product was used as DNA template, along with 2μl of enzyme, 1μl of enzyme specific buffer and 7μl of ultra pure water in a 20μl RFLP reaction. The complete product was run on a 2% agarose gel or a 6% denaturing polyacrylamide gel, depending on the expected size of the fragments. RFLP bands were compared by eye against a 1kb standard ladder.

2.5.1 Restriction Fragment Length Polymorphism of the control region and surrounding tRNA genes

The primers used were *12SAR-3'* and *Dloop*. The annealing temperature of the PCR reaction was 50°C and the extension time, at a temperature of 72°C, was 2 minutes. PCR amplified DNA was digested according to the manufacturers instructions (Strata-

[3]Conor McCarthy, Chromas 1.45. Available at: http://www.technelysicum.com.au/chromas.html

gene™and GIBCOBRL™) with the following 5 restriction enzymes: *Alu*I, *Cfo*I, *Rsa*I, *Hae*III and *Taq*I. A separate digestion reaction was used for each restriction enzyme. Restriction fragments were run out on an ethidium bromide stained 2% agarose gel, or 6% denaturing polyacrylamide gel, and visualized under UV light. Gels were scored by hand.

2.5.2 Restriction fragment analysis

A five letter composite mtDNA genotype, indicating the fragment pattern for each restriction enzyme, was created for each individual. The data was then inputted into the Arlequin software package (Schneider *et al.*, 1997) for statistical analysis.

2.6 Microsatellite DNA analysis

2.6.1 PCR of microsatellite loci

Individuals were screened for variation at a total of seven variable loci (named Sof 1 to Sof 7) previously isolated and characterised for *Sepia officinalis* by Shaw & Perez-Losada (1999). An annealing temperature of 51-61°C for 30 cycles was used for Sof 1, 2 and 7, while a temperature of 55-65°C, also for 30 cycles was used for Sof 3, 4, 5 and 6. The annealing temperature starts at the lower temperature and steps up by 1°C during each cycle of the reaction until reaching the highest temperature. Such a system ensures strong initial binding of primers to template, followed by increasingly specific extension and sequence replication as the temperature rises.

The reaction mixes contained 20ng of template DNA, 1.5-2.5mM MgCl, 0.2mM of each nucleotide, 0.2μM of each primer (forward primer 5' end-labelled with a Cy5 fluorescent dye group), 0.2 U[4] Taq polymerase (Bioline UK™) with the manufacturers supplied 1x NH_4 buffer, in a final reaction volume of 10μl.

[4]Unit (U) definition: One unit incorporates 10nmol of deoxyribonucleotide into acid-precipitable material in 30 minutes at 74 °C.

2.6.2 Microsatellite screening

Amplified products were resolved on a 6% denaturing polyacrylamide gel run on an ALF-express™ (Pharmacia Biotech) automated sequencer. Product sizes were determined by comparison with an internal series of standard size markers using FRAGMENT MANAGER v1.2 software (Pharmacia Biotech).

2.6.3 Microsatellite analysis

Genotypes at all pairs of loci were tested for genotypic disequilibrium and deviations from Hardy Weinberg using exact tests with significance determined by a Markov chain method. Levels of nonrandom association of alleles within samples were also estimated for each locus by calculating F_{IS} (Weir & Cockerham, 1984) with the Genepop package (GENEPOP v3.1d, Raymond & Rousset 1995). F_{IS}, the inbreeding coefficient, is a general measure of inbreeding and quantifies the level of reduction in heterozygozity due to inbreeding within a subpopulation.

Differentiation between samples can be estimated using one of two main alternative models of microsatellite evolution, the infinite alleles model (IAM, Kimura & Crow 1964) or a stepwise mutation model (SMM, Kimura & Ohta 1978) (see section 1.6.4). Two phase models (see section 1.6.4), which are intermediate between IAM and SMM, were not used in this thesis. Given that Sof 3 displays two parallel sets of allelic arrays differing by one base pair (see section 5.3.1) it is likely that a non-SMM would better reflect the data. Therefore non-SMM based analyses (F_{ST}) were used preferentially, though for comparison and because mutational mechanisms are still a matter of argument, SMM statistical analyses (R_{ST} and $(\delta\mu)^2$) were also used.

2.7 Numerical methodology 1: Intra population

2.7.1 Nucleotide Diversity

Sequence data

A convenient measure of polymorphism for sequence data is Π, the average number of nucleotide differences between two sequences randomly chosen from the population.

However Π is not convenient for making comparisons across loci because it is dependent on the length of the sequence, L. To remove this drawback it is necessary to simply standardise Π by L. That is,

$$\pi = \Pi / L$$

Nucleotide diversity, denoted π, is the number of nucleotide differences per site between two randomly chosen sequences in a sample (Li, 1997). Put another way, it is the probability that two randomly chosen homologous nucleotides are different. It is equivalent to the gene diversity at the nucleotide level (i.e. gene diversity divided by the number of base pairs for sequence data, or loci for microsatellite data).

$$\hat{\pi} = \sum_{i=1}^{k} \sum_{j<i} p_i p_j \hat{d}_{ij},$$

where $\hat{d}_{ij}$ is an estimate of the number of mutations having occurred since the divergence of haplotypes i and j, k is the number of haplotypes, and p_i is the frequency of haplotype i. The variance is as given in Tajima (1983).

Microsatellite data

Same as with sequence data except that it is called the average gene diversity over loci, and assumes no recombination and selective neutrality (see section 1.6.3).

RFLP data

Nucleotide diversity with RFLP data is the probability that two randomly chosen homologous nucleotides are different and is equivalent to the gene diversity at the nucleotide level Tajima (1983).

$$\hat{\pi}_n = \frac{\sum_{i=1}^{k} \sum_{j<i} p_i p_j \hat{d}_{ij}}{L}$$

Notably, the assumption of no recombination and selective neutrality still stand.

2.7.2 Theta (θ)

Theta (θ) is an estimation of the product of the effective population size (N_e) and the mutation rate (u). Again, in estimating θ there is an assumption that the sequences used are under no selective constraints, so that the sequence variation is not affected by natural selection (see section 1.6.3). There has been much recent work on ways to improve the estimates of θ (for example Kuhner *et al.* 1995 use a maximum likelihood approach while Fu 1994 has proposed an approximate approach that is computationally less demanding than the likelihood approach).

Estimation via nucleotide diversity (π)

The first estimate of θ used for sequence data is calculated with the use of $\hat{\pi}$, the nucleotide diversity (section 2.7.1). Under the infinite site model (section 1.6.5) and the assumption of random mating (section 1.6.1), Watterson (1975) showed that the infinite site equilibrium relationship is given between the mean number of pairwise differences and θ equals the expectation (E) of $\hat{\pi}$:

$$\theta_{ND} = E(\hat{\pi})$$

The subscript $_{ND}$ denotes nucleotide diversity. An advantage of the method is that it is independent of sample size.

Estimation via segregating sites (S)

The second estimate of θ used in this project is also based on the infinite site equilibrium relationship (Watterson, 1975). θ is taken from the relationship between the number of segregating sites (S), and sample size (n) and for a sample of non-recombining DNA is given as:

$$\theta_{SS} = \frac{S}{a_1}$$

where

$$a_1 = \sum_{i=1}^{n-1} \frac{1}{i}.$$

The subscript $_{SS}$ denotes segregating sites. The variance is given by Tajima (1989a). Watterson's estimator is apparently more statistically efficient than Tajima's estimator by virtue of its smaller variance. However the estimation is affected by sample size.

2.8 Numerical methodology 2: Inter population

2.8.1 Analysis of Molecular Variance (AMOVA)

The genetic structure of populations as tested by analysis of molecular variance (AMOVA) is based on the analysis of variance framework as initially defined by Cockerham (1969, 1973) and extended by others (e.g. see Weir & Cockerham, 1984). The AMOVA is essentially similar to other approaches based on analyses of variance of gene frequencies, but it takes into account the number of mutations between molecular haplotypes (Schneider *et al.*, 1997).

Formally, in the haploid case, it is assumed that the i-th haplotype frequency vector from the j-th population in the k-th group is a linear equation of the form:

$$x_{ijk} = x + a_k + b_{jk} + c_{ijk}$$

where the vector x is the unknown expectation of x_{ijk} , averaged over the whole study. The effects are a for group, b for population and c for haplotypes within a population within a group. They are assumed to be additive, random and independent, and have the associated variance components σ_a^2 , σ_b^2 and σ_c^2 respectively. The total molecular variance (σ^2) is the sum of the associated variance components (i.e. the sum of variances resulting from differences among haplotypes within a population (σ_c^2), the sum of variances resulting from differences among haplotypes among different populations within a group (σ_b^2), and those resulting from differences among the groups of populations (σ_a^2)).

The researcher can define a population hierarchy that will be tested. For example, within this project there is the hierarchy of tests between LMEs, followed by tests of populations sampled within those LMEs. The hierarchical analysis of variance partitions the total variance into components that are intra-individual differences, inter-individual differences and inter-population differences. The variance components are used to calculate the fixation indices as originally defined by Wright (1951, 1965) in terms of inbreeding coefficients.

In the case of a simple hierarchical genetic structure consisting of haploid individuals within populations, the algorithm leads to a fixated index of F_{ST} (section 2.8.2), absolutely identical to the weighted average F statistic over loci defined by Weir & Cockerham (1984). The F-statistic analogs produced by AMOVA are often designated as Φ-statistics.

The significance of the fixation indices are tested using a permutation approach wherein haplotypes, individuals or populations, among individuals, populations or groups of populations are permutated and statistics recomputed to get their null distribution. The pattern of permutations performed depend on the tested statistic and the given hierarchical design. Under this procedure neither the normality assumption nor the assumption of equality of variance, as required under the usual analysis of variance tests, are needed.

2.8.2 F_{ST} pairwise differences

F statistics quantify the inbreeding effect of population substructure. Classically, it is the reduction in heterozygosity expected with random mating at any one level of a population hierarchy relative to another, more inclusive level of the hierarchy. The symbol for a fixation index is F embellished with subscripts denoting the levels of the hierarchy being compared. F_{ST}, the fixation index, measures all effects of the population substructure and is the most used in this project. F statistics are a good measure of short-term genetic distances between populations, and use the assumptions of the infinite alleles model (section 1.6.4). They are calculated through the AMOVA (section 2.8.1). Common pitfalls of all F_{ST} type estimates of genetic differentiation and migration is that they assume *(i)* that subpopulation sizes are identical, and *(ii)* that migration between subpopulations is symmetrical. Also, sampling variance is not usually separated from natural variance in allele frequencies. Coalescence analysis overcomes some of these problems, though is not used in this project because its development is still in its infancy (Page & Holmes, 1998).

Although F_{ST} has a theoretical minimum of 0 (indicating no differentiation) and 1 (indicating fixation of alternative alleles in the compared populations), the observed maximum is usually much less than 1. Wright (1978) suggested the following qualitative guidelines for the interpretation of F_{ST};

- 0 - 0.05 may be considered as indicating little genetic differentiation,
- 0.05 - 0.15 may be considered as indicating moderate genetic differentiation,
- 0.15 - 0.25 may be considered as indicating great genetic differentiation,
- above 0.25 may be considered as indicating very great genetic differentiation.

The null distribution of pairwise F_{ST} values, under the hypothesis of there being no difference between any pair of populations, is obtained by permuting haplotypes between the populations. The p value of the test is simply the proportion of permutations leading to an F_{ST} value larger than or equal to the observed one.

An estimation of the number of migrants between pairs of populations can be made once a calculation of F_{ST} is available. It is assumed that the two populations, of size N, exchange a fraction m of migrants, and that the mutation rate u is negligible compared to the migration rate m. In such a case the relationship at equilibrium between migration and drift is (Schneider *et al.*, 1997):

$$F_{ST} = \frac{1}{2M + 1}$$

Therefore M, which is the absolute number of migrants exchanged between the two populations can be given by:

$$M = \frac{1 - F_{ST}}{2F_{ST}}.$$

2.8.3 R_{ST} pairwise differences

R_{ST} is equivalent to F_{ST} but instead uses the stepwise mutation model (section 1.6.4) as its basis and is calculated according to Slatkin (1995):

$$R_{ST} = \frac{\overline{S} - SW}{\overline{S}}$$

Where $\overline{S}$ is twice the estimated variance in allele size across populations, and SW is twice the estimated variance in allele size within each population.

This original calculation assumes *(i)* equal sample size and *(ii)* that all loci have equivalent variances. These assumptions are not met for the data in this project. The program RstCalc (Goodman, 1997) deals with these sources of bias.

The problem of differing sample sizes is solved by using conventional statistical approaches, whereby:

$$\rho = \frac{Sb}{(Sb + SW)}$$

where Sb is the component of variance that is between populations.

The problem of differences in variance is solved by standardising the data set before testing R_{ST} pairwise differences. The transformation procedure involves standardising the entire data set so that alleles are expressed in terms of standard deviation from the global mean, rather than repeat unit number (giving each locus a global mean allele size of zero and a standard deviation of 1).

2.8.4 $(\delta\mu)^2$ pairwise differences

Another measure of genetic distance based on the stepwise mutation model is $(\delta\mu)^2$ (Goldstein *et al.*, 1995). The distance measurement allows estimation of population separation time independent of population size, using only the mutation rate. Case studies of $(\delta\mu)^2$ used on known phylogenies of primates, show that it performs well in determining relationships amongst the primates species, but less well in determining relationships between closely related human populations. However it accurately reconstructed the split between African and non-African human populations.

Using the assumption that each population is at equilibrium for mutation and drift, the variances within each population do not, on average, grow with time. The linear growth in the average squared distance (ASD) results from the changing means. The measurement of $(\delta\mu)^2$ focuses on this difference. By averaging first within populations, this distance achieves a standardisation with respect to the variation within populations without the need to estimate additional parameters (Goldstein *et al.*, 1995):

$$(\delta\mu)^2 = (\mu_A - \mu_B)^2$$

where μ_A and μ_B are the the means of allele size in population A and B respectively.

Goldstein *et al.* (1995) show that $(\delta\mu)^2$ is linear with time, having a slope equal to twice the average mutation rate across loci.

2.8.5 Kimura's Two-parameter model

The two-parameter model described by Kimura (1980) is a model of the dynamics of nucleotide substitution. It is used to estimate the sequence divergence between two

nucleotide sequences. It takes into account the likelihood that transitions are going to be more common than transversions, and thus the transitional and transversional rates of substitution for each nucleotide differ (see the text by Li, 1997, page 62 for a more detailed description of the model).

2.9 Numerical methodology 3: Neutrality tests

2.9.1 Tajima's D and Fu & Li's estimator

Two common measures of nucleotide polymorphism are Π, heterozygosity based on the number of variable nucleotide positions among all pairs of sequences in the sample, and S, heterozygosity based on the number of variable positions (i.e. segregating sites) in the aligned samples of sequences. The differences between these two measures are the basis for the Tajima (1989b) test of neutrality. Figure 2.2 shows graphically the relationship between the nucleotide sequence and the corresponding tree, along with the relative values of D.

In a random sample of sequences drawn from a population at equilibrium for mutation and drift, Π should equal S (Tajima, 1989b). However, because Π is sensitive to the frequency of a given type of sequence, and S is not, differences in the measurement of Π and S occur. When Π is greater than S, Tajima's D is positive and balancing selection or a mixture of two distinct populations is possible. On the other hand, when S is greater than Π, a selective sweep or a recent population bottleneck are possible explanations (Rand, 1996).

The test statistic used is:

$$D = \frac{\hat{\theta}_\pi - \hat{\theta}_S}{\sqrt{V(\hat{\theta}_\pi - \hat{\theta}_S)}}$$

where V is variance, $\hat{\theta}_\pi = \hat{\pi}$ and $\hat{\theta}_S = S/\sum_{i=0}^{n-1}(1/i)$ and S is the number of segregating sites in the sample (Tajima, 1989b).

A similar test statistic proposed by Fu & Li (1993) similarly reflects on the shape of the gene tree, and compares the relative numbers of internal (occurring within the tree)

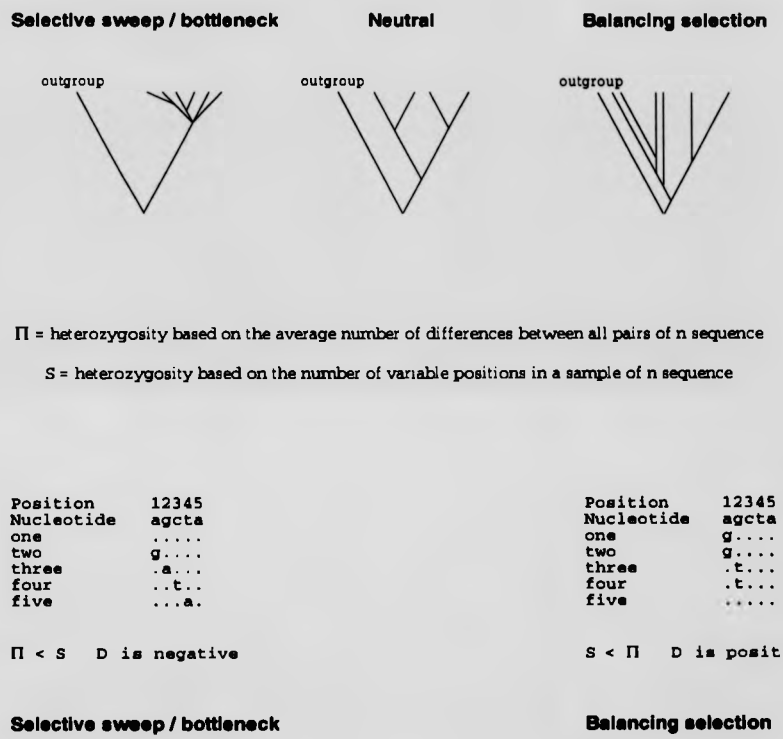

Figure 2.2: Schematic diagram of Tajima's *D*. In a random sample of sequences drawn from a population at equilibrium for mutation and drift, Π should equal *S*, as shown in the central tree. Because Π is sensitive to the frequency of a given type of sequence, and *S* is not, differences in the two measurements occur. When *S* is greater a selective sweep or a recent population bottleneck are possible explanations, as shown on the left of the figure. Alternatively, when Π is greater balancing selection or a mixture of two distinct populations is possible, as shown on the right of the figure. Figure adapted from Rand (1996)

and external (occurring at the edge of a tree) mutations. However Tajima's D is used preferentially in this thesis because of its more common use in the literature.

2.9.2 The McDonald-Kreitman Test

McDonald & Kreitman (1991) proposed a simple method to compare the patterns of within and between species polymorphism at synonymous and non-synonymous (termed replacement) sites. Assuming that the same coding region between species have the same evolutionary history, and assuming that polymorphisms should arise according to the rules of neutral evolution, then the ratio of replacement to synonymous substitutions within species should be the same as those between species. A significant difference between the two ratios can therefore be used to reject the neutral mutation hypothesis. The probability of a significant difference is given by Fisher's test on the ratios of replacement to synonymous substitutions between and within species (see Sokal & Rohlf, 1981).

2.10 Numerical methodology 4: Tree reconstruction

Principles of the tree reconstruction methods have been given in section 1.6.9. Described here are details of the methodology of each tree building algorithm used in the project. All trees, except the minimum spanning trees described in section 2.10.3, were generated using the PHYLIP (Felsenstein, 1989) package of phylogenetic software programs. In the case of UPGMA a preliminary step is a creation of a distance matrix for all pairs of operational taxonomic units (OTUs) to be used in the tree building process. Such matrices (e.g. F_{ST}) are described in section 2.8.2. For an excellent review of all tree reconstruction methods, and for greater detail and worked examples of the methodology, see Li (1997).

2.10.1 UPGMA

UPGMA employs a sequential clustering algorithm. Firstly, the two OTUs closest to one another (i.e. have the shortest genetic distance) are identified and combined as a

new, single OTU. This new OTU is termed a composite OTU and a new distance matrix is computed with the new distance involving composite OTUs being simply the average of the constituent OTUs that form the composite OTU. In general, the distance between two clusters of OTUs X and Y is given by (Li, 1997);

$$d_{XY} = \sum_{i,j} d_{ij}/(n_X n_Y)$$

where the summation is over every i in cluster X and every j in cluster Y and n_X and n_Y are the numbers of OTUs in X and Y respectively.

Using the new group of OTUs, the next pair with the highest similarity is identified and formed as a composite OTU. This process continues until there are only two OTUs left. Thus the topological relationships are built up in a stepwise manner, in order of decreasing similarity.

2.10.2 Fitch

The Fitch method starts with the aligned nucleotide sequences. Informative sites are chosen (i.e. those that have at least two different kinds of nucleotides at the site, each of which is present at least twice in the population). All possible trees connecting the sequence of informative sites are generated and the number of minimum substitutions required for each one is inferred.

Every OTU and group of OTUs is then removed and re-added at its most parsimonious position until no further changes occur. This method improves the result, since the position of every species is reconsidered, though it causes a drastic increase in computation (Felsenstein, 1989).

The Fitch trees are not bootstrapped because of *(i)* the computational time to create each tree and *(ii)* the closeness of all the OTUs leads to considerable homoplasy, hence bootstrap values are very low and not particularly informative.

2.10.3 Molecular variance parsimony (Minimum spanning trees)

Trees of equally parsimonious solutions compete with one another using a set of criteria other than mere tree length, drawn from population analysis of haplotype frequencies and geographic location. Each competing tree is first translated into a matrix of evolutionary (patristic) distances. From each competing matrix AMOVA statistics are calculated to incorporate haplotypic frequency and geographic information as further tree choice criteria. Finally, null values of the three statistics are calculated by sampling from the space of all the minimum spanning trees (MSTs), which are the equally parsimonious trees from which to choose. A heuristic procedure is then implemented that searches for ever increasingly better trees, so that a global optimum is found, ultimately defining a set of excellent trees (termed a minimum spanning network) that are both highly correlated to each other and to the defining data set, using a cophenetic correlation analog (Sokal & Rohlf, 1962). Further details on the method of tree construction according to molecular variance parsimony can be found in Excoffier & Smouse (1994).

Chapter 3

Molecular Genetic Analysis of *Trachurus* spp.

3.1 Preamble

Trachurus spp. have been introduced in section 1.5.1 on page 26. Chapter 3 presents the results of genetic analysis of the *Trachurus* species collected from the Gulf of Guinea, Benguela LME and the UK. Two regions of the mtDNA were analysed; a partial sequence of the control region/12S rRNA gene, and a partial sequence of the cytochrome *b* gene. Restriction fragment length polymorphism (RFLP) analysis was used to investigate variation of the control region/12S rRNA sequence, while sequence analysis was used for the partial sequence of the cytochrome *b* gene. RFLP results were only powerful enough to display significant variation across LMEs and lacked detail for within LME analysis. Sequence results from the partial cytochrome *b* gene sequence display strong temporal structure among sampled populations within the Gulf of Guinea, far outweighing spatial variation between sample sites. There is also some evidence for sex biased dispersal. Interpretations of these results are provided and discussed.

3.2 Sample collection

Rough locations of the sample sites for all species are given in Chapter 2 (page 46). As displayed on page 46 the three regions from where samples were obtained are: the Gulf of Guinea (7 sample sites); the Benguela LME (3 sample sites from off the Namibian coastline and 1 sample from off the South African coast - the South African samples were actually from the Agulhas LME, but for convenience will be grouped with the Benguela samples); and the UK (1 sample from off Plymouth, on the south-west coast of England). Table 3.1 gives details of the sample size, co-ordinates, depth (when known) and date of collection for all sample sites in this study. A brief station description is also given for a few stations.

The first two letters of the station name denotes the species, i.e. tt for *Trachurus trachurus* and *T. trecae*. The numbers denote the location of the station (displayed in Figures 3.1 and 3.2). Samples from outside the Gulf of Guinea are further identified; "nam" denotes all those from the Benguelan LME ("nam" denotes Namibia, as samples were supplied by a collaborator from Namibia) and "plym" indicates the sample from the UK (denoting Plymouth as they were supplied by a colleague from the Marine Biological Association at Plymouth).

The station description briefly draws attention to any exceptional circumstances at a station that will be referred to later in this chapter. The sample taken from tt70 in the Gulf of Guinea was from a very large catch of *T. trecae* that was registered as an acoustic signature on board the sampling vessel. Samples from stations tt90 and tt104 showed bimodal length frequency plots (see Figure 3.3).

Figure 3.1 displays the stations for all samples taken in the Gulf of Guinea whilst on board the RV Dr. Fridtjof Nansen. Details of the cruise and the catching gear are given in section 2.2.3. Although the *Trachurus* catch was not predictably regular, effort was made to take samples from across the whole region covered by the cruise. i.e. from the Liberia/ Côte d'Ivoire border to the Benin/Nigeria border. Such efforts were broadly successful given the locations of samples tt99 and tt44, virtually on the borders of the eastern and western limits of the survey's range. For catches that included a range

Station	sample size	Co-ordinates & Depth	Station Description	Date
tt44	41	5°31'N 0°12'E 105m	-	May 1999
tt51	80	5°20'N 0°11'W 60m	-	May 1999
tt70	63	4°29'N 2°10'W 111m	acoustic sig.	May 1999
tt84	23	4°30'N 6°57'W 62m	-	May 1999
tt97	40	4°56'N 5°22'E 91m	-	May 1999
tt99	60	4°59'N 4°58'E 85m	bimodal l.f.	May 1999
tt104	50	5°04'N 4°34'E 85m	bimodal l.f.	May 1999
ttnam1	50	22°20'S 12°57'E	-	June 1997
ttnam2	50	19°02'S 12°26'E	-	June 1997
ttnam3	25	17°15'S 11°03'E	offshore	June 1997
ttnam4	25	17°17'S 11°29'E	onshore	June 1997
ttnam5	50	33°58'S 26°23'E	adults	Sep/Oct 1997
ttplym	30	unknown	-	Pre 1998

Table 3.1: Shows sample size, co-ordinates, depth and description of each station used for collection of *Trachurus* spp. from all stations. Sample locations tt44 to tt104 are from the Gulf of Guinea (*T. trecae*); those from ttnam1 to ttnam5 are from the Benguelan LME (*T. trachurus*); while ttplym is the sample from the UK (*T. trachurus*). Co-ordinates are given in degrees south or north and east or west. Depth is measure in meters. See text for more details regarding station names and descriptions and sample sizes.

of body lengths (Figure 3.3 gives length frequency data for Gulf of Guinea *T. trecae* samples), individuals were taken from across the range to allow for investigation into differences between age groups in the same shoal (or at least the same catch).

Sample locations of Benguela LME *T. trachurus* catches are given in Figure 3.2. The samples ttnam1 to ttnam4 were taken on the Dr. Fridtjof Nansen cruise of June 1997, while the single sample from South African waters, ttnam5, was taken on the Benefit cruise of September and October 1997. Length frequency histograms are given in Figure 3.4. Immediately notable, when compared to the Gulf of Guinea samples, is that those samples of *Trachurus* spp. taken from the Benguelan LME are of a larger size (see section 1.5.1). The largest adults come from ttnam5, off the coast of South Africa, with a length range of 34 to 48cm fork length.

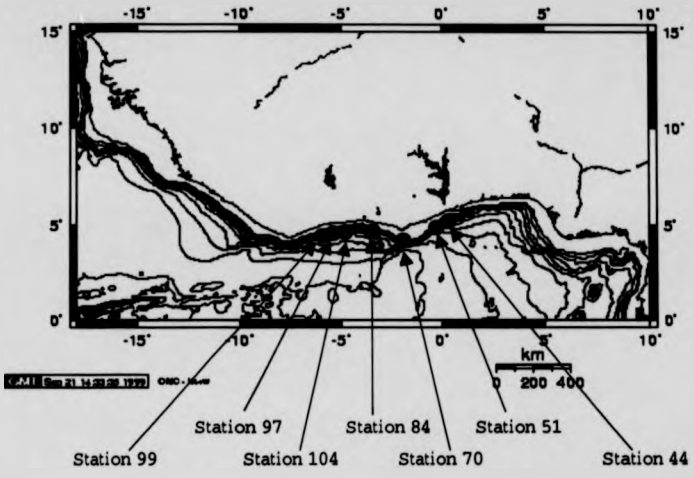

Figure 3.1: Diagram of collecting sites for *Trachurus trecae* along the west African coast of the Gulf of Guinea. All samples were collected on board the RV Dr. Fridtjorf Nansen. The bathymetry is derived from the ETOPO5 data set, available from the National Geophysical Data Centre.

3.3 The control region, tRNA(*Phe*) and 12S rRNA

The partial sequence of the control region, tRNA(*Phe*) and 12S rRNA genes (a diagram of the mitochondrial genome can be found on page 41) was sequenced. RFLP analysis of the region followed to gain initial information on the nature of molecular diversity within and between the sampled LMEs. The partial sequence has been submitted to Genebank[1] and can be viewed with the accession code AF271658. It serves as a resource from which to design primers for this highly variable region for future population genetic studies.

3.3.1 Partial sequence of the control region, tRNA(*Phe*) and 12S rRNA genes

Figure 3.5 shows a partial sequence of the control region, tRNA(*Phe*) and 12S rRNA genes for *T. trecae* from the Gulf of Guinea. Restriction sites are shown for *Cfo*l,

[1] http://ncbi.nlm.nih.gov

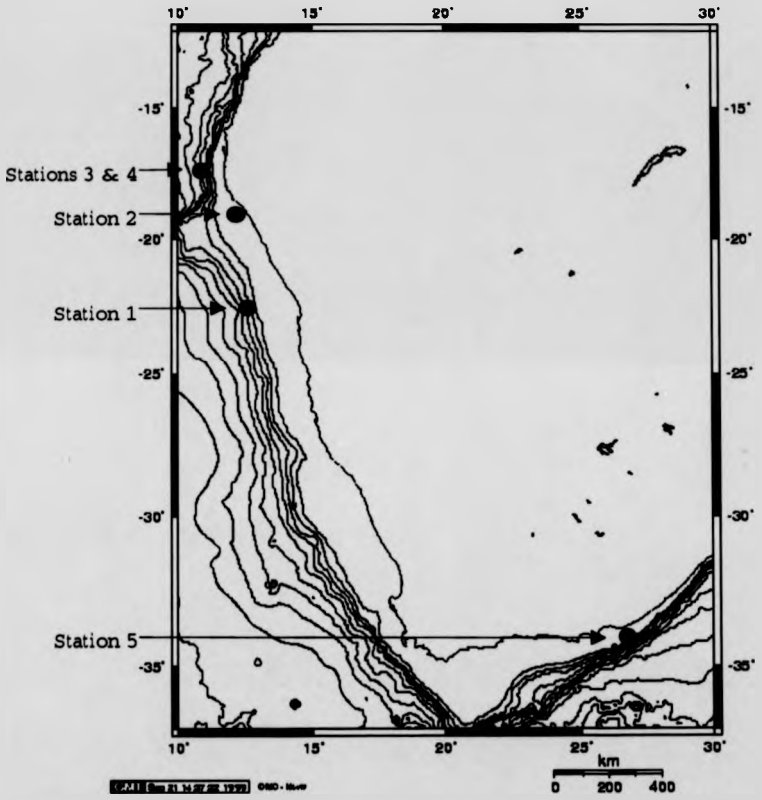

Figure 3.2: Diagram of collecting sites for *Trachurus trachurus* along the southwest African coast of the Benguela LME. All samples were collected by Ekkehard Klingelhoeffer. The Lüderitz upwelling region is from 25°S to 31°S (see section 1.3.4). The bathymetry is derived from the ETOPO5 data set, available from the National Geophysical Data Centre.

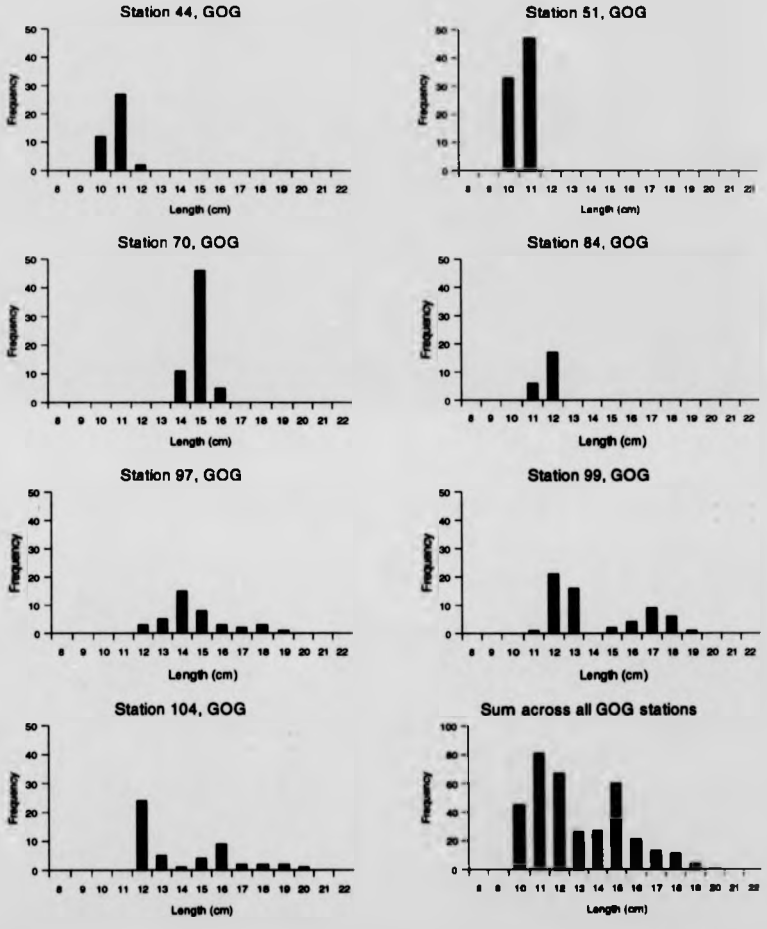

Figure 3.3: Length frequency data for Gulf of Guinea catches of *Trachurus trecae*. All axes are the same. The Gulf of Guinea samples are smaller than those of the Benguela LME (see section 1.5.1). Note that stations 99 and 104 have a bimodal length frequency; i.e. there are two distinct length cohorts.

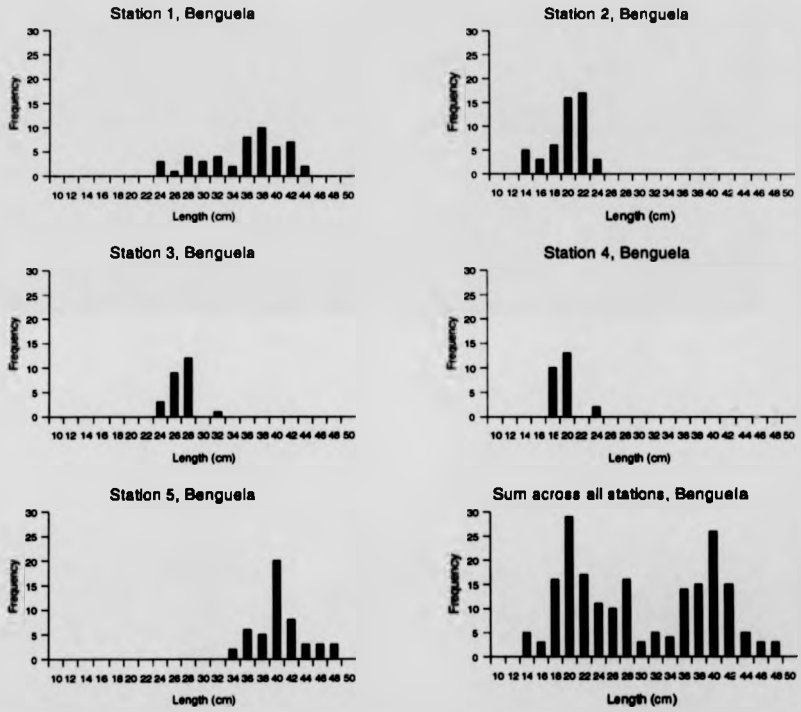

Figure 3.4: Length frequency data for Benguela catches of *Trachurus trachurus*. All axes are the same. When compared to the Gulf of Guinea samples the Benguelan LME samples are of greater length (see section 1.5.1). The largest adults come from station 5, off the coast of South Africa, and south of the Lüderitz upwelling region, with a length range of 34 to 48cm fork length.

sample location	sample size	number of polymorphic sites	nucleotide diversity
tt70	63	2	0.008
tt104	24	2	0.057
ttnam2	27	7	0.227
ttnam4	10	0	0
ttnam5	32	4	0.088
ttplym	10	0	0

Table 3.2: Sample location, sample size, the number of polymorphic sites, and nucleotide diversity calculated for all *Trachurus* spp. samples analysed with RFLP of the partial sequence of the control region, tRNA(*Phe*) and 12S rRNA genes.

*Hae*III and *Rsa*I, along with rough locations of the primers used to sequence the region. Further sequences of the control region that can be aligned with the sequence presented here can be found in Lee *et al.* (1995), while the complete sequence of the rainbow trout, *Oncorhynchus mykiss*, is presented in Zardoya *et al.* (1995).

3.3.2 Molecular diversity and estimates of theta (θ)

The 1739bp region was restricted with a suite of restriction enzymes (*Alu*I, *Cfo*I, *Rsa*I, *Hae*III and *Taq*I) for a total of 166 individuals of *Trachurus* spp. Of these, 87 were from across two sites in the Gulf of Guinea, 37 were from Namibian waters, 32 from off the coast of South Africa, and 10 from the UK. Only those restriction enzymes than revealed polymorphism are used in the present analysis, namely: *Hae*III, *Rsa*I, and *Cfo*I. Table 3.2 gives the sample locations and sample sizes used; information on the amount of genetic diversity found is given by the number of polymorphic sites and nucleotide diversity for each sample.

Sample sizes varied from 10 to 63 individuals, and likewise the amount of variation found fluctuated greatly, from 0 polymorphic sites (and thus also a value of 0 for nucleotide diversity) to 7 polymorphic sites and a corresponding nucleotide diversity of 0.227. Such results most probably arise from *(i)* the differences in samples sizes, *(ii)* the low number of restriction enzymes used and *(iii)* the relatively poor information gleaned from RFLP studies.

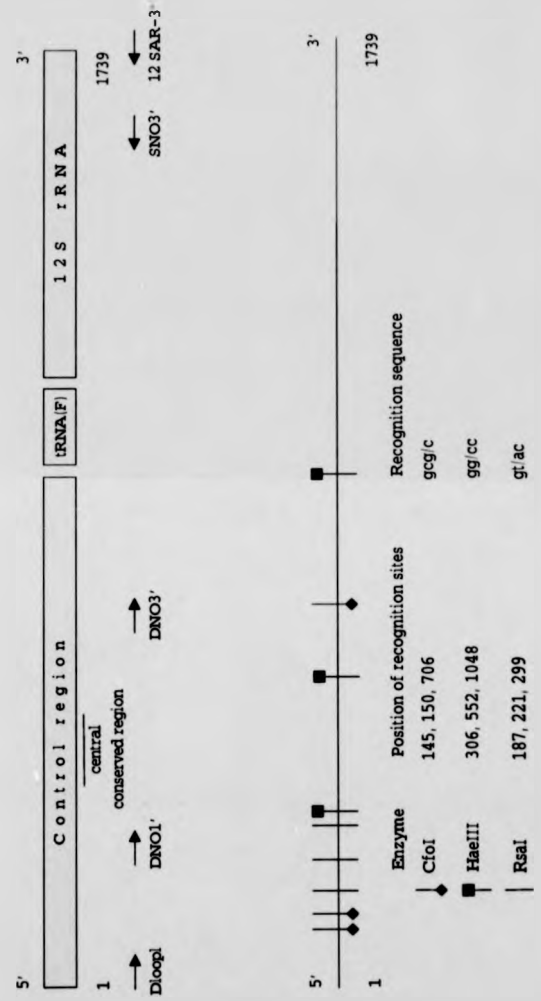

Figure 3.5: Genetic map, restriction sites and primer sites of the control region, tRNA(*Phe*) and 12S rRNA genes of *T. trecae*. Subsequent primers were designed using the Primer3 Software. Details of the primers used can be found in section 2.4.3 on page 53. The three restriction enzymes shown are all four cutters and were used in restriction fragment length polymorphism (RFLP) analysis of the sequence - see text for details. Full sequence can be found in Genebank with the accession code AF271658.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among LMEs	2	65.506	0.702	71.38
Among populations within LMEs	3	2.373	0.021	2.1
Within populations	160	41.723	0.261	26.52
Total	165	109.602	0.983	
Fixation index			F_{ST}: 0.735	$p = 0.00$

Table 3.3: Patterns of variation of a 1739bp fragment of the mitochondrial control region, tRNA(*Phe*) and 12S rRNA genes of *Trachurus* spp. from the Gulf of Guinea and Benguela LMEs, and an outgroup from the UK, designated for this test as a third LME with a single population. Over 71% of the total variation can be explained by variance between LMEs.

3.3.3 Analysis of molecular variance (AMOVA)

An analysis of molecular variance (AMOVA) was run on the restriction site data to quantify the amount of variation between and within the sampled LMEs. The results are given in Table 3.3. Over 70% of the variation results from differences between LMEs, with just over 2% resulting from variation between samples within LMEs. The rest of the variation is within samples. The large variation between the LMEs is notable, and supports the taxonomic distinctiveness attributed to the Gulf of Guinea samples, *T. trecae*, and samples from the Benguela LME and the UK, *T. trachurus* (but see section 3.5.4).

3.3.4 Pairwise comparisons

The F_{ST} pairwise comparisons for all pairs of samples are given in Table 3.4. No significant differences are found between any populations within LMEs. However, as expected highly significant differences are found between all pairs of samples taken from differing LMEs. It is notable that, while not statistically significant, there is greater difference between the pair of Gulf of Guinea samples than between any of the pairs of samples from the Benguela LME (compare with Table 3.7).

	tt70	tt104	ttnam2	ttnam4	ttnam5	ttplym
tt70		0.100	0.639	0.947	0.788	0.982
tt104	-		0.456	0.754	0.636	0.902
ttnam2	***	***		0.073	0.032	0.660
ttnam4	***	***	-		-0.001	1
ttnam5	***	***	-	-		0.863
ttplym	***	***	***	***	***	

Table 3.4: Matrix of pairwise F_{ST} values (above diagonal) and associated p values (below diagonal) for *Trachurus* spp. samples from the Gulf of Guinea, Benguela LME and the UK. Molecular marker used is RFLP of the partial sequence of the control region, tRNA(*Phe*) and 12S rRNA genes. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$

3.4 Cytochrome *b*

3.4.1 Test of cytochrome *b* as a population genetic marker

To investigate the potential of the cytochrome *b* nucleotide sequence to be used as a marker in taxonomic studies a "Consense" tree was generated in PHYLIP (Felsenstein, 1989) from across a range of species in the order Perciformes (sequences taken from the National Centre for Biotechnology Information website,[2] Entrez database; see section C in the appendices for more details on the method used). Figure C.1, in the appendices on page 198, shows that while there is high confidence in assignation of closely related species, or populations in a species, confidence limits from the bootstrapping process are lower for higher taxonomic levels. Such a pattern of confidence limits (particularly the 100 for *Dicentrarchus* samples taken from the Atlantic and Mediterranean) suggests *(i)* a relatively high mutation rate and *(ii)* that cytochrome *b* may be a suitable molecular marker for intraspecific variation.

3.4.2 Molecular diversity and estimates of theta (θ)

The 211 base pair sequence of the cytochrome *b* region was sequenced for a total of 243 *Trachurus* spp. individuals. Of these, 156 were from the Gulf of Guinea, 71 were from the Benguela LME, and 16 from the UK. Across all populations there was a total of 47 polymorphic sites and a total of 88 haplotypes; 62 haplotypes exclusively from the Gulf of Guinea, and the remaining 26 from the combined Benguela LME and UK populations.

[2] http://ncbi.nlm.nih.gov

Transitions outweighed transversions by about 2 to 1 (a total of 42 transitions to 23 transversions). Within the Gulf of Guinea the ratio of transitions to transversions was higher (around 2.5:1) than in the Benguela and UK populations (around 1:1).

When assigned as a coding region, the 211 base pair sequence displays a total of 51 synonymous sites and 159 non-synonymous sites (the final base pair is not part of a completely sequenced codon, so is omitted). Within the Gulf of Guinea samples there were 32 mutations found at synonymous sites and 14 mutations at non-synonymous sites. While the populations of the Benguela LME and UK samples showed a total of 17 mutations at synonymous sites and 10 mutations at non-synonymous sites.

Table 3.5 lists the number of individuals sequenced from each sampling station, and for each sample location gives the number of polymorphic sites, the nucleotide diversity, the value of Tajima's D, and two estimates of θ (see Chapter 2 for methodology). Stations 99 and 104 from the Gulf of Guinea showed a distinct bimodal length frequency distribution (see page 74) and therefore both stations have been split into a small individuals (e.g. tt99s) and big individuals (e.g. tt99b) class, denoted by the "s" or "b" suffix, to allow intra shoal comparisons. The remaining stations did not show such distinct bimodality and so are treated as a single sample per station. The sample locations "gog all" and "nam all" give estimations of nucleotide diversity and θ for all populations of the Gulf of Guinea and Benguela samples respectively, while "tt all" gives estimates from all sampled *Trachurus* spp. individuals.

Nucleotide diversity varied considerably between localities, with a trend of higher diversities in the Gulf of Guinea, as reflected by a comparison of nucleotide diversity for "gog all" and "nam all". Figure 3.6 shows the comparisons of diversity in the Gulf of Guinea and Benguela samples, with 50% and 95% confidence limits. A Kruskall-Wallace test of the nucleotide diversity between Gulf of Guinea and Benguela LMEs did not show statistically significant results ($\chi^2 = 4.376$, $df = 2$, $p = 0.1$), however the p value of 0.1 indicates variation that may be found to be significant if looked for in a designed experiment. The highest diversity was found in two samples, tt99b from the Gulf of Guinea, and the outgroup sample ttplym from the UK.

sample location	sample size	no. of polymorphic sites	nucleotide diversity	Tajima's D	θ_{ND} (via π)	θ_{SS} (via S)
tt44	21	9	0.005	-1.921	1.086	2.502
tt51	40	18	0.009	-1.834	1.855	4.232
tt70	25	11	0.008	-1.500	1.610	2.913
tt99s	24	14	0.010	-1.557	2.083	3.749
tt99b	17	11	0.011	-1.145	2.250	3.254
tt104s	14	5	0.003	-1.889	0.714	1.572
tt104b	15	7	0.007	-1.084	1.505	2.153
gog all	156	40	0.008	-2.255	1.741	7.113
ttnam2	20	7	0.003	-2.121	0.700	1.973
ttnam3	18	10	0.009	-1.286	1.869	2.907
ttnam4	13	3	0.003	-0.814	0.718	0.967
ttnam5	20	3	0.001	-1.723	0.300	0.846
nam all	71	16	0.004	-2.130	0.907	3.311
ttplym	16	10	0.011	-0.842	2.333	3.014
tt all	243	47	0.013	-1.846	2.832	7.745

Table 3.5: Sample location, sample size, the number of polymorphic sites, nucleotide diversity, Tajima's D and two estimates of θ calculated for all sampled locations of *Trachurus* spp. (see text for more details).

Values of Tajima's D gives information on the shape of the gene tree for each sample, with progressively negative values representing more shallow trees typical of a population bottleneck or a recent selective sweep (see page 64). An interesting trend is the significant correlation found in the Gulf of Guinea samples wherein the samples of smaller individuals (using the median length) have a more negative value of D than do samples consisting of the larger individuals (product-moment correlation coefficient D: $r = 0.93$, $df = 4$, $p = 0.01$).

The two estimates of θ (the product of the mutation rate and the effective population size, see page 58) also showed a trend between the two LMEs, with the Gulf of Guinea having a higher overall value of θ than the Benguela LME (Kruskall Wallace: $\chi^2 = 4.473$, $df = 2$, $p = 0.1$), though there was considerable variation between localities. The difference between the two alternative estimates of θ given (calculated via π and S respectively) can largely be explained by the variation in sample sizes, as θ_{SS} is sensitive to sample size changes while θ_{ND} is not (see page 58).

Figure 3.6: Nucleotide diversity in Gulf of Guinea and Benguela LME *Trachurus* spp. samples. The boxes represent 50% confidence limits while the top most and bottom most error bars represent 95% limits. Graph produced in STATA™.

3.4.3 Minimum spanning haplotype tree

The 88 haplotypes found in the total sample of *Trachurus* spp. were related by a minimum spanning tree as shown in Figure 3.7 (see page 68 for methodology). Two distinct clusters are marked as *(i) T. trachurus* (those individuals from the Benguela LME and the UK) and *(ii) T. trecae* (individuals from the Gulf of Guinea). A pair of attached haplotypes are separated by a single nucleotide difference, unless dashes are present across the line attaching the two haplotypes. Pairs of haplotypes that are separated by two or more nucleotide differences are marked with a corresponding number of dashes. No haplotypes were shared between *T. trachurus* and *T. trecae*. The large circles representing haplotypes 1, 63 and 66 show the most common haplotypes for the Gulf of Guinea, Benguela and UK populations respectively. The common haplotypes 1, 63 and 66 can be accessed from Genebank under accession numbers AF271653, AF271654, and AF271655 respectively[3].

Using the two parameter model of Kimura (1980) (methodology on page 63) the

[3]http://ncbi.nlm.nih.gov

percentage sequence divergence between the common haplotypes of the Benguela LME (haplotype 63) and the Gulf of Guinea (haplotype 1) is 1.44%. Between the Benguela LME and the UK (haplotype 66) the sequence divergence has the expectedly lower value of 0.48%, while between the UK and the Gulf of Guinea the sequence divergence is 1.93%.

The haplotype tree for each of the two clusters is dominated by one central haplotype and a star shaped pattern of closely related haplotypes, most often of one nucleotide difference. A star shaped pattern such as this indicates a shallow gene tree for each population and is typical of negative Tajima's D values as seen in Table 3.5. The star pattern in most particularly noticeable in *T. trecae* from the Gulf of Guinea and may be indicative of either demographic or historical effects, such as a founder effect or a population bottleneck, or alternatively may result from selection (see Chapter 6 for further discussion).

Relationships between haplotypes were further investigated by the construction of a Fitch Margolish tree of all haplotypes from all samples of *Trachurus* spp. individuals (Figure 3.8). As expected, the *T. trachurus* haplotypes show a distinctive cluster amongst the more shallow *T. trecae* haplotypes, supporting the two star-like clusters in Figure 3.7.

3.4.4 Analysis of molecular variance (AMOVA)

Genetic divergences between samples were tested with AMOVA. Results are given in Table 3.6 for three arrangements of AMOVA: *(i)* tests of all samples, *(ii)* of samples from within the Gulf of Guinea only, and *(iii)* for samples from within the Benguela LME only. F_{ST} values and their corresponding p values are given for each of the three AMOVAs (see page 59 for methodological details).

Between the Gulf of Guinea, the Benguela LME and the UK samples (the UK sample is designated as a third LME) over 62% of the variation is explained due to differences between the LMEs, and only just over 1.8% due to differences among populations within the LMEs (comparable to results due to RFLP analysis of the control region, tRNA(*Phe*)

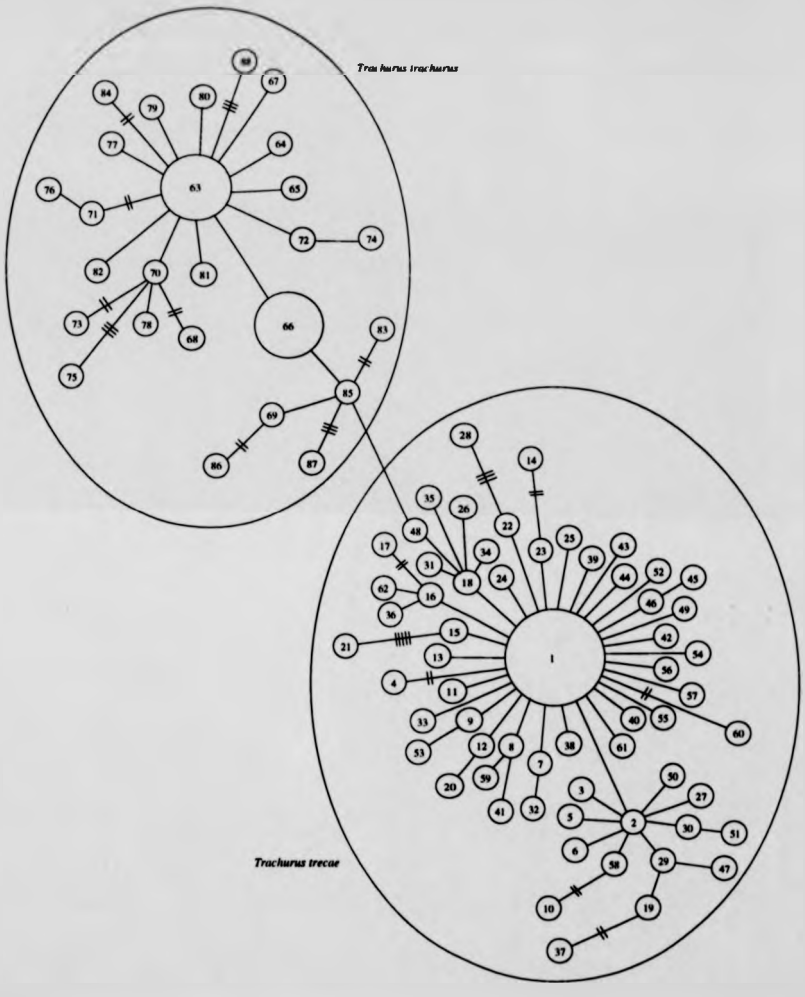

Figure 3.7: Haplotype tree of the cytochrome *b* region in *Trachurus* spp. No haplotypes were shared between *T. trecae* (from the Gulf of Guinea) and *T. trachurus* (from the Benguela LME and UK) samples. A pair of attached haplotypes are separated by a single nucleotide difference, unless dashes are present across the line attaching the two haplotypes. Pairs of haplotypes that are separated by two or more nucleotide differences are marked with a corresponding number of dashes. The large circles representing haplotypes 1, 63 and 66 show the most common haplotypes for the Gulf of Guinea, Benguela and UK populations respectively, and can be found in Genebank with the accession numbers AF271653, AF271654, and AF271655 respectively.

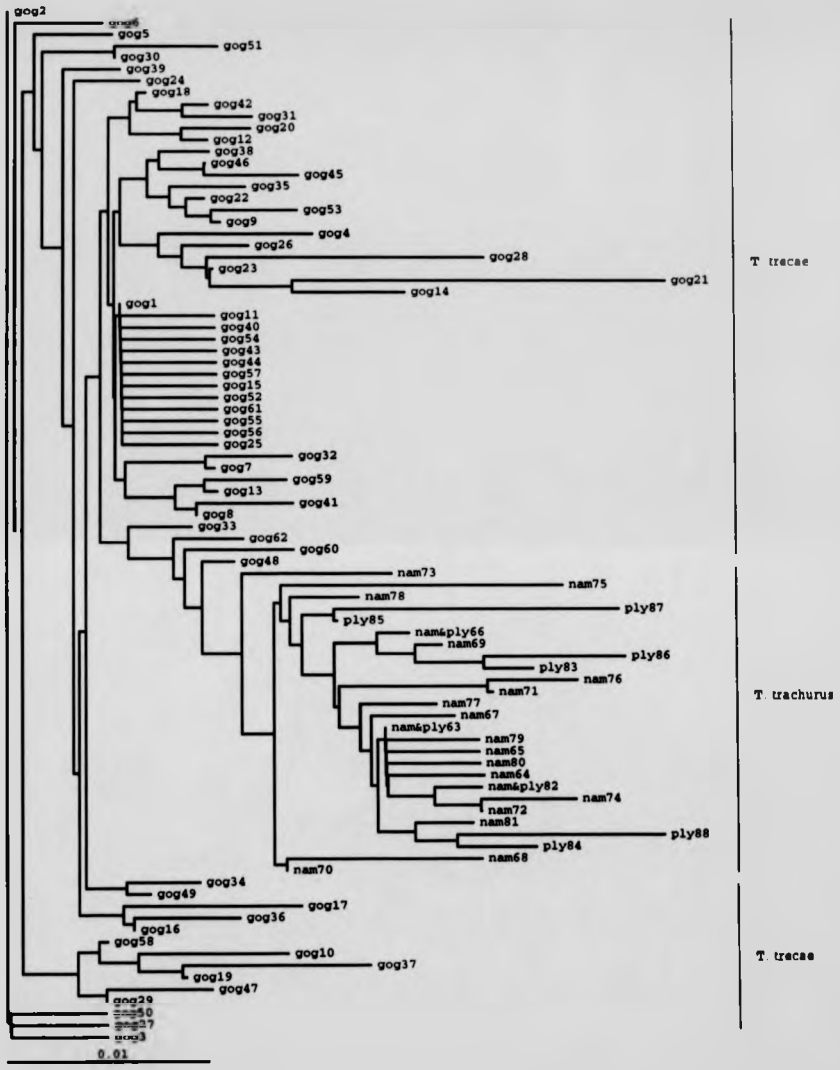

Figure 3.8: Fitch Margolish tree of mtDNA haplotypes from all samples of *Trachurus* spp. from the Gulf of Guinea, Benguela LME and the UK. All distances are additive according to percentage sequence divergence. The haplotypes from *T. trachurus* and *T. trecae* species are shown. See text for more details and section 2.10.2 for details of methodology.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
		for all samples		
Among LMEs	2	158.52	1.29	62.37
Among populations within LMEs	9	13.45	0.04	1.8
Within populations	231	170.68	0.74	35.84
Total	242	342.65	2.06	
Fixation index			F_{ST}: 0.642	$p = 0.000$
		within Gulf of Guinea only		
Among populations	6	11.411	0.049	5.61
Within populations	149	123.487	0.829	94.39
Total	155	134.897	0.878	
Fixation index			F_{ST}: 0.056	$p = 0.000$
		within Benguela LME only		
Among populations	3	2.036	0.013	2.93
Within populations	67	29.697	0.443	97.07
Total	70	31.732	0.457	
Fixation index			F_{ST}: 0.029	$p = 0.0154$

Table 3.6: Patterns of variation of a 211bp fragment of the mitochondrial cytochrome *b* gene of *Trachurus* spp. for three arrangements of AMOVA: *(i)* tests of all samples, *(ii)* of samples from within the Gulf of Guinea only, and *(iii)* for samples from within the Benguela LME only. F_{ST} values and their corresponding p values show that over all samples over 62% of the variation is explained due to differences between the LMEs. Within the Gulf of Guinea over 5% of variation is due to differences between samples, and nearly 3% inter-sample variation is found in the Benguela LME.

and 12s rRNA partial sequence on page 78). The remaining variation is due to differences within the individual samples. Such a result, reflecting deep divergence between the Gulf of Guinea and elsewhere, is both highly significant and expected, given that no haplotypes are shared between the *T. trecae* of the Gulf of Guinea and the *T. trachurus* of the Benguela LME and UK populations (see Figure 3.7).

Within the Gulf of Guinea nearly 95% of the variation was found within samples, with only 5% resulting from inter sample variation. The F_{ST} value of 0.056 reflects a degree of little to moderate genetic differentiation (following Wright 1978, see page 61) and is highly significant ($p = 0.000$).

While there is less variation between samples within the Benguela LME (2.93%) than between samples within the Gulf of Guinea, the p value (0.015) indicates the liklihood

that there too is non-random mating between sampled populations.

3.4.5 F_{ST} Pairwise comparisons

Pairwise F_{ST} comparisons are shown in Table 3.7. The figures above the diagonal give the actual F_{ST} pairwise value, while the asterisks below the diagonal give the level of significant difference between pairs of samples (methodological details are on page 61). As expected from the AMOVA results (page 86) and haplotype tree (page 84) there is a block of highly significant pairwise differences between the Gulf of Guinea samples and all other samples. The outgroup from the UK shows a highly significant difference from all other samples.

Figure 3.9 shows the significant differences between the Gulf of Guinea and elsewhere with a plot of genetic distance (measure by F_{ST} pairwise values) against geographic distance (measured as approximate number of miles between each pair of samples). The very distinct samples, with high F_{ST} values, in the centre of the plot are comparisons of Gulf of Guinea samples to samples from either the Benguela LME or the UK. The pairs of samples separated by the greatest geographic distance, but that display relatively low F_{ST} values, are plots of Benguela LME samples with the sample from the UK. Note that if the distinct samples in the centre of the plot are removed, the slope of the line connecting the remaining pairwise comparisons indicates some evidence of isolation by distance between Benguela and UK *T. trachurus* samples.

However, in addition to inter LME variation, pairs of intra LME samples also show significant levels of variation. There are significant differences between ttnam5, the adult population off the coast of South Africa (from the Agulhas LME), and samples ttnam3 and ttnam4 from Namibian waters (the Benguela LME). None of the pairwise sample comparisons from within the Benguela LME proper (i.e. samples ttnam2-ttnam4) showed significant differences. Within the Gulf of Guinea LME most of the significant variation involves two samples: tt70 and tt99b. While showing similarity to one another, samples tt70 and tt99b display varying degrees of significant differences to all other samples. It is notable that those stations with a bimodal length frequency, i.e.

	tt44	tt51	tt70	tt99s	tt99b	tt104s	tt104b	ttnam2	ttnam3	ttnam4	ttnam5	ttplym
tt44		0.017	0.104	0.011	0.089	0.018	0.067	0.759	0.629	0.736	0.807	0.649
tt51	-		0.058	0.009	0.080	0.011	0.019	0.663	0.576	0.632	0.692	0.598
tt70	**	**		0.091	0.015	0.122	0.138	0.713	0.603	0.683	0.753	0.587
tt99s	-	-	***		0.061	0.038	0.035	0.670	0.563	0.629	0.707	0.586
tt99b	**	***	-	*		0.149	0.121	0.674	0.552	0.629	0.720	0.531
tt104s	-	-	***	*	***		0.091	0.804	0.646	0.791	0.865	0.660
tt104b	*	-	***	-	**	**		0.723	0.580	0.692	0.779	0.592
ttnam2	***	***	***	***	***	***	***		0.027	0.015	0.000	0.218
ttnam3	***	***	***	***	***	***	***	-		0.019	0.051	0.137
ttnam4	***	***	***	***	***	***	***	-	-		0.064	0.226
ttnam5	***	***	***	***	***	***	***	-	**	*		0.274
ttplym	***	***	***	***	***	***	***	***	***	***	***	

Table 3.7: Matrix of pairwise F_{ST} values (above diagonal) and associated p values (below diagonal) for *Trachurus* spp. samples from the Gulf of Guinea, Benguela LME and UK. Molecular marker used is partial sequence of the cytochrome *b* gene. * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

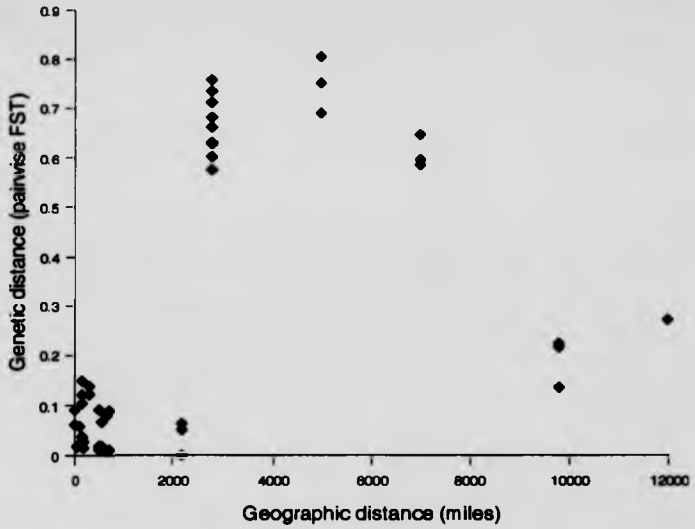

Figure 3.9: Plot of genetic against geographic distance (data from the pairwise comparisons in Table 3.7) for all samples of *Trachurus* spp. The distribution clearly shows the Gulf of Guinea populations, in the centre of the plot, as being distinctly different from the UK and Benguela LME populations. The samples at the bottom right of the plot, i.e. large geographic seperation but relatively low genetic distance, are pairs of Benguela/UK samples.

tt99 and tt104, display significant variation between the two length classes.

Figure 3.10 shows a UPGMA tree (see page 66 for methodology) using the pairwise F_{ST} samples from Table 3.7. The two clear clusters of *T. trachurus* and *T. trecae* are easily visible. In addition, within the Gulf of Guinea the younger fish form one cluster of samples (samples tt44, tt51, tt99s and tt104s), while those from an older age class form a second, though less distinct, cluster (samples tt70, tt99b and tt104b). There is no evidence of clustering according to geographic closeness, as even samples of differing lengths from the same trawl show significant differences. Therefore, rather than clustering as a result of geographical proximity, samples from within the Gulf of Guinea cluster according to their similar length (and consequently, similar age).

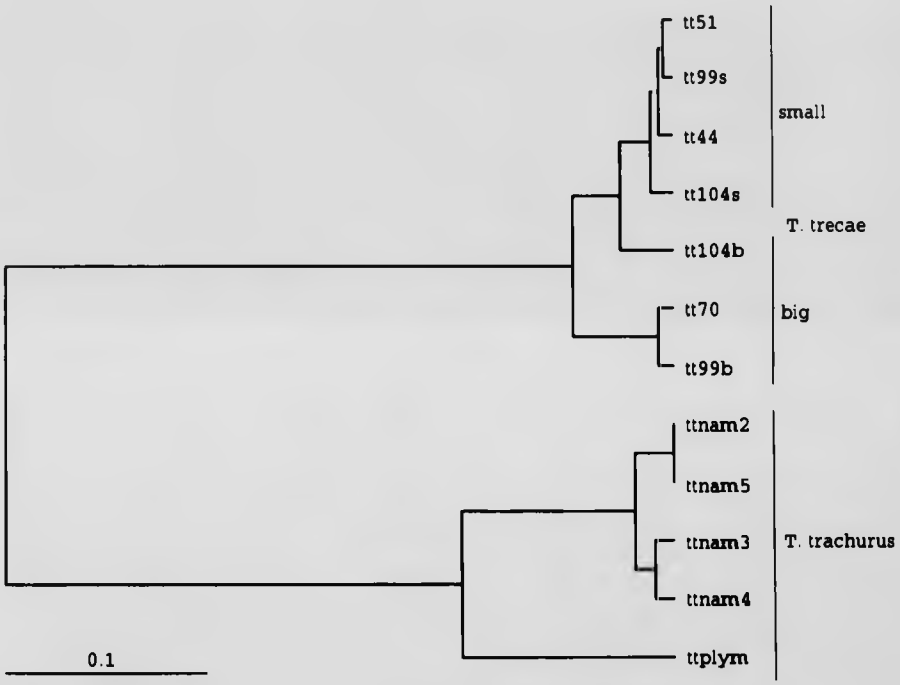

Figure 3.10: UPGMA plot of the F_{ST} pairwise comparisons displayed in Table 3.7. The two distinct clusters of *T. trachurus* and *T. trecae* are marked. Within the *T. trecae* cluster of the Gulf of Guinea there is a further noticeable clustering of samples of small individuals (tt44, 51, 99s and 104s) and a second, though less distinct, cluster of samples of large individuals (tt70, 99b and 104b).

3.4.6 A test of the importance of temporal structure against spatial structure

To test the importance of this apparent temporal structure (shown in section 3.4.5) against any signs of geographic structuring, AMOVAs were performed to test alternative, proposed population structures. The population structures chosen to be tested represent alternative possibilities according to whether age or geographic proximity drives the genetic structure of the *T. trecae* populations studied in the Gulf of Guinea. The population structures tested were:

- Spatial structure: simple classification into a population of those samples taken from east of Cape Three Points (tt44 and tt51) and those from the west of Cape Three Points (tt99 and tt104 - both small and big individuals). The sample tt70 was not used as it is from just off Cape Three Points itself and so not easily assigned (see below).

- Temporal structure (2 divisions): simple classification into a "small" (i.e. consists of small fish) population (tt44, 51, 99s, 104s) and a "big" (i.e. big fish) population (tt70, 99b, 104b).

- Temporal structure (3 divisions): as with two division but with the large classification split further into two size groups (effectively "big" and "bigger" fish).

- Temporal structure (tt70 separate): as with two divisions but with tt70 taken out of the large division and place on its own as a separate population (recall from Table 3.1 that station tt70 was a large haul of *Trachurus* spp.).

The results shown in Table 3.8 clearly indicate that virtually none of the genetic variation is explicable by spatial population structuring (0.12% variation between the two populations each side of Cape Three Points), indicating high levels of gene flow across the Gulf of Guinea. In contrast the temporal structure is highly significant and explains from a little over 3% to nearly 5% of the total variation depending on the temporal structure tested. The highest variance is explained when tt70 is considered as a sepa-

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
		spatial structure		
Among populations	1	0.931	0.001	0.12
Within populations	129	111.466	0.864	99.88
Total	130	112.397	0.865	
Fixation index			F_{ST}: 0.001	$p = 0.33$
		temporal structure (2 divisions)		
Among populations	1	3.453	0.036	4.04
Within populations	154	131.444	0.854	95.96
Total	155	153.404	1.009	
Fixation index			F_{ST}: 0.04	$p = 0.0001$
		temporal structure (3 divisions)		
Among populations		4.036	0.029	3.24
Within populations	153	130.862	0.855	96.76
Total	155	134.897	0.884	
Fixation index			F_{ST}: 0.032	$p = 0.001$
		temporal structure (tt70 separate)		
Among populations	2	5.237	0.043	4.82
Within populations	153	129.660	0.847	95.18
Total	155	134.897	0.890	
Fixation index			F_{ST}: 0.048	$p = 0.0000$

Table 3.8: Using AMOVA to test spatial vs. temporal population structuring in *T. trecae* populations in the Gulf of Guinea. Virtually none of the genetic variation is explicable by spatial population structuring (0.12% variation). In contrast the temporal structure is highly significant and explains from a little over 3% to nearly 5% of the total variation depending on the temporal structure tested (see text for more details).

rate sampled population from the assigned "small" (tt44, 51, 99s, 104s) and "big" (tt99b, 104b) divisions.

To test any effect that tt70 may have on spatial structure the analysis was repeated with tt70 alternatively designated to the samples to the east and to the west of Cape Three Points. When allocated to the east of Cape Three Points (i.e. the Ghana side) F_{ST} was 0.003 ($p = 0.17$); when allocated to the west (i.e. the Côte d'Ivoire side) F_{ST} was 0.007 ($p = 0.08$). Neither result is statistically significant.

3.4.7 Sex based dispersal

Mt DNA is often perceived as being limited by its maternal inheritance, in that no inferences can be made about male dispersal. However, O'Corry-Crowe *et al.* (1997) argue otherwise. They argue that because only those dispersers who are still living are available for sampling, only recent male dispersion (and only those who have survived) can reduce the degree of mtDNA differentiation. In contrast to males, female dispersers and their offspring are available for sampling and mtDNA analysis, leading to a far greater homogenizing effect on mtDNA variation over time than that of male dispersers. Therefore, if mtDNA results show that between population differentiation is greater among females than males it is conservative to infer that the relative dispersal rate of males between populations is greater than that of females.

The only Gulf of Guinea population that has good data on the gender of the sampled *Trachurus* individuals is tt70. To investigate whether any evidence is available to test whether gene flow across the Gulf of Guinea is mediated predominately by one sex or the other, sample tt70 was split into male and female subsamples and F_{ST} pairwise comparisons were performed between the gender-based tt70 subsamples and the other populations of the Gulf of Guinea. The results, shown in Table 3.9, give comparisons of F_{ST} values and estimates of number of migrants (for methodology of F_{ST} see page 61). Please note that the sample sizes of female and male individuals are very small (N = 5 and 16 respectively), and that there is a male to female ratio of 3:1. Such a small sample size will inevitably introduce error to the results.

The results indicate a very definite trend towards greater gene flow resulting from the dispersal of males rather than females. Consistently the F_{ST} values of tt70 male pairwise comparisons with other Gulf of Guinea samples are around one tenth of the corresponding F_{ST} values involving tt70 females. The estimated number of migrant figures should be taken as a guide only, given the assumptions made in the methodology (see section 2.8.2). They do however give some indication of the order of magnitude of the number of dispersing individuals. It is interesting to note that no similar such trend is apparent with *T. trachurus* populations in the Benguela LME.

	tt70 female	tt44	tt51	tt99s	tt99b	tt104s	tt104b
				F_{ST} *values*			
tt70 female (N=5)	-	0.488	0.287	0.322	0.139	0.604	0.435
tt70 male (N=16)	0.172	0.049	0.018	0.04	0.012	0.062	0.086
				Estimated no. of migrants			
tt70 female (N=5)	-	0.52	1.24	1.05	3.11	0.33	0.65
tt70 male (N=16)	2.41	9.75	27.74	12.08	41.31	7.58	5.28

Table 3.9: Matrix of F_{ST} values and estimated number of migrants with *T. trecae* sample tt70 split into tt70 female and tt70 male, and samples tt44, 51, 99s, 99b, 104s and tt104b from the Gulf of Guinea. A trend is apparent of greater gene flow resulting from the dispersal of males rather than females. The F_{ST} values of tt70 male pairwise comparisons are around one tenth of the corresponding F_{ST} values involving tt70 females. The estimated number of migrant figures give some indication of the order of magnitude of migration. Note also that the number of female migrants increases with age (see text for more details).

A second feature that is apparent from Table 3.9 is that the number of female migrants increases with age. For the two sample stations where bimodal length frequencies were found, tt99 and tt104, the number of female migrants from the tt70 gender based subsamples were greater when compared to the samples of older individuals from station 99 or 104 (i.e. tt99b and tt104b) than from the samples of younger individuals (tt99s and tt104s). However, even with this increase with age, male dispersal still predominates.

3.5 Discussion

3.5.1 Pattern of genetic variation in *Trachurus* spp.

The variation observed in this study is primarily at synonymous third position codons and thus is likely to be silent as there would be no effect on protein structure (a key feature of the neutral theory described in section 1.6.3, but see references therein), though there would be an effect on which tRNAs are used in protein synthesis. Any selection that is occurring is probably small (but see section 6.2). Given the effective neutrality of the variation, the distribution of polymorphisms mostly results from population structuring and the reproductive ecology of individuals and populations. However, the whole mitochondrial chromosome behaves as a single locus because there is usually no

recombination and therefore total linkage (but see Hey, 2000). Selection at other sites on the mitochondrial chromosome could therefore affect the calculation of coalescence times and change the estimates of population parameters. Results from the RFLP analysis will not be specifically discussed because *(i)* they broadly support the sequencing studies and *(ii)* it is of low sensitivity so says nothing of within LME variation.

3.5.2 Relationships between and within LMEs

The Gulf of Guinea samples differ markedly from the Benguela and UK samples. While there is little evidence for any consistent population subdivision within the Benguela LME, somewhat surprising given the samples are taken from either side of the central upwelling region at Lüderitz (see Figure 3.2 on page 73), the fixation index for the Gulf of Guinea samples indicates small to moderate population differentiation, indicating non-random mating between samples.

Estimates of nucleotide diversity and θ follow a similar pattern, with Gulf of Guinea samples showing in general higher levels of diversity than their Benguela LME counterparts. However there were two noticeable examples against the general trend, samples tt104 and ttnam3. The former sample had a relatively low nucleotide diversity while the latter a distinctly high diversity. The general pattern however could be indicative of a number of events. The estimates of θ indicate that the effective population size of the Gulf of Guinea population is greater than that of the Benguela LME (assuming the same mutation rate in all samples), which may also be reflected in the observed higher diversity of Gulf of Guinea samples. Alternatively, greater heterogeneity of the physical and oceanographic environment of the Gulf of Guinea LME over the Benguela LME could lead to a greater genetic diversity in the population(s) that reside there.

Within the groups of samples from both the Gulf of Guinea and Benguela LMEs there is no indication of isolation by distance. In fact, within the Gulf of Guinea samples there is far stronger evidence of temporal structuring, leading to a cluster of age groups in Figure 3.10.

3.5.3 Temporal variation and sex biased dispersal

As mentioned above, within the Gulf of Guinea samples there is strong evidence of temporal structuring, illustrated as a clustering of age groups in Figure 3.10. Such clustering is so marked that contemporary age classes from opposite ends of the Gulf of Guinea show far greater similarities to one another than to the other age class from the same trawl (i.e. shown in stations 99 and 104, where there were two length classes in one trawl). Such a pattern could arise from various aspects of fish behaviour or natural selection. One explanation is the sweepstakes reproductive strategy (Cushing, 1972) whereby, in theory, the surviving offspring in a given locality may come from a few lucky parents. Such a theory therefore also explains the often found small effective population size, N_e, as well as why cohorts of year classes share genetic similarity (Hedgecock, 1994; Ruzzante *et al.*, 1996). Alternatively the pattern of results could be interpreted as evidence for more than one relatively discrete reproductive season per year (e.g. as in *T. trachurus* off Agulhas Bank, South Africa; Hecht, 1990), and that spawning adults remain loyal to a certain season in time, thus resulting in discrete age classes. Natural selection could play a part whereby each cohort progressively becomes genetically more similar as the fish in the cohort, from a single spawning event, may share similar environmental conditions, with correspondingly similar selection pressures, leading to a genetically similar surviving population (Johnson & Black, 1982, 1984). Further discussion regarding temporal variation can be found in section 6.3.1.

Because only recent male dispersion can reduce the degree of mtDNA differentiation between male samples, if results show that between population differentiation is greater among females than males it is conservative to infer that the relative dispersal rate of males between populations is greater than that of females (O'Corry-Crowe *et al.*, 1997). The results of section 3.4.7 therefore indicate greater gene flow resulting from the dispersal of males rather than females. A second observation is that the number of female migrants increases with age, though male dispersal still predominates.

3.5.4 Are *T. trecae* and *T. trachurus* different species?

Sequence divergence between Gulf of Guinea *T. trecae* and Benguela LME *T. trachurus* for the cytochrome *b* gene is 1.44%, while between *T. trecae* and *T. trachurus* from the UK the divergence is 1.93%. Following Avise & Walker (1999), who compared fragments of the cytochrome *b* gene across 1,832 species, clades interpreted as intraspecific phylogenetic groups were distinguished consistently by at least 0.6% sequence divergence in mtDNA. Assigned sister species however showed, roughly nine times out of ten, mtDNA sequence divergences of greater than 2%. Consequently, following the argument of Avise & Walker (1999), the amount of divergence between between the Gulf of Guinea and Benguela LMEs does not convincingly support separate species, though does support distinctive, geographically orientated, intraspecific mtDNA phylogroups.

However, the fragment of the cytochrome *b* gene used is only small (211bp) and as such does not necessarily give a representative view of the entire cytochrome *b* gene or the mtDNA genome. In addition, morphological rather than molecular differences have been used to assign the current taxonomic status of separate species, and this study takes no account of them. There is also plenty of molecular evidence in this thesis which strongly supports the argument that *T. trecae* and *T. trachurus* are separate species, most particularly: *(i)* the very high AMOVA results in Tables *3.3* and *3.6*; *(ii)* there are no shared haplotypes between *T. trecae* and *T. trachurus* samples (Figure 3.7); and *(iii)* the greatest pairwise distances are between Gulf of Guinea (*T. trecae*) samples and *T. trachurus* samples from both the Benguela LME and the UK, *either side* of the Gulf of Guinea (Figure 3.9). Given the current taxonomic position of *T. trecae* and *T. trachurus* being separate species, there is insufficient evidence to argue that they are conspecific. The precautionary principle in fisheries management is to assume that species are separate taxonomic units unless there is strong evidence to the contrary.

The divergence of 0.48% between *T. trachurus* from the Benguela LME and *T. trachurus* from the English Channel indicates merely intraspecific geographic grouping, being of around the 0.6% sequence divergence typical of intraspecific mtDNA phylogroups. It does not support the designation of *T. trachurus* off South Africa and Namibia as a

separate species: *T. capensis* (the Cape horse mackerel).

3.5.5 Management implications

Broadly the pattern indicates that management of *Trachurus* spp. along both the Benguela and Gulf of Guinea LMEs needs to be directed at a regional, rather than a national level. There is no clear evidence for more than one population from the Liberia/Côte d'Ivoire border to the Benin/Nigeria border, or from the Namibia/Angola border to the waters of the Republic of South Africa. Management for classified "demersal" fisheries, as *Trachurus* spp. are often thought of, is often based around the nation state and national waters. However, *Trachurus* populations at least do not seem to take notice of such borders and regularly migrate large distances for feeding and spawning.

Chapter 4

Molecular Genetic Analysis of *Pagellus bellottii*

4.1 Preamble

Pagellus bellottii has been introduced in section 1.5.2 on page 28. Chapter 4 presents results of genetic analysis of *P. bellottii* from the central and western regions of the Gulf of Guinea. A partial sequence of the 12S rRNA gene of the mtDNA genome was used for the analysis. Sequence analysis of the mtDNA region revealed little variation across the Gulf of Guinea LME, though there was some evidence of isolation by distance, primarily displayed by the sample from the western region. Four individuals taken from a single trawl in the central region of the Gulf of Guinea showed marked and highly significant difference from all other samples. Alternative interpretations of the results are provided and discussed.

4.2 Sample collection

Rough locations of where samples for all species were taken is given on page 46. A more detailed map with sample locations is given in Figure 4.1. Table 4.1 gives the station name, sample size, co-ordinates, depth and date collected for all samples. Also given is a brief station description where relevant.

Station	sample size	Co-ordinates & Depth	Station Description	Date
pbgu15	41	unknown	-	Sep/Oct 1998
pbgu25	80	unknown	-	Sep/Oct 1998
pb28	48	6°08'N 2°11'E 60m	bimodal l.f.	May 1999
pb62	52	4°21'N 1°23'W 80m	-	May 1999
pb66	28	4°42'N 1°47'W 47m	-	May 1999
pb67	62	4°31'N 1°47'W 62m	bimodal l.f.	May 1999
pb84	50	4°30'N 6°57'W 62m	-	May 1999
pb105	40	5°08'N 4°10'W 84m	-	May 1999
pb106	40	5°11'N 4°11'W 48m	-	May 1999

Table 4.1: Sample size, co-ordinates, depth and description of each station used for collection of *Pagellus bellottii* from the Gulf of Guinea. Samples from stations 15 and 25, off the coast of The Republic of Guinea, were collected by Eric Morize. All other samples were collected on the RV Dr. Fridtjof Nansen cruise.

The sample names follow the convention of the previous chapter (see page 70). The first two letters, "pb", denote the species, *Pagellus bellottii*, while the numbers indicate the sampling site from which the sample was taken. Samples pbgu15 and pbgu25 are from the national waters of The Republic of Guinea in the western part of the Gulf of Guinea and are denoted by the additional letters "gu."

The two regions where samples have been obtained are the central region of the Gulf of Guinea, collected on the Nansen cruise (see page 48), and the western part of the Gulf of Guinea. As with previously described collections on the Nansen cruise, when catches were made that showed a range of length classes, individuals were sampled from across that range allowing intra shoal (or at least intra trawl) comparison. Effort was made to take samples from across the whole range of the sampled region of the Gulf of Guinea.

Length frequency histograms for Gulf of Guinea stations are shown in Figure 4.2, and for the Republic of Guinea samples, in the west, in Figure 4.3. Distinctive bimodal length frequency histograms are apparent for samples pb28 and pb67, and thus as with *Trachurus* spp. they were split into subsamples of small (denoted by the suffix "s") and big (denoted by the suffix "b") fish (e.g. pb28s and pb28b respectively).

Rijavec (1973) found that the length frequency curves of *P. bellottii* off Ghana were almost always unimodal, suggesting that the shoals were grouped by age. However the length frequency histograms of Figure 4.2 are not dominated by unimodal curves and

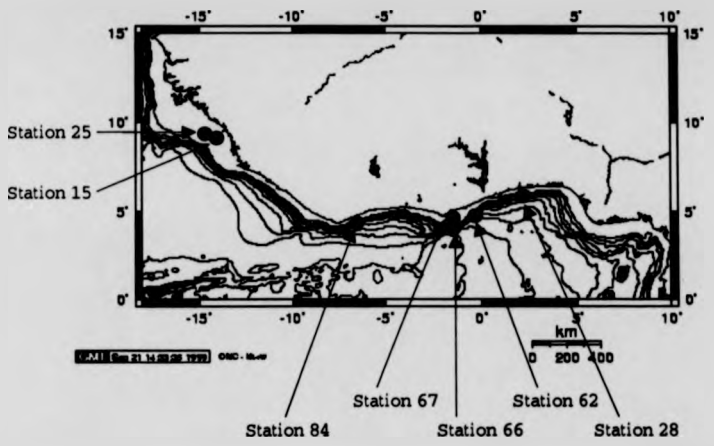

Figure 4.1: Diagram of collecting sites for *Pagellus bellottii* along the west African coast of the Gulf of Guinea. Samples from the western region were collected by Eric Morize. All other samples were collected on board the RV Dr. Fridtjorf Nansen. The bathymetry is derived from the ETOPO5 data set, available from the National Geophysical Data Centre.

thus the shoals are not grouped by age (though it is possible that, at stations such as 28 and 67, two shoals of differing sized fish happened to be caught in the same trawl).

4.3 The control region, tRNA(*Phe*) and 12S rRNA

Part of the control region, tRNA(*Phe*) and 12S rRNA genes were sequenced (a diagram of the mitochondrial genome can be found on page 41). The 1161 base pair partial sequence has been submitted to Genebank[1] and can be viewed with the accession code AF271659. It serves as a resource from which to design primers for this highly variable region for future population genetic studies. An important resource with regard to studies on the control region in *Pagellus* is Ostellari *et al.* (1996), in which specific primers are published for the control region of *Pagellus bogaraveo*. Further sequences of the control region that can be aligned with the sequence presented here can be found in Lee *et al.* (1995), while the complete sequence of the rainbow trout, *Oncorhynchus*

[1] http://ncbi.nlm.nih.gov

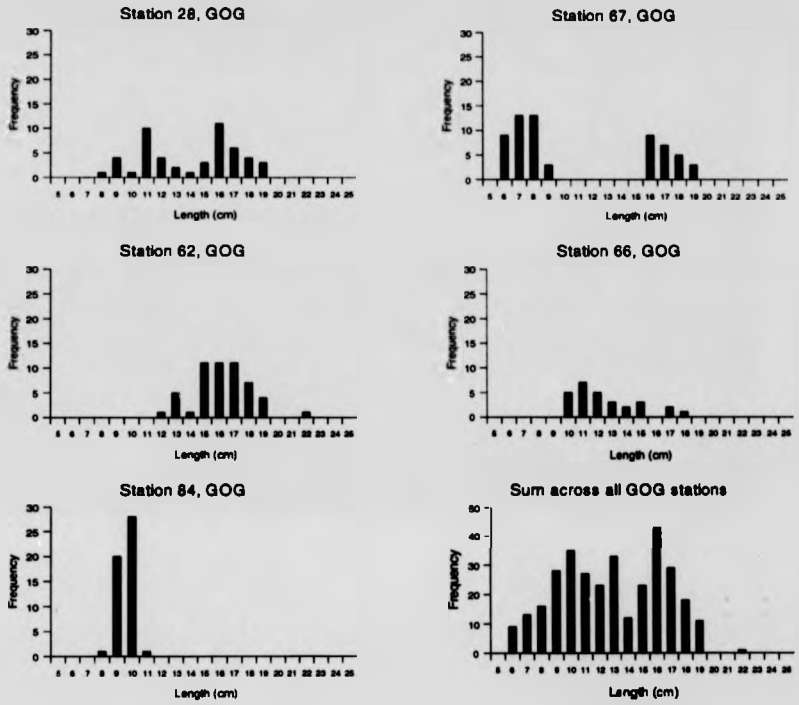

Figure 4.2: Length frequency histograms for *Pagellus bellottii* from the Gulf of Guinea as collected on the RV Dr. Fridtjorf Nansen cruise. All axes are the same except for the final histogram, which has a Y axis of 50 instead of 30. Distinctive bimodal length frequency histograms are apparent for samples pb28 and pb67, and thus they were split into subsamples of small and big fish.

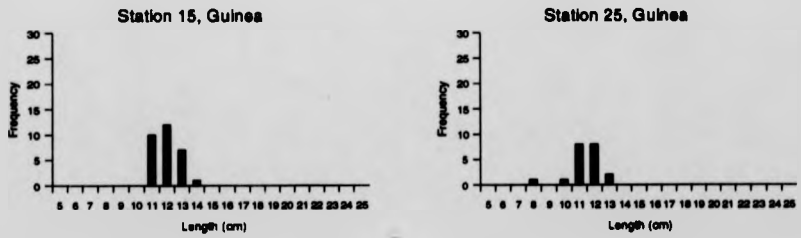

Figure 4.3: Length frequency histograms for *Pagellus bellottii* from the Gulf of Guinea as collected by Eric Morize. All axes are the same.

mykiss, is presented in Zardoya *et al.* (1995).

A diagram of the region is shown in Figure 4.4. Primers used to sequence the region are shown along with cutting points for three restriction enzymes, *Cfo*I, *Hae*III, and *Rsa*I.

4.4 12S rRNA

4.4.1 Molecular diversity and estimates of theta (θ)

In total 88 *P. bellottii* individuals were sequenced for the 289 base pair 12S rRNA sequence. Over all samples there were 22 polymorphic sites and a total of 21 haplotypes were defined. There were 14 observed transitions to 11 observed transversions. Notably, there was also an insertion in four individuals, all from sample pb67b, at position 67 of the sequence. The 12S rRNA gene does not code for a protein so there are no synonymous or non-synonymous mutations to compare.

Table 4.2 gives the sample size analysed for each sample location, along with details on the number of polymorphic sites, nucleotide diversity (for methodology see page 57), Tajima's D (see page 64), and two estimates of θ (see page 58). As shown in Figure 4.2, samples tt28 and tt67 show distinct bimodal length frequency distributions and have been split into subsamples. The italicised row simply labelled as "*(-ins)*" is data for pb67b but with the four individuals with insertions removed (thus the sample size falls from 14 to 10). Finally, the sample "pb all" gives details for all *P. bellottii* samples when treated as a single population.

As with *Trachurus* spp. (see page 81) there was considerable variation in values of nucleotide diversity and θ across all samples. The highest value, by some margin, of nucleotide diversity and θ was found at sample pb67b. However, if the four individuals with the insertion at position 67 are removed from the analysis (denoted in Table 4.2 as *"(-ins)"*) the nucleotide diversity falls to a value of 0.002, in line with the other samples, while the two estimates of θ (via π and via S) become 0.707 and 0.556 respectively. The variation between the two values of θ can largely be explained by variation in sample sizes between the various locations (see page 58 for methodological details)

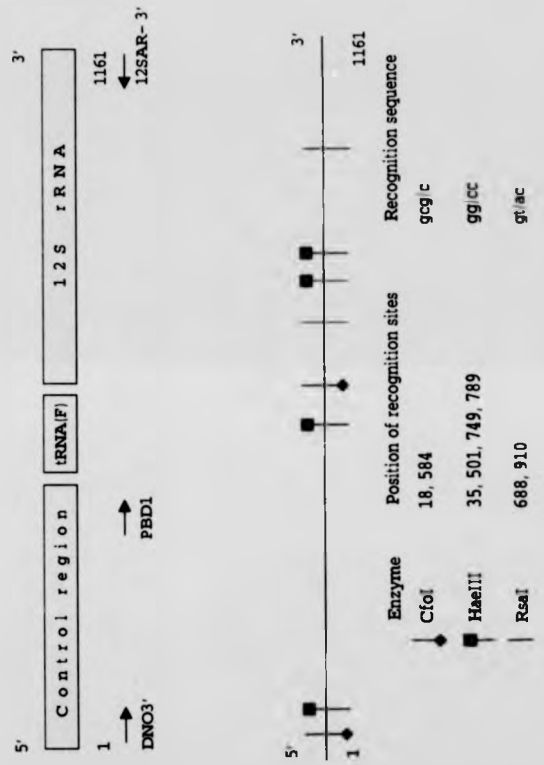

Figure 4.4: Genetic map, restriction sites and primer sites of the control region, tRNA(*Phe*) and 12S rRNA genes of *P. bellottii*. Primers were designed using the Primer3 Software found online. Details of the primers used can be found in section 2.4.3 on page 53. The three restriction enzymes shown are all four cutters. Full sequence can be found in Genebank with the accession code AF271659.

sample location	sample size	no. of polymorphic sites	nucleotide diversity	Tajima's D	θ_{ND} (via π)	θ_{SS} (via S)
pb28s	11	3	0.002	-1.600	0.545	1.024
pb28b	19	7	0.004	-1.311	1.216	2.003
pb67s	21	3	0.002	-0.677	0.619	0.834
pb67b	14	10	0.013	1.080	3.626	2.830
(-ins)	*10*	*2*	*0.002*	*-0.691*	*0.556*	*0.707*
pb84	15	7	0.005	-1.259	1.400	2.153
pbgu	8	7	0.007	-1.359	1.929	2.700
pb all	88	22	0.006	-1.830	1.590	4.159

Table 4.2: Sample location, sample size, the number of polymorphic sites, nucleotide diversity, Tajima's D and two estimates of θ calculated for all sampled locations of *P. bellottii*. Note the effect of the 4 individuals with inserts on the values of pb67b when they are included and when the are removed, denoted by "*(- ins)*". Without the 4 individuals, pb67b falls into line with the other samples for measurements such as nucleotide diversity and Tajima's D.

Values of Tajima's D are all negative, except for pb67b. However, if once more we remove the four individual's sequences that carry the insertion the value of D falls to -0.691, very similar to the sample of smaller individuals from the same catch, pb67s. Other than the case of pb67, all other values of Tajima's D are very similar, probably indicating either an expanding population after demographic effects, such as a population bottleneck or a founder effect, or because of recent selection pressures.

4.4.2 Minimum spanning haplotype tree

The 21 haplotypes found across all populations of *P. bellottii* were related by a minimum spanning tree (see page 68 for methodology) as shown in Figure 4.5. Links between pairs of haplotypes that have no dashes indicate a difference of one nucleotide substitution. Those haplotypes that are separated by two or more substitutions are indicated by a corresponding number dashes. The large circle corresponding to haplotype 1 indicates the most commonly found haplotype in all samples. As with *Trachurus* spp. (see page 84) there is a predominantly star shaped appearance to the tree, as predicted by the strongly negative values of Tajima's D in Table 4.2.

Most noticeable are the outlying haplotypes 13 and 14. These contain the insertion at position 67 and are only found in four individuals from sample tt67b (note that the

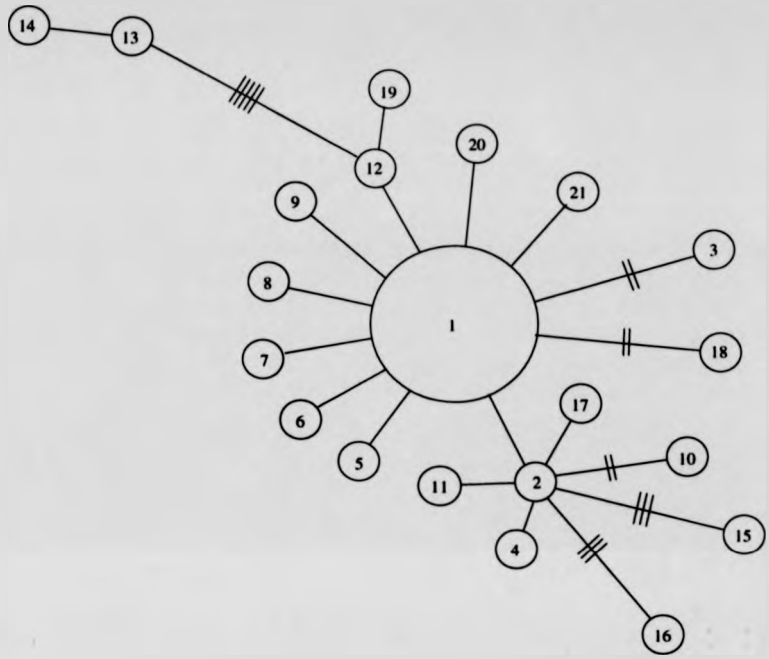

Figure 4.5: Minimum spanning tree for all haplotypes from *P. bellottii* populations from the Gulf of Guinea. A pair of attached haplotypes are separated by a single nucleotide difference, unless dashes are present across the line attaching the two haplotypes. Pairs of haplotypes that are separated by two or more nucleotide differences are marked with a corresponding number of dashes. Sequences of the haplotypes 1 and 13 can be found in Genebank with the accession numbers AF271656 and AF271657 respectively.

rest of the sequence of these four individuals also differs markedly from haplotype 1 and the rest of the samples). They account for much of the variation within and between samples. Sequences of the haplotypes 1 and 13 can be found in Genebank with the accession numbers AF271656 and AF271657 respectively[2].

A Fitch Margolish tree was constructed from all 21 haplotypes found in *P. bellottii* samples from the Gulf of Guinea (Figure 4.6). Similar to the minimum spanning tree shown in Figure 4.5, most of the haplotypes form a shallow structure, separated by just one or two substitutions. The haplotypes with insertions at position 67 (haplotypes number 13 and 14) stand clearly removed from all other haplotypes.

[2]http://ncbi.nlm.nih.gov

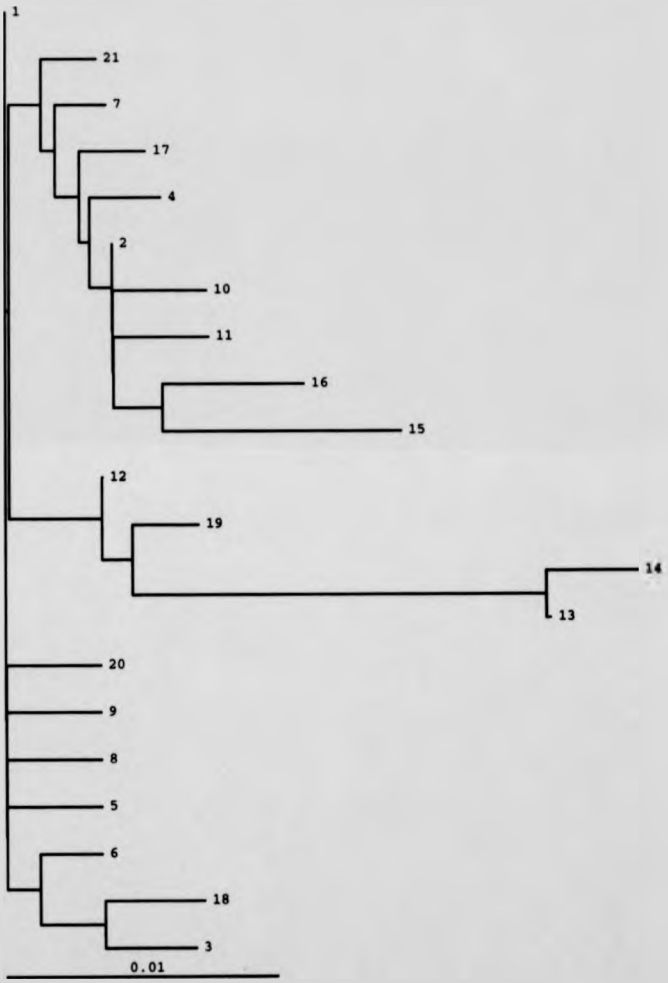

Figure 4.6: Fitch Margolish tree of all 21 haplotypes of *P. bellottii* from the Gulf of Guinea. All distances are additive according to percentage sequence divergence. See text for more details and section 2.10.2 for details of methodology.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
		for all samples		
Among populations	5	9.161	0.076	9.46
Within populations	82	59.987	0.732	90.54
Total	87	69.148	0.808	
Fixation index			F_{ST}: 0.095	$p = 0.002$
		all samples minus insertions		
Among populations	5	2.799	0.004	0.88
Within populations	78	38.915	0.499	99.12
Total	83	41.714	0.503	
Fixation index			F_{ST}: 0.009	$p = 0.286$

Table 4.3: Patterns of variation of a 288-bp fragment of the mitochondrial 12S rRNA gene of *P. bellottii* from the Gulf of Guinea. Two population structures were tested *(i)* for all samples and *(ii)* all samples minus the four individuals with insertions. For all samples, nearly 10% of the total variation is due to differences between samples. However when the four samples from pb67b with insertions are removed all significant variation is lost. Only 0.88% is left explicable due to variations between samples.

4.4.3 Analysis of Molecular Variance (AMOVA)

Genetic divergence between samples were investigated with an AMOVA (methodology on page 59). Two population structures were tested in the first instance, reflecting the outstanding influence that haplotypes 13 and 14 have on the genetic analyses. The two tests were:

- For all samples: all samples as designated in Table 4.2.
- All samples minus insertions: as with all samples but with the four individuals with insertions at position 67 removed from the analysis.

The results are given in Table 4.3. Nearly 10% of the total variation results from differences between populations across all samples, with an associated p value of 0.002 indicating a highly significant probability of non-random mating. However, it is immediately apparent that when the four samples from pb67b with insertions are removed, as in the second AMOVA shown in Table 4.3, all significant variation is lost. Only 0.88% is left explicable by variation between populations. The second AMOVA therefore shows no evidence of any population structuring across the whole of the Gulf of Guinea.

	pb28s	pb28b	pb67s	pb67b	pb84	pbgu
pb28s		0.043	0.029	0.154	0.027	0.013
pb28b	0.043		-0.010	0.182	-0.012	0.065
pb67s	0.029	-0.010		0.207	-0.012	0.063
pb67b	-0.026	-0.035	-0.039		0.147	0.055
pb84	0.027	-0.012	-0.012	-0.025		0.032
pbgu	0.013	0.065	0.063	0.022	0.032	

Table 4.4: Matrix of Pairwise F_{ST} values for all samples of *P. bellottii* from the Gulf of Guinea (above diagonal) and for all samples but without the four individuals with the insertions from pb67b (below diagonal). See text for more details.

4.4.4 F_{ST} pairwise comparisons

Pairwise F_{ST} comparisons are shown in Table 4.4. The figures above the diagonal give F_{ST} values for all samples, while those below the diagonal are the same samples but with the four individuals with insertions at position 67 removed from the analysis (as with the second AMOVA described in section 4.4.3).

Figure 4.7 shows a UPGMA tree using the F_{ST} pairwise comparisons of Table 4.4, taken from above the diagonal (i.e. values from all samples). Plainly the sample tt67b forms a distinct outgroup. Amongst the other samples, pb28s and pbgu form a separate cluster from the remaining samples.

However, if the four individuals with insertions from sample pb67b are removed and placed in a separate population of their own (labelled as "pb67ins"), it becomes quite apparent that they now form a highly distinct cluster (Figure 4.8). Sample pb67b, now without the four samples with insertions, returns to a position that falls well within the cluster of all other samples of *P. bellottii* from the Gulf of Guinea.

Leaving pb67ins out as a separate cluster, the remaining samples do not show clustering along temporal lines, as was found in *T. trecae* (see section 3.4.6), or along geographic lines, at least not in the central region of the Gulf of Guinea. However, the one sample that does lie somewhat removed from the rest of the (non-insert) samples, is pbgu. The western sample, pbgu, from The Republic of Guinea is of course geographically distant from the other samples, all of which are from the central Gulf of Guinea (see Figure 4.1). However, the sample size is small (see Table 4.2) and sampling bias may have caused its relative genetic isolation.

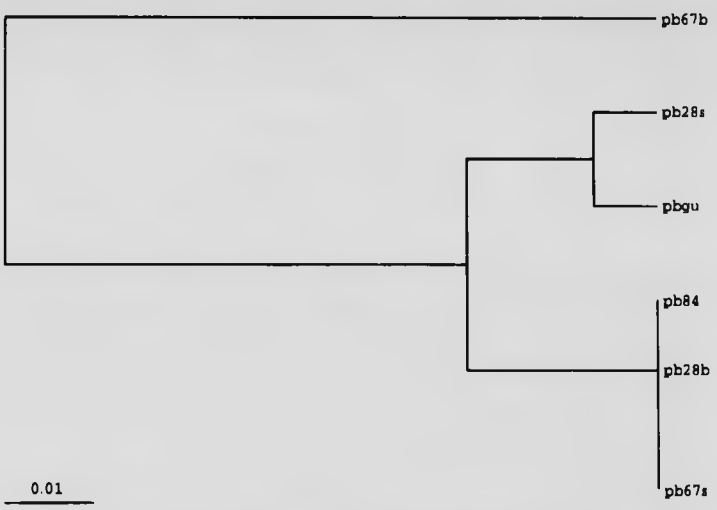

Figure 4.7: UPGMA tree of *P. bellottii* from F_{ST} pairwise comparisons (above diagonal) of Table 4.4. The sample tt67b forms a distinct outgroup when the four individuals with insertions are included (but see Figure 4.8).

4.5 Discussion

4.5.1 Pattern of genetic variation in *P. bellottii*.

The 12S rRNA gene used in the study of *P. bellottii* does not code for a protein and therefore has no reading frame and no synonymous and non-synonymous sites. The whole mitochondrial chromosome behaves as a single locus because there is usually no recombination and therefore total linkage (but see Hey, 2000); selection at other sites on the mitochondrial chromosome could therefore affect the calculation of coalescence times. An important point to be made here is that all samples sizes are small and hence results have to be viewed with extreme caution.

4.5.2 Relationships within LMEs

The amount of nucleotide diversity was small for *P. bellottii*, though there was a notable exception of four individuals from sample pb67b. These four individuals not only had an insertion in the 12S rRNA sequence, leading to a sequenced length of 289 base pairs,

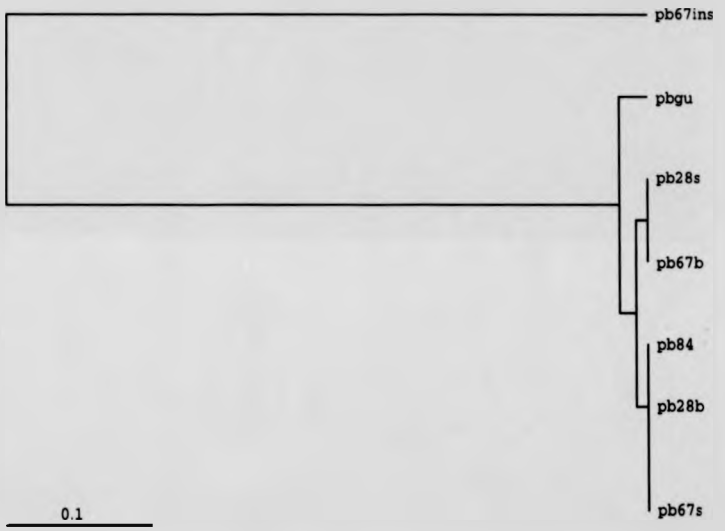

Figure 4.8: UPGMA tree of *P. bellottii* from pairwise comparisons when pb67ins is considered as a separate population (denoted pb67ins). When compared to Figure 4.7 it is immediately apparent that the four individuals with insertions caused sample pb67b to be an outgroup. If the four individuals with insertions are removed to a separate population, as in this figure, pb67b falls into line with the rest of the samples.

but also varied at a number of other nucleotide sites in the sequence (haplotypes 13 and 14 in Figure 4.5). Aside from the four exceptional individuals in pb67b there was little change in nucleotide diversity across samples. Values of Tajima's D, once the four exceptional individuals are accounted for, are found to be all negative, indicating a shallow branching tree indicative of a recent selective sweep or population bottleneck.

The pairwise F_{ST} values show no evidence of samples from the same trawl being more closely related to one another than to any other samples (see section 6.5.1 for a discussion on the genetic relationships within shoals, and compare with section 3.5.3). Equally there is little evidence of the structuring according to age class observed in *T. trecae*. Overall, the results show evidence of high gene flow throughout the sampled region supporting the view that there is one stock across the whole of the Gulf of Guinea (following Koranteng & Pitcher, 1987). There appears to be a small amount of isolation by distance given that The Republic of Guinea sample falls outside of the other Gulf of Guinea samples (see Figure 4.8). However, the small samples sizes, particularly of the Guinea sample (just 8 individuals), means that it cannot be deemed significant.

4.5.3 Four individuals - who are they?

The amount of variation in the four individuals in pb67b is clearly shown by the AMOVA results in Table 4.3. With the four individuals included there is a high probability of subpopulation structuring within the Gulf of Guinea, while without them there is no subpopulation structuring. What exactly these individuals are and where they are from is a question difficult to answer from such a small sample set. It is possible that *(i)* they are a rare, distinct subpopulation from within the Gulf of Guinea; *(ii)* they are *P. bellottii* migrants from a region outside of the LME; or *(iii)* they are another closely related species wrongly assigned as *P. bellottii* (see also section 6.5.3). Some suggestion as to other species of the same genera that could be mistaken for *P. bellottii* include *Pagellus erythinus* (found as far south as the Cape Verde Islands), *Pagellus bogaraveo* and *Pagellus acarne* (found as far south as the Canaries Islands); the presence of any of which in the Gulf of Guinea is yet to be confirmed (Anon, 1981). The relevant

sequences of the common haplotype and an extended haplotype can be accessed on the internet[3] and used in further studies of *P. bellottii*. Only through sample analysis of more individuals can identification of these four samples begin to be realised.

4.5.4 Management implications

As with *Trachurus* spp. in the previous chapter, management of *P. bellottii* needs to be directed at a regional, rather than a national level. There is no clear evidence for more than one population from Republic of Guinea waters to the Benin/Nigeria border. Although a recognised demersal species, with relatively little migratory potential, *P. bellottii* displays enough gene flow across the region to prevent the creation of genetically distinct subpopulations. Such a result may in part be due to the eggs or larvae being pelagic and dispersive, however, little is known of the early juvenile stages of *P. bellottii*. Good governance therefore needs to be aware of the wide population range of this and probably many other demersal species in the Gulf of Guinea.

[3]The common haplotype (haplotype 1) is accession number AF271656, while the extended haplotype (haplotype 13) is AF271657 - see figure 4.5. Database at http://ncbi.nlm.nih.gov

Chapter 5

Molecular Genetic Analysis of *Sepia* spp.

5.1 Preamble

Sepia spp. has been introduced in section 1.5.2 on page 28. Chapter 5 presents results of genetic analysis of *Sepia* spp. from the central and western regions of the Gulf of Guinea. In addition data is presented from analyses of *Sepia officinalis* taken from the east and west coasts of Spain. Microsatellite analysis of the same seven loci were used for all samples. Results indicate very little variation across the Gulf of Guinea LME, though there was evidence of temporal variation at one locus. Alternative analytical methods of quantifying the genetic distance between samples were used, using differing assumptions of the pattern of mutation at microsatellite loci. While all methods of analysis indicated highly significant differences between LMEs, there were some differences in the amount of variation estimated within the Gulf of Guinea LME. Alternative interpretations of the results are provided and discussed.

5.2 Sample collection

Chapter 2, page 46, gave rough locations of where samples for all species were taken. The two regions from which *Sepia* spp. samples have been obtained for this project are

the central (collected on the Nansen cruise, section 2.2.3) and western regions (collected on the Antea cruise, section 2.2.2) of the Gulf of Guinea. Table 5.1 gives details of all sample sizes, the co-ordinates and depth, and date of sampling for all stations from which *Sepia* spp. samples were taken.

The naming follows the same convention set out for *Trachurus* spp. on page 70. The first two letters indicate the probable species, *Sepia officinalis*, while the numbers indicate the station samples were collected from. For samples sogu93 and sogu94, from The Republic of Guinea, the further letters "gu" denote Guinea. Individuals were grouped into three main samples designated as:

1. GTB (for Ghana, Togo and Benin, i.e. samples from the east of Cape Three Points)
2. CdI (for Côte d'Ivoire, i.e. samples collected between Cape Three Points and Cape Palmas)
3. GUI (for Guinea, as collected by Eric Morize).

The sample designation column in Table 5.1 denotes the sample to which each station was designated. Grouping of samples in this fashion was required as most individual sample sizes were very small, because of small catches on the Nansen cruise (for details of the cruise see page 48). Figure 5.1 shows the locations of the sampling sites. The "meta" samples of GTB, CdI and GUI group together quite naturally either side of major physical features. Cape Three Points clearly separates samples GTB and CdI, while the samples from Guinea (GUI) form a third sample a substantial distance away from those collected on the Nansen cruise. These three samples will collectively be referred to as the west African samples.

Data from two outgroups from either side of the Iberian Peninsula are also used (data supplied by P. W. Shaw, University of Hull) to provide an outgroup to the results from the west African samples (GTB, CdI, GUI). The two outgroup samples are designated RIV and AL (denoting their sampling sites, Riveira and Alicante respectively) and henceforth will collectively be referred to as the Spanish samples. Riveira is off the

Station	sample size	Co-ordinates & Depth	Sample Designation	Date
sogu93	20	unknown	GUI	Sep/Oct 1998
sogu94	17	unknown	GUI	Sep/Oct 1998
so26	5	6°12'N 2°37'E 42m	GTB	May 1999
so27	7	6°08'N 2°35'E 71m	GTB	May 1999
so28	1	6°10'N 2°11'E 60m	GTB	May 1999
so35	6	6°03'N 1°25'E 45m	GTB	May 1999
so36	16	6°00'N 1°24'E 56m	GTB	May 1999
so37	47	5°53'N 1°02'E 25m	GTB	May 1999
so78	12	5°00'N 3°00'W 41m	CdI	May 1999
so97	4	4°57'N 5°22'W 91m	CdI	May 1999
so102	3	5°09'N 4°34'W 33m	CdI	May 1999
so105	12	5°08'N 4°10'W 83m	CdI	May 1999
so107	6	5°12'N 4°11'W 28m	CdI	May 1999
so109	1	5°07'N 3°45'W 41m	CdI	May 1999
so110	2	5°03'N 3°45'W 75m	CdI	May 1999

Table 5.1: Sample size, co-ordinates and depth of each station used for collection of *Sepia* spp. from the Gulf of Guinea. Samples from stations 93 and 94, off the coast of The Republic of Guinea, were collected by Eric Morize. All other samples were taken on the RV Dr. Fridtjof Nansen cruise. Letters in the "sample designation" column denote the "meta" samples to which individuals from each station were designated (see text).

north west corner of Spain while the Alicante samples are from the Mediterranean coast of Spain, near the Balearic Islands.

Four individuals from the west African samples showed highly atypical allele sizes compared to other samples from the region and thus have been placed in a separate group simply designated as "odd". In addition to their atypical allele sizes they also all failed to amplify at loci Sof 3, 4 and 6. Figure 5.2 shows length frequency histograms for Gulf of Guinea samples collected on the Nansen cruise, while Figure 5.3 shows *weight* frequency histograms for those samples collected from Guinean waters. Length frequency data were not collected on the the Antea cruise in Guinean waters so weight frequency data is used instead. Note that unlike the fish species of the previous two chapters there are no occurrences of distinct bimodal or unimodal length frequency distributions. *Sepia* have a life span of only one year (page 30) resulting in no overlapping year classes and no clearly unimodal or bimodal length frequency distributions.

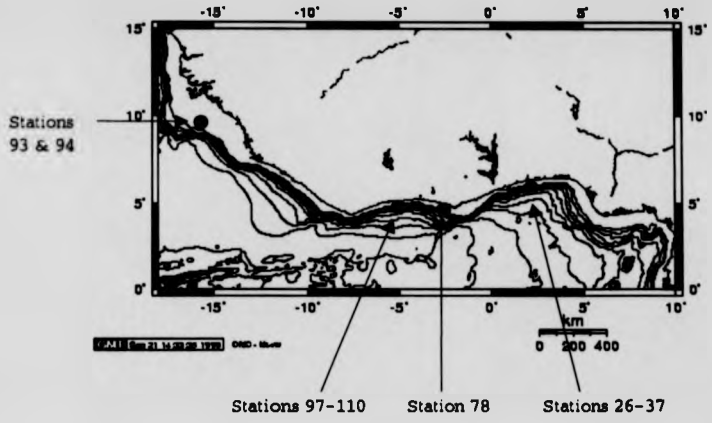

Figure 5.1: Diagram of collecting sites for *Sepia* spp. along the west African coast of the Gulf of Guinea. Samples from the western region were collected by Eric Morize. All other samples were collected on board the RV Dr. Fridtjof Nansen. The bathymetry is derived from the ETOPO5 data set, available from the National Geophysical Data Centre.

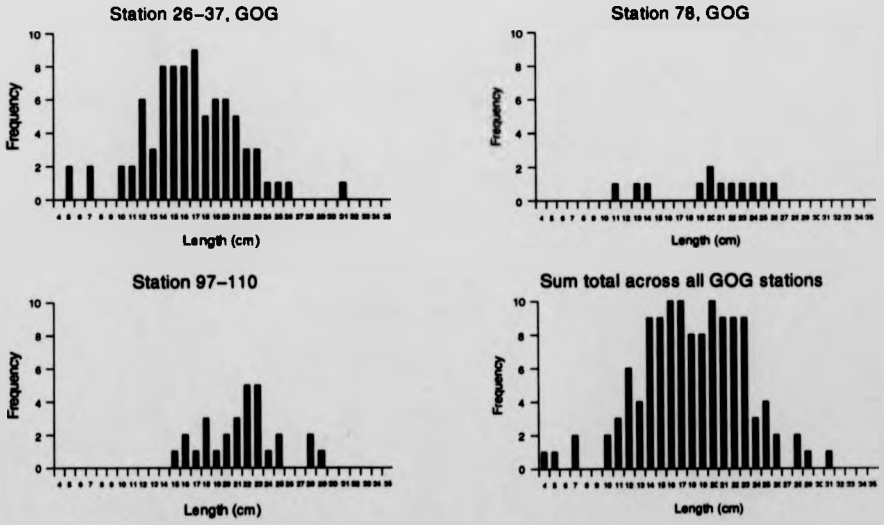

Figure 5.2: Length frequency histograms for *Sepia* spp. from the Gulf of Guinea as collected on the RV Dr. Fridtjof Nansen cruise. All axes are the same. *Sepia* have a life span of only one year resulting in no overlapping year classes and no clearly unimodal or bimodal length frequency distributions.

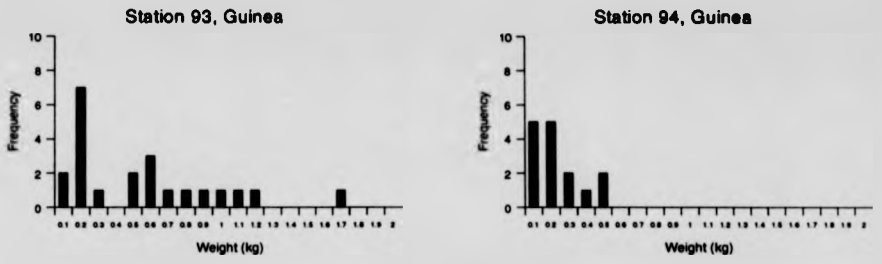

Figure 5.3: Weight frequency histograms for *Sepia* spp. from the Gulf of Guinea as collected by Eric Morize. All axes are the same. Please note however that as length data was not available the X axis is weight (kg) rather than length.

5.3 Microsatellite analysis

5.3.1 Molecular diversity

Table 5.2 displays information for all loci across all samples: Allele sizes, number of alleles (N_a), estimation of F_{IS} and the probability of the locus meeting Hardy Weinberg expectations (p) (see section 1.6.2). Differences in product sizes were consistent with all alleles at all loci representing a simple 3bp repeat motif, except for locus Sof 3. As described by Shaw & Perez-Losada (1999) and Perez-Losada *et al.* (2000) Sof 3 displays two parallel sets of allelic arrays differing by 1 bp: a "long" series and a "short" series of alleles. Such a difference is most likely to have arisen through the loss or gain, in some individuals, of a single nucleotide from the non-repeat region either side of the microsatellite. For ease these "long" and "short" series were combined.

Exact tests for genotypic disequilibrium (most likely resulting from linkage of alleles) between pairs of loci within samples resulted in 4 significant values ($p < 0.05$), after the Bonferroni sequential test, out of a table of 45 comparisons. The significant values were found for loci pairs Sof 1-Sof 3, Sof 1-Sof 4, Sof 3-Sof 6 and Sof 1-Sof 6, all in sample CdI. Therefore overall there is little evidence for a high degree of linkage disequilibrium between most of the screened loci, though a concentration of significant results is found in sample CdI, most particularly with loci Sof 1, 3 and 6.

Average gene diversity over all loci, shown in Figure 5.3 (see page 57 for methodol-

Locus	Samples				Mean (SD)
	GTB	CdI	GUI	odd	
Sof1 (ATT)					
Allele size	211-271	217-271	217-253	253-253	211-274
N_a	20	17	9	1	8.54
F_{IS}	0.018	0.030	0.200	-	0.100
p	0.290	0.060	0.227	-	0.120
Sof2(AAT)					
Allele size	130-139	130-139	130-136	157-187	130-187
N_a	4	3	2	6	1.71
F_{IS}	0.123	0.085	-0.110	-0.143	0.130
p	0.513	0.810	1.000	1.000	0.230
Sof3(AAT)					
Allele size	169-220	175-214	190-206	-	169-220
N_a	17	14	5	0	7.87
F_{IS}	0.351	0.506	0.434	-	0.080
p	0.000	0.000	0.153	-	
Sof4(ATT)					
Allele size	99-153	102-156	102-159	-	99-159
N_a	18	14	8	0	7.83
F_{IS}	0.482	0.347	0.492	-	0.080
p	0.000	0.000	0.007	-	
Sof5(ATT)					
Allele size	111-114	111-120	111-114	114-198	111-198
N_a	2	3	2	7	2.65
F_{IS}	0.217	0.149	0.407	-0.043	0.190
p	0.060	0.569	0.341	1.000	0.400
Sof6(AAT)					
Allele size	227-278	236-278	239-266	-	227-278
N_a	16	13	10	0	6.81
F_{IS}	0.182	0.175	0.023	-	0.090
p	0.000	0.004	0.817	-	
Sof7(ATT)					
Allele size	165-177	171-174	171-174	171-174	165-177
N_a	4	2	2	2	0
F_{IS}	-0.09	-	-	-0.5	0.29
p	1	-	-	1	0

Table 5.2: Levels of genetic variation observed at seven microsatellite loci from *Sepia* spp. samples taken from the Gulf of Guinea: Allele size (in base pairs); number of alleles (N_a); Weir & Cockerham (1984) estimate of F_{IS} and the p value of the test using the Guo & Thompson (1992) Markov chain algorithm. The p value is associated with the H_o (i.e. Hardy Weinberg equilibrium - see page 32).

Sample location	Sample size	Average gene diversity
GTB	81	0.705
CdI	40	0.663
GUI	9	0.593
odd	4	0.798
RIV	66	0.754
AL	65	0.614

Table 5.3: Average gene diversity and sample size for *Sepia* spp. samples from west Africa and Spain. Note the differing sample sizes.

ogy), for all populations displays little variation between samples, both west African and Spanish. The average gene diversity per locus per population (Figure 5.4) also shows relatively little variation between populations, though there is considerable variation between loci. Sof 1, 3, 4 and 6 all display high levels of gene diversity for most populations (values of around 0.9), while Sof 2, 5 and 7 show more variable levels of gene diversity (as low as 0.03 for CdI at loci Sof 7).

Between populations there are three particular features that are of note:

1. In general, variations in diversity between populations is split between west African and Spanish samples (e.g. at loci Sof 1, 3, 4 and 6 Spanish populations have lower diversity, while at Sof 2 and 5 they have greater diversity).

2. The loci for which there is data for the sample "odd" (Sof 2, 5 and 7), reveal greater similarity to the Spanish samples than to the west African samples.

3. For loci Sof 1, 5 and 7 the gene diversity of AL is notably lower than RIV.

5.3.2 Analysis of molecular variance (AMOVA)

To investigate the amount of genetic variation both among west African samples, and between them and the Spanish samples, an AMOVA (see page 59 for methodology) was used to reveal the pattern of genetic variation. For the AMOVA, the west African samples are all located within one LME, while the two Spanish samples, one being from the Atlantic coast and one from the Mediterranean, are allocated as being from separate LMEs (the Iberian Coastal LME and Mediterranean Sea LME, following Sherman 1994).

	GTB	CdI	GUI	odd	RIV	AL
Sof1	0.93	0.94	0.94	-	0.80	0.54
Sof2	0.39	0.41	0.50	0.88	0.83	0.84
Sof3	0.91	0.90	0.88	-	0.80	0.81
Sof4	0.92	0.92	0.98	-	0.68	0.82
Sof5	0.36	0.41	0.38	0.96	0.70	0.29
Sof6	0.91	0.91	0.91	-	0.82	0.87
Sof7	0.20	0.03	0.14	0.50	0.65	0.13

Table 5.4: Gene diversity per locus and population for *Sepia* spp. samples from west Africa and Spain. Three observations are made: *(i)* variations in diversity between populations is split between west African and Spanish samples, *(ii)* The sample "odd" shows greater similarity to the Spanish samples than to the west African samples and *(iii)* for loci Sof 1, Sof 5 and Sof 7 the gene diversity of AL is notably lower than RIV.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among LMEs	3	210.496	0.623	30.86
Among populations within LMEs	2	0.88	-0.015	-0.720
Within populations	524	738.810	1.410	69.870
Total	529	950.187	2.018	
Fixation index			F_{ST}: 0.301	$p = 0.000$

Table 5.5: Patterns of variation over all microsatellite loci of *Sepia* spp. from west African samples (GTB, CdI and GUI as one LME), and Spanish samples (Spanish populations assigned to separate LMEs). The F_{ST} of 0.3 indicates very great genetic differentiation amongst all groups. The negative value (effectively zero) for percentage of variation among populations within LMEs displays a lack of differentiation within the Gulf of Guinea.

The results of the AMOVA across all samples are shown in Table 5.5. The among samples, within LMEs component of genetic variation, -0.015, indicates no genetic variation within the west African samples. However there is very significant variation between the three LMEs ($F_{ST} = 0.3$).

Therefore, as with *Trachurus trecae* and *Pagellus bellottii* there is very little population differentiation evident within the Gulf of Guinea samples, but very significant divisions between the Gulf of Guinea and European populations.

	GTB	CdI	GUI	odd	RIV	AL
GTB		-0.002	0.002	0.444	0.351	0.413
CdI	-0.002		-0.007	0.451	0.343	0.413
GUI	-0.008	-0.015		0.335	0.296	0.405
odd	0.444	0.451	0.335		0.110	0.286
RIV	0.248	0.249	0.223	0.110		0.279
AL	0.255	0.250	0.249	0.286	0.179	

Table 5.6: Pairwise estimates of F_{ST} for all populations of *Sepia* spp. (i.e. west African and Spanish samples). Below diagonal are pairwise estimates using all loci, above diagonal shows pairwise estimates using only loci Sof 1, 2, 5 and 7. Loci not used, Sof 3, 4 and 6, are those that show signs of possible linkage disequilibrium.

5.3.3 F_{ST} pairwise comparisons

Pairwise F_{ST} estimates for all pairs of samples are given in Table 5.6. The below diagonal figures are pairwise F_{ST} comparisons for all samples using all loci. Above diagonal figures are pairwise F_{ST} comparisons using only loci Sof 1, 2, 5 and 7. The loci not used, Sof 3, 4 and 6, are those which show signs of possible linkage disequilibrium (see Table 5.2 on page 119).

Figure 5.4 shows a UPGMA tree (see page 66 for methodology) calculated from the F_{ST} pairwise comparisons displayed in Table 5.6. Deep divergence is clearly shown between west African and Spanish populations. Within the two clusters, further divergence is seen between the Spanish populations, as described in detail in Perez-Losada *et al.* (2000), but no divergence is found between the Gulf of Guinea populations. As with results for *T. trecae* and *P. bellottii* in the previous chapters, *Sepia* spp. displays no evidence of spatial population structure within the Gulf of Guinea.

As with the data on genetic diversity in Table 5.4, the four individuals classified as "odd" from west Africa show greater similarity to the Spanish samples than to other west African samples, and as such clearly cluster with the Spanish populations (most notably the RIV sample, from the Atlantic coast of Spain).

5.3.4 R_{ST} pairwise comparisons

Table 5.7 gives the results from tests of R_{ST} pairwise comparisons (based on the single stepwise mutation model - SMM). The results indicate that the west African samples are

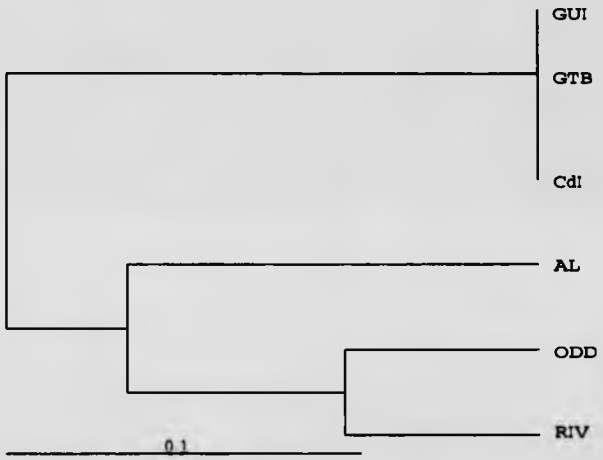

Figure 5.4: UPGMA tree of *Sepia* spp. F_{ST} pairwise comparisons from Table 5.6. Divergence is clearly shown between west African and Spanish populations. Within the two clusters, further divergence is seen between the Spanish populations, but none is found between the Gulf of Guinea populations. Individuals classified as "odd" from west Africa show greater similarity to the Spanish samples than to other west African samples.

	GTB	CdI	GUI	odd	RIV	AL
GTB		0.001	0.0567	-	0.635	0.636
CdI	0.013		0.0219	-	0.651	0.641
GUI	-0.019	-0.015		-	0.645	0.605
odd	0.719	0.740	0.737		-	-
RIV	0.792	0.818	0.810	0.199		0.406
AL	0.848	0.896	0.883	0.305	0.534	

Table 5.7: Pairwise estimates of R_{ST} for all populations of *Sepia* spp. (i.e. west African and Spanish samples). Below diagonal are pairwise estimates using only loci Sof 2, 5 and 7 (i.e. the loci for which data is available for the "odd" sample), while above diagonal shows pairwise estimates using all loci. Over all loci the west African samples are similarly distant to the two Spanish samples and the GUI sample from the western Gulf of Guinea shows some variation from the two central Gulf of Guinea samples. Loci Sof 2, 5 and 7, shows the "odd" sample closest to RIV on the Atlantic coast of Spain.

similarly distant to the two Spanish samples AL and RIV. It is notable that under the SMM model the GUI sample from the western Gulf of Guinea shows some variation from the two central Gulf of Guinea samples (GTB and CdI). In comparison, the non-SMM F_{ST} results displayed in Table 5.6 show no such differentiation.

Results of R_{ST} pairwise comparisons using only loci Sof 2, 5 and 7 (i.e. loci for which there is data for the "odd" sample), show the west African samples GTB and CdI sharing a slightly closer affinity to the RIV sample, taken from the Atlantic coast. However, the effect is very small. The "odd" sample also shows a more notable affinity to RIV than to AL. However, the small sample size of "odd" may mean that sampling error has a significant effect (with reference to sample size in microsatellite studies see Ruzzante, 1998).

5.3.5 $(\delta\mu)^2$ pairwise comparisons

Table 5.8 gives values for $(\delta\mu)^2$ for all loci, without the population "odd", above the diagonal. The "odd" sample was not included in the analysis of all loci as they did not successfully amplify for loci Sof 3, 4 and 6. Results using only loci Sof 2, 5 and 7 (loci for which there is data available for the "odd" sample) are shown below the diagonal.

Unlike the R_{ST} pairwise comparisons, the results across all loci indicate that the west African samples are slightly more closely related to the Spanish sample AL than

	GTB	CdI	GUI	odd	RIV	AL
GTB		0.156	2.062	-	46.438	45.357
CdI	0.006		1.278	-	45.833	43.636
GUI	0.054	0.037		-	45.263	38.519
odd	97.686	97.694	100.787		-	-
RIV	82.760	82.996	86.091	3.250		7.893
AL	65.895	66.324	69.387	17.094	6.200	

Table 5.8: Pairwise estimates of $(\delta\mu)^2$ for all populations of *Sepia* spp. (i.e. west African and Spanish samples). Below diagonal are pairwise estimates using only loci 2, 5 and 7, while above diagonal shows pairwise estimates using all loci. Results across all loci indicate that the west African samples are more closely related to the Spanish sample AL, the Mediterranean coast sample. However, the effect over all loci is very small. In common with results from R_{ST} pairwise comparisons, there is some differentiation between central and western Gulf of Guinea samples. Results using only loci 2, 5 and 7 also show the west African samples are more closely related to the Spanish sample AL. In agreement with R_{ST} and F_{ST}, the "odd" sample shows a closer affinity to RIV.

RIV, i.e. the sample taken from the Mediterranean coast. However, the effect over all loci is again very small. Although $(\delta\mu)^2$ is not thought to be so good at distinguishing closely related populations, it is worth noting that, in common with results from R_{ST} pairwise comparisons, it indicates greater difference between central Gulf of Guinea samples (GTB and CdI) and the western Gulf of Guinea sample (GUI), than was found with F_{ST} pairwise comparisons. However, as with the sample "odd" the small sample size of GUI may mean that sampling error has a significant effect.

Results using only loci Sof 2, 5 and 7 (loci for which there is data available for the sample designated "odd"), shown below the diagonal in Table 5.8, the results show again AL (the Mediterranean sample) as being closer to the west African samples. Also it is interesting to note that, this time in agreement with the R_{ST} and F_{ST} measurements, the "odd" sample shows a far closer affinity to RIV than AL (again, with reference to sample sizes in microsatellite studies, see Ruzzante, 1998).

5.3.6 Temporal variability

To investigate whether there was any evidence of genetic variation between varying length classes of the west African *Sepia* sp. samples, individuals from populations GTB and CdI were re-assigned to one of four new samples on the basis of length. The four

samples contained individuals with a mantle length of:

- up to 13cm mantle length
- from 14 to 18cm mantle length
- from 19 to 23cm mantle length
- greater than 24cm mantle length

These length classes were chosen somewhat arbitrarily, with the aid of the length frequency histograms on page 117. The assigned populations of GTB and CdI were retained to the extent that for each length class, e.g. the shortest length class (up to 13cm mantle length), there were two subsamples, one of small *Sepia* individuals from GTB and one of small *Sepia* individuals from CdI. In total, therefore, there are 8 samples in the new length based classification.

An AMOVA was then run to investigate the amount of variation between these four length classes. The results, shown in Table 5.9, indicate a small but significant ($p = 0.023$) structure among the length classes. Certainly the above population structure based on size classes explained more of the genetic variation than did the geographic population structure within the Gulf of Guinea (for comparison see Table 5.5 on page 121).

Note that the relevant test statistic is F_{CT} because we are testing the probability by permuting populations among groups (σ_a^2 in section 2.8.1) as opposed to the usual case in this thesis of permuting haplotypes among populations among groups (σ_c^2 in section 2.8.1, giving rise to the F_{ST} test statistic). This is because the variation we are interested in is between length classes, rather than amongst populations within length classes.

Following the result of the AMOVA in Figure 5.9 pure length classes are only used for the following analyses. For example, all *Sepia* individuals of up to 13cm mantle length are placed in one length class, and are now *not* subdivided according to whether they are from the geographic samples of GTB or CdI.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among classes	3	7.787	0.018	0.93
Among populations within classes	4	6.895	-0.010	-0.51
Within populations	234	454.748	1.943	99.58
Total	241	469.430	1.952	
Fixation index			F_{CT}: 0.009	$p = 0.023$

Table 5.9: Patterns of variation over all microsatellite loci of *Sepia* sp. from the Gulf of Guinea when split into 4 length classes, defined in the text (up to 13cm; 14 to 18cm, 19 to 23cm; 24cm and over). The fixation index of 0.009 indicates low, but significant, genetic differentiation amongst all groups. More variation is explained due to length classes than to geographic differentiation.

To investigate further which loci may be causing this genetic variance amongst age classes F_{ST} pairwise comparisons were run for each pair of length classes for each loci. The results are shown in Table 5.10, and graphically displayed in Figure 5.5. Immediately apparent from Figure 5.5 is that effectively all of the variation observed between length classes is from one locus: Sof 5. The other loci show no relationship between length class and genetic distance (measured as F_{ST}). To test the significance of the relationship an AMOVA was used to assign the amount of variation resulting from differences between length classes in loci Sof 5. The significant result found nearly five percent of the variation resulted from variation between length classes (F_{ST}: 0.046, $p = 0.032$).

Upon examination it is found that locus Sof 5 is dominated by two allele lengths: 111bp and 114bp. It is a shift in the relative amounts of these two allele lengths that cause the observed variances. Figure 5.6 is a schematic diagram of the change in frequency of the 111bp allele, correlated with the sample's length class. The individuals from the youngest length class, i.e. those with up to 13cm mantle length, have a frequency of 0.86 for the 111bp allele, which falls to a frequency of 0.54 for the longest length class, i.e. individuals of 24cm mantle length and over.

5.3.7 Distribution of allele sizes

The allele size ranges found for all samples at each loci are shown in Figure 5.7. They further illustrate the similarity between west African "odd" individuals and the Spanish

	length class (cm)		
	1 to 13	14 to 18	19 to 23
Sof1			
14 to 18	0.008		
19 to 23	-0.004	-0.000	
24 +	-0.022	0.014	0.006
Sof2			
14 to 18	-0.013		
19 to 23	-0.004	0.010	
24 +	-0.017	-0.016	-0.020
Sof3			
14 to 18	-0.000		
19 to 23	0.003	-0.004	
24 +	0.014	-0.002	0.010
Sof4			
14 to 18	0.001		
19 to 23	0.010	0.014	
24 +	0.001	0.009	0.010
Sof5			
14 to 18	-0.014		
19 to 23	0.032	0.015	
24 +	0.202	0.161	0.049
Sof6			
14 to 18	0.003		
19 to 23	-0.002	0.005	
24 +	0.026	-0.003	-0.020
Sof7			
14 to 18	-0.018		
19 to 23	-0.015	0.003	
24 +	-0.037	-0.027	-0.008

Table 5.10: F_{ST} pairwise comparisons between length classes of *Sepia* sp. from the Gulf of Guinea for loci Sof 1 to Sof 7. Length classes are in cm and refer to the mantle length of the *Sepia* individual. See text for more details of length classes used. The only locus for which there appears to be a relationship is Sof 5. See Figure 5.5 for a graphical representation of the above data.

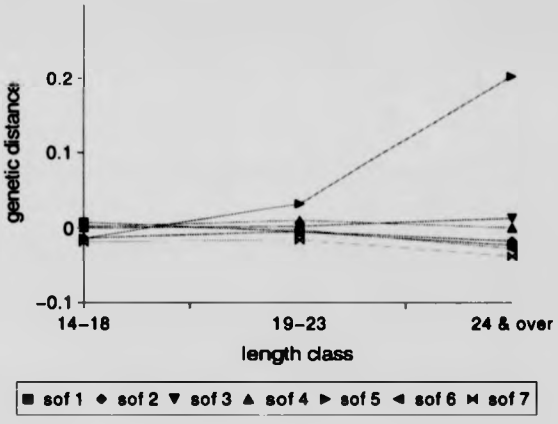

Figure 5.5: Plot of the correlation between length class and genetic distance (data from Table 5.10) for all loci of *Sepia* sp. from the Gulf of Guinea. Locus Sof 5 clearly has a positive correlation while all other loci show no relationship. All of the variation observed between length classes, as shown in Figure 5.9, is due to one locus: Sof 5. See text for more details of length class construction.

samples (for previous evidence of similarity see Tables 5.4, 5.6, 5.7, and 5.8, and Figure 5.4 and associated text). In particular, loci Sof 2 and Sof 5 display far smaller allele sizes in west African samples than in both the Spanish samples and the "odd" west African sample.

5.4 Discussion

5.4.1 Pattern of genetic variation in *Sepia* spp.

Departures from Hardy-Weinberg expectations, as measured by F_{IS}, were found at two loci within most samples (Sof 3 and 4, see Table 5.2), and observed occasionally at other samples and loci. The most likely cause of such results are null alleles[1] (Perez-Losada *et al.*, 2000). Positive F_{IS} samples can also be explained by a degree of inter-sampling genetic differentiation (i.e. a Wahlund effect, wherein the average homozygozity decreases because subpopulations are sampled together). *Sepia* exhibit complex mating behaviour

[1]Null alleles are alleles that are non-amplifying, most often because of mutation at one or both primer sites

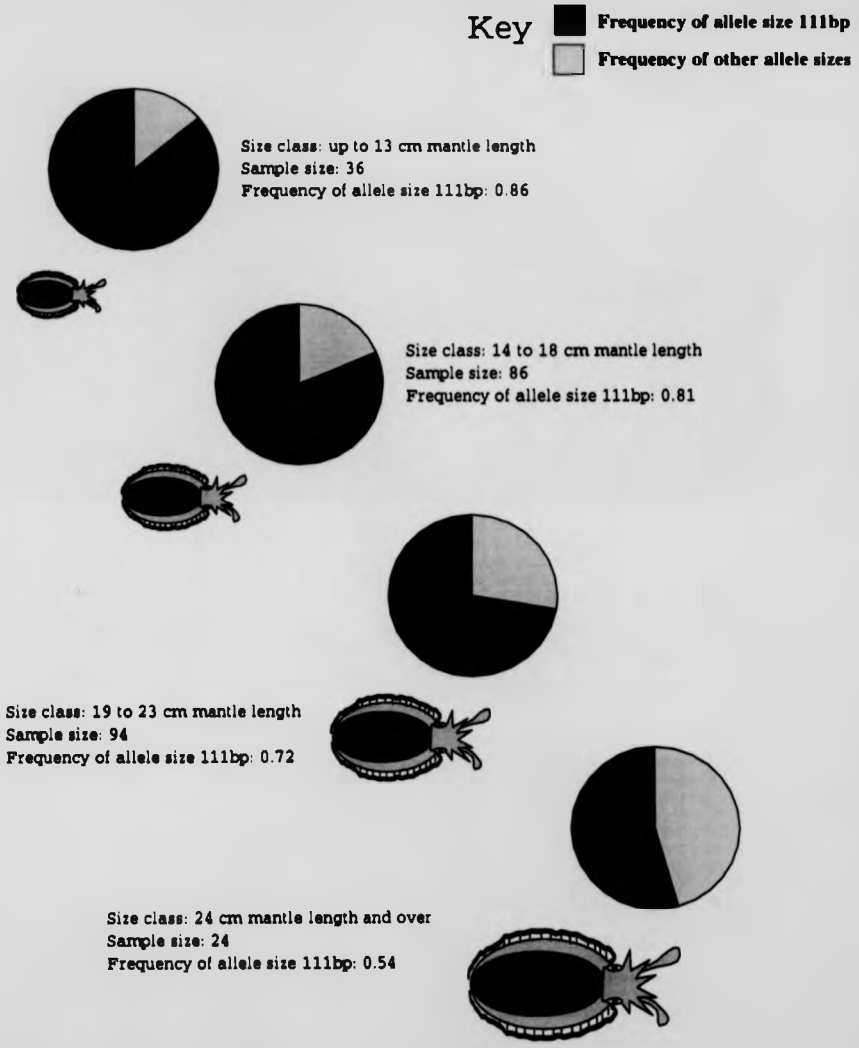

Figure 5.6: Schematic diagram of the change in allele frequency of Sof 5 across age groups in *Sepia* sp. individuals from the Gulf of Guinea. Locus Sof 5 is dominated by two allele lengths: 111bp and 114bp. A shift is apparent in the relative amounts of these two allele lengths, and causes the variance seen in Figure 5.5.

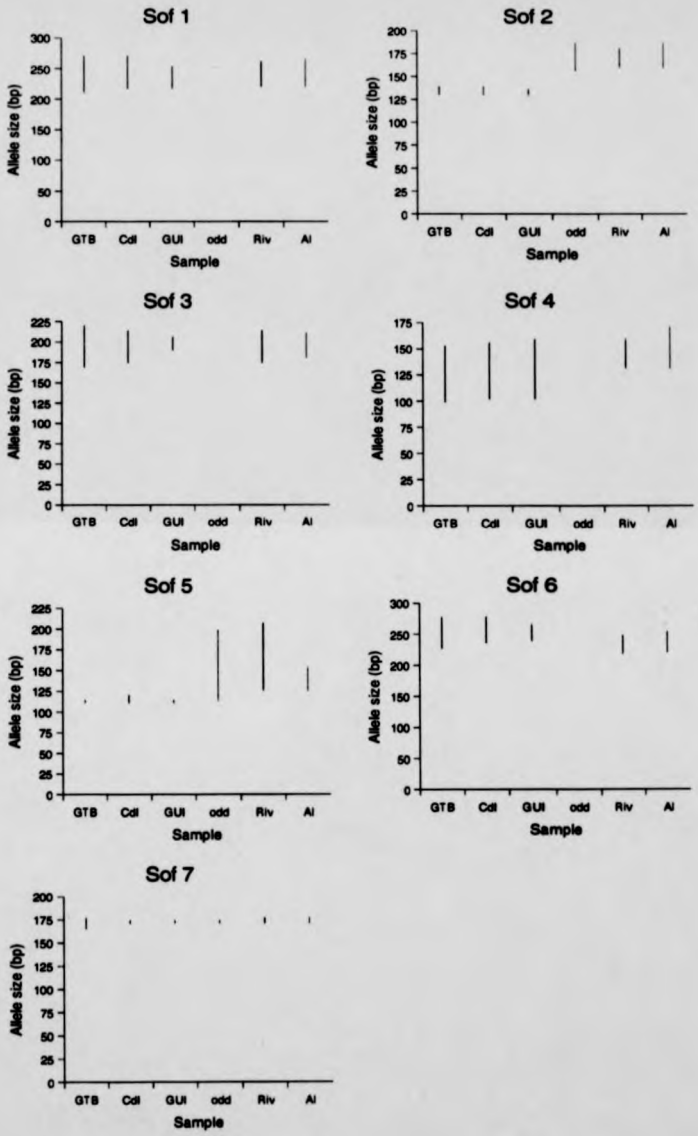

Figure 5.7: Allele sizes found in analysis of *Sepia* spp. microsatellites loci Sof 1-7, from the west African and Spanish samples. Allele sizes are given in base pair length of the PCR product.

whereby males compete for females (Hanlon & Messenger, 1996) and hence mating is not considered random (i.e. non-assortative), possibly also leading to the positive measures of F_{IS}.

Linkage disequilibrium was found only in CdI, most noticeably with Sof 1, 3 and 6. The possible linkage should be born in mind, though when removed from analysis of population substructure they did not significantly bias patterns of genetic relatedness (see Table 5.6). Perez-Losada *et al.* (2000) state that it is unlikely that the loci are physically linked.

5.4.2 Relationships between and within LMEs

High levels of polymorphism were displayed at all seven loci. The results tie in well with the three previous studies of microsatellite variation in cephalopods (see Shaw & Perez-Losada, 1999; Shaw *et al.*, 1999; Perez-Losada *et al.*, 2000) which also found markedly higher molecular diversity with microsatellites than with other previously used markers (e.g. see Boyle & Boletzky, 1996). Generally, levels of diversity within west African and Spanish populations were similar. However there were fluctuations in diversity across samples and loci, most noticeable in Sof 1, 2 and 4. Spanish populations displayed noticeably higher levels of diversity than their west African counterparts at Sof 2, while at Sof 1 and 4 the west African samples showed greater diversity.

Population differentiation is marked across all populations. There is a deep evolutionary division between European and west African samples. However there appears to be no population subdivision within the Gulf of Guinea as measured with F_{ST}, shown in Figure 5.4. Calculations of genetic distance show slightly less genetic differentiation between the west African samples and the Spanish sample from the Atlantic coast (RIV), than to the sample from the Mediterranean (AL).

R_{ST} pairwise comparisons (based on the single stepwise mutation model - SMM) however show the GUI sample, from the western Gulf of Guinea, as being somewhat more distinct from the two central Gulf of Guinea samples (GTB and CdI). The west African samples GTB and CdI share a very slightly closer affinity to the RIV sample,

taken from the Atlantic coast, while the "odd" sample shows a more notable affinity to RIV than to AL. However (Ruzzante, 1998) notes that when R_{ST} is used samples sizes should be about equal, or else they should all be moderately large ($N \geq 50$); conditions which are not met with regards the "odd" sample.

With $(\delta\mu)^2$ pairwise measurements, unlike the R_{ST} pairwise comparisons, the results indicate that the west African samples are slightly more closely related to the Mediterranean Spanish sample AL, rather than the Atlantic sample RIV. However, the effect over all loci is again very small. $(\delta\mu)^2$, in common with results from R_{ST}, also indicates some variation between central Gulf of Guinea samples (GTB and CdI) from western Gulf of Guinea samples (GUI) (though Ruzzante (1998) showed that $(\delta\mu)^2$ is biased at small sample sizes). Results using only loci 2, 5 and 7 show again AL (the Mediterranean sample) as being closer to the West African samples. Also in agreement with R_{ST} and F_{ST} measurements, the "odd" sample shows a far closer affinity to RIV than AL.

Overall therefore there appears to be about equal genetic distance between the west African samples and the two Spanish samples. The four "odd" individuals however consistently show some evidence of closer relationship to the Atlantic *Sepia* samples from RIV, using all pairwise distance measures. Measures using the stepwise mutation model, $(\delta\mu)^2$ and R_{ST}, also show evidence of genetic difference between the western Gulf of Guinea sample and the central samples, though F_{ST}, using the infinite alleles model (IAM), found no such difference. The appropriate mutation model (i.e. IAM or SMM) is probably species specific (O'Connell & Wright, 1997) and as yet it is unknown which is more suitable for *Sepia* spp. O'Connell & Wright (1997) recommend that given the uncertainty regarding the relative role of mutation models at present it is safer to take the conservative approach and use conventional F-statistics.

5.4.3 Who are the "odd"?

The pattern that emerges from all pairwise comparisons is that the "odd" population displays closer affinity to Spanish samples than to west African samples, both in measurement of genetic distance and in comparative sizes of alleles. In particular the "odd"

sample shows greatest similarity to RIV, the Spanish sample on the Atlantic coast, rather than AL from the Mediterranean. However, the sample size of merely four individuals is obviously insufficient to say much of them, apart from the fact they appear to be at least another subspecies. It could be hypothesised, by their relatedness to Spanish samples, that they are *S. officinalis officinalis* samples, while the mainstay of the Gulf of Guinea samples are the subspecies *S. officinalis hierredda*. However, if the odd samples are true *S. officinalis officinalis* then all loci would be expected to amplify successfully, not just loci Sof1, 2 5 and 7.

Therefore, as with *P. bellottii* in the previous chapter, there are a number of possibilities: *(i)* they are a rare, distinct subpopulation from within the Gulf of Guinea; *(ii)* they are *S. officinalis* migrants from a region outside of the LME; *(iii)* they are a separate subspecies, less well represented in the Gulf of Guinea than the common *S. officinalis hierredda*; or *(iv)* they are another closely related species wrongly assigned as *S. officinalis* (possibilities include *Sepia orbignyana* and *Sepia elegans*). With the current data it is not possible to distinguish between the alternatives (see also section 6.5.3), though loss of microsatellite loci would usually indicate a separate species.

5.4.4 Is *Sepia* spp. in the Gulf of Guinea *Sepia officinalis*?

S. officinalis is often named *S. officinalis hierredda* in the Gulf of Guinea, and as such classified as a distinct subspecies of *S. officinalis*. True *S. officinalis*, e.g. the Spanish samples, are thus classified as *S. officinalis officinalis*. Samples from the Gulf of Guinea do indeed show distinct genetic subdivision from the Spanish samples. However, regional subpopulations are common in most species (Avise & Walker, 1999), without the additional classification of new subspecies. Therefore the question is whether there is enough variation within *Sepia officinalis* to warrant a new subspecies, *S. officinalis hierredda*, in the Gulf of Guinea.

Microsatellites are not the ideal marker to use for such questions of speciation, given their high mutation rate and still relatively unknown mutation mechanism (particularly at speciation). Sequencing studies are far more common (see Avise & Walker, 1999),

particularly of mtDNA, and as such easier to analyse. Hence the microsatellite results presented here are by themselves insufficient to answer the question whether or not *Sepia* in the Gulf of Guinea is merely a geographically isolated population of *S. officinalis*, or whether they deserve to be designated as a new species/subspecies.

5.4.5 Is there selection at locus Sof 5?

Locus Sof 5 showed low gene diversity in the west African samples (see Table 5.4) and was found to be dominated by two allele lengths, 111bp and 114bp. Figure 5.6 clearly shows the change in the relative frequencies of these two alleles, correlated to the mantle length of the sampled *Sepia* individuals. Whilst such a change may result from chance (significance tests are not truly significant as there was not a proposed test of a hypotheses, merely a chance observation), or because of low N_e, there is also the possibility that selection is operating at that locus.

Selection has been proposed previously with regard to microsatellite DNA though mostly on the basis of mutational bias across species (e.g. Dover, 1995; Rubinsztein *et al.*, 1995). The function(s), if any, of microsatellite DNA are unknown, and they are currently often referred to as "selfish DNA" (Majerus *et al.*, 1996). However a number of human diseases, for example Huntington's disease, are known to result from abnormal expansion of a block of triplets most often, but not always, in the reading frame of a gene (Willems, 1994). Therefore the length of the repeat may be acted on by selection (though most of the diseases associated with triplet repeats in humans manifest themselves late in life, after reproduction; Majerus *et al.* 1996). As such it is not possible to propose what selection, if any, may be occurring; however further studies using the primer sets should be aware of the possible temporal selection at locus Sof 5.

However, an important point to be considered is that some *Sepia officinalis*, mostly males, may live up to 3 years and obtain mantle lengths of over 40cm (Roddy Williamson, pers. comm.). Hence the largest individuals (in Figure 5.6) may be more than one year old and therefore from a previous year class, and hence differing allele frequencies could be a result of drift (i.e. selective sweepstakes, see section 6.5.4) rather than selection. It

is noticeable that the biggest change in relative frequencies of the two alleles at Sof 5 is in the largest length class (24cm mantle length and over).

5.4.6 Management implications

As with the previous two chapters, on *Trachurus* spp. and *P. bellottii*, there is no clear evidence for subpopulations within the sampled region of the Gulf of Guinea, from The Republic of Guinea to the Benin/Nigeria border. Although a demersal species with relatively little migratory potential and extremely short life span, the results indicate sufficient rates of gene flow across the region to prevent the creation of genetically distinct subpopulations. Therefore, as with the other two species in the project, management should be focused at the regional level of the Gulf of Guinea.

However, following the precautionary principle, as introduced on page 97, and given the very small sample size from the western Gulf of Guinea, it can be argued that *Sepia* from the western subsystem should be managed separately (until or unless further evidence of high gene flow between the central and western subsystems is presented).

Chapter 6

Comparative Molecular Ecology and Evolution

6.1 Preamble

Some comparative issues that stretch across the individual species chapters are explored. The first section tests the neutrality of markers, and leads to the conclusion that selection pressures do appear and that they vary across regions. The nature of the variation at each marker is explored along with the feasibility of future mixed stock analyses. The temporal variation and number of single haplotypes warns that mixed stock analysis would be a very problematic and unsafe process with the current markers. The partial sequence of the control region, tRNA(*Phe*) and 12S rRNA genes of *P. bellottii* and *Trachurus* spp. are compared. The second part of the chapter deals with ecology and behaviour, including genetical relationships within and between shoals, recruitment and metapopulations. Finally there is some discussion of what can be learned about the Gulf of Guinea as a whole.

6.2 Tests of neutrality

6.2.1 Tajima's and Fu and Li's estimators

The methodology of Tajima's test statistic D (Tajima, 1989b) is given in section 2.9.1. The value of D reflects the shape of the gene tree and whether there are signs of either balancing selection or a relatively recent selective sweep or bottleneck. The negative values of D indicate a selective sweep or bottleneck and consequent non-neutral mutation ($p < 0.01$) for both *T. trachurus* and *T. trecae* in African waters, as well as *P. bellottii*.

The test statistic of Fu & Li (1993) similarly reflects on the shape of the gene tree. As with Tajima's D, the results of the test indicate a high probability of non-neutral mutation ($p < 0.02$) for both *T. trachurus* and *T. trecae* in African waters. The test statistics complement one another and indicate a selective sweep or a population bottleneck.

However, while there is no doubt about the shape of the tree, there is a problem with the interpretation of these test results. While the high number of recent mutations could result from a selective sweep, it could alternatively be explained not by deviations from predictions of neutrality, but rather by demographic conditions such as population bottlenecks or founder effects. A test of neutrality that could give some indication as to the more likely causal factor is the McDonald & Kreitman (1991) test.

6.2.2 McDonald and Kreitman test

The methodology of the McDonald & Kreitman (1991) test is given on page 66 in section 2.9.2. As it compares synonymous and non-synonymous sites it cannot be used on the 12S rRNA gene used for the genetic analysis of *P. bellottii*, as it does not code for a protein and thus has no synonymous or non-synonymous sites. However, it can be used with regards to the cytochrome *b* gene used for analyses of *T. trecae* and *T. trachurus*. A test was run between *T. trecae* and *T. trachurus*, however because there were zero fixed substitutions between the two species the contingency table, and hence test statistic, could not be calculated. Therefore a data set of seven individuals of *Tropheus moorii*

	Differences	
Type of change	Fixed	Polymorphic
Synonymous	16	36
Replacement	2	23

Table 6.1: McDonald Kreitman test: Gulf of Guinea. Numbers of replacement and synonymous fixed differences between species and of polymorphisms witin species. Assuming that the same coding region between species have the same evolutionary history, and assuming neutral evolution, the ratio of replacement to synonymous substitutions within species should be the same as those between species. A significant difference can therefore reject the neutral mutation hypothesis.

	Differences	
Type of change	Fixed	Polymorphic
Synonymous	26	16
Replacement	2	15

Table 6.2: McDonald Kreitman test: Benguela LME. Numbers of replacement and synonymous fixed differences between species and of polymorphisms witin species. Compare with Table 6.1

(a cichlid) was taken from the National Centre for Biotechnology Information website[1] and used as the second data set for each species of *Trachurus*. The results are given in Tables 6.1 and 6.2.

The p value of Fishers exact test for the Gulf of Guinea and Benguela LME samples both indicate significance (Gulf of Guinea: $p = 0.04$; Benguela LME: $p = 0.00$). Of course, neither of these results can claim statistical significance as they are not results from a planned experiment to test the neutral theory of evolution. Nevertheless, the low p values are difficult to explain under the assumption of complete neutrality for cytochrome b in *Trachurus* spp., particularly the very low p value from the Benguela LME.

Therefore results combined from Tajima's D (Tajima, 1989b), Fu and Li's Test (Fu & Li, 1993) and the McDonald-Kreitman test (McDonald & Kreitman, 1991) indicate results that do not conform to neutral evolution. However, whether or not this was because of selection pressures or demographic events could not be investigated by the first two tests. The McDonald-Kreitman test however indicates that at least some selection appears to be occurring, and that there is greater evidence of selection in *T. trachurus*

[1] http://ncbi.nlm.nih.gov

than *T. trecae*.. The differences between the two species indicates the likelihood that alternative selection pressures are operating in the two ecosystems (also note the differing ratios of transitions to transversions of *T. trecae* and *T. trachurus* in section 3.4.2 on page 79).

Such a result of selection appearing to occur in mtDNA is not altogether unexpected, given that a number of other studies have observed selection (see Ballard & Kreitman, 1995, for a review), and that the whole genome acts as a single locus because there is no recombination (but see Hey, 2000). Indeed, a possible cause for the temporal variation observed in *T. trecae* is alternative selection pressures between different annual and seasonal cohorts, as explored in section 6.5.4.

6.3 Molecular markers and repeat sampling

An important feature of any competent molecular marker for population studies is that they can be used repeatedly on the same population and, vagaries of sampling error aside, give the same answer. In addition, given the number of sub-stocks that often make up an exploited stock, the application of mixed stock analysis (MSA) is becoming of greater significance (Utter & Ryman, 1993; Brown *et al.*, 1996).

To manage fish populations in a viable manner through the application of MSA it is necessary to have stable molecular markers at your disposal. Prior to conducting MSA it is therefore necessary to investigate qualities of the data generated by a molecular marker; qualities such as distribution, reliability and stability of the revealed polymorphism. It is also important to have a stable temporal population structure. This section reviews whether the current results bode well for future use of MSA in the Gulf of Guinea with the markers and species used in this project.

6.3.1 Temporal variation

Specific questions regarding temporal variation in any study of molecular ecology, whether using MSA or not, include (from Brown *et al.*, 1996):

- Are stocks variable across years?

- Does sample size influence sampled yearly variation?
- Does variation between years contribute to the total genetic variation of variation among stocks?
- Given the temporal variation we encounter, what is the best method for incorporating year to year variation into the MSA exercise?

Significant genetic heterogeneity was shown for *T. trecae* from the Gulf of Guinea (see section 3.5.3). Pairwise analysis of *P. bellottii* (without including the four individuals with insertions - see section 4.4.1) revealed that the second most distinct pairwise difference was between pb28s and pb28b (i.e. samples of small and large individuals from the same trawl). Such a result leads one to consider that had greater numbers of individuals been analysed a similar, temporally dominated pattern as found in *T. trecae*, may also have become apparent in *P. bellottii*. *Sepia* populations have to be sampled over several years to measure temporal variation, given their short life span. This was not the sampling regime of the present study and consequently *Sepia* will not be discussed further in this section.

Therefore with regards to the question of whether there is temporal variation, the answer with respect to *T. trecae*, the one species for which there is enough data to make a judgement, is yes. In *T. trecae* the amount of temporal variation far outweighs the amount of spatial variation. Such temporal variation makes MSA extremely difficult to use as it relies on stability through time to allow for assignation of haplotype frequencies to stocks. The current results from *T. trecae* presented in this project does not allow for confident MSA using cytochrome *b*.

6.3.2 The suitability of the molecular markers

The presence or absence of an individual variant (termed a singleton, i.e. appearing only once in a sample) is of little use for stock discrimination or MSA. However, continued sampling, or pooling of samples, may mean that what were rare, uninformative singletons, become more valuable twin haplotypes. Interestingly Brown *et al.* (1996)

found that the proportions of singletons remained constant, regardless of sample size. Therefore a molecular marker that produces a high proportion of singletons is not a powerful one for MSA.

Cytochrome *b* and *Trachurus* spp.

Figures 6.1 and 6.2 show the frequency of haplotypes in the cytochrome *b* gene fragment from samples of the Gulf of Guinea, Benguela and UK populations of *Trachurus* spp. The proportion of singletons is quite high (around 20-40% of the total number of mutations) for samples from both populations, indicating its lack of suitability for MSA (compare to Brown *et al.*, 1996). Alternatively, the excess of rare singletons could could result from a relatively recent expansion in population size, and that the populations sampled are significantly displaced from genetic equilibrium (Pogson *et al.*, 1995).

Although perhaps counter intuitive, RFLP data would probably perform better at MSA than direct sequencing studies precisely because of its lower resolution. However, the few restriction enzymes used on *Trachurus* samples in this study did not give enough detail to separate populations satisfactorily within ecosystems, thus there is actually too little detailed information from the data. MSA relies on data with not too much, but also not too little, molecular detail.

A striking difference between *T. trecae* samples and *T. trachurus* samples is the variation in the number of "shared others". These are haplotypes that are shared by two or more samples and occur twice or more within at least one population (to separate them from shared singletons, which only occur once within a sample). None of the *T. trachurus* populations within the Benguela LME share any haplotypes except the common haplotype and a few singletons. In fact, apart from the singletons, all of the samples from the Benguela LME are dominated by the common haplotype. The exception is the UK population, whose dominant common haplotype is shared with, but uncommon in, the Benguela samples.

In contrast, the populations of *T. trecae* from the Gulf of Guinea have a number of "shared others." Also the number of unique singletons is high within the Gulf of

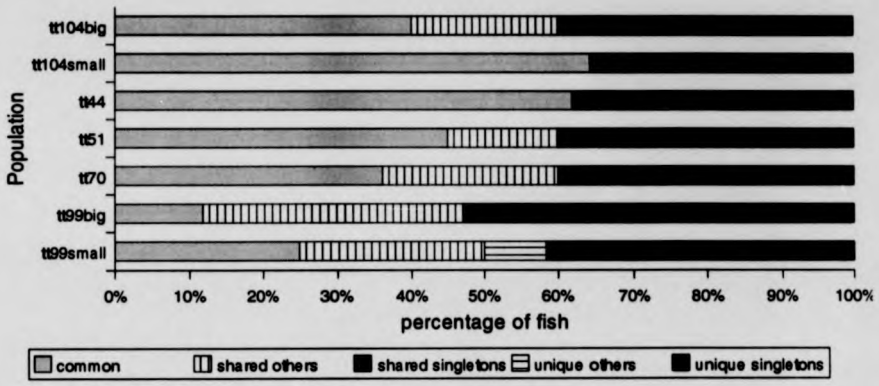

Figure 6.1: Distribution of shared and unshared haplotypes among *Trachurus trecae* Gulf of Guinea populations

Guinea samples, with even the lowest percentage of unique singletons being around 25%. However, it is the distribution of the "shared others" that allows interpretation of the population structure in the region, and not the unique singletons which although adding to diversity, do not aid in studies of population structure.

12S rRNA and *Pagellus bellottii*

Figure 6.3 shows the frequency of haplotypes for *P. bellottii* populations in the Gulf of Guinea; in this case the molecular marker used was a fragment of the 12S rRNA gene. Overall the 12S rRNA gene fragment produced fewer unique singletons than was found from the cytochrome *b* gene fragment used for the *Trachurus* populations. The "unique others" found in pb67b involve the individuals with insertions at base pair position 67. The sample pbgu, from Guinea, is noticeable in its very high instance of unique singletons (50%), though the sample size is of course very small (only eight individuals), increasing the liklihood of sampling error.

6.3.3 To pool or not to pool

The two key features required for successful use of mixed stock analysis (MSA) have been examined in this section, i.e.: *(i)* temporal variation, and *(ii)* the suitability of

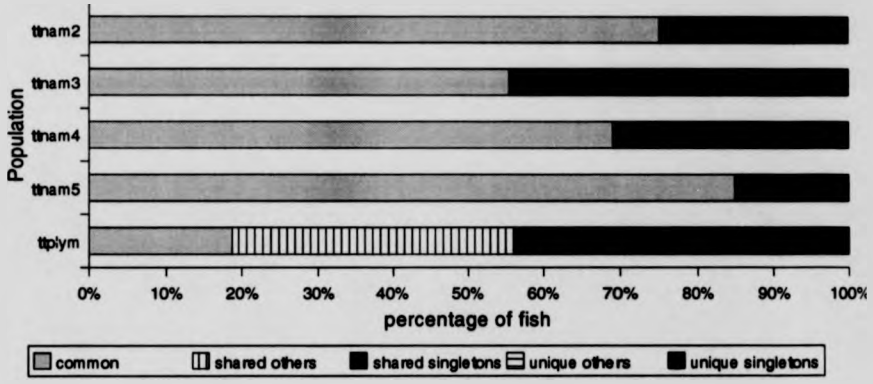

Figure 6.2: Distribution of shared and unshared haplotypes among *Trachurus trachurus* Benguela LME populations

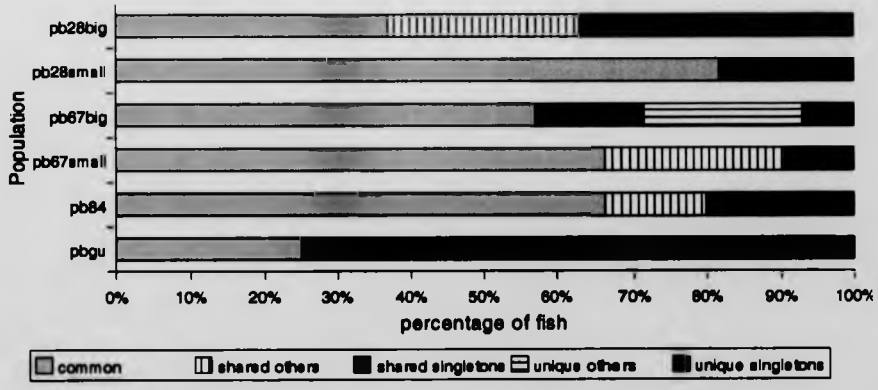

Figure 6.3: Distribution of shared and unshared haplotypes among *Pagellus bellottii* Gulf of Guinea populations

the molecular markers. It is apparent that temporal variation is present in the Gulf of Guinea populations sampled, and that it is greater in effect than spatial variation. In addition to this, the molecular markers used, particularly on *Trachurus* spp., revealed a relatively high number of unique singletons. Such haplotypes, while adding to the overall measured diversity of the population, do not aid MSA as they fail to supply useful information for stock assignation. It is therefore apparent that with the excessive detail of the markers, and degree of temporal variation found in the sampled populations, it is not safe to pool samples in an attempt to calculate MSA.

6.4 Comparison of control region, tRNA (*Phe*) and 12S rRNA between *T. trecae* and *P. bellottii*

The sequences of part of the control region, tRNA(*Phe*) and 12S rRNA genes of *T. trecae* and *P. bellottii* can be found in Genebank[2] with the accession codes AF271658 and AF271659 respectively. Primarily they are intended to be used as a resource for future work on the species, allowing design of further PCR primers.

The highest homology between the two species is 95.59%, found in the 68 base pair tRNA gene. A slightly lower 85.96% homology is found in the 641 base pair 12S rRNA genes. There is only 50.69% homology in the 434 base pair control region, though that is not including insertions which are common in the control region and which when accounted for include a further 56 base pairs (35 insertions in *T. trecae* and 21 insertions in *P. bellottii*). Using a dot plot (wherein two compared sequences are plotted against one another on an X - Y axis) it is apparent that the only repeat units are those found in the 12S rRNA genes and as such are detailed below in section 6.4.1.

The region sequenced for *P. bellottii* is shorter than the *T. trecae* region. The *P. bellottii* region sequence starts in the "central conserved region", a feature well marked in the *T. trecae* sequence along with a number of other teleost groups (Lee *et al.*, 1995). Insertions are by far most common in the control region (49 insertions), as is expected from other work on the control region (for example see Lee *et al.*, 1995); by comparison

[2]http://ncbi.nlm.nih.gov

Pagellus bellottii

ccccgctttcactggggttggttaccccgctatccctggggttagata
ccccgctttcactggggttggttaccccgctatccctggggttagata
ccccactatcgctggggttagataccccactataactggggttagata
ccccactatcgctggggttagataccccactataactggggttagata
ccccactataactgggattagataccccactatcactggg

Trachurus trecae

ccccactatacctgggattagataccccactataactgggattagata
ccccactatacctgggattagataccccactataactgggattagata
ccccactataactgggattagataccccactataactgggattagata
ccccactataactgggattagataccccactataactgggattagata
ccccactataactgggattagataccccactataactggg

Table 6.3: Alignment of 12S rRNA repeat sequence in *T. trecae* and *P. bellottii*. The alignment is high, with 90.95% homology. Much of the variation that does occur between the two species can be found in the first 96 base pairs of the repeat sequence.

there are no insertions in the tRNA genes, and just seven in the 12S rRNA region. Such frequent insertions and deletions make the region unsuitable for inter-specific study, though the high variability might be valuable for intra-specific work.

6.4.1 The repeat sequence homology

The repeat sequence homology of the sequenced 12S rRNA gene is shown in Table 6.3. The homology is high (90.95%) and results from functional restraints. Although the high homology may indicate the region as suitable for the design of universal primers, the repeated sequence would potentially encourage repeat primer binding. Such behaviour would lead to multiple length PCR products, making sequencing and/or RFLP analysis difficult. Consequently, the repeat sequence is not likely to allow the design of efficient PCR primers.

6.5 Molecular ecology and fish behaviour

Although the detail of the molecular markers, and the temporal variation of the sampled *T. trecae* populations may not aid the use of mixed stock analysis (see section 6.3), they may be useful to the study of fish behaviour. This section explores some aspects of shoaling and reproductive behaviour that may be illuminated through the use of

molecular markers, as well as what such markers may tell us of longer term evolutionary relationships.

6.5.1 The genetic makeup of shoals

An introduction to the shoaling and schooling behaviour of fish is given on page 24. The results from the study of *T. trecae* in the Gulf of Guinea allow some interesting observations with regard to shoaling behaviour. The most striking result found was the often significant pairwise differences between different age classes within the same shoal, found both with *T. trecae* and *P. bellottii* samples. Taken at face value, the results immediately dispel any notion that shoals consist of genetically homogenous individuals, and consequently kin based altruistic selection (as reviewed in Pitcher & Parish, 1993) is not a possibility. However, for a number of reasons the situation is likely to be more complex.

Firstly, the samples, although coming from the same trawl, do not necessarily consist of individuals from the same shoal. The demersal trawl was run for 15 minutes at each station; in addition, depending on the depth of the trawl it may take up to 15-20 minutes for the trawl to be lowered and raised. During such time, it is possible (though unlikely) that more than one shoal of the same species was sampled. One way to ensure that this is not the case is the use of new technology, whereby it is possible to film the gates of the trawl during its time underwater and physically see the fish as they enter the trawl nets. This allows the user to record whether one or more shoals appear to have been caught. However, no such technology was used on the Nansen cruise. Acoustic methods can also be used in conjunction with trawling to sample discrete shoals (e.g. with the use of an echo sounder such as the Simrad EK500). Though it might be the case that the catch consisted of two quite separate shoals the chances are against it. Therefore it is assumed for the following discussions that, if two age classes were sampled from a single trawl, they were taken from a single shoal.

Secondly, although the age class samples from each trawl probably came from the same shoal it is unknown how permanent, or temporary, the shoal was. If the shoal

was a feeding assemblage it is quite possible that it consists of two or more smaller shoals that would split up during, for example, predator threat. Therefore although the shoal appears genetically diverse it is possible that it is a temporary aggregation of individuals, and that with regard to reproductive and migratory behaviour the separate constituents of the shoal would divide to form smaller schools of closely related fish. Results indicating such behaviour in the Atlantic mackerel were recently published by Nesbo *et al.* (2000). They showed that no genetic structuring was found among shoals of individuals outside of the spawning areas, but that population structuring *was* found between shoals sampled from spawning grounds. In such a case the alternative spawning shoals mix for feeding, but remain faithful to spawning locations. Section 1.4.4 contains a further example of similar behaviour in the Atlantic cod, where genetic division was found at spawning sites even though intermingling of adult stocks occurred during winter feeding (Bentzen *et al.*, 1996; Ruzzante *et al.*, 1996).

Therefore the genetic diversity of the shoals found does not in itself rule out the possibility that natal homing occurs in tropical species. Such an investigation would require the collection of spawning fish from defined and known spawning grounds. Although such work may be possible for well funded and already relatively well understood temperate fisheries off Europe and north America, it is difficult to sample accurately in relatively unknown waters such as the Gulf of Guinea.

The differences amongst age classes within shoals, therefore, are not in themselves enough to deny the possibility of kin relatedness among shoaling fish. The marine environment is a dynamic and complex system (e.g. see Roughgarden *et al.*, 1988), and most likely many of the shoals are temporary structures of aggregations of individuals. Within these larger shoals however there are quite possibly smaller subunits of kin related fish that form distinct schools at certain periods of their life history, most obviously during spawning. The next section looks at what light can be shone upon these relationships within and among shoals.

6.5.2 Relationships within and between shoals

In both fish species there was a pronounced trend to highly negative values of Tajima's *D*. The value of *D* reflects the shape of the gene tree, as explained in the methodology section 2.9.1, and graphically on page 65. The negative value could also result from selection or demography, as discussed in section 6.2.1. However, there are also more subtle variations of *D* within shoals (see also section 6.5.5).

As noted on page 81, there is a correlation in *T. trecae* between the median length of the age class and the value of Tajima's *D* (product-moment correlation coefficient *D*: $r = 0.93$, $df = 4$, $p = 0.01$). The more negative values of *D* are consistently found in the younger age classes, indicating an aggregation of closely related individuals, at most varying by just one base pair difference in the cytochrome *b* gene fragment used in the sequencing studies. The samples of older age classes on the other hand have less negative values of *D*, which reflects a tree with deeper branch structure.

One possible explanation for such a pattern is that the older age classes consist of groups of fish from various spawning events having joined, however temporarily, to make mixed shoals for feeding behaviour, mutual safety etc. The younger fish on the other hand have not yet had time to mix with conspecifics of a similar age, and thus are still relatively closely related groups of individuals during feeding and all other behaviour. This pattern, of closely related cohorts of juvenile fish, supports the selective sweepstakes of Cushing (1972), as developed by Hedgecock (1994) and Li & Hedgecock (1998) (see section 6.5.4).

A population structure such as this requires migration between shoals as the fish age, leading to a deeper tree structure for groups of older fish. Section 3.4.7 gives some evidence as to who these fish may be. Population subdivision, when the sex of the fish are taken into account as given in Table 3.9, is far more pronounced between samples of females than between samples of males. It is therefore not unreasonable to propose (following the argument of O'Corry-Crowe *et al.*, 1997) that the young males are more likely to stray between shoals than are females, leading to the pattern of variation in Table 3.9, although it appears that the number of female strays/migrants increase

slightly with age.

6.5.3 Exceptional individuals

Both *P. bellottii* (section 4.5.3) and *Sepia* spp. (section 5.4.3) samples contained four individuals with highly atypical sequences/allele sizes. Within the discussion of the relevant Results Chapters various options as to who these individuals are have been put forward; ranging from being a rare, distinct subpopulation, to being a closely related species which has been mistakenly attributed as either *P. bellottii* or *S. officinalis* respectively. In the case of *P. bellottii* the four individuals were found clustered together in one sample, while with *Sepia* they were distributed throughout the Gulf of Guinea.

Both cases act as a warning and raise the issue of sampling error in a study such as this. Sample sizes were smaller than would be preferred, because of problems of sampling and time limitations, though any sensible sample size will only ever cover the tiniest proportion of the total population. Finding four individuals in one sample of *P. bellottii* raises the question of what if a handful of other, similarly exceptional individuals where found in another sample of *P. bellottii*? In such a case conclusions would be drawn about a more complex population structure of *P. bellottii* in the Gulf of Guinea. The handful of exceptional individuals therefore enlarge the phase space of possibilities, while not giving us sufficient enough information to define our position in it.

6.5.4 Recruitment

A striking pattern that was mentioned with regard to pooling samples in section 6.3 is the temporal variation between length classes, regardless of their geographical proximity. A picture arises of related cohorts of fish, typical of a length class, having greater similarity with geographically distant fish of the same length than with neighbouring fish of different length (see Arnason *et al.* 2000 for similar results using cytochrome *b* sequence variation in Atlantic cod, and Ruzzante *et al.* 1996 for similar results with microsatellites). Such a pattern could appear broadly for three reasons:

1. Selection is operating on each age class as it develops through it's own unique series of environments, because of slight variations year to year of the physical, biological and oceanographical environment (Koehn & Williams, 1978; Johnson & Black, 1982, 1984). Each year class therefore has had slightly different selection pressures, resulting in preferred survivability of a number of suitably adapted individuals unique to that year class.

2. Selective sweepstakes is a term introduced by Cushing (1972) to describe the lottery of spawning adults in the marine environment. Each spawner has the potential to contribute massively to the next generation, because of the usually enormous number of eggs and sperm ejaculated during spawning. Given the fluctuations of the marine environment, there is a very uneven contribution to the number of successful offspring, with a handful of lucky spawners parenting most of the next generation (Hedgecock, 1994).

3. Alternatively the pattern of results could be interpreted as evidence for more than one relatively discrete reproductive season per year (e.g. as in *T. trachurus* off Agulhas Bank, South Africa; Hecht, 1990), and that spawning adults remain loyal to a certain season in time, thus resulting in discrete age classes. The number of discrete reproductive seasons per year would therefore dictate the number of genetically distinct cohorts in the population. Accurate ageing (e.g. with the use of otoliths) would be needed to test this hypothesis.

Therefore is it most likely a result of chance (i.e. genetic drift), selection or reproductive behaviour? There are criticisms of all explanations. Although the selective sweepstakes theory is appealing because of the obvious potential for hugely variable numbers of offspring from individual to individual, research has shown that more often than not a number of experienced spawners all do relatively well, particularly compared to the greater number of first time, inexperienced spawners (Buckley *et al.*, 1991; Hutchings *et al.*, 1999). The spawning process is grossly oversimplified if described as a lottery. Subtle shoaling behaviour has been noted for a number of fish (for an example see Barber *et al.*, 1998) and spawning is a time of great behavioural complexity. There is

no evidence to support the commonly held misconception that it is a totally random process, with all participants having an equal chance of fathering or mothering most of the next generation.

In addition, even if a great number of eggs were fertilised from a very small subset of the population, the daily death rate of eggs kept in even the most perfect conditions is massive (Cohen, 1977). Very soon, 30,000 apparently viable eggs can drop down to 300, and that is not even taking into account predation and the vagaries of currents and tides that could end up sweeping any number of pelagic eggs and larvae out to sea.

However, although the spawning event is not the lottery as is often perceived, the pelagic phase can introduce chance. Currents and tides are not always predictable and no doubt numerous otherwise viable eggs and larvae are swept out sea (Norcross & Shaw, 1984). However, quantifying the effect is very difficult and research has shown that the subtle vertical migration of young larvae ensures that they are not always as defenceless and vulnerable to oceanographic conditions as may initially appear (Boehlert & Mundy, 1988; Miller, 1988; Suthers & Frank, 1991; Sabates & Pilar Olivar, 1996). In addition, many species with pelagic larvae take every measure to ensure that their young offspring do not utilise their potential mobility (Johannes, 1978). Because pelagic larvae *could* be used for dispersal and migration, research had often assumed that their role was for just such behaviour. However this is by no means always the case (Todd 1998; Todd & Thorpe 1998; but also see Economou 1991).

The case for selection is likewise open to criticism. Firstly there is no evidence that selection can occur across the whole genome, rather than on specific genes or packages of genes. The theory of neutrality was put forward as an alternative to selectionist ideas that could bend and shape themselves to fit any scenario on offer (see section 1.6.3). There is also a problem with the notion that an area the size of the Gulf of Guinea could have suitably similar selection pressures across its whole area to create a whole generation of fish across the region with relatively similar genotypes (Cowen *et al.*, 1993; Sabates & Pilar Olivar, 1996). Many variations that could cause such selective pressures are often localised, and the juvenile fish are often separated in mangroves or estuaries

along the coastline that act as nursery areas (see section 1.3 for an introduction to the Gulf of Guinea and its nursery grounds). It is therefore quite difficult to see how selection would create such cohorts of genetically related, similarly aged fish. However, increased use of remote sensing data will allow further insight into the relationships between local and regional oceanographic variability and the genetic variability of marine populations (Roughgarden *et al.*, 1991). Remote sensing data is currently being collected for the Gulf of Guinea as part of the Gulf of Guinea Sustainable Fisheries Project (Anon, 2000, and see section 7.5).

Finally there is the chance that loyalty to reproductive seasons leads to the observed pattern of temporal variation in *T. trecae*. Hecht (1990) presented evidence for two spawning peaks during the reproductive season of *T. trachurus* off South Africa, extending from June to November (supported by the observations of bimodal length frequency distributions of Kerstan (1995) in several of the age-groups of *T. trachurus* in the same region). It it therefore likely that more than one reproductive season may apply to *T. trecae* in the Gulf of Guinea (particularly related to the two upwelling seasons in the central subsystem; see section 1.4.2)

However, although the opportunity arises, because of the liklihood of more than one reproductive season per year in the Gulf of Guinea, it does not follow that spawning individuals necessarily remain loyal to one season per year. Although a great deal has been written about the possibility of spatial loyalty to spawning grounds (Iles & Sinclair, 1982; Sinclair, 1988; McQuinn, 1997, and see section 1.4.4) and the possible mechanisms involved (Heard, 1991; Sandercock, 1991; Cury, 1994; McQuinn, 1997) rather less has been said about loyalty to temporal seasons; though McQuinn (1997) does propose that, after adoption of a spawning site in both space and time, Atlantic herring (*Clupea harengus*) remain loyal to that season throughout adult life (though for genetic subdivision to occur the loyalty would have to be natal, not adopted). To test whether cohorts of fish from different reproductive seasons of the same year are genetically differentiated it is necessary to *(i)* collect good age data (e.g. from otoliths or length frequency distributions) and *(ii)* sample when the fish are still juvenile as older fish are more difficult to

assign accurately to one cohort or another (because of possible selection and the effects of the environment on growth throughout a lifetime).

6.5.5 Metapopulations and evolutionary relationships

There is a need to judge molecular genetic results as a combination of two timetables; the evolutionary and the ecological. Most of the discussion in this project has been concerned with the effect of fish behaviour such as reproduction and migration, and thus is dominated by effects of the ecological timescale. However, underneath such short term fluctuations in population dynamics and molecular patterns there are longer term trends that can tell us something about the evolutionary relationships of populations.

Across all populations of all species the value of Tajima's D is strongly negative. Such a consistent finding, in association with the fact that estimates of N_e from genetic measures are nearly always orders of magnitude lower than estimates of N_e from stock analysis or other predictions of census size (Avise *et al.*, 1988; Avise, 1992; Frankham, 1995; Jorde & Ryman, 1996; Miller & Kapuscinski, 1997; Chikhi *et al.*, 1998), highlights the long term instability of marine environments. It appears that population extinction, or at least drastic drops of population numbers, are frequent for many marine species (Chikhi *et al.*, 1998). Bartley *et al.* (1992) found that the $N_e : N$ ratio (i.e. the ratio of the effective population size, N_e, to the estimated total population size, N) for the Chinook salmon (*Oncorhynchus tshawytscha*) was 0.04, emphasising that only a tiny proportion of the adult population actually contributes to the next generation.

With the vagaries of such an unstable environment, the metapopulation model (page 24 and Levins 1968) is an attractive theory to help describe long term population dynamics. The model describes the diversity found between cohorts of fish in the Gulf of Guinea as resulting from the sampling of constituent populations of the Gulf of Guinea metapopulation. Identity of the populations would result from loyalty to breeding grounds, even though the adult populations regularly mix to form feeding shoals. However there would be enough vagrancy to ensure that if, because of fluctuations in oceanographic conditions, a local population was rendered extinct, colonists from other local populations

would soon reinstate a population (Frank, 1992; Planes *et al.*, 1996; McQuinn, 1997). A metapopulation such as this would result in diversity because of natal fidelity, but there is enough gene flow resulting from vagrancy to ensure that the local populations do not substantially diverge genetically from one another. It would also result in low values of N_e and negative values of Tajima's D, because of *(i)* the frequent local extinctions followed by a founder effect, and *(ii)* if not complete population extinction, then frequent population crashes, leading to population bottlenecks.

Such a conceptual model of populations in the marine environment incorporates thoughts on the stock concept in fisheries (see section 1.4.3 for an introduction to the concept). It is often noted that the term "stock" is used with a wide range of definitions, depending on who is defining and why (Carvalho & Hauser, 1994). It is for such a reason that I have tried to avoid using the term, except in its loosest sense, during the writing of this thesis. However the metapopulation concept is a useful term as it reflects the dynamic nature of marine populations as opposed to the original typological and rather static view of the stock as based around the Harden-Jones (1968) triangle of migration.

Nations often have a political interest in dividing fisheries into neat and locally manageable stocks. As such management decisions are seldom, if ever, taken exclusively on the biology and ecology of the species in question (Carvalho & Hauser, 1994), it is important to be aware of the political and social factors involved in the governance of the region. In the case of demersal fish in the Gulf of Guinea, management has hitherto been based around individual nations states (M. A. Mensah, pers. comm.), mostly because of political considerations. Therefore a result that emphasises that such an approach is, at least biologically, unreasonable implies that management of the fisheries in the region needs to change. The markers used, at least for *Trachurus* spp. and *Sepia* spp., are known to be sensitive enough to reveal heterogeneity at a fine scale such as in the Gulf the Guinea. Indeed, *T. trecae* did show significant heterogeneity, but only temporal rather than spatial. Therefore the results presented are not a lack of results, rather they are results that emphasise *(i)* the need to change management of demersal fisheries in the Gulf of Guinea and *(ii)* the need to revise the idea of typological stocks and instead

replace it with the dynamical approach of Smith *et al.* (1990) and Dizon *et al.* (1992), and possibly incorporate the metapopulation model of Levins (1968) (as presented in McQuinn 1997). However, the key to further changes in the management of fisheries in the region will have to be political and social, rather than biological.

6.6 Molecular ecology and the Gulf of Guinea

The genetic analysis of marine populations in the Gulf of Guinea is still in its infancy. Projects such as this serve as preliminary analyses and together will hopefully help generate more informed questions for future work. There have most likely been other studies of West African marine fish by local scientists (Dr. E. Abban, pers. comm.), however these only rarely reach the scientific literature. The only other genetic analysis of species in the area of the Gulf of Guinea, that I am aware of, is Chikhi (1995), who worked on *Sardinella aurita* and *S. maderensis* for his PhD thesis. The results of his study echoes those presented here, in that little genetic variation was found within the Gulf of Guinea, and any variation that was found was mostly explained by temporal genetic variation.

Taken together, the work on five species, *S. aurita*, *S. maderensis* (both from Chikhi, 1995), *T. trecae*, *P. bellottii* and *S. officinalis* all fail to show any significant spatial subdivision within the Gulf of Guinea. Consequently it is conservative to infer that there are no physical or oceanographic features in the sampled regions of the Gulf of Guinea that promote population subdivision. There is widespread mobility along the east-west coastline of the central region, even for demersal species such as *Sepia* and *P. bellottii* which are not reported to have high dispersal ability when compared to the three other powerful swimming fish species studied.

However, to offer a word of warning, a key feature of all work done thus far on the Gulf of Guinea has concentrated on the central upwelling region of Côte d'Ivoire to Benin. A few samples from the eastern upwelling region were featured in this project, however sample numbers were small. One reason why work is concentrated on the central region is the relative safety of working in the region. In particular, the nation

states of Sierra Leone and Liberia are not ideally suited to scientific research. Also the central upwelling region is dominated by the coastlines of Côte d'Ivoire and Ghana, two nations (particularly Ghana) with large fishing fleets and biologically rich waters, and with vested interests to help fund and host researchers in marine and fisheries biology. However, it is unfortunate that more samples have not been studied from the western and eastern subsystems, and as such that will have to be the domain of future work.

The small number of samples that were analysed for this project from the eastern subsystem (all from the waters of the Republic of Guinea) did not show significant difference from samples in the central upwelling system. Thus the notion that although the subsystems feature different oceanographic characteristics (see section 1.3), such features do not in themselves add much of a barrier to gene flow. There may possibly be some evidence of isolation by distance, however, small sample sizes makes testing the hypothesis with the current data impossible.

Work on *S. aurita*, *S. maderensis*, *T. trecae* and *S. officinalis* have all shown very striking differentiation when comparing individuals from within the Gulf of Guinea to those of ecosystems of southern Europe (*S. aurita*, *S. maderensis* and *S. officinalis*) or south and southwest Africa (*T. trecae*). As noted in the introductory chapter (see section 1.3) the Gulf of Guinea is a unique ecosystem because of its thermal stratification, upwelling and east-west orientation. Comparisons across LMEs do not at the moment indicate very much gene flow across the borders of the Gulf of Guinea LME, and interesting work would be to sample either side of the borders to see if a sharp genetic continuity exists or if rather there are signs of genetic introgression (see section 7.4).

Current work does not rule out the possibility that the regions to the east of the two main capes may be important for population structuring. If, as seems apparent, there is high mobility across the region, most probably to follow food resources (Chikhi, 1995), it is still feasible that constituent schools of fish return faithfully to particular spawning grounds (Frank, 1992; Ruzzante *et al.*, 1996; Nesbo *et al.*, 2000). However, there is still as yet no evidence for such behaviour (see section 7.4).

Chapter 7

Discussion and Conclusions

7.1 Preamble

Chapter 1 served to introduce the Gulf of Guinea and the three species under study, along with some key concepts in molecular ecology and fisheries biology. Chapter 2 detailed the materials and methods used, covering fieldwork, laboratory analyses, and numerical and statistical methodology. Chapters 3 to 6 are the Results chapters: chapters 3, 4 and 5 covered the molecular genetic analysis of *Trachurus* spp., *Pagellus bellottii* and *Sepia* spp. respectively, while chapter 6 covers issues across the previous three results chapters, such as neutrality, the potential of mixed stock analysis in the region, comparisons of the control region, shoaling and schooling behaviour and general trends found in the Gulf of Guinea.

This concluding chapter will first present a summary of the results from the four results chapters, as organised into species and themes apparent across species. Major conclusions are then drawn and from these suggestions for future work are made. To conclude the chapter the work is placed in the context of the Gulf of Guinea Sustainable Fisheries Project of the European Union, from which the work was funded, and indicates how molecular ecology may make a contribution to the good governance of the region.

7.2 Summary of results

7.2.1 *Trachurus* spp.

Trachurus spp. have been introduced in section 1.5.1 on page 26. Two regions of the mtDNA were analysed: a partial sequence of the control region/12S rRNA gene, and a partial sequence of the cytochrome *b* gene. Restriction fragment length polymorphism (RFLP) analysis was used on the control region/12S rRNA sequence, while sequence analysis was used on the partial sequence of the cytochrome *b* gene. RFLP results were only powerful enough to display variation across LMEs and lacked detail for within LME analysis. The rest of the *Trachurus* spp. summary concentrates on results from the sequencing studies of the cytochrome *b* gene.

Relationships between LMEs

The Gulf of Guinea samples differed markedly from Benguela LME samples. There was little evidence for any population subdivision within the Benguela LME, surprising given the Lüderitz upwelling region between samples. The Gulf of Guinea samples displayed small, but significant, population differentiation.

From the estimates of θ there is some evidence that the effective population size of the Gulf of Guinea population is higher than that of the Benguela LME. Alternatively the higher estimates of θ could result from greater heterogeneity of the physical and oceanographic environment of the Gulf of Guinea LME.

Variation within LMEs

Within both the Gulf of Guinea and Benguela LMEs, samples displayed no indication of isolation by distance, though there may be evidence of isolation by distance between *T. trachurus* samples from the UK and south-west Africa (Figure 3.9). Within the Gulf of Guinea samples there is substantial evidence of temporal structuring, revealed as a clustering of age groups in phylogenetic trees, irrespective of geographic proximity. Such a pattern possibly results from *(i)* a sweepstakes reproductive strategy whereby the surviving offspring in a given locality may come from a few lucky parents, *(ii)* differing

selection pressures for each age class resulting in the observed temporal patterning, or *(iii)* more than one discreet reproductive season that adults remain loyal to (see section 7.2.4).

Are *T. trecae* and *T. trachurus* different species?

The amount of divergence (following Avise & Walker, 1999) between the Gulf of Guinea and Benguela LME common sequences does not support separate species, however there is substantial other molecular evidence that does support separate species (e.g. no shared haplotypes and very high F_{ST} values between samples). Hence, following the precautionary principle, the molecular evidence overall supports the notion of separate species. The lower divergence between *T. trachurus* from the Benguela LME and *T. trachurus* from the UK indicates merely intraspecific geographic grouping and not the presence of a further species, *T. capensis* (the Cape horse mackerel).

7.2.2 *Pagellus bellottii*

Pagellus bellottii has been introduced in section 1.5.2 on page 28. A partial sequence of the 12S rRNA gene of the mtDNA genome was used for the analysis.

Nucleotide diversity was low for *P. bellottii*, though with an exception of four individuals from pb67b that showed high sequence variation from all other samples, including an insertion (haplotypes 13 and 14 in Figure 4.5). Values of Tajima's D, once the four exceptional individuals are accounted for, were negative, indicating a probable selective sweep or recent population bottleneck.

Spatial population structure in the Gulf of Guinea

The was no evidence of same trawl samples being more similar to one another than to other samples, though equally there was little evidence of the temporal structure observed in *T. trecae*. The results indicate high gene flow throughout the sampled region, supporting the view that there is one stock across the whole of the Gulf of Guinea (Koranteng & Pitcher, 1987). However, there appeared to be a small amount

of isolation by distance, though sample sizes were very small and the results were not significant.

7.2.3 *Sepia* spp.

Sepia spp. have been introduced in section 1.5.3 on page 29. Seven microsatellite loci were used for analyses. Departures from Hardy-Weinberg expectations were found at two loci, probably resulting from null alleles (Perez-Losada *et al.*, 2000). Four west African samples showed very uncharacteristic allele sizes and were designated as a separate sample called "odd."

Relationships between West African and Spanish populations

High levels of polymorphism were displayed at all seven loci within west African and Spanish samples. There was highly significant genetic division between Spanish and west African samples, though there was no further subdivision within the west African samples. Generally, levels of diversity within west African and Spanish populations were similar.

Who are the "odd"?

The "odd" sample, of four west African individuals, displayed closer affinity to Spanish, rather than African, samples. However, the sample size of merely four individuals is insufficient to uncover their identity.

Is *Sepia* spp. in the Gulf of Guinea *Sepia officinalis*?

Microsatellite results are by themselves insufficient to answer the question confidently. Mitochondrial DNA and morphological characteristics would be more suitable measures of possible speciation.

7.2.4 Comparative molecular ecology and evolution

Neutrality

Estimators of Tajima (1989b) and Fu & Li (1993) for neutrality reflect the shape of the gene tree. The test statistics complement one another and indicate a shallow, branching tree. However, the shape of the gene tree could originate from either to selection or demography.

With the use of the McDonald & Kreitman (1991) test there was evidence that selection was occurring for the cytochrome *b* marker. Results indicate that differing selection pressures are operating in the Gulf of Guinea and Benguela LMEs.

Mixed stock analysis

Little temporal variation and suitable molecular markers are necessary for mixed stock analysis (MSA). However, for *T. trecae* the amount of temporal variation far outweighed the amount of spatial variation. Such temporal variation makes MSA extremely difficult to use as it relies on stability through time to allow for assignation of haplotype frequencies to stocks (Utter & Ryman, 1993; Brown *et al.*, 1996).

In addition to this, the molecular markers used, particularly on *Trachurus* spp., revealed a relatively high number of unique singletons. Such haplotypes, while adding to the overall measured diversity of the population, do not aid MSA as they fail to supply useful information for stock assignation (Brown *et al.*, 1996). Therefore, with *T. trecae* and with the markers used, it is not safe to pool samples in an attempt to calculate MSA.

Molecular ecology and fish behaviour

Significant pairwise differences between different age classes within the same shoal were found with *T. trecae* and *P. bellottii* samples. However, although coming from the same trawl the samples do not necessarily come from the same shoal; also it is unknown how permanent or temporary the shoal was. Therefore although the shoal appears genetically diverse it is quite possible that it is a temporary aggregation of individuals, and that it

consists of smaller subunits of genetically related individuals.

A correlation in *T. trecae* between the median length of the age class and the value of Tajima's *D* reveals that the older age classes have a gene tree with a deeper branch structure. Possibly the older age classes consist of groups of fish from various spawning events have joined temporarily for feeding and/or safety, while the younger fish have not yet had time to mix with conspecifics of a similar age. Results indicate that young males are more likely to stray between shoals than females.

The temporal variation found in *T. trecae* can appear because: *(i)* selection is operating on each age class as it develops though its own unique series of environments (Koehn & Williams, 1978; Johnson & Black, 1982, 1984), or *(ii)* drift resulting from "selective sweepstakes" whereby because of chance, there is an uneven contribution to the number of successful offspring, with a handful of fish parenting most of the next generation (Cushing, 1972; Hedgecock, 1994), or *(iii)* because of spawning adults remaining loyal to a certain season in time (as in Hecht, 1990), thus resulting in discrete age classes.

It appears that population extinction, or at least drastic drops of population numbers, are frequent for many marine species (Avise *et al.*, 1988; Bartley *et al.*, 1992). With the vagaries of such an unstable environment, the metapopulation model (Levins, 1968) is an attractive proposition to describe long term population dynamics. It would also explain low values of N_e and negative values of Tajima's *D*, because of: *(i)* the frequent local extinctions followed by a founder effect, and *(ii)* if not complete population extinction, then frequent population crashes, leading to population bottlenecks.

Molecular ecology and the Gulf of Guinea

The only other genetic analysis of species in the area of the Gulf of Guinea is Chikhi (1995), who worked on *Sardinella aurita* and *S. maderensis*. Taken together, the work on five species, *S. aurita*, *S. maderensis*, *T. trecae*, *P. bellottii* and *S. officinalis* all fail to show any significant spatial subdivision within the Gulf of Guinea. However all work has concentrated on the central upwelling region of Côte d'Ivoire to Benin (apart from

the small samples analysed herein from the western subsystem). There may possibly be some evidence of isolation by distance within the Gulf of Guinea. Comparisons across LMEs do not indicate very much gene flow beyond the borders of the Gulf of Guinea LME.

The regions to the east of the two main capes may be important for population structuring. It is feasible that constituent schools of fish return faithfully to particular spawning grounds. However, there is still as yet no evidence for such behaviour in the Gulf of Guinea.

7.3 Conclusions

Several broad conclusions can be drawn. They are:

1. Within the Gulf of Guinea there are no population subdivisions within the species studied, indicating that there is high mobility across the region. However there is some evidence for heterogeneity that may indicate natal homing.

2. Gulf of Guinea populations of *T. trecae* and *Sepia* sp. show very highly significant genetic variation from non-Gulf of Guinea populations, supporting the notion that the Gulf of Guinea is a unique and well defined large marine ecosystem.

3. *T. trecae* and *T. trachurus* are confirmed as separate (though very closely related) species, while the taxonomy of *Sepia* is not best answered by microsatellite data. *Trachurus* samples from the south-west of Africa are not considered distinct enough to be named as a separate species (*T. capensis*), but rather are a regional population of *T. trachurus*.

4. *T. trecae* shows strong evidence of temporal variation, whereby similarly aged samples have great genetic similarity regardless of geographic proximity. *S. officinalis* showed temporal variation at one locus.

5. Shoals of both *T. trecae* and *P. bellottii* show significant genetic heterogeneity, indicating that they are not aggregations of closely related individuals. However,

there is the possibility that they are temporary aggregations for (e.g.) feeding purposes, and that they may split into more closely related schools of fish for (e.g.) spawning behaviour.

6. In *T. trecae* there is some evidence for greater intra stock migration by males rather than females.

7. There are signs of differing selective pressures at the cytochrome *b* gene in *T. trecae* and *T. trachurus*.

7.4 Future work

There are many directions future work could take, and some have been highlighted in discussions in the text. Broadly I wish to concentrate on six possible avenues of future work: *(i)* testing the temporal variation found in *T. trecae*; *(ii)* within Gulf of Guinea population structuring and natal homing; *(iii)* the relationship of the Gulf of Guinea to its neighbouring large marine ecosystems; *(iv)* the identification of the "odd" *Sepia* sp. and *P. bellottii* individuals; *(v)* more sample collection of *P. bellottii*, most particularly from the Canary Current LME and European waters; and *(vi)* further work on the taxonomy of *Trachurus* spp. and *Sepia* spp.

Testing the temporal variation found in *T. trecae*

The temporal variation found in this project is highly significant and very striking. However, further sample collection and analysis should follow that explicitly tests the hypothesis that temporal variation is more significant than spatial variation. Temporal variation has been found for a number of species (e.g. see Chikhi 1995; Jorde & Ryman 1996; Miller & Kapuscinski 1997; Arnason *et al.* 2000, but see also Tessier & Bernatchez 1999 and references therein) but has often been written off as being of little importance. However, the amount and regularity of such variation being found suggests that it reflects significant behavioural or ecological patterns in a number of species and as such should be considered.

Certain predictions regarding the cause of the temporal variation can be tested. If the temporal variation results from the "selective sweepstakes" of Cushing (1972) then specific cohorts of new recruits, to the extent that they represent a small subset of the spawning population, should have *lower* genetic diversity than the adult population as a whole (Hedgecock, 1994; Li & Hedgecock, 1998). If on the other hand selection is occurring (in the manner of Koehn & Williams, 1978; Johnson & Black, 1982, 1984) then specific cohorts should have *greater* genetic diversity than the adult population. If the differentiation is because of loyalty to reproductive seasons, then (e.g.) summer spawners of one year will show similarity to summer spawners of the following year, but differ to (e.g.) spring spawners.

However, the most difficult aspect in testing the above predictions is the collection of suitable sample sets. Juveniles are difficult to sample because *(i)* it is difficult to know where they are (good places to start include beach seines and lagoons) and when to sample, and *(ii)* juvenile fish are more difficult to identify in the field.

Within Gulf of Guinea population structuring and natal homing

Current results do not rule out the possibility that the regions to the east of the two main capes may be important for population structuring. If, as seems apparent, there is high mobility across the region, most probably to follow food resources, it is still feasible that constituent schools of fish return faithfully to particular spawning grounds (as found in Atlantic cod by Bentzen *et al.*, 1996). The only way to test such a hypothesis would be to sample spawning fish populations and, if possible, juvenile fish in nursery grounds (as in Nesbo *et al.*, 2000). The previous paragraph highlights some difficulties of sampling juveniles.

The relationship of the Gulf of Guinea to neighbouring large marine ecosystems

Work on *S. aurita*, *S. maderensis*, *T. trecae* and *S. officinalis* have all shown very striking differentiation when comparing individuals from within the Gulf of Guinea to those of ecosystems of southern Europe (*S. aurita*, *S. maderensis* and *S. officinalis*) or southern

and western Africa (*T. trecae*). Future work could be to sample either side of the borders to see if a sharp genetic continuity exists, such as that seen for example in the Gulf of Mexico (Avise, 1992) or the Straits of Gibraltar (Perez-Losada *et al.*, 2000), or if there are signs of genetic introgression.

One problem with such a project would be to decide where exactly the border is. Section 1.3 introduces the Gulf of Guinea and highlights how borders between the neighbouring LMEs vary from season to season, as they are tidal ranges or fronts caused by varying oceanographic processes (Sherman, 1993). The current work with remote sensing data in the Gulf of Guinea (Anon, 2000) promises to help define in greater detail where these fronts are and how they vary seasonally.

The identification of the "odd" *Sepia* sp. and *P. bellottii* individuals

Within the Gulf of Guinea *Sepia* samples there are four individuals who, because of their extraordinary allele sizes, were classified together and termed "odd". The "odd" sample displays closer affinity to Spanish than west African samples. Are they *Sepia officinalis* samples, while the mainstay of the Gulf of Guinea samples are the subspecies *Sepia officinalis hierredda*? If the "odd" samples are true *Sepia officinalis* why do not all loci amplify successfully (as in Perez-Losada *et al.*, 2000), rather than just Sof 1, 2, 5 and 7? Likewise, four individuals of *P. bellottii* were distinct from all other individuals, their contribution to genetic variation totally dominating the rest of the samples. As with the "odd" sample their origin and identity is currently unclear.

Once further samples are collected it will be necessary to use the same molecular markers to allow comparison between sample sets. The primers used on *P. bellottii* are given in this project (page 53), and sequences of the 12S rRNA gene is available via Genebank[1]. For *Sepia* spp. the microsatellite primers are described in Shaw & Perez-Losada (1999) and results in this thesis include information such as allele sizes (page 131) that can be used for comparison.

[1]The common haplotype (haplotype 1) is accession number AF271656, while the extended haplotype (haplotype 13) is AF271657 - see Chapter 4 for more details. Database at http://ncbi.nlm.nih.gov

More sample collection of *P. bellottii*

The assemblages of Longhurst (1969) introduced in section 1.3.2 suggest that the sub-thermocline fauna of the Gulf of Guinea is more similar to European waters than to neighbouring tropical waters. An investigation of the relationships between samples from the Gulf of Guinea, the Canary Current LME and European/Mediterranean waters would offer some valuable insight into the evolutionary relationships not only of *P. bellottii* populations, but also to the assemblage structures and relationships along the northwest African coast.

Pagellus bellottii is distributed from the Straits of Gibraltar to Angola, and around the Canary Islands. It is also found in the southwestern Mediterranean (Whitehead *et al.*, 1986). Given the similarity of Gulf of Guinea sub-thermocline waters to European waters it would be interesting to see if samples from these two regions were genetically and/or morphologically more similar to one another than to samples from the Canary Current LME, presumably because of either similar selection pressures (but see also Chikhi *et al.*, 1997), or a recent range expansion (i.e. a founder effect).

Two valuable primer resources for *Pagellus*, and particularly *Pagellus bogaraveo*, are Ostellari *et al.* (1996) (primers in the control region of *Pagellus bogaraveo*) and Stockley *et al.* (2000) (ten microsatellite primer sets for *Pagellus bogaraveo*). Primers from both papers could be tested on *P. bellottii* samples.

Further work on the taxonomy of *Trachurus* spp. and *Sepia* spp.

There were highly significant variations between Gulf of Guinea populations of *Trachurus* sp. and *Sepia* sp. and non-Gulf of Guinea populations of the two species. Such a result is unsurprising, given the unique physical characteristics of the Gulf of Guinea. However, from this study there is not enough variation to justify taxonomic distinctiveness (following Avise & Walker, 1999). Further work, both molecular and morphological, could shine further light on the taxonomic identity of these and other Gulf of Guinea populations, as the taxonomic comparisons in this thesis are based on limited molecular data.

7.5 Gulf of Guinea Sustainable Fisheries Project

This study was funded by the European Union. It fell under the aegis of the EU project entitled: Impacts of Environmental Forcing on Marine Biodiversity and Sustainable Management in the Gulf of Guinea (Anon, 2000). As such the two main goals of the project are: *(i)* to assess the impacts of upwelling and other forms of environmental forcing on marine biodiversity and the dynamics of artisanal and industrial fisheries and *(ii)* to develop and implement an information and analysis system for the sustainable management and governance of fisheries resources (FIAS), thereby applying knowledge from goal *(i)* to real world problems.

The project targeted six measurable objectives. These were:

1. Identification of the spatio-temporal scales over which environmental forcing can be detected using remotely sensed and field data derived from a variety of local, regional and international sources.

2. Identification of biophysical habitats along the Guinean continental shelf critical to marine fisheries and biodiversity.

3. Generation of testable hypotheses regarding causal relationships between environmental forcing factors and inter-annual variability in ecosystem behaviour, using output from coupled models and observational records.

4. Investigation and parameterisation of the relationships identified in (3) through field studies and experiments with emphasis on the temporal and spatial patterning of physical events (seasonal sequences, resonant behaviour and episodic variability) and the consequent sensitivity of biological systems to changes in hydrodynamic processes.

5. Implementation of the BDRM (regional marine database) and FIAS with a Gulf of Guinea focus in research institutes in Ghana and selected centres in the region.

6. Establishment of an information network for researchers in European Member states and in the Gulf of Guinea region.

This project on the population structure of three important species provides valuable information in the pursuit of various objectives, most notably objective numbers 1, 3 and 4.

Objective 1: It has aided in the identification of widespread movement along the Gulf of Guinea, even of demersal species which previously were thought to maintain more localised populations, such as *P. bellottii* and *S. officinalis*. Thus signs of environmental forcing, with respect to the genetic identity of spatially localised sub-populations, is not apparent across the length of the sampled region of the Gulf of Guinea (other than possibly through dispersal by currents).

Objective 3: The heterogeneity found points to there being at least semi-distinct sub-populations in the region that may well be based around natal homing to particular spawning grounds along the coastline, such as regions to the east of the two main capes (Cape Palmas and Cape Three Points). This is a testable hypothesis of the kind to be identified in objective 3. To test this hypothesis it would be necessary to catch spawning individuals and, if possible, identify separate nursery grounds and sample juveniles from there. It would also aid in the identification of critical biophysical habitats mentioned in objective 2.

Objective 4: The temporal variation found in *T. trecae* has highlighted the need for investigation into the relationship between the temporal patterns of physical events, such as the seasonal upwelling, and the interaction of the biological systems. The relationship of the genetic variation of cohorts of juveniles from particular seasons, to the physical characteristics of those seasons, is of great interest not only in the Gulf of Guinea but for all fishery ecosystems. The process of recruitment and the question of "who are the spawners?" is of key importance to the understanding, and good governance, of all fisheries.

Of course the successful and productive management and governance of a fishery needs a great deal more than a scientific base. The major problems are often those of communication, consensus and, more often than not, effective enforcement. However, it is

hoped that by providing a preliminary study on the genetic variation of some commercially important species in the region, further questions on the biophysical relationships in the region will be more focused, more testable and thus more useful to scientists and managers alike.

Bibliography

Anon, 1981. FAO species identification sheets: fishing areas 34, 47 (in part) (E.C. Atlantic). FAO, Rome.

Anon, 1992. FAO yearbook fishery statistics of catches and landings. FAO, Rome.

Anon, 1994. Aquaculture development and research in sub-Saharan Africa Cifa technical paper, No. 23, suppl. FAO, Rome.

Anon, 1999. Surveys of the fish resources of the western Gulf of Guinea: survey of the pelagic and demersal resources, 19 April - 6 May, 1999. NORAD - FAO/UNDP Project, GLO 92/013; Cruise reports of the RV Dr.Fridtjof Nansen.

Anon, 2000. EC INCO-DC project: Impacts of environmental forcing on marine biodiversity and sustainable management in the Gulf of Guinea. Contract No. ERBIC18CT960094.

Arnason, E., Petersen, P. H., Kristinsson, K., Sigurgislason, H. & Palsson, S., 2000. Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod from Iceland and Greenland. *Journal of Fish Biology*, **56**:409–430.

Avise, J. C., 1992. Molecular population structure and the biogeographic history of a regional fauna: a case history with lessons for conservation biology. *Oikos*, **63**:62–76.

Avise, J. C., 1994. *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York. 511pp.

Avise, J. C., Marin Ball, R. & Arnold, J., 1988. Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial

DNA lineages and inbreeding theory for neutral mutations. *Molecular Biology and Evolution*, **5**:331–344.

Avise, J. C. & Walker, D., 1999. Species realities and numbers in sexual vertebrates: Perspectives from an asexually transmitted genome. *Proceedings of the National Academy of Sciences, USA*, **96**:992–995.

Ballard, J. W. O. & Kreitman, M., 1995. Is mitochondrial DNA a strictly neutral marker? *Trends in Ecology and Evolution*, **10**:485–488.

Barber, I., Downey, L. C. & Braithwaite, V. A., 1998. Parasitism, oddity and the mechanism of shoal choice. *Journal of Fish Biology*, **53**:1365–1368.

Bartley, D., Bagley, M., Gall, G. & Bentley, B., 1992. Use of linkage disequilibrium data to estimate effective population size of hatchery and natural fish populations. *Conservation Biology*, **6**:365–375.

Bentzen, P., Taggart, C. T., Ruzzante, D. E. & Cook, D., 1996. Microsatellite polymorphism and the population structure of the Atlantic cod (*Gadus morhua*) in the northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**:2706–2721.

Bernatchez, L. & Martin, S., 1996. Mitochondrial DNA diversity in anadromous rainbow smelt, *Osmerus mordax*, mitchell: a genetic assesment of the member-vagrant hypothesis. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**:424–433.

Binet, D. & Marchal, E., 1993. The Large Marine Ecosystems of shelf areas in the Gulf of Guinea: Long term variability induced by climatic changes. In Sherman *et al.*, editor, *Large Marine Ecosystems - Stress Mitigation and Sustainability*. American Association for the Advancement of Science. 376pp.

Boehlert, G. W. & Mundy, B. C., 1988. Roles of behavioural and physical factors in larval and juvenile fish recruitment to estuarine nursery areas. *American Fisheries Society Symposium*, **3**:51–67.

Bowen, B. W., Meylan, A. B., Perran Ross, J., Limpus, C. J., Balzas, G. H. & Avise, J. C., 1992. Global population structure and natural history of the green turtle (*Chelonia mydas*) in terms of matriarchal phylogeny. *Evolution*, **46**:865–881.

Boyle, P. R. & Boletzky, S., 1996. Cephalopod populations: definitions and dynamics. *Philosophical Transactions of the Royal Society of London, series B.*, **351**:985–1002.

Brown, B. L., Epifanio, J. M., Smouse, P. E. & Kobak, C. J., 1996. Temporal stability of mtDNA haplotype frequencies in American shad stocks: to pool or not to pool across years. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**:2274–2283.

Buckley, L. J., Smigielski, A. S., Halavik, T. A., Caldarone, E. M., Birns, B. R. & Laurence, G. C., 1991. Winter flounder *Pseudopleuronectes americanus* reproductive success. II. Effects of spawning time and female size on size, composition and viability of eggs and larvae. *Marine Ecology Progress Series*, **74**:125–135.

Cantatore, P., Roberti, M., Pesole, G., Ludovico, A., Milella, F., Gadaleta, M. N. & Saccone, C., 1994. Evolutionary analysis of cytochrome b sequences in some perciformes: evidence for a slower rate of evolution than in mammals. *Journal of Molecular Evolution*, **39**:589–597.

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

Chikhi, L., 1995. *Differenciation genetique chez* Sardinella aurita *et* S. maderensis. *Allozymes et ADN mitochondrial*. Ph.D. thesis, Universite Pierre et Marie Curie, Paris VI. 236pp.

Chikhi, L., Agnese, J. F. & Bonhomme, F., 1997. Marked mitochondrial DNA differences between Mediterranean and Eastern Atlantic populations of the round sardinella *Sardinella aurita*. *Academie des Sciences, Paris, Sciences de la vie*, **320**:289–297.

Chikhi, L., Bonhomme, F. & Agnese, J. F., 1998. Low genetic variability in a widely distributed and abundant clupeid species, *Sardinella aurita*. *Journal of Fish Biology*, **52**:861–878.

Cockerham, C. C., 1969. The variance of gene frequencies. *Evolution*, **23**:72–83.

Cockerham, C. C., 1973. Analysis of gene frequencies. *Genetics*, **74**:679–700.

Cohen, J., 1977. *Reproduction*. Butterworths, London. 356pp.

Cole, J., 1999. Environmental conditions, satellite imagery, and clupeoid recruitment in the northern Benguela upwelling system. *Fisheries Oceanography*, **8**:25–38.

Cowen, R. K., Hare, J. A. & Fahay, M. P., 1993. Beyond hydrography: can physical processes explain larval fish assemblages within the middle Atlantic bight? *Bulletin of Marine Science*, **53**:567–587.

Cury, P., 1994. Obstinate nature: an ecology of individuals. thoughts on reproductive behaviour and biodiversity. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**:1664–1673.

Cushing, D., 1995. *Population production and regulation in the sea: a fisheries perspective*. Cambridge University Press. 354pp.

Cushing, D. H., 1972. The production cycle and the numbers of marine fish. *Symposium of the Zoological Society of London*, **29**:213–232.

DiRienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M. & Freimer, N. B., 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences, USA*, **91**:3166–3170.

Dizon, A. E., Lockyer, C., Perrin, W. F., Demaster, D. P. & Sisson, J., 1992. Rethinking the stock concept: A phylogeographic approach. *Conservation Biology*, **6**:24–36.

Dover, G., 1995. Slippery DNA runs on and on... *Nature Genetics*, **10**:254–256.

Economou, A. N., 1991. Is dispersal of fish eggs, embryos and larvae an insurance against density dependence. *Environmental Biology of Fishes*, **31**:313–321.

Excoffier, L. & Smouse, P. E., 1994. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: Molecular variance parsimony. *Genetics*, **136**:343–359.

Excoffier, L., Smouse, P. E. & 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.

Felsenstein, J., 1989. Phylip – phylogeny inference package (version 3.2). *Cladistics*, **5**:164–166.

Fischer, K. M., 1998. Expanded (CAG)n, (CGG)n and (GAA)n trinucleotide repeat microsatellites, and mutant purine synthesis and pigmentation genes cause schizophrenia and autism. *Medical hypotheses*, **51**:223–233.

Fitch, W. M., 1977. On the problem of discovering the most parsimonious tree. *American Naturalist*, **111**:223–257.

Frank, K. T., 1992. Demographic consequences of age-specific dispersal in marine fish populations. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**:2222–2231.

Frankham, R., 1995. Effective population size/adult population size ratios in wildlife: a review. *Genetic Research*, **66**:95–107.

Fu, Y. H., 1994. A phylogenetic estimator of effective population size or mutation rate. *Genetics*, **136**:685–692.

Fu, Y. H. & Li, W.-H., 1993. Statistical tests of neutrality of mutations. *Genetics*, **133**:693–709.

Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L. & Feldman, M. W., 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences, USA*, **92**:6723–6725.

Goodman, S. J., 1997. RST CALC: A collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Molecular Ecology*, **6**:881–885.

Guo, S. W. & Thompson, E. A., 1992. Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics*, **48**:361–372.

Hanlon, R. T. & Messenger, J. B., 1996. *Cephalopod behaviour*. Cambridge University Press. 232pp.

Harden-Jones, F. R., 1968. *Fish Migration*. Edward Arnold, London. 375pp.

Hartl, D. H. & Clark, A. G., 1997. *Principles of Population Genetics*. Sinauer Associates, Sunderland, Massachusetts. 542pp, 3rd edition.

Heard, W. R., 1991. Life history of pink salmon (*Oncoryynchus gorbuscha*). In Groot, C. & Margolis, L., editors, *Pacific salmon life histories*, pages 121–230. University of British Columbia Press, Vancouver, B.C.

Hecht, T., 1990. On the life history of cape horse mackerel (*Trachurus trachurus capensis*) off the south east coast of South Africa. *South African Journal of Marine Science*, **9**:317–326.

Hedgecock, D., 1994. Does variance in reproductive success limit effective population sizes of marine organisms? In Beaumont, A. R., editor, *Genetics and Evolution of Aquatic Organisms*, pages 122–134. Chapman and Hall, London.

Hedgecock, D., Hutchinson, E. S., Li, G., Sly, F. L. & Nelson, K., 1994. The central stock of northern anchovy (*Engraulis mordax*) is not a randomly mating population. *CalCOFI Report*, **35**:121–136.

Hey, J., 2000. Human mitochondrial DNA recombination: can it be true? *Trends in Ecology and Evolution*, **15**:181–182.

Hoeh, W. R., Blakley, K. H. & Brown, W. M., 1991. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. *Science*, **251**:1488–1490.

Houghton, R. W., 1983. Seasonal-variations of the subsurface thermal structure in the Gulf of Guinea. *Journal of Physical Oceanography*, **13**:2070–2081.

Hutchings, J. A., Bishop, T. D. & McGregor-Shaw, C. R., 1999. Spawning behaviour of Atlantic cod, *Gadus morhua*: evidence of mate competition and mate choice in a broadcast spawner. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**:97–104.

Iles, T. & Sinclair, M., 1982. Atlantic herring: stock discretness and abundance. *Science*, **215**:627–633.

Johannes, R. E., 1978. Reproductive strategies of coastal marine fish in the tropics. *Environmental Biology of Fishes*, **3**:65–84.

Johnson, M. S. & Black, R., 1982. Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology*, **70**:157–164.

Johnson, M. S. & Black, R., 1984. Pattern beneath the chaos: the effect of recruitment on genetic patchiness in an intertidal limpet. *Evolution*, **38**:1371–1383.

Jorde, P. E. & Ryman, N., 1996. Demographic genetics of brown trout (*Salmo trutta*) and estimation of effective population size from temporal change of allele frequencies. *Genetics*, **143**:1369–1381.

Karlouriga, C. & Economidis, P. S., 1997. Spawning frequency and batch fecundity of horse mackerel *Trachurus trachurus* (L.) in the Saronikos Gulf (Greece). *Journal of Applied Ichthyology*, **13**:97–104.

Kerstan, M., 1995. Ages and growth-rates of Agulhas Bank horse mackerel *Trachurus trachurus capensis* - comparison of otilith ageing and length frequency analyses. *South African Journal of Marine Science*, **15**:137–156.

Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature*, **217**:624–626.

Kimura, M., 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**:893–903.

Kimura, M., 1971. Theoretical foundation of population level at the molecular level. *Theoretical population biology*, **2**:174–208.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**:111–120.

Kimura, M. & Crow, J. F., 1964. The number of alleles that can be maintained in a finite population. *Genetics*, **49**:725–738.

Kimura, M. & Ohta, T., 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences, USA*, **75**:2868–2872.

King, J. L. & Jukes, T. H., 1969. Non-darwinian evolution: Random fixation of selectively neutral mutations. *Science*, **164**:788–798.

Kingsford, M. J., 1990. Linear oceanographic features: a focus for research on recruitment processes. *Australian Journal of Ecology*, **15**:391–401.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X. & Wilson, A. C., 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences, USA*, **86**:6196–6200.

Koehn, R. K. & Williams, G. C., 1978. Genetic differentiation without isolation in the American eel, *Anguilla rostrata* II. Temporal stability of geographic patterns. *Evolution*, **32**:624–637.

Koranteng, K. A., McGlade, J. M. & Samb, B., 1996. Review of the Canary Current and Guinea Current Large Marine Ecosystems. In *ACP-EU Fisheries Research Initiative. Second Dialogue Meeting, Dakar, 22-26 April, 1996*.

Koranteng, K. A. & Pitcher, T. J., 1987. Population parameters, biannual cohorts, and assesment in the *Pagellus bellottii* (Sparidae) fishery off Ghana. *Journal du Conseil*, **43**:129–138.

Kreitman, M., 1996. The neutral theory is dead. long live the neutral theory. *BioEssays*, **18**:678–683.

Kuhner, M. K., Yamato, J. & Felsenstein, J., 1995. Estimation of effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics*, **140**:1421–1430.

Laegdsgaard, P. & Johnson, C. R., 1995. Mangrove habitats as nurseries: unique assemblages of juvenile fish in subtropical mangroves in eastern Australia. *Marine Ecology Progress Series*, **126**:67–81.

Lee, W.-J., Conroy, J., Huntting Howell, W. & Kocher, T. D., 1995. Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**:54–66.

Lessios, H. A., 1992. Testing electrophoretic data for agreement with Hardy-Weinberg expectations. *Marine Biology*, **112**:517–523.

Levins, R., 1968. *Evolution in changing environments: Some theoretical explorations*. Princeton University Press. 120pp.

Levinton, J. S., 1995. *Marine Biology: function, biodiversity, ecology*. Oxford University Press, New York.

Li, C. C., 1988. Pseudo-random mating populations. In celebration of the 80th anniversary of the Hardy-Weinberg law. *Genetics*, **119**:731–737.

Li, G. & Hedgecock, D., 1998. Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variance in reproductive success. *Canadian Journal of Fisheries and Aquatic Scienes*, **55**:1025–1033.

Li, W. H., 1997. *Molecular Evolution*. Sinauer Associates, Sunderland, Massachusetts. 487pp.

Lipinski, M. R., Compagno Roeleveld, M. A. & Augustyn, C. J., 1992. First study on the ecology of *Sepia australis* in the Southern Benguela ecosystem. *The Veliger*, **35**:385–395.

Longhurst, A. R., 1969. Species assemblages in tropical demersal fisheries. In *Proceedings of the symposium on oceanography and fisheries resources in the tropical Atlantic*.

Longhurst, A. R. & Pauly, D., 1987. *Ecology of tropical oceans*. Academic Press, London. 407pp.

Lu, G., Li, S. & Bernatchez, L., 1997. Mitochondrial DNA diversity, population structure and conservation genetics of four native carps within the Yangtze River, China. *Canadian Journal of Fisheries and Aquatic Scienes*, **54**:47–58.

Magurran, A. E., 1993. Individual differences and alternative behaviours. In Pitcher, T. J., editor, *Behaviour of Teleost Fishes*, pages 441–477. Chapman and Hall, London, 2nd edition.

Majerus, M., Amos, W. & Hurst, G., 1996. *Evolution: The Four Billion Year War*. Longman, Addison Wesley Longman Ltd., Harlow. 340pp.

Mann, K. H. & Lazier, J. R. N., 1991. *Dynamics of Marine Ecosystems*. Blackwell Scientific Publications, Boston. 466pp.

Margush, T. & McMorris, F. R., 1981. Consensus n-trees. *Bulletin of Mathematical Biology*, **43**:239–244.

Maxim, C., 1995. Horse mackerel and false scad stock assessment and catch projections, CECAF Division-4.1.3 and Division-34.3.1. *Scientia Marina*, **59**:611–627.

McDonald, J. & Kreitman, M., 1991. Adaptive protein evolution at *adh* locus in *Drosophila*. *Nature*, **351**:652-654.

McQuinn, I. H., 1997. Metapopulations and the altantic herring. *Reviews in Fish Biology and Fisheries*, **7**:297–329.

Meylan, A. B., Bowen, B. W. & Avise, J. C., 1990. A genetic test of the natal homing versus social faciliation models for green turtle migration. *Science*, **248**:724–727.

Miller, J. M., 1988. Physical processes and the mechanisms of coastal migrations of immature marine fishes. *American Fisheries Society Symposium*, **3**:68–76.

Miller, L. M. & Kapuscinski, A. R., 1997. Historical analysis of genetic variation reveals low effective population size in a northern pike (*Esox lucius*) population. *Genetics*, **147**:1249–1258.

Naish, K. A., Hecht, T. & Payne, A. I. L., 1991. Growth of Cape horse mackerel *Trachurus trachurus capensis* off South Africa. *South African Journal of Marine Science*, **10**:29–35.

Nei, M., 1975. *Molecular population genetics and evolution*. North-Holland, Amsterdam. 288pp.

Nei, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York. 512pp.

Nesbo, C. L., Rueness, E. K., Iverson, S. A., Skagen, D. W. & Jakobsen, K. S., 2000. Phylogeography and population history of Atlantic mackerel (*Scomber scombrus* L.): a genealogical approach reveals genetic structuring among the eastern Atlantic stocks. *Philosophical Transactions of the Royal Society of London, series B.*, **267**:281–292.

Norcross, B. L. & Shaw, R. F., 1984. Oceanic and estuarine transport of fish eggs and larvae: a review. *Transactions of the American Fisheries Society*, **113**:153–165.

O'Connell, M. & Wright, J. M., 1997. Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries*, **7**:331–363.

O'Corry-Crowe, G. M., Suydam, R. S., Rosenberg, A., Frost, K. J. & Dizon, A. E., 1997. Phylogeography, population structure and dispersal patterns of the beluga whale *Delphinapterus leucas* in the western Nearctic revealed by mitochondrial DNA. *Molecular Ecology*, **6**:955–970.

Ohta, T., 1996. The current significance and standing of neutral and nearly neutral theories. *BioEssays*, **18**:673–677.

Ohta, T. & Kimura, M., 1973. A model of mutation appropiate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetic Research*, **22**:201–204.

Olson, D. E., Schupp, D. & Macins, V., 1978. An hypothesis of homing behaviour in walleyes as related to observed patterns of passive and active movement. *American Fisheries Society Special Publications*, **11**:52–57.

Ostellari, L., Bargelloni, L., Penzo, E., Paternello, P. & Paternello, T., 1996. Optimization of single-strand conformation polymorphism and sequence analysis of the mitochondrial control region in *Pagellus bogaraveo* (Sparidae, Teleostei): rationalized tools in fish population biology. *Animal Genetics*, **27**:423–427.

Owens, D. W., Grassman, M. & Hendrickson, J. R., 1982. The imprinting hypothesis and sea turtle reproduction. *Herpetologica*, **38**:124–135.

Page, R. D. M. & Holmes, E. C., 1998. *Molecular Evolution: a phylogenetic approach*. Blackwell Science, Oxford. 346pp.

Palumbi, S. R., 1996. Nucleic acids II: The Polymerase Chain Reaction. In Hillis, D. M., Moritz, C. & Mable, B. K., editors, *Molecular Systematics*. Sinauer Associates, Sunderland, Massachusetts. 655pp, 2nd edition.

Perez-Losada, M., Shaw, P. W., Guerra, A., Carvalho, G. R. & Sanjuan, A., 2000. Fine population structuring in the cuttlefish *Sepia officinalis* (Mollusca: Cephalopoda) around the Iberian Peninsula revealed by microsatellite DNA markers. *Marine Biology*.

Picaut, J., 1983. Propagation of the seasonal upwelling in the eastern equatorial Atlantic. *Journal of Physical Oceanography*, **13**:18–37.

Pilar Olivar, M. & Shelton, P. A., 1993. Larval fish assemblages of the Benguela current. *Bulletin of Marine Science*, **53**:450–474.

Pitcher, T. J. & Parish, J. K., 1993. Functions of shoaling behaviour in teleosts. In Pitcher, T. J., editor, *Behaviour of Teleost Fishes*, pages 363–439. Chapman and Hall, 2nd edition.

Planes, S., Galzin, R. & Bonhomme, F., 1996. A genetic metapopulation model for reef fishes in oceanic islands: the case of the surgeonfish, *Acanthurus triostegus*. *Journal of Evolutionary Biology*, **9**:103–117.

Pogson, G. H., Mesa, K. A. & Boutilier, R. G., 1995. Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics*, **139**:375–385.

Quinn, T. P. & Busack, C. A., 1985. Chemosensory recognition of siblings in juvenile coho salmon (*Oncorhynchus kisutch*). *Animal Behaviour*, **33**:51–56.

Rand, D. M., 1996. Neutrality tests of molecular markers and the connection between DNA polymorphism, demography and conservation. *Conservation Biology*, **10**:665–671.

Rand, D. M., Dorfsman, M. & Kann, L. M., 1994. Neutral and non-neutral evolution of drosophila mitochondrial DNA. *Genetics*, **138**:741–756.

Raymond, M. & Rousset, F., 1995. Genepop (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**:248–249.

Reynolds, J., Weir, B. S. & Cockerham, C. C., 1983. Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, **105**:767–779.

Rijavec, L., 1973. Biology and dynamics of *Pagellus coupei*, *Pagrus ehrenbergi* and *Dentex canariensis* in Ghana waters. Scientific documents of the Oceanography Research Centre, Abidjan.

Roper, C. F. E., Sweeney, M. J. & Nauen, C. E., 1984. Cephalopods of the world. an annotated and illustrated catalogue of species of interest to fisheries. FAO Fisheries Synopsis (125). 277pp.

Roughgarden, J., Gaines, S. & Possingham, H., 1988. Recruitment dynamics in complex life cycles. *Science*, **241**:1460–1466.

Roughgarden, J., Running, S. W. & Matson, P. A., 1991. What does remote sensing do for ecology? *Ecology*, **72**:1918–1922.

Rozas, J. & Rozas, R., 1999. DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**:174–175.

Rubinsztein, D. C., Amos, W. & Leggo, J., 1995. Microsatellites and generally longer in humans compared to their homologues in non-human primates: evidence for directional evolution at microsatellite loci. *Nature Genetics*, **10**:337–343.

Ruzzante, D. E., 1998. A comparison of several measures of genetic distance and population structure with microsatellite data: bias and sampling variance. *Canadian Journal of Fisheries and Aquatic Scienes*, **55**:1–14.

Ruzzante, D. E., Taggart, C. T. & Cook, D., 1996. Spatial and temporal variation in the genetic composition of a larval cod (*Gadus morhua*) aggregation: cohort contribution and genetic stability. *Canadian Journal of Fisheries and Aquatic Scienes*, **53**:2695–2705.

Sabates, A. & Pilar Olivar, M., 1996. Variation of larval fish distributions associated with variability in the location of a shelf-slope front. *Marine Ecology Progress Series*, **126**:253–259.

Saitou, N. & Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**:406–425.

Sandercock, F. K., 1991. Life history of coho salmon (*Oncoryynchus kisutch*). In Groot, C. & Margolis, L., editors, *Pacific salmon life histories*, pages 397–445. University of British Columbia Press, Vancouver, B.C.

Sanjuan, A., Perez-Losada, M. & Guerra, A., 1996. Genetic differentiation in three *Sepia* species (Mollusca: Cephalopoda) from Galician waters (north west Iberian Peninsula. *Marine Biology*, **126**:253–259.

Schneider, S., Kueffer, J. M., Roessli, D. & Excoffier, L., 1997. *Arlequin ver 1.1: A software for population genetic data analysis*. Genetics and Biometry Laboratory, University of Geneva Switzerland. http://lgb.unige.ch/arlequin.

Shaw, P. W. & Perez-Losada, M., 1999. Polymorphic microsatellites in the common cuttlefish *Sepia officinalis*. *Molecular Ecology*, **In Press**:In Press.

Shaw, P. W., Pierce, G. J. & Boyle, P. R., 1999. Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. *Molecular Ecology*, **8**:407–417.

Sherman, K., 1993. Large Marine Ecosystems as global units for marine resources management - an ecological perspective. In Sherman *et al.*, editor, *Large Marine Ecosystems - Stress, Mitigation and Sustainability*. American Association for the Advancement of Science. 376pp.

Sherman, K., 1994. Sustainability, biomass yields and health of coastal ecosystems: an ecological perspective. *Marine Ecology Progress Series*, **112**:277–301.

Simpson, J. H. & Hunter, J. R., 1974. Fronts in the Irish Sea. *Nature*, **250**:404–406.

Sinclair, M., 1988. *Marine Populations: An essay on Population Regulation and Speciation*. Washington Sea Grant/University of Washington Press. 252pp.

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

Smith, P. J., Jamieson, A. & Birley, A. J., 1990. Electrophoretic studies and the stock concept in marine teleosts. *Journal du Conseil*, **47**:231–245.

Smouse, P. E., 1998. To tree or not to tree. *Molecular Ecology*, **7**:399–412.

Sokal, R. R. & Rohlf, F. J., 1962. The comparison of dendrograms by objective methods. *Taxon*, **11**:33–40.

Sokal, R. R. & Rohlf, F. J., 1981. *Biometry*. W.H. Freeman, San Francisco. 859pp.

Stockley, B. M., Rogers, A. D., Iyengar, A., Menezes, G., Santos, R. & Long, A., 2000. Ten microsatellite loci isolated and developed for the blackspot seabream, *Pagellus bogaraveo* (Brunnich 1768). *Molecular Ecology*, **9**:999–1000.

Suthers, I. M. & Frank, K. T., 1991. Comparative persistence of marine fish larvae from pelagic versus demersal eggs off southwestern nova scotia, canada. *Marine Biology*, **108**:175–184.

Tajima, F., 1983. Evolutionary relationships of DNA sequences in finite populations. *Genetics*, **105**:437–460.

Tajima, F., 1989a. The effect of change in population size on DNA polymorphism. *Genetics*, **123**:597–601.

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

Tessier, N. & Bernatchez, L., 1999. Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* l.). *Molecular Ecology*, **8**:169–179.

Thompson, J., Higgins, D. & Gibson, T., 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.

Tilot, V. & King, A., 1993. A review of the subsystems of the Canary Current and Gulf of Guinea Large Marine Ecosystems. IUCN Marine Programme. 50pp.

Todd, C. D., 1998. Larval supply and recruitment of benthic invertebrates: do larvae always disperse as much as we believe? *Hydrobiologia*, **376**:1–21.

Todd, C. D.and Lambert, W. J. & Thorpe, J. P., 1998. The genetic structure of intertidal populations of two species of nudibranch molluscs with planktotrophic and pelagic lecithotrophic larval stages: are pelagic larvae "for" dispersal? *Journal of experimental marine biology and ecology*, **228**:1–28.

Toreson, R., 1986. Length and age at maturity of Norwegian spring spawning herring for the year classes 1958-1961 and 1973-1978. ICES, C.M. H:42.

Utter, F. & Ryman, N., 1993. Genetic markers and mixed stock fisheries. *Fisheries*, **18**:11–21.

Utter, F. M., Campton, D., Grant, S., Milner, G. B., Seeb, J. & Wischard, L., 1980. Population structures of indigenous salmonid species of the Pacific Northwest. In McNeil, W. J. & Hinsworth, D. C., editors, *Salmonid ecosystems of the Pacific Northwest*, pages 285–304. Oregon State University Press, Corvallis.

Van Havre, N. & Fitzgerald, G. J., 1988. Shoaling and kin recognition in the three-spined stickleback (*Gasterosteus aculeatus*, L.). *Biology of Behaviour*, **13**:190–201.

Watterson, G. A., 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**:256–276.

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

Whitehead, P. J., Bauchot, M. L., Hureau, J. C., Nielsen, J. & Tortonese, E., 1986. *Fishes of the North-Eastern Atlantic and the Mediterranean*. UNESCO, Paris. 1473pp.

Willems, P. J., 1994. Dynamic mutations hit double figures. *Nature Genetics*, **8**:213–215.

Wirgin, I. I. & Waldman, J. R., 1994. What DNA can do for you. *Fisheries*, **19**:16–27.

Wright, S., 1951. The genetical structure of populations. *Annual 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, S., 1978. *Evolution and the genetics of populations. Volume 4. Variability within and among natural populations*. University of Chicago Press.

Zardoya, R., Garrido-Pertierra, A. & Bautista, J. M., 1995. The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. *Journal of Molecular Evolution*, **41**:942–951.

Zhao, J. L. & Li, S. F., 1995. Isozymes study on population divergence of silver carp, bighead carp, grass carp and black carp in the middle and lower regions of Chang jiang River. *Journal of Fisheries in China*, **20**:104–110.

Zouros, E., Pogson, G. H., Cook, D. I. & Dadswell, M. J., 1995. Apparent selective neutrality of mitochondrial DNA size variation: a test in the deep sea scallop *Placopecten magellanicus*. *Evolution*, **46**:1466–1476.

Appendix A

List of common names

All species listed in the text are given here, alphabetically, with their common English names in parentheses. Some species or genera which do not have common English species names have the common English family name instead.

Acanthocybium solandri (wahoo)

Archocentrus (a cichlid)

Argyrosomus regius (meagre)

Balistes capriscus (grey triggerfish)

Boops boops (bogue)

Brachydeuterus auritus (bigeye grunt)

Caquetaia (a cichlid)

Channa (snakehead)

Clupea harengus (Atlantic herring)

Cynglossus spp. (tonguesoles)

Decapterus rhonchus (false scad)

Dentex (a seabream)

Dentex angolensis (Angola dentex)

Dicentrarchus (bass)

Engraulis anchorita (Southwest Atlantic anchovy)

Engraulis encrasicolus (European anchovy)

Engraulis mordax (northern anchovy)

Epinephelus aeneus (white grouper)

Eretmodus (a cichlid)

Etheostoma (darter)

Ethmalosa fimbriata (bonga shad)

Euthynnus alletteratus (Atlantic little tunny)

Galeoides decadactylus (lesser African threadfin)

Geryon maritae (deep-sea red crab)

Haplochromis (a cichlid)

Herotilapia (a cichlid)

Ilisha africana (West African ilisha)

Katsuwonus pelamis (skipjack tuna)

Macroramphosus spp. (snipefish)

Merluccius capensis (shallow water cape hake)

Neogobius (a goby)

Oncorhynchus gorbascha (pink salmon)

Oncorhynchus kisutch (coho salmon)

Oncorhynchus mykiss (rainbow trout)

Oncorhynchus tshawytscha (chinook salmon)

Oreochromis (a cichlid)

Osmerus mordax (rainbow smelt)

Pagellus acarne (Axillary seabream)

Pagellus bellottii (red pandora)

Pagellus bogaraveo (blackspot seabream)

Pagellus erythrinus (common pandora)

Pagrus (a seabream)

Parachromis (a cichlid)

Paracubiceps ledanoisi (silver-rag driftfish)

Parapenaeus longirostris (deepwater rose shrimp)

Pentheroscion mbizi (blackmouth croaker)

Perca (perch)

Plectorhynchus mediterraneus (rubberlip grunt)

Pomadasys incisus (bastard grunt)

Pomadasys jubelini (sompat grunt)

Pomatomus saltatrix (bluefish)

Pseudotolithus senegalensis (cassava croaker)

Pseudotolithus typus (longneck croaker)

Pterygosquilla armata capensis (a stomatopod)

Sarda sarda (Atlantic bonito)

Sardina pilchardus (European pilchard; Sardine)

Sardinella aurita (round sardinella)

Sardinella maderensis (madeiran sardinella)

Scomber japonicus (chub mackerel)

Scyris alexandrinus (alexandria pompano)

Sepia australis (Austalian cuttlefish)

Sepia elegans (elegant cuttlefish)

Sepia officinalis (common cuttlefish)

Sepia orbignyana (pink cuttlefish)

Solea spp. (soles)

Sparus caeruleostictus (bluespotted seabream)

Spathodus (a cichlid)

Stizostedion vitreum (walleye)

Tanganicodus (a cichlid)

Thunnus albacares (yellowfin tuna)

Thunnus obesus (bigeye tuna)

Tilapia (a cichlid)

Tomocichla (a cichlid)

Trachurus (horse mackerel)

Trachurus trachurus (Atlantic horse mackerel)

Trachurus trecae (Cunene horse mackerel)

Tropheus moorii (a cichlid)

Appendix B

Translation code for vertebrate mtDNA

The translation code for vertebrate mtDNA is given in Table B.1. The code is taken from The Genetic Codes website hosted by the National Centre for Biotechnology Information[1]. The site gives the translation codes for a variety of genomes of vertebrate and invertebrate species. Differences of the vertebrate mtDNA code from the standard eukaryote code are as follows:

	mtDNA	Standard
AGA	* Ter	R Arg
AGG	* Ter	R Arg
ATA	M Met	I Ile
TGA	W Trp	* Ter

[1] http://www.embl-heidelberg.de/JaMBW/2/3/TranslationTables.html

1st	2nd T	2nd C	2nd A	2nd G	3rd
T	F phe	S Ser	Y Tyr	C Cys	T
	F phe	S Ser	Y Tyr	C Cys	C
	L Leu	S Ser	* Ter	W Trp	A
	L Leu	S Ser	* Ter	W Trp	G
C	L Leu	P Pro	H His	R Arg	T
	L Leu	P Pro	H His	R Arg	C
	L Leu	P Pro	Q Gln	R Arg	A
	L Leu	P Pro	Q Gln	R Arg	G
A	I Ile	T Thr	N Asn	S Ser	T
	I Ile	T Thr	N Asn	S Ser	C
	M Met	T Thr	K Lys	* Ter	A
	M Met	T Thr	K Lys	* Ter	G
G	V Val	A Ala	D Asp	G Gly	T
	V Val	A Ala	D Asp	G Gly	C
	V Val	A Ala	E Glu	G Gly	A
	V Val	A Ala	E Glu	G Gly	G

Table B.1: The translation code for vertebrate mtDNA

Appendix C

Test of cytochrome *b* as a marker

As introduced in section 3.4.1, to investigate the potential use of cytochrome *b* as a marker, a Consense tree was generated across a range of species in the order Perciformes (sequences from the the National Centre for Biotechnology Information website,[1] Entrez database). The sequences were chosen as a variety of cytochrome *b* sequences from the large number of such entries in the database, given the common use of cytochrome *b* for assessing intraspecific variation.

The twenty fish were chosen due to their availability in the database. They are: *Channa* (snakehead), *Dentex* (a seabream), *Dicentrachus* (bass - Mediterranean and Atlantic), *Etheostoma* (darter), *Neogobius* (a goby), *Pagrus* (a seabream), *Perca* (perch), *Trachurus* (horse mackerel), *Tilapia*, *Oreochromis*, *Haplochromis*, *Tanganicodus*, *Eretmodus*, *Spathodus*, *Archocentrus*, *Parachromis*, *Tomocichla*, *Caquetaia* and *Herotilapia* (all cichlids).

Firstly, the dataset of cytochrome *b* genes is edited to contain only the particular 211 base pair length of the gene that was used within this project. This allows direct comparison with the *Trachurus* spp. individuals used in chapter 3. The data is bootstrapped 100 times using the program "SeqBoot" in PHYLIP (Felsenstein, 1989) to generate 100 alternative datasets. The PHYLIP program "DNAdist" then computes distance matrices for each of the 100 datasets. Using the distance matrices for each of the 100 datasets, 100 trees are computed with the neighbour-joining method, as described

[1] http://ncbi.nlm.nih.gov

in Saitou & Nei (1987), using the program "Neighbor" in PHYLIP. Finally, the Consense tree was built using PHYLIP "consensus" program, which computes consensus trees by the majority-rule consensus tree method (Margush & McMorris, 1981).

The Consense tree in Figure C.1 shows, as explained in section 3.4.1, that while there is high confidence in assignation of closely related species, or populations in a species, confidence limits from the bootstrapping process are markedly lower for higher taxonomic levels. Such a pattern of confidence limits (particularly the 100 for *Dicentrarchus* samples taken from the Atlantic and Mediterranean) suggests *(i)* a relatively high mutation rate and *(ii)* that cytochrome *b* may be a suitable molecular marker for intraspecific variation. However, the lower confidence limits of the higher taxonomic levels indicate that it is less likely to be useful for comparisons across families or higher taxonomic levels.

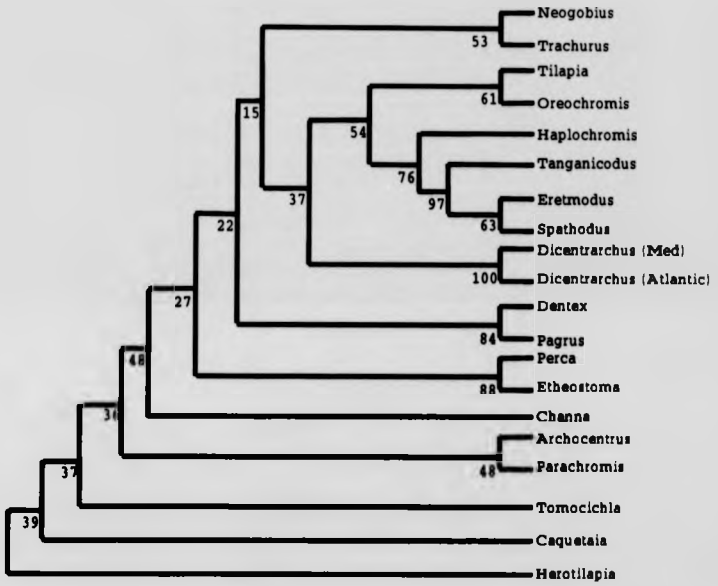

Figure C.1: Consense tree for a number of species from the order Perciformes. The numbers at the forks indicate the number of times the group consisting of the species which are to the right of that fork occurred among the trees, out of 100 trees. The high confidence for the famiy nodes indicates that the sequence can be used for differentiating closely related species, or populations of a species, but is of little use for building a tree of broader taxonomic orders (see text).

Appendix D

Translated sequence of *T. trecae* cytochrome *b*

A translated sequence of *T. trecae* cytochrome *b* partial sequence as used for sequence analysis is given in Table D.1. The vertebrate mitochondrial code used for translation of the codons to amino acids is given in Table B.1. The sequence translated is from *Trachurus trecae* individual 2, sampled at station tt99 in the Gulf of Guinea.

```
ACAGGTCTTTTCCTAGCTATACACTACACCTCGGACATCGCAACCGCCTTTAC
T--G--L--F--L--A--M--H--Y--T--S--D--I--A--T--A--F--T-

ATCCGTAGCACACATCTGCCGGGACGTAAACTACGGCTGACTTATCCGCAATA
-S--V--A--H--I--C--R--D--V--N--Y--G--W--L--I--R--N--M

TGCACGCCAACGGCGCCTCCTTCTTTTTCATTTGCATTTACCTTCACATCGGC
--H--A--N--G--A--S--F--F--F--I--C--I--Y--L--H--I--G--

CGAGGCCTTTACTACGGCTCATACCTCTACAAAGAAACCTGAAACACCGGGG
R--G--L--Y--Y--G--S--Y--L--Y--K--E--T--W--N--T--G--
```

Table D.1: A translated partial sequence of cytochrome *b*. The translation table for vertebrate mitochondrial DNA can be found in the appendices. The sequence translated here is from *Trachurus trecae* individual 2, sampled at station tt99 in the Gulf of Guinea.