Title: " Population Genetics Of The Prawn *Penaeus indicus* (Crustacea: Decapoda) In The Western Indian Ocean".

Author: Jacopo Querci
Supervisor: Prof. A.T. Forbes

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Life and Environmental Sciences, University of Natal, Durban, from JUNE 1998 to JULY 2000 under the supervision of Professor A.T. Forbes and Dr. C. O'Ryan (Department of Molecular Biology University of Cape Town).

Unless otherwise stated in the aknowledgements section, these studies represent original work by the author and have not been submitted in any form to another University.

Date 24/07/03

Signed

THE EXPERIMENTER

"It's an experience like no other experience I can describe, the best thing that can happen to a scientist, realizing that something that's happened in his or her mind exactly corresponds to something that happens in Nature. It's startling every time it occurs. One is surprised that a construct of one's own mind can actually be realized in the honest to goodness world out there. A great shock, and a great great joy."

Leo Kadanoff

Acknowledgements

This thesis would have never been written were it not for the support and encouragement of many friends and colleagues. Special thanks goes to those who helped me in the lab mastering all the techniques used in this research particularly Dr. Geeta Heik and Ms Marla Hughes and those who read and commented upon drafts, Dr. Giancarlo Contrafatto, Dr. Colleen O'Ryan, Dr. Jackie Bishops. To my brother Pierfrancesco Querci and to my friend Elisabeth Ocko, I owe special thanks for the numerous hours of conversation and time spent to homogenate the muscle tissue of more than 300 prawns in Durban. I am grateful to Prof. Amornrat Phandora from the University of Songala Thailand who provided the COI sequences from *P. merguensis* and useful information for the amplification of Pi06 locus. I owe particular thanks to Dr. Colleen O'Ryan, of the department of Molecular biology of the University of Cape Town for making up her mind to co-supervise this project the minute I appeared in the tea-room of her department. I owe particular thanks to Prof. A. T. Forbes and Dr. Giancarlo Contrafatto for the support, patience (I will learn English spelling and grammar I promise) and great encouragement during these five years in South Africa.

I owe most to my Parents and my brother Pierfrancesco (great surfer), my biggest supporter, fiercest critic and greatest friend when I was away from home. Thanks to Mauro Piga for all those nights spent looking at "Er Monnezza" and talking about great waves of the Indian Ocean. Special thanks to Piero Querci for the great cooking in Umhlanga rock's flat in 1999, hope you enjoyed Pa. Lastly I must thank "Nature" for constantly reminding me of what is really important in life.

Abstract

This research programme assessed the genetic structure and degree of polymorphism of the prawn *Penaeus indicus* (Crustacea: Decapoda) from five populations in the western Indian Ocean off South Africa, Mozambique, Tanzania, Madagascar and Oman. Four molecular techniques were applied: allozyme electrophoresis, RAPD (Random amplification of polymorphic DNA), DNA sequencing of the autosomal locus Pi06 and the mtDNA locus Cytochrome oxidase I (COI) which expressed the degree of genetic diversity at different molecular levels, within these populations. A total of 25 loci were screened by allozyme electrophoresis. The LDH, MDH, LGG, MPI loci showed a low mean number of alleles (<2) and population structure (F_{st} = 0.024). RAPD using 20 different primers revealed a pattern with higher diversity and, although within population structuring was high (Analysis of Molecular Variance: 97 %), a significant genetic differentiation among Oman, Tanzania and South Africa samples was detected (p<0.05).

Genetic diversity was further tested by sequence polymorphism using the Pi06 locus and the COI mtDNA locus. The autosomal locus Pi06 was generated from a banding pattern of RAPD primer OPC-6. The 500 bp insert was then cloned into a bacterial host (*E. coli*: DH5α lacZΔM13 competent cell) and sequenced to ascertain PCR misincorporation artefact and allele variation. A specific parameter arbitrarily named private nucleotide index was calculated as a frequency of nucleotides belonging to only one population and then compared with nucleotide diversity of mtDNA COI sequences results. The level of polymorphism was unexpectedly higher in Pi06 than in COI, and the private nucleotide index suggested a separation of Tanzania and Madagascar samples from the Oman, Mozambique and South Africa. Contrary to other research (Baldwin *et al.*, 1998), a low level of polymorphism was reported not only at allozyme level but also in the mtDNA sequence data. COI sequences from *P. indicus* analysed by "analysis of molecular variance" showed an F_{st} of 0.082 with no significant separation among the populations sampled. These results were then compared with 26 different samples of *Penaeus merguensis*, collected from the east and west coast of Thailand peninsula. The hypothesis tested, assuming selective neutrality for COI, was that no genetic subdivision was expected within ecologically similar species unless different effective population sizes and/or historical events affected these species. Sequences of the COI locus in *P. merguensis* revealed a much higher polymorphism and a significant population structuring

(Analysis of molecular variance, $F_{st} = 0.27$, $p < 0.05$) within a shorter geographical distance (hundreds of km). Mismatch distribution analysis revealed also a different evolutionary history between the two species. A strict unimodal distribution characterised East African samples of *P. indicus* whereas several peaks were recorded in the Thailand populations of *P. merguensis*. It is suggested that the low genetic variation among populations of *P. indicus* along the East coast of Africa could have been caused by a strong bottleneck during the last Pleistocene glaciations (12 000-10 000 years before present). If the COI locus is selectively neutral, the difference in the degree of genetic polymorphism between Thailand and East Africa is consistent with a westward migration route of decreasing genetic diversity which has its centre of origin in the East Indies triangle (IWP) as proposed by Briggs (1999), or by a subsequent migration towards the Indonesian peninsula from a refuge in the Pacific ocean (Benzie *et al*, 2002). Although high-resolution molecular techniques were involved, shallow subdivision among wide geographically separated samples was observed. In the light of these results calls for management caution, when issuing harvest quotes in trawling fisheries, cannot be ruled out. Furthermore the low level of genetic diversity of *P. indicus*, in the Western Indian Ocean also needs to be taken into consideration in relation to programmes of prawn aquaculture along the East Coast of Africa.

LIST OF CONTENTS

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER ONE

INTRODUCTION

LIST OF FIGURES

LIST OF TABLES

CHAPTR TWO
ALLOZYMES

LIST OF FIGURES

LIST OF TABLES

CHAPTER THREE
MOLECULAR ANALYSIS (RAPD)

LIST OF FIGURES

LIST OF TABLES

CHAPTER FOUR
SEQUENCING OF NUCLEAR AND mtDNA (PI06 AND COI)

LIST OF FIGURES

LIST OF TABLES

CHAPTER FIVE

SYNTHESIS

Chapter one

Introduction

This study presents a biochemical and molecular survey of the widespread species *Penaeus indicus* Milne Edwards, 1837 (Crustacea: Decapoda) collected from different geographical sites along the western border of the Indo-West Pacific geographical region (IWP). This investigation is the first attempt to characterise the genetic structure of *Penaeus indicus* along the East Coast of Africa. The main objectives of this thesis were to collect genetic information for the local identification of this species as well as to close the informational gap, both, genetic as well as phylo-geographic, in the data collected from the East and the West IWP regions (see folded map: Appendix).

1.1 Geographic distribution of species

The Indo-West Pacific (IWP) geographical regions enclose an enormous area extending longitudinally more than halfway around the world and encompassing more than 60° of latitude. The biota is extremely diverse, incorporating habitats such as coral reef, mangrove forests, seagrass beds, wetlands, open ocean, and deep-sea up-welling systems (UNEP, 1985a). This extensive and diverse area includes approximately 6 570 000 km^2 of shelf habitat (less than 200m in depth) and hosts more than 6000 species of molluscs, 3000 species of Crustacea (873 belonging to the Penaeidae), 800 species of echinoderms, 500 species of hermatypic corals and 4000 species of fishes (Briggs, 1995). One of the most interesting features of the IWP is that, despite a basic homogeneity caused by the occurrence of many wide ranging species, there are great differences in species diversity amongst the various part of the region.

The majority of the tropical marine families have their greatest concentration of species within a comparatively small triangle formed by the Philippines, the Malay Peninsula and New Guinea (Fig 1.1). As one moves towards the periphery of this hypothetical triangle, there is a notable decrease in species diversity that appears to be correlated with distance (Briggs, 1999). This centre of high diversity in the East Indies owes its origin to the Tethys Sea (Briggs, 1999), the ocean that was situated between the northern and southern continents

during the early Tertiary period. At first the area of greatest species diversity, estimated from fossil mollusc

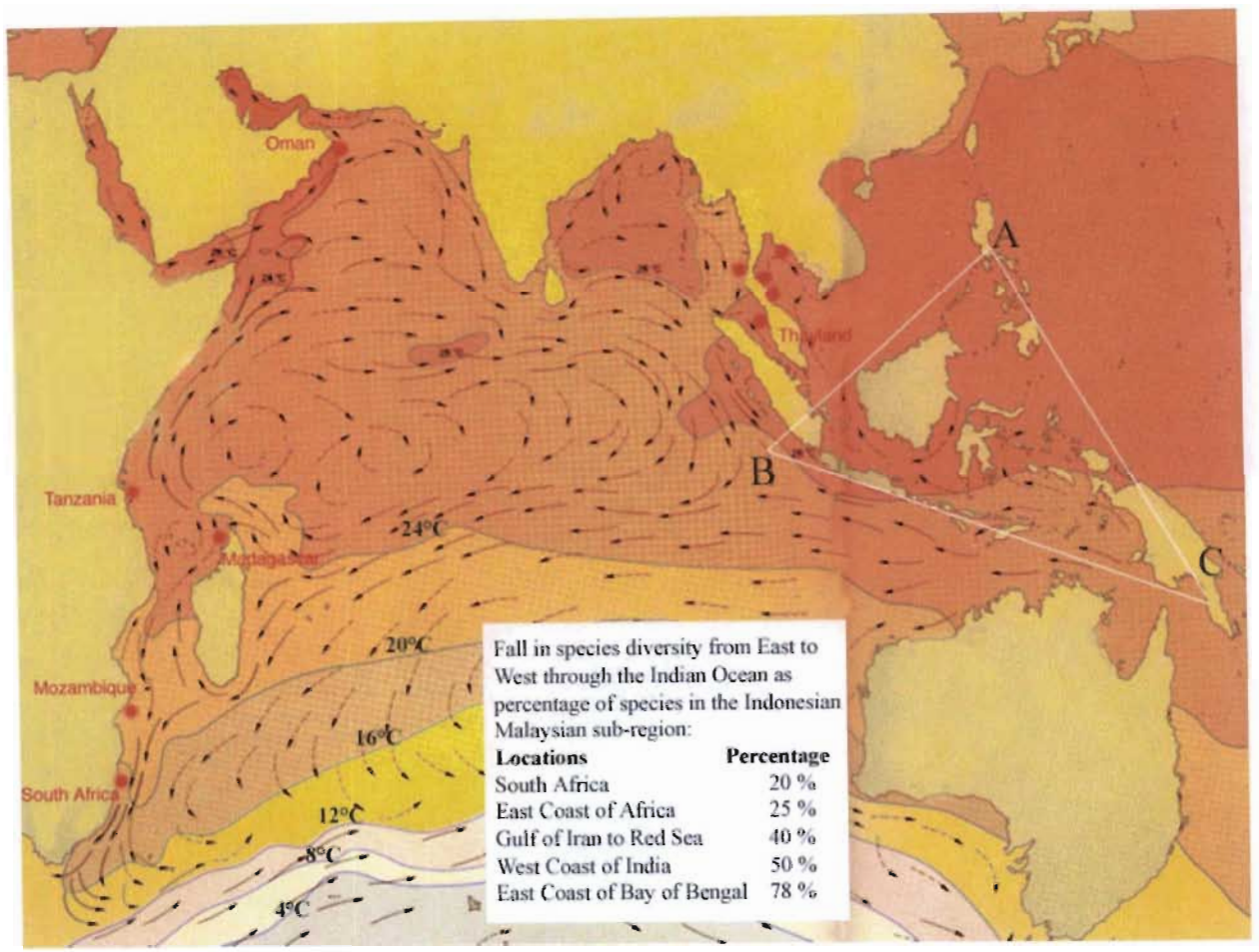


Figure 1.1 – Current patterns in the Indian Ocean (modified from Atlante Zanichelli) during the winter Monsoon season. Triangle **ABC** delimits the high species diversity described by Briggs (1995). Red dots indicate the sampling location for *P. indicus* in Africa and *P. merguensis* in Thailand.

assemblages appeared to have extended from Europe to North Africa and to India (Piccoli *et al.*, 1987). Fossil material from Java indicates the building up of assemblages in the East Indies, beginning in the middle Eocene epoch, and that this expansion of species diversity was completed by the end of the Miocene (Briggs, 1999). Briggs stressed the possibility that the major events responsible for the eastward or westward migration from this diversity centre were the decline in global temperature, starting from the middle Eocene, and the collision of Africa and Eurasia that caused the elimination of the Tethys Sea and established the Mediterranean sea. The time elapsed since these major evolutionary events, approximately 10

million years (MY), allowed the establishment of certain biogeographic patterns of species distribution within the Indo West Pacific region that appear to be of evolutionary importance. As illustrated in figure 1.1, there is a decrease in species diversity that is negatively correlated with distance from the Indo-West Pacific region (IWP). In general, the decrease in species diversity from the IWP region to the East African coasts, is more pronounced than that found in other directions (Dall *et al.*, 1990). The general pattern of diversity in the IWP may be illustrated by reference to the species distribution within the damselfish (Pomacentridae), the molluscan family (Strombidae), and the decapods family (Penaeidae). In all these families, the patterns of distribution are remarkably similar despite the different life cycles of these organisms (Allen, 1975; Abbott, 1960; Dall *et al.*, 1990; Veron, 1995). Whatley (1987) argues that several genera from the East Indies triangle were mostly Cenozoic, but those from Polynesia originated in the late Jurassic to upper Cretaceous; ostracod species have been dispersing outward from the East Indian / South Pacific region since the Miocene. These authors identified westwards dispersal flows, through the Indian Ocean to East Africa, by studying the history and biogeography of different species. Data supporting this hypothesis come from recent works on the genetic structure of widespread Indo West Pacific species (Benzie and Williams, 1997; Palumbi, 1997; Baldwin *et. al.*, 1998; Amornrat *et al.*, 1999). These authors provide evidence for the existence of routes of genetic exchange and decreasing gradients of genetic diversity that parallel the drop of species numbers across the Pacific. The different genetic markers used in these investigations, such as mitochondrial DNA (mtDNA) and allozymes, indicated the general area of the East Indies triangle as a place of origin for inter-oceanic radiation of species (Bowen and Grant, 1997; Nichida and Lucas, 1988). Dominant species that have been able to spread over large geographic areas probably arose in such places. However these large-scale dispersal events may partially explain the geographical species partitioning seen in the IWP. Small, isolated populations are capable of rapid speciation; they possess limited genetic variation and are possibly subject to genetic drift and inbreeding depression that may impair their successful distribution (Avise, 1994). In addition, there is empirical evidence that, species formed in peripheral locations have difficulties in expanding their ranges against a gradient of increasing diversity. For example, the IWP is a highly diverse bio-geographic region compared with East Africa or the Eastern Pacific (Leis, 1984; Emerson, 1991). Bio-geographical barriers, such as deep water and ocean currents, between these areas are formidable hurdles for numerous species of shallow water invertebrates and fishes. Nonetheless, a variety of IWP species have managed to cross these natural barriers and become established in the eastern or western part of the IWP triangle. Conversely there is almost no evidence of successful migrations in the opposite direction,

despite the fact that, in certain cases, ocean currents have been shown to carry larval stages to the outermost areas of a distribution range (Scheltema, 1988). Examples of this process can be found around the Cape of Good Hope, and in the Mediterranean sea, where a few tropical species have managed to migrate, but no reverse dispersal has yet been recorded (Briggs, 1995). These cases support the hypothesis that it is extremely difficult for marine species originating from areas of low diversity to colonise areas of greater species diversity. This issue, however, is controversial and the debate is still far from settled.

In spite of the above-mentioned debate, the East Indies triangle is a species rich environment, and several biotic and evolutionary factors could drive these species distribution phenomena (Ekman, 1953; Ladd, 1960). For example, extinction patterns may follow a predictable sequence of events: the evolution of an ancestral species in the IWP, the subsequent dispersal over large geographic areas, the initiation of extinction in the centre of origin and the final replacement of the evolved species following the onset of extinction. The widespread geographic patterns observed today would be created by repetitions of the replacement and extinction processes over millions of years (Palumbi, 1997). Ultimately, the Cenozoic world-wide rise and fall of sea levels is thought to have been responsible for the making and breaking of numerous barriers that evidently promoted allopatric speciation within the IWP region (Springer and Williams, 1990; Paulay, 1997).

The distribution of any marine species is strictly related to multiple ecological as well as physical factors in the environment around them (Briggs, 1984). Therefore, the studies cited above are indicative of barriers to various types of *Penaeidae* distributions that are not always easy to predict.

1.2 Ecological Factors Influencing the Distribution of *Penaeus* spp.

Dall (1990) described the Penaeidae as conforming to the general malacostracan morphology plan (Fig 2.1). They are laterally compressed, elongate decapods, with a well-developed abdomen adapted for swimming. In the Penaeidae the head and thorax are fused into the cephalothorax, which is completely covered by the carapace. The rostrum is always prominent, with a high median blade bearing dorsal teeth (7-8 in *P. indicus*) as well as ventral teeth in some genera. The thorax has three pairs of maxillipeds and five pairs of pereiopods; the first three pairs are used for feeding and the last two for walking. *Penaeus indicus* is normally taxonomically diagnosed by the apical tuft of setae on the propodus of adult male 3rd

maxilliped which is about equal in length to the dactyl; thelycal flaps corrugated medially; rostrum of adult sinuous, blade low (Dall *et al.*, 1990).

1.2.1 Life Cycle

In order to elucidate the main factors that can contribute to the dispersion and consequent distribution of these marine organisms, one should analyse in detail the different life cycles that characterise the genus being considered. It is still not clear which of the many extant penaeid life cycles most resemble the common one of the ancestral *Penaeus*. Much of the fossil record shows the presence of Penaeidae in marine shales (Glaessner, 1969). This may suggest that they lived in muddy substrata in inshore waters, as most extant species do and, therefore, it is reasonable to infer that the life history of modern forms of *Penaeus* are similar to those of the ancestral ones. For all known members of the family, the sequences of the larval stages are similar: they have a planktonic larva, with several naupliar, protozoea, mysis and postlarval stages followed by juvenile and adult stages. The different life cycles present in this species can be identified based on the particular habitat chosen by the postlarval stage. In general, four types are identified (Fig1.2): entirely estuarine, estuarine-inshore-offshore, inshore-offshore and predominantly off-shore (Kutkuhn, 1966a).

Entirely estuarine life cycles appear to be restricted to the smaller species of *Metapenaeus* particularly *M. bennetae*, *M. conjunctus*, *M. elegans*, *M. moyebi*, and sometimes *M. brevicornis* (Miquel, 1982). The postlarvae tend to migrate upstream into water of lower salinity and, as they grow, the juveniles move progressively towards areas of higher salinity in the lower estuary.

More usually, spawning takes place offshore, the depth varying with the species. Planktonic stages migrate inshore towards the end of larval development, when the postlarvae settle on their preferred nursery grounds. Preference for a particular substratum or vegetation type appears to be the dominating factor that governs where postlarvae will settle, although salinity preferences have also been claimed to be important (Staples and Vance, 1985). The postlarval preferences for inshore habitats determine the characteristic of the cycle. For example, in the estuarine-inshore-offshore type of life cycles, *Penaeus* postlarvae settle in mangrove-lined, muddy estuaries and may ascend rivers for as far as 85 kilometres, where salinity is very low (Staples, 1980a). The juveniles of these species are mostly euryhaline and only when they reach about half adult length do they leave the estuary for offshore waters. *Penaeus artemisia,*

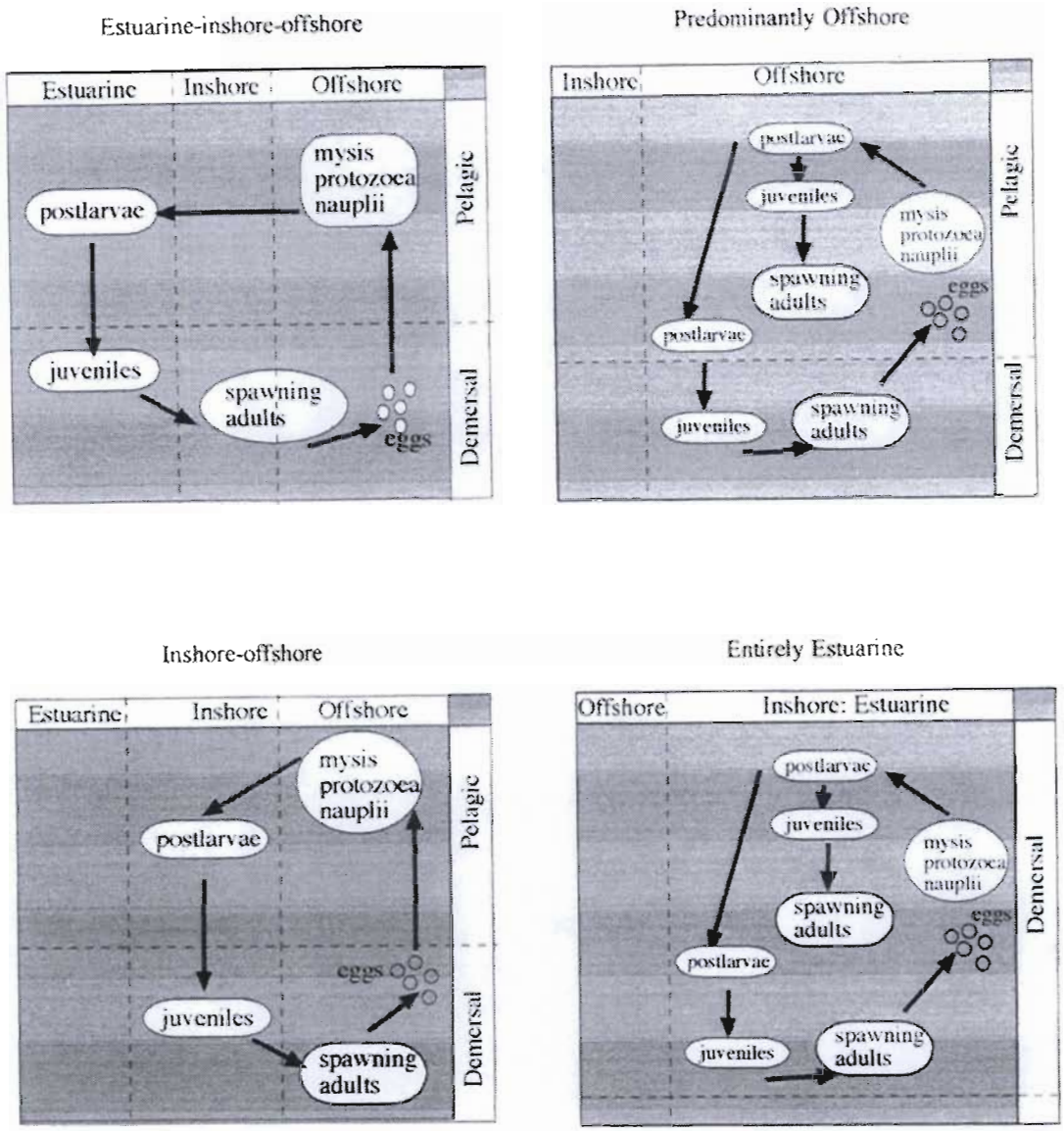


Figure 1.2 – Types of penaeid life cycles. See text for explanation. Modified from Dall *et al.* (1990)

P. atypopenaeus, P. macropetesma, P. trachypenaeus, P. metapenaeopsis and *P. parapenaeopsis*, prefer seagrass beds or sandy mud as nursery grounds (Hall, 1962), to which they migrate during the postlarvae stage. In all *Penaeus* species the pre-adults move offshore to pelagic waters. A completely pelagic life cycle is present in some species such as *P. pelagopenaeus, P. parapenaeus, P. penaeopsis* where (Kensley *et al.*, 1987) both larval stage and juvenile live in off shore waters. Within this context, it is noteworthy that the genus

Parapenaeopsis possesses the largest recorded eggs of all Penaeidae, suggested to be an adaptation to fully pelagic waters (Tom *et al.*, 1988).

Given this wide variability in life cycles and habitat, it is not surprising that the penaeid prawns exhibit a rather complex seasonal life-history pattern (Garcia, 1985). Although one can recognise different life-history stages within the genus *Penaeus* that depend on local environmental as well as ecological conditions, it has been found that temperature and rainfall are the major factors contributing to the length of the cycle (Garcia, 1988). Close to the equator, most penaeids appear to spawn throughout the year. Their reproductive cycles are affected by rainfall regimes, often associated with seasonal monsoon winds or temperature changes brought about by shifts in winds, especially in regions adjacent to continental land masses (Dall *et al.*, 1990). However two main yearly spawning periods have been observed in a number of tropical *Penaeus* and *Metapenaeus* species, including several populations close to the equator (Hall, 1962; Staples and Rothlisberg, 1980). Typically, these occur from September to November, and between March and May. These periods are characterised by decreased wind speed and currents during the tropical inter-monsoon season or, at slightly higher latitudes, by rising or falling temperatures. At higher latitudes (Tropical and Subtropical) bimodal spawning and recruitment are also common but, seasonal rainfall and lower winter temperatures often result in one or other of the generations being dominant in the offshore phase (Garcia, 1985). For example, in some species such as *Penaeus notialis* and *Penaeus indicus*, spring spawning is generally greater and more consistent than the autumn spawning (Garcia, 1977). Some authors suggest that greater springtime availability of phytoplankton and higher temperature of shallow waters in the inshore nursery grounds would help the degree of survival of larvae and juveniles. Additionally, Garcia (1977) noticed that the seasonality of offshore catch per unit effort (CPUE) tended to be bimodal in tropical waters, becoming unimodal at higher latitudes both to the south and the north. In all the countries examined, the period of higher CPUE appeared to coincide with the main period of rainfall. In summer the possible barriers to dispersion of Penaeidae populations can include the following aspects.

Temperature; the Penaeidae are, predominantly tropical stenotherm. This means that they can withstand only a limited range of temperature, with few species thriving below a minimum of 15°C. Thus, cold inshore and offshore upwellings, such as along the Cape of Good Hope coast and the Somali coast, could be considered barriers to the dispersion of larval stages.

Ocean currents; the pelagic larval stages (Fig 1.2) make most Penaeidae species susceptible to the influence of current flows moving in unfavourable directions.
Coastal environments such as a desert with high inshore salinity, or rocky shores with deep water inshore may hinder the dispersal of certain species particularly during the post-larval and juvenile stages.

1.3 Phylogeny and Biogeography of the Marine Prawn *Penaeus*

In light of the above discussion on the general pattern of species distribution in the IWP region, I will now introduce the possible processes that may have contributed to the distribution and evolution of the Penaeidae species. Prawns of the genus *Penaeus* constitute a diverse and abundant group of benthic taxa found in the tropical and subtropical waters around the world.

Decapod fossils from the early Tertiary have been found in Antarctica (Feldman and Zinsmeister, 1984a,b). These findings, together with temperature data values reconstructed from geological studies, indicate that a penaeid fauna might have existed in these waters, at least in the Cretaceous. However, geologically more recent fossil finds suggest that such a fauna could not have persisted (Zinsmeister, 1982) because as Australasia moved northward, there were extreme environmental changes (Shackleton and Kennett, 1975). The temperature fell to less than 10°C as the ocean became continuous around Antarctica about 40 MYBP, and continued to fall through the remainder of the Tertiary period. Later still, during the further northward shifting of Australasia, water temperatures rose again to over 20°C (30 MYBP) and the endemic fauna was replaced by warm water invading species from the Indo-Pacific. Thus it is very unlikely that any penaeid species of Australasia are relicts from Gondwana (Dall *et al.*, 1990).

The Penaeidae have been considered as the most primitive group of the decapoda because of the nauplius larval stage, morphological features and the estimated age of their fossil record. The earliest Penaeidae fossils were found in deposits from the Triassic period (Calman, 1909). Unequivocal *Penaeus sp.*, have also been found in Jurassic shales (Woods, 1925) and became more common in the Cretaceous, with a record from India dating to the lower Tertiary (Glaessner, 1969). Therefore, taking the fossils records into account, the genus *Penaeus* appears to be the oldest of the penaeidae. Unfortunately, no penaeids have yet been found in more recent deposits, so there is no paleontological indication as to when existing penaeids

may have arisen. Consequently, estimation of the times of divergences of the various taxa has to be made by other techniques, such as molecular ones (Dall *et al.*, 1990).

Molecular investigations using approaches such as mitochondrial DNA (mtDNA) and microsatellites (Benzie *et al.*, 2002) were able to confirm the hypothesis that the genus arose in the Indo- Pacific (Dall *et al.*, 1990). This hypothesis was originally formulated on the basis that biogeographic centres of origin have the highest species diversity and the deepest morphological differentiation (Briggs, 1995). This scenario is supported by results obtained from the analysis of the COI (Cytochrome Oxidase I) mtDNA locus in the genus *Penaeus* (Baldwin *et al.* 1998). Populations from the Indo-Pacific that have been sampled share a number of haplotype sequences unique to this area, although they show the highest mtDNA diversity of all Penaeids (Balwin *et al.*, 1998). The relationships among *Penaeus* species, as resolved by mtDNA studies, are similar to those reported for genera of tropical marine organisms, where a gradient of diversity is commonly related to lineages radiating eastward and/or westward from the IWP (Briggs, 1984; Bowen *et al.*, 1998).

The above patterns are generally explained by the widely invoked mechanisms of vicariance (Fig 1.3). Under vicariance interpretations, related populations, or taxa, became separated when more or less continuous ranges of ancestral forms were sundered by environmental events (Nelson and Rosen, 1981), such as the break-up of landmasses or the physical subdivision of a body of water that split populations of aquatic life-forms (Avise, 1986). Under dispersalist interpretations, a taxonomic assemblage came to occupy its present range through active or passive dispersal from one or more ancestral centres of origin (Briggs, 1984). The relative role of these two processes has been the subject of strong debate in recent decades, with vicariance prevailing in the 1970s and '80s (Avise, 2000), but with more evidence for dispersal accumulating since the introduction of molecular evolutionary analyses (Rosenblatt and Waples, 1986; Bowen and Grant, 1997) although Knowlton (1993) suggests a greater role for vicariance.

The relationships of eastern Pacific and western Atlantic forms bear the imprint of a vicariant separation. However, older vicariance events, such as the closure of the Tethys sea by the collision of Africa and Eurasia (approximately 20 MYBP) and opening of the Atlantic Ocean (about 10 MYBP), predate the deepest lineages within *Penaeus*. For the spread of *Penaeus* during Tertiary and Pleistocene periods, from the Western Pacific to the Eastern Pacific and from the Indian Ocean into the Eastern Atlantic Ocean, a dispersal mode has been strongly

indicated. Clearly, both vicariance and dispersal have a role in shaping the biogeographic history of *Penaeus* (Baldwin *et al.*, 1998). However more molecular data are necessary, in particular, more species need to be analysed and a less patchy population sampling is necessary in order to avoid overestimating the evolutionary distances of very closely related morphological species.

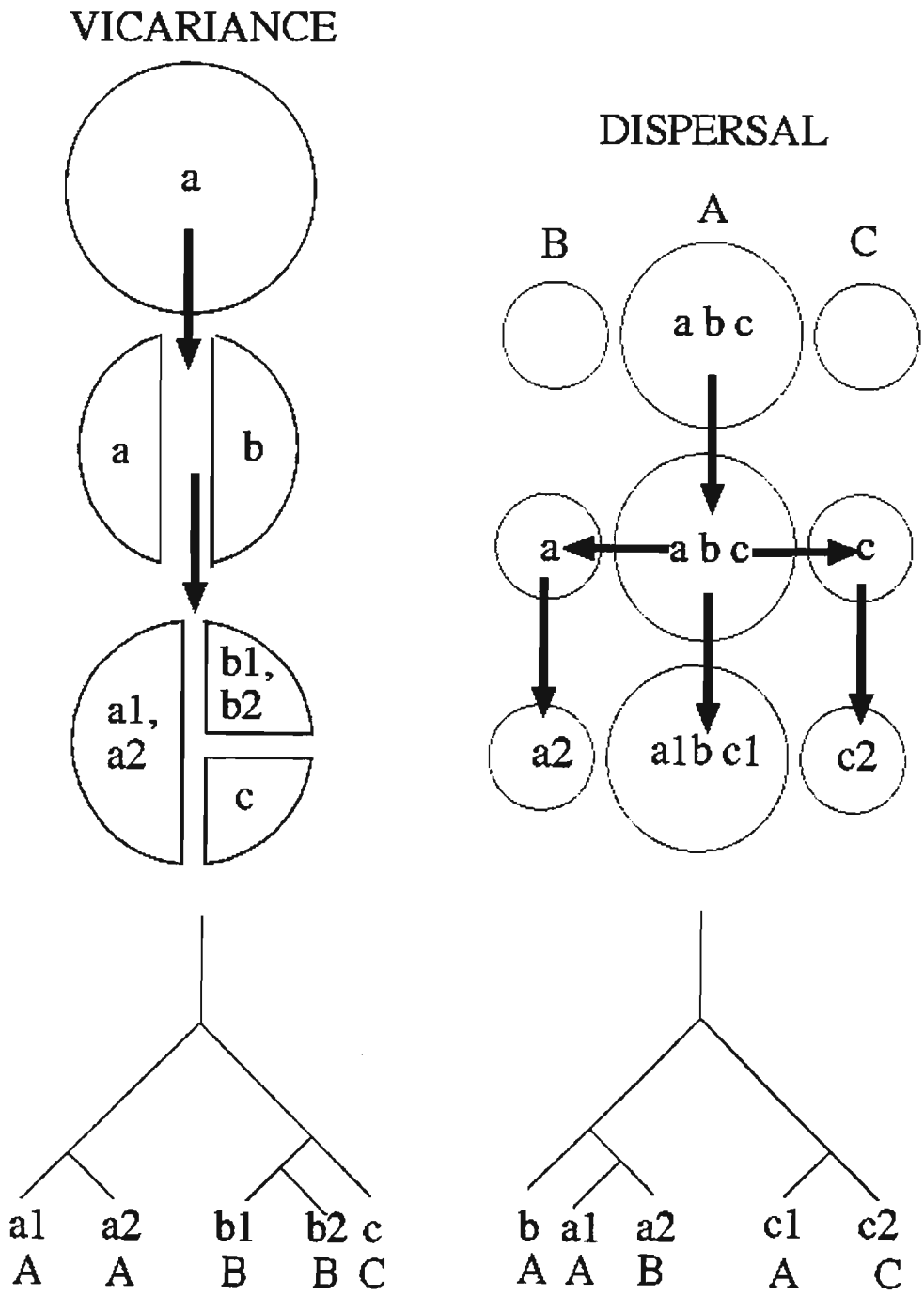


Figure 1.3 - Phylogenetic relationships of spatially disjunctive populations or species under vicariance and dispersal hypotheses. Lowercase letters represent taxa; uppercase letters, geographic areas. Modified from Avise (2000)

1.4 Economical importance of *Penaeus sp.*

The genus *Penaeus* is the largest of the family Penaeidae, and contains the greatest number of species of commercial importance (Table 1.1) to aquaculture and the fisheries industry. The total world catch of prawns in 1990 (FAO, 1990) was 2.1 million metric tons, 41 % of which was fished in the Indo West Pacific region, where penaeids are the most valuable of all marine species caught (FAO, 1990). According to this source 11 % of the world total volume was caught in the Western Indian Ocean, and two % of this volume was harvested from Tanzania, Kenya, Mozambique, Madagascar and South Africa. Along the East Coast of Africa the prawn fishery can be considered multispecies, with several species being harvested in different quantities throughout the year by local trawlers. A total of 107 species belong to the genus *Penaeus* (Table 1.1), 81 of which are found in the diverse habitats of the Indo West Pacific region. Although several of these are harvested in various regions along the coast of the Indo Pacific, not all are considered valuable resources. Commercially the most important species are *P. monodon* Fabricius 1798, *P. japonicus* Bate 1888, *P. merguiensis* De Man 1888, *Penaeus indicus*, *Metapenaeus monoceros*. Trawlers harvest all of these at different times of the year normally along bank formations outside river basins and at maximum depths of 90-100 m where these species spend their adult stage (Fig. 1.4).

Penaeid prawns are also an important part of the aquaculture industry in many areas of the world with increasing emphasis being placed on pond and raceway culture recently. Advances in laboratory controls of maturation have allowed the selection of populations for commercial production (Fast and Lester, 1992). These industries have developed rapidly over the past 30 years and are dominated by four species: *Penaeus monodon*, (60 % of world production); *Penaeus vannamei* Boone 1931, (27 %); *Penaeus stylirostris*, Stimpson, 1874 (three %); *Penaeus japonicus*, (less than eight %). Several other species account for the remaining two % of world production, prominent among which are the following species: *Penaeus indicus*; *Penaeus merguiensis* and *P. penicillatus*, Alcock, 1905 (Benzie, 2000). Of the 20 species whose commercial usefulness was investigated from 1990 to 2000, those mentioned here were considered the most profitable (Benzie, 2000).

Table 1.1 - Species of *Penaeus* of interest to fisheries in IWP. Range of depth and habitat preferences are also specified. Modified from FAO (1990)

Species name	*Depth range*	*Habitat/Life cycle*
Penaeus africanus (Balss, 1913)	28m	Marine – estuarine
Penaeus stenodactylus (Stimpson, 1860)	11 – 27m	Bottom mud
Penaeus barbatus (De Haan, 1844)	20- 70m	Marine bottom mud or rocky
Penaeus lamellatus (De Haan,1844)		
Penaeus novae-guineae (Haswell,1879)	11-30m	Marine
Penaeus philippii (Bate, 1881)	18-22m	Marine Bottom mud
Penaeus affinis (H.Milne Edwards, 1837)	69-73m	Marine
Penaeus brevicomis (H. Milne edwards,1837)	0-30m	Estuarine and Marine
Penaeus dobsoni (Miers, 1378)	0	Shallow brackish and salt water
Penaeus ensis (De Haan, 1844)	44-50m	Marine
Penaeus intermedius (Kishinouye,1900)	13-33m	Marine bottom sand
Penaeus jouyneri (Miers, 1880)	10-20m	Marine Bottom mud
Penaeus lysianassa (De Man, 1888)	5-9m	Marine
Penaeus macleayi (Haswell,1879)	22-37m	Estuarine
Penaeus monoceros (Fabricius, 1798)	170m	Marine
Penaeus fissurus (Bate, 1881)	50-274m	Marine
Penaeus rectacutus (Bate,1888)	180-750m	Marine
Penaeus chinensis (Osbeck,1765)	90-180m	Marine
Penaeus indicus (H. M. Edwards,1837)	2-90m	Estuarine
Penaeus merguiensis (De Man,1888)	10-45m	Estuarine
Penaeus penicillatus (Alcock,1905)	2-90m	Marine
Penaeus japonicus (Bate,1888)	1-90m	Marine
Penaeus canaliculatus (Olivier,1811)	33-46m	Marine
Penaeus latisulcatus (Kishinouye,1896)	0-80m	Marine bottom mud
Penaeus longistylus (Kubo,1943)	35-55m	Marine
Penaeus marginatus (Randall,1840)	0-300m	Estuarine
Penaeus plebejus (Hess,1865)	2-160m	Estuarine
Penaeus esculentus (Haswell,1379)	16-22m	Marine
Penaeus monodon (Fabricius, 1798)	0-110m	Estuarine
Penaeus semisulcatus (De Hann,1844)	2-130m	Estuarine
Penaeus anchoralis (Bate,1881)	12-52m	Marine

1.5 Genetic variability in *Penaeus sp.*

Most of the genetic studies on this species have been directed at determining stock structure for fisheries management purposes but, in addition, they have provided useful information on the genetic diversity available in nature and for the future planning of the source of brood stock for closed-cycle breeding programs (Hedgecock and Malecha, 1991).

Much of the early work using allozyme markers generally emphasised the lack of significant geographical structure in the wild populations of many *Penaeidae* species (Hedgecock, 1986; Hedgecock *et al.*, 1982). The information revealed by these studies is highly variable with respect to heterozygosity and sample size for both number of loci and number of individuals

sampled (Benzie, 2000). The number of individuals sampled for any given locus ranges from one to more than 200, but most fall within the average of 20-60 per population. Similarly, the number of loci sampled ranges from 1 to 40, but most studies used between 15 and 35. Most of the investigations appear to have been concentrated in Australia, followed by Hong Kong, the Gulf of Mexico, the east coast of the USA and Thailand. Other regions sampled, usually only one region for any given species, have been Japan, south-east Asia, and south-east Africa. However, in south-east Africa, sampling has been patchy, and probably not consistent with the real distribution range of the species under investigation (Table 1.2).

Table 1.2 - Published biochemical and molecular studies of the genera *Metapenaeus* and *Penaeus*. The species surveyed in the present research, *Penaeus indicus*, is highlighted. Modified from Benzie (2000).

Species	Geographical extent of study	Type of marker	Nº of Populations Sampled
M. affinis	*Hong Kong*	*Allozymes*	*1*
M. bennettae	East Australia	Allozymes	1-6
M. eboracensis	North Australia	Allozymes	1
M. endeavouri	North and West Australia	Allozymes	2-6
M. ensis	North and East Australia, Hong Kong, Japan	Allozymes/RAPDs	1-3
M. insolitus	North Australia	Allozymes	1
M. joyneri	Hong Kong	Allozymes	1
M. macleayi	East Australia	Allozymes	3- 11
P. aztecus	Gulf of Mexico South Carolina	Allozymes/Microsatellites	1
P. brasiliensis	Gulf of Mexico	Allozymes	1
P. chinensis	Hong Kong	Allozymes	1
P. duorarum	Florida Gulf of Mexico	Allozymes/Microsatellites	1-4
P. esculentus	North East and West Australia	Allozymes	2
P. indicus	**Kenya**	**Allozymes**	**1**
P. japonicus	Japan Hong Kong	Allozymes/RAPDs	1-4
P. kerathurus	Adriatic France Spain Italy Tunisia	Allozymes	1-6
P. latisulcatus	North West and South Australia	Allozymes	1-8
P. longistylus	North and East Australia	Allozymes	2
P. merguensis	North and East Australia Hong Kong	Allozymes	1-2
P. monodon	South Africa Thailand Hong Kong North West East Australia	Allozymes/Microsatellites/ mtDNA/RAPDs	1-13
P. notialis	Cuba	Allozymes	1
P. penicillatus	Hong Kong	Allozymes	1
P. plebejus	East Australia	Allozymes	4-11
P. semisulcatus	North East Australia Japan Hong Kong	Allozymes	1-2
P. setiferus	Texas Louisiana Gulf of Mexico	Allozymes/Microsatellites	1-4
P. stylirostiris	Mexico Equador	Allozymes	3
P. vannamei	Mexico Panama Ecuador	Allozymes	2-3

Early work noted that populations separated by thousand of kilometres showed no significant differences in allele frequencies (Lester, 1979; Mulley and Latter, 1980; Forbes and Demetriades, 1999; Benzie *et al.*, 1992). These data all support the view that prawn populations are panmictic and this conclusion is reinforced by tagging data from several species (Dall *et al.*, 1990), which show considerable movement by individual prawns, of hundreds of kilometres over a few months. The possibility has been considered that this lack of spatial differentiation among wild populations was the result of the low resolution of the technique itself (Benzie, 2000). This prompted the utilisation of potentially more variable markers, that are more sensitive to recent evolutionary time-frames, such as, mtDNA, microsatellites, as well as RAPD (Random Amplification of Polymorphic DNA) and direct sequencing of specific target loci both nuclear as well as mitochondrial.

Thus far, the available literature includes, three reports using mtDNA (Benzie, 1993; Benzie, 2000; Klinbunga *et al.*, 1998), three reports using RAPD (Meruane *et al.*, 1997, 1998; Tassanakajon *et al.*, 1997b), and six using microsatellites (Ball *et al.*, 1998; Tassanakajon *et al.*, 1998a, b; Broeker *et al.*, 1999). Microsatellites show a far higher proportion of polymorphic loci than that from some reported allozyme studies, although it will be some time before realistic estimates are available, given the difficulty of isolating loci from prawns for which reliable primers can be designed. Similarly, RAPD data show levels of polymorphism that are higher than those of allozymes, and available mtDNA data, confirm a general pattern of higher polymorphism than that detected by allozymes (Tassanakajon *et al.*, 1998a; Broker *et al.*, 1999). For example, allozyme, microsatellite and mtDNA markers generally demonstrate significant genetic differences between Andaman Sea and Gulf of Thailand populations of *P. monodon* (Tassanakajon, *et al.*, 1998b). Similarly, western Australian populations of *P. monodon* were significantly different from those on the east coast of Australia (Benzie *et al.*, 2000; Brooker *et al.*, 1999).

It appears that in *Penaeus*, high DNA sequence divergence between species may be masked at the level of amino acid sequence. These findings parallel the striking morphological and ecological similarities observed among divergent members of this genus (Palumbi and Benzie, 1991). Palumbi and Benzie (1991) proposed two possible explanations for the apparent differences in molecular and morphological evolution: the rate of DNA evolution, particularly mitochondrial DNA, might be accelerated in the shrimp, or the rate of morphological divergence might be slow, possibly due to stabilising selection on morphological characters. Furthermore, the conclusions regarding the genetic structuring of

Penaeus spp., are based on surveys of few species, mostly those of major economic interest such as *P. monodon* (Ko *et al.*, 1983; Sodsuk *et al.*, 1992; Benzie *et al.*, 1992; Forbes *et al.*, 1999; Benzie, 2000) and *P. japonicus* (Meruane *et al.*, 1997; Tam and Chu, 1993; Taniguchi and Han, 1989), and lack of information is evident in respect of other species within the IWP regions. Thus the controversy regarding the evolution and population structure of this widespread genus is far from being resolved. It is necessary to keep collecting data from wild populations from different and widely separated locations. The use of different markers in order to resolve the rate of divergence at which morphological and molecular characteristic evolve in this genus, is also highly recommended.

1.6 Neutral Theory and Molecular methods

Studies that use molecular markers to address questions in ecology and conservation biology often assume a strictly neutral model of molecular evolution as the basis for analysing and interpreting the results. It is therefore worthwhile to review briefly the status of the neutral theory after nearly 20 years of studies of DNA sequence variation among and within species (Ford, 2002; Kreitman and Akashi, 1995; Hughes, 1999).

The strictly neutral theory proposes that the vast majority of new mutations fall into one of two categories: deleterious and selectively neutral (Kimura, 1983). Deleterious mutations are expected to be eliminated rapidly due to natural selection against them, and therefore presumably contribute little to variation among and within species. On the other hand, mutations that are selectively equivalent to the allele(s) already present in the population, are expected to have dynamics governed by genetic drift, and make up the vast majority of the observed variation both, within and among samples. Beneficial mutations are expected to be extremely rare and to contribute little to observed patterns of DNA sequence variation, although they must be involved in evolutionary success. Overall, patterns of DNA sequence variation generally support one key aspect of the neutral theory, that many of the observed differences within and between populations are non-adaptive (Kimura, 1983; Hughes, 1999). For example, there is a broad negative correlation between the functional importance of a nucleotide site and its substitution rate. Additionally, a vast part of the genomes of many organisms contain non-coding DNA that serves no known purpose, and most mutations in these areas are presumably neutral (Ford, 2002).

A second key aspect of the neutral theory, that most variation has dynamics governed predominantly by genetic drift, is not supported by observed patterns of DNA variation (Begun and Aquadro, 1992; Kraft *et al.*, 1998). Instead correlation has been found between recombination and variation, and this fact has important implications for the neutral theory. It means that a gene's recombinational environment is a large contributing factor in determining patterns of variation within species. Therefore even if much of the observed variation is selectively neutral, the dynamics of this neutral variation could well be governed more by linkage between loci, than by genetic drift (Ford, 2002).

In recent years, largely due to advances in DNA sequencing techniques and to statistical methods (Kreitman and Akashi, 1995; Yang and Bielawski, 2000), the proportion of genes known to be subjected to positive selection has increased (Chalesworth and Mc Vean, 2001; Fay *et al.*, 2001; Wang *et al.*, 1999). For example, genes that code for enzymes involved in energy metabolism form a group where statistical evidence for positive selection has been found (Eanes, 1999). The previously cited authors warned that routinely assuming selective neutrality for allozyme and mtDNA variation was problematic, due to considerable evidence for natural selection effects on these types of genetic markers. When conducting quantitative analyses such as estimating divergence times, rates of gene flow or effective population sizes, it is particularly worth avoiding assumptions of strict neutrality, when in fact, variation is affected by selection.

In summary the neutralist *vs.* selectionist debate of the 1970s and 1980s has died away (Hey, 1999); it has been replaced by a more complicated view in which selection appears to leave deep footprints in the genome. In this case the strictly neutral theory, rather than being a simple explanation for patterns of genetic diversity, has instead become the primary null hypothesis used to test for the effects of natural selection (Ford, 2002).

1.6.1 Coalescent Theory

The stochastic process known as coalescence has played a central role in population genetics since the early 1980, and results based on coalescent assumptions are now routinely used to analyse DNA sequence polymorphism data (Nordborg, 2001). The coalescent theory provides a relatively simple and powerful tool for exploratory data analysis through the generation of simulated data. Comparison of observed data with data simulated under

various assumptions could give considerable insight. The coalescent was described by Kingman (1982a, b, c) but discovered independently also by Hudson (1983) and Tajima (1983). The basic assumption of the model is that since selectively neutral variants by definition do not affect reproductive success, it should be possible to separate the neutral mutation process from the genealogical process. As a consequence the evolutionary dynamics of neutral allelic variants can be modelled through so-called gene dropping (mutation dropping). Allelic states are assigned to the original generation in a suitable manner, and the lines of descent then simply follow forward in time. In particular the allelic states of any group of individuals can be generated by assigning an allelic state to their most recent common ancestor (MRCA) and then dropping mutations along the branches of the genealogical tree that leads to them. Most of the genealogical history of the population is then irrelevant. Basically it is possible to model the genealogy of a group of individuals backward in time without worrying about the rest of the population; in fact each individuals (assuming selectively neutrality) in a generation can be viewed as picking its parent at random from the previous generation. The realisation that a pattern of neutral variation observed in a population can be viewed as results of random mutations on a random tree is a powerful one, which profoundly affects the way we think about data (Stephens, 2001). In this general framework it is then possible to insert a different set of biological, as well as evolutionary phenomena, which can than be treated as a simple linear change in the time scale of the coalescent. The bad news is that biological and evolutionary phenomena will never be amenable to inference based on polymorphism data alone. For example, molecular ecologists in general are interested in how populations are genetically subdivided, or vary in size, and these processes cannot generally be modelled as a linear change in the time scale of the coalescent approach (Stephens, 2001).

1.6.2 Analysis of Population Subdivision: inference methods

Species, or populations, usually do not constitute a single panmictic unit where individuals breed at random over the all species range (Excoffier, 2001). In the marine realm especially, broadcast spawners such as the penaeid species, are astoundingly fecund. As individuals, they face exceptional challenge in matching reproductive activity to extremely variable oceanographic conditions. Reproductive success is dependent on a combination of gamete production, fertilisation, larval development, larval settlement and recruitment into the adult population (Hedgecock *et al.*, 1982; Li and Hedgecock, 1998). Large variances in the viability

of such an organism, following these life stages are also expected on a generation basis and may be recorded as two types of molecular observations:
(*i*) chaotic patchiness, defined as stochastic variation in genotypic frequencies over small spatial scales; (*ii*) low levels of genetic diversity, implying recent coalescent times in gene genealogies relative to what might have been suggested from the often very large populations typical of these marine species.

These biologically driven features, if not properly identified, can affect interpretation of the observed pattern of genetic diversity. For example, over 80 years ago Wahlund (1928) showed that hidden subdivisions lead to an apparent excess of homozygotes. At a molecular level, especially, hidden population subdivision can greatly increase the number of segregating sites within samples, while local inbreeding could result in the opposite effect (Tajima, 1989b). In a population made up of demes interconnected by migrations, the mean number of pairwise differences (or the number of heterozygous sites within individuals) only depends on the sums of deme sizes, irrespective of the migration pattern (Slatkin, 1987). In statistical terms population structure introduces some correlation or covariance between genes taken from different subdivision levels. Therefore, population genetic studies attempt to track, measure and test the presence of internal subdivisions, on the basis of observed correlations in the samples. In order to do so, inference moment methods such as those devised by Wright (1969), called F-statistics, which partition heterozygote deficiency into within and between population components, have been developed. This permits an assessment of the levels of structuring in samples of natural populations. The fixation index, one of the F-statistics parameters, includes a measure of heterozygote deficit within populations (F_{IS}), whereas a measure of these parameters among populations (F_{ST} and F_{IT}) expresses the global deficiency of heterozygotes.

By far the most common inference methodology considers only squares of allele frequencies or equivalent frequencies of identical pairs of genes (Analysis of Variance: ANOVA). The analysis of genetic subdivision under the analysis of variance framework is based on Cockerham's fixation index (1969; 1973) that showed how it is possible to decompose the total variance of gene frequencies into variance components associated with different subdivision levels within the ANOVA of gene frequencies. Cockerham also showed that these variance components and F-statistics were a different, but equivalent, parameterisation of correlations of genes. The model that is assumed is the existence of several demes of finite population sizes that have diverged simultaneously from an ancestral

population that was in Hardy-Weinberg equilibrium. Since the time of divergence, the demes have remained separate and have been exposed to the same conditions. Samples from different demes are thus expected to differ from each other because of the sampling process of individuals within demes (statistical sampling) and because of the stochasticity in the evolutionary process between populations (genetic sampling). The major interest is in defining the relations (correlation, covariance) between genes found in the different levels of the hierarchy. Finally the procedure estimates the correlation of gene frequencies at different levels of subdivision via the method of moments by equating the mean squares expressed in terms of allele and genotype frequencies to their expectations in terms of correlation parameters. It should be noted, however, that the inferences stemming from this methodology do not represent real biological patterns encountered in natural population.

Another inference methodology, likelihood analysis, often used in genetics has been only recently applied to genetically structured populations (Beerli and Felsentein, 1999). While the F_{st} approach is conceptually straightforward, it does not claim efficiency, in contrast to likelihood methods (Excoffier, 2001). Several methods have been developed to compute the likelihood for various models of population structure and different types of data with different mutation models. These methods are based on 'coalescent arguments (see section 1.6), that is to say, they derive the probability of the sample from consideration of the sequence of events that relates the individuals in the sample to their common ancestor (MRCA) (Hudson, 1990; Nordborg, 2001; Stephens, 2001). These methods do not attempt to find an explicit expression for the likelihood, or to generate samples for different values of the parameters. Rather, they are based on simulation and involve importance-sampling algorithms.

Finally, an empirical approach to infer the general demographic history of a population from gene tree data involves the examination of two types of different measures of haplotype variation (Grant and Bowen, 1998). Haplotype diversity[1] (h) condenses information on the numbers of frequencies of different alleles at a locus, regardless of their sequence relationships. Nucleotide diversity[2] (π) is a weighted sequence divergence between individuals in a population, regardless of the number of different haplotypes. Intuitively, a population with low π and low h may have experienced a prolonged or severe demographic

[1] diversity ($h = 1 - \sum f_i^2$, where f_i is the frequency of the i haplotype)

[2] ($\pi = \sum f_i f_j p_{ij}$, where p_{ij} is the sequence divergence between the i^{th} and j^{th} haplotype)

bottleneck in recent time. Conversely, high values of π and h are expected signatures from stable populations with large long-term effective population size (N_e); or they also might be observed in admixed samples of individuals from historically sundered populations. High h and low π suggest rapid population growth from an ancestral population with small N_e, provided that the time was sufficient for recovery of haplotype variation via mutation yet too short for accumulation of large sequence differences. Conversely, low π and high h could result from a transient bottleneck in a large ancestral population because an extremely short crash can eliminate many haplotypes without necessarily impacting π severely (Nei *et al.*, 1975). Low π and high h also might reflect an admixture of samples from small, geographically subdivided populations (Bowen and Grant, 1997).

In the present study, the analysis of molecular variance (AMOVA) framework has been applied to DNA molecular data, while the Wright F-statistic approach, was use for inference on population structure for the allozyme data. As described in the previous paragraph, comparisons of different genetic diversity measures were performed to determine empirically the possible causes of population differentiation.

1.7 Analytical Model

Following on the analytical framework briefly outlined above (chapter 1.6); it is possible to implement the following experimental steps while attempting to explain the genetic variation in natural populations of *Penaeus indicus*.

Observation: *Penaeus indicus* is a widely dispersed species in the IWP ocean. Its life cycle is type two (Fig.1.3): marine and estuarine with high potential for dispersal. However the geographical separation of the sampling locations, could suggest, at least for some samples, an isolation by distance model. Physical as well as biological factors could influence local retention and survival of larvae, at least in some areas, thus preventing long-distance dispersal and hence triggering genetic isolation.

Model 1: Hardy–Weinberg equilibrium (HWE). Under the assumption of random mating, absence of selection and genetic drift and presence of non-overlapping generations, one would expect the frequencies of different alleles at unlinked and codominant loci to be in HWE.

Model 2: Wright (1943) Island model. Under the assumption of many finite sub-populations as sources of migrants between widely separated sites, one would evaluate the Fixation Index F_{is}, F_{it} and F_{st} as coefficients of inbreeding within and between populations respectively.
Null hypothesis: no allele or haplotype frequency changes are expected when comparing populations along the east coast of Africa.

In the process of testing hypotheses, however, efforts should be made to avoid the following two pitfalls. Firstly, a representative sample of individuals from a natural population must be randomly selected.
Secondly, a sample of gene loci representative of the whole genome of the organism under study must be carefully identified.

Once the distribution of the allele frequencies is measured, they can be compared with those expected from the model under test using a statistical test such as which would allow acceptance or rejection of the null hypothesis. In such a case, it is possible to give a genetic explanation of the observations that have been made. This applies in the event that one already knows the expected distribution of the variables being tested. In certain situations, however, one may know very little of the expected distribution of the data, and hence fail to meet the assumptions required for customary statistical tests. In such instances, non-parametric randomisations test, have been found to be powerful tools. A test of this type involves three steps:
Consider an observed sample of frequencies as one of many possible, but equally likely, different outcomes that could have arisen by chance alone.
Enumerate the possible outcomes that could be obtained by randomly rearranging the frequencies.
On the basis of the resulting distribution of outcomes, decide whether the single outcome observed is improbable enough to warrant rejection of the null hypothesis.

Sometimes determining all the possible outcomes of a given distribution is feasible and exact randomisation tests, such as Fisher's exact test or the Markov chain method (MCM) can be performed. The latter is a sample test that can be used in the event that the possible enumeration of outcomes is too large. In this case one takes a random sample, computes the statistics, empirically finds a distribution of these statistics, and decides on the significance of the original observation (Sokal and Rohlf, 1981).

Allozyme methods and molecular approaches such as microsatellites, mtDNA-RLFP, and direct sequencing of nuclear and mtDNA (Cytochrome Oxidase I, COI, and internal transcriber spacer, ITS) seem to violate the rules of scientific design listed above. This is because they rely mostly on a low number of loci, sometime single genes, and on the efficient work of primers and enzymes, for a successful identification of genetic variability (Nei, 1984; Beamount and Nichols, 1994).

A further source of possible error is the fact that loci to be investigated are not chosen at random but typically their final selection is dictated mainly by the possibility of unambiguously scoring the results. Besides, financial considerations usually argue against the choice of a number of loci high enough to be representative of the entire genome. In these events therefore, these statistical devices can be of invaluable help in reducing the possibility of erroneously rejecting a null hypothesis. Nevertheless, in molecular ecology and population genetics, genetic structure is influenced by several factors such as population sizes, inbreeding typology, geographical distribution of populations and individuals, mutation rates, migration rates, and last but not least, natural selection (Hartl and Clark, 1997). It is almost impossible to be able to predict the synergistic effects of all these natural processes so as to allow the construction of elementary models, which can allow reliable inferences of population structure using only the relevant features of a system. Thus, it is important to compare results obtained from different molecular techniques as well as outcomes from inference methods with empirical predictions based on particular evolutionary scenarios (Heydrick, 1999).

1.8 Purpose of the research and layout of the thesis

The research presented here describes the molecular record of *Penaeus indicus* in the west Indian Ocean. The present study was designed to examine the population genetic of *P. indicus*, the most abundant and economically important species along the east coast of Africa, in order to gain insights into the biogeography and genetic structure of this widespread genus in the western Indian Ocean. To accomplish this, populations of *P. indicus* were sampled along the East coast of Africa and in Madagascar, and examined by different DNA molecular techniques as well as allozyme electrophoresis. The hypothesis of Hardy–Weinberg equilibrium was tested assuming a panmictic population over the entire sampling range of the species, and in the case of molecular data, a Wright–Fisher model of neutral selection was tested based on a classical as well as coalescent approach. Comparison among the different molecular techniques was performed for the main purpose of establishing the resolution at which possible population diversity can be detected and to verify if inter-population genetic diversity was consistent across all nuclear and mtDNA loci. Furthermore, as a preliminary comparison of population structure and genetic diversity, data on the phylogenetically close species *P. merguiensis* are presented in the form of mtDNA sequences collected from different locations in the Andaman Sea and in the gulf of Thailand[3]. The aim of this data set was to compare the genetic variability present at the Cytochrome Oxidase I (COI) between morphologically and ecologically similar species. *Penaeus indicus* and *P. merguensis* are in fact phylogenetically recognised sister taxa (Baldwin *et al.*, 1998) and interesting evolutionary hypotheses could be tested comparing the genetics of these two species. Furthermore, as stated by Avise (2000), genealogical concordance across co-distributed species presumably would reflect shared historical evolutionary elements. If this hypothesis were to be rejected, several evolutionary scenarios regarding the colonisation of the east coast of Africa could become relevant and would provide significant information to the understanding of speciation as well as dispersal mechanisms in this genus.

Therefore, this investigation is detailed in the chapters that follow, whereby chapter one provides a theoretical framework to the analytical methods (experimental design) used to infer population structure, while the chapter that follows introduces, describes and discusses the genetic variation, as measured by allozyme polymorphism, found in *P. indicus* from east Africa. Likewise chapter three describes the variability of genomic DNA as detected by

[3] kindly provided by the University of Songala in Thailand

polymerase chain reaction of random primers, while chapter four details sequencing analysis of both, nuclear and mitochondrial DNA. A synthesis and comments on the significance of these aspects of *P. indicus* genetic variation is presented in chapter five.

Chapter Two

Allozymes

2.1 Introduction

As mentioned in chapter one (Table 1.2, page 14), population genetic studies have traditionally been based on the analysis of iso-enzyme allelic variation. Investigations of allozymes by starch gel electrophoresis have proved informative in elucidating genetic structure and phylogenetic history of many marine invertebrates (Beamount, 1994). The aim was to understand how various isozymes could reveal genomic variability in *Penaeus indicus* under the assumption of dealing with panmictic populations. However, as noted by Hillis and Moritz (1996), and by Contrafatto *et al.* (1994), allozyme electrophoresis appears to be most informative when elucidating inter-specific relationships. Hence, this methodology was chosen for the present study as a means of testing the Hardy-Weinberg equilibrium and as a useful means of providing overall background information on *Penaeus indicus*. The intention was to subsequently integrate this information with more sensitive molecular methods, which are detailed in the next two chapters.

2.2 Materials

The populations and specimens used for the analysis of allozyme variation were also utilised for the extraction of genomic and mitochondrial DNA detailed in the next two chapters.

2.2.1 The Populations Studied

Specimens of *Penaeus indicus* (Fig. 2.1) were obtained pre-frozen from commercial fisheries operating on the Sofala banks, off the Zambesi river (20° 00' S, 35° 00' E), in Mozambique and from Tanzania and Oman (the exact localities in the latter two countries are unknown). Specimens from Mahajanga (15°30' S, 46°30' E) (Madagascar) were obtained fresh from local commercial trawlers. Specimens from St. Lucia (28°30' S, 32°30' E) in KwaZulu-Natal (South Africa) were collected using a beam trawl and immediately frozen on dry ice (-70°C), before being delivered to the laboratory of the School of Life and Environmental Sciences, University of Natal. Small portions of the abdominal pleopods were stored in an ultra deep

freezer (Nuaire, UK) at –78°C. A detailed list of the 261 *Penaeus indicus* specimens studied, and the localities is provided in table 2.1

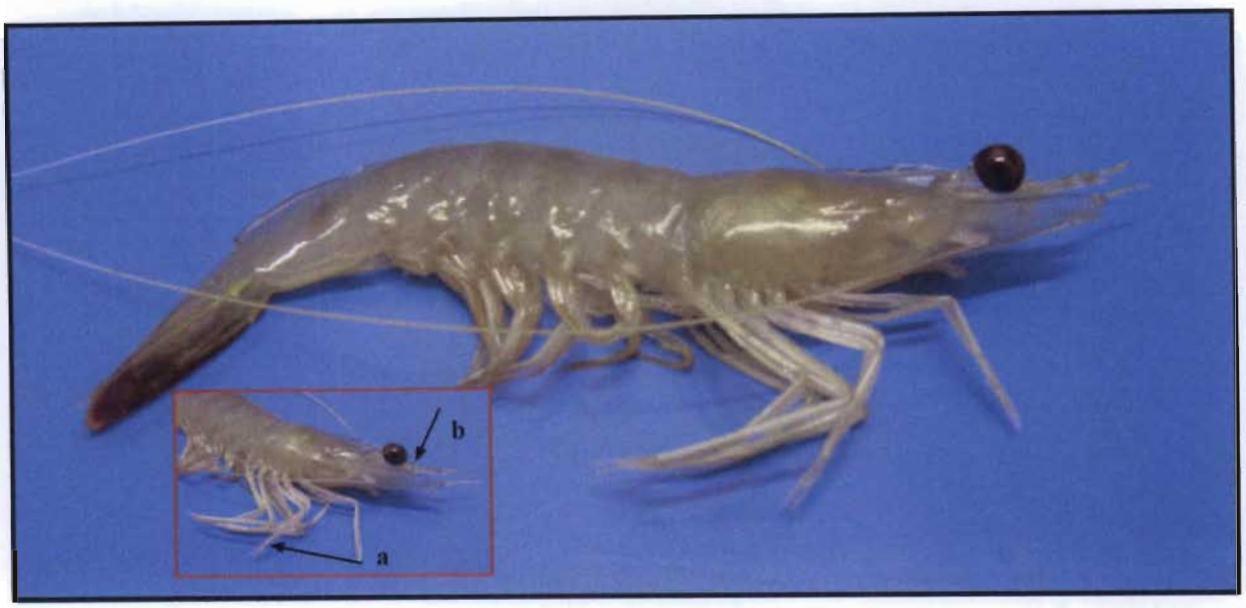


Fig 2.1 - Digital photograph of specimen of *Penaeus indicus* collected in South African waters. a: 3rd maxilliped; b: rostrum.

Table 2.1 - Sampling sites and number of specimens collected for the five samples of *Penaeus indicus* surveyed in this investigation.

Pop ID	Site of collection	N° of individuals
1 South Africa	St. Lucia	50
2 Mozambique	Maputo	51
3 Tanzania	Dar es Salam	50
4 Madagascar	Nosi Bè	50
5 Oman	Muscat	60

2.3 Methods

Extracts of abdominal muscle (2:1; tissue mass: grinding buffer volume) were prepared by weighing approximately 4 g of tissue, adding 2 ml of 0.04 % aq. ß-Mercaptoethanol, and then homogenising the tissue with a glass pestle in a small plastic mortar kept on ice. Ground glass was sprinkled on each sample in order to assist tissue disruption. The samples of homogenised muscle tissue were transferred to 50 ml centrifuge tubes and stored on ice in the cold (5°C) until centrifugation at 8000g for 30 min at 4°C (Beckman Model CP centrifuge). The supernatant was then transferred to a 1,5 ml Eppendorff tubes and stored at -78°C for further analysis. Twenty-six loci (Table 2.2) were tested to determine which enzyme was active enough to allow reliable scoring and which electrophoretic media and buffers yielded the best resolution. Tissue homogenates were subjected to electrophoresis on non-commercial starch gels, using four buffer types, TC7, TEC 7.9, TEB 8.4, LIOH which had been tested successfully with *P. monodon* tissue (Ballment *et al.*, 1994; see Appendix IIA).

Table 2.2 - Enzymes used in this investigation, their enzyme nomenclature, reference number (E. C.) numbers and activity observed during a preliminary survey of appropriate buffer systems.

Enzyme	*Full Name*	*E.C.*	*Buffer*	*Activity during survey*
AAT	Aspartate aminotransferasi	2.6.1.1	S/TEB	Adequate for population survey
ACO	Aconitate hydratase	4.2.1.3	S/TC7	Activity insufficient for reliable scoring
ADH	Alcohol dehydrogenase	1.1.1.1	S/TEB	Adequate for population survey
AK	Adenylate kinase	2.7.4.3	S/TEB	Adequate for population survey
ALD	Fructose-bisphosphate aldolase	4.1.2.13	S/TEB	Adequate for population survey
CK	Creatine kinase	2.7.3.2	S/TC7	Resolution inadequate and activity weak
FBP	Fructose bisphosphatase	3.1.3.11	S/TC7	Adequate for population survey
FUM	Fumarate hydratase	4.2.1.2	S/TC7	Adequate fro population survey
GDH	Glutamate dehydrogenase	1.4.1.3	S/TEB	Resolution too poor for reliable scoring
G3PD	Glycerol-3-phosphatedehydrog.	1.1.1.8	S/TEB	Resolution too poor for reliable scoring
GPI	Glucosephosphate isomerase	5.3.1.9	S/TEC	Adequate for population survey
HK	Hexokinase	2.7.1.1	S/TEC	Adequate for population survey
IDH	Isocitrate dehydrogenase	1.1.1.42	S/TEC	In locus IDH2 resolution too poor
LDH	L-Lactate dehydrogenase	1.1.1.27	S/TEC	Adequate for population survey
MDH	Malate dehydrogenase	1.1.1.37	S/TC	Adequate for population survey
ME	Oxaloacetate-decarboxylase	1.1.1.40	S/TC	Adequate for population survey
MPI	Mannose-6-phosphate isomerase	5.3.1.8	S/TEC	Adequate for population survey
LA	Peptidase (Leucylalanine substr.)	3.4.11/13	S/LI	Adequate for population survey
LGG	Peptidase (Leucylglycylglycine)	3.4.11/13	S/LI	Adequate for population survey
LP	Peptidase (Leucylproline substr.)	3.4.11/13	S/LI	Adequate for population survey
LT	Peptidase (Leucyltyrosine substr.)	3.4.11/13	S/LI	Adequate for population survey
LV	Peptidase (Leucylvaline substr.)	3.4.11/13	S/LI	Resolution too poor for reliable scoring
PGD	Phosphogluconate dehydrogenase	1.1.1.44	S/TC	Adequate for population survey
PGM	Phosphoglucomutase	5.4.2.2	S/TC	Adequate for population survey
SOD	Superoxide dismutase	1.15.1.1	S/TEB	Adequate for population survey
XDH	Xanthine oxidase	1.1.1204	S/TEB	Adequate for population survey

Gels were prepared at a 13 % concentration with all buffers used, liquefied, degassed and finally poured into a horizontal Perspex mould (210 mm X 136mm X 8mm) covered with a heavy glass plate and allowed to set overnight at room temperature. The glass cover was then removed, the gel trimmed to remove starch which had overflowed onto the sides of the mould, and cut across its width, 4 cm from one of the short sides. Pieces of chromatography paper wicks (10mm X 3mm, Whatman No. 3) were inserted in the Eppendorf tubes to absorb part of each homogenate and then inserted next to one another along the cut in the gel. Gels were then placed on trays of buffer (in a cold room) at 5-8 °C and cuts of buffer-soaked sponge-cloth were used as electrophoretic wicks. Voltage varied with the type of buffer uses (see Appendix IIA for details on media and volts applied). After five hours, gels were removed, placed on a large sheet of perspex and sliced into four or five sheets of 1.4 mm thickness. In order to produce an even surface on each slice, a slight downward pressure was applied to the gel during slicing by positioning a piece of glass-plate on the top of the gel. Slices were obtained by drawing through the gel a 0.5 mm diameter aluminium wire, kept under tension, and resting on rulers placed on both side of the gel.

Staining solutions were prepared using standard recipes (see Appendix IIB modified from Harris and Hopkinson, 1976). These were applied to the slices as solutions in molten agar (10 ml of stain solution plus 10 ml of aqueous 1.5 % agar, at 60 °C, per gel slice) or poured on the gel in a shallow container (50 ml of buffer solution per gel slice). After staining with the appropriate substrate for each enzyme tested, a varying number of bands appeared in the lanes corresponding to the individual extracts.

These bands or isozymes may correspond to the following interpretations:

1. Products of the expression of several genes at several loci;
2. Products of the expression of several alleles of a given gene at a given locus;
3. Molecules produced by conformational changes of a given protein molecule;
4. Molecule synthesised by a given gene or group of genes, which have undergone various post-translational transformations.

Finally, in order to allow reproducible scoring and to identify less intense bands, electromorphs showing enzymatic activity were photographed with a digital camera model 244 (Sony, Japan) and each image was analysed using an image editor (Adobe Photoshop 5.0). With this software, each picture, representing a particular locus, was converted from 24-bit colour to greyscale, adjusted in contrast and exposure after which a zymogram was constructed.

2.3.1 Data analysis

As described in the previous section, electrophoresis of 26 enzymes (Table 2.2, page 27) was carried out on 14 % non-commercial starch gels[1]. Details of electrophoresis and staining techniques are given in Appendix IIB. Zymograms were constructed to determine how many chromosomal genes, or rather how many loci, were responsible for the observed pattern, and how this pattern could be explained in terms of individual genotypes. Loci were then subdivided into monomorphic and polymorphic, using the frequencies of different alleles of each type. Loci were considered polymorphic if more than one genotype was present according to Clark (Heydrick, 1999). Although polymorphic loci are generally considered more informative as they allow testing hypotheses on evolutionary forces, monomorphic loci were not neglected because they provide an estimate of the real population heterozygosities (Nei, 1987). Allele frequencies were estimated from zymograms assuming codominant expression and selective neutrality. To compute basic population genetic summary statistics such as, allele frequency, gene diversity, genetic distance, F-statistics and multilocus structure, the computer programmes POPGENE version 3.1 (Yeh and Yang, 1999) and F-STAT version 2.9.3.1 (Goudet, 1995) were chosen. Hardy-Weinberg equilibrium was tested by computing expected genotypic frequencies, under the assumption of random mating, using the algorithm of Levene (1949). Significance for Hardy-Weinberg equilibrium at each locus was tested by chi-square (χ^2). Estimates of F_{IS} as a measure of heterozygote deficiency or excess (Wright, 1978) were also computed.

Genetic distances were computed using Nei's unbiased (1978) estimates of heterozygosity and were used to carry out cluster analysis by the Unweighted Pair Group Method with Averages (UPGMA) which was presented graphically in the form of phenograms. Fixation index values across samples (F_{st}, F_{is} and F_{it}) were calculated following Nei (1987) and Weir and Cockerham estimators (1984) of gene diversities and differentiation as well as to test for genotypic disequilibrium. The statistic used to test for linkage disequilibrium between pairs of loci was the log-likelihood ratio. The *p*-value of the test is obtained as follows. Genotypes at any two loci are randomly associated a number of times and the statistic is recalculated on the randomised data set. The *p*-value is then estimated as the proportion of values, calculated from the randomised data sets that are greater or equal to those observed. The parameters H_O,

[1] Personal communication, Dr. G. K. Campbell, School of Life and Environmental Sciences University of Natal, Durban

H_S, H_t, D_{st}, D_{st}', H_t', G_{st}, G_{st}' and G_{is} (Nei, 1973) were estimated for each locus and overall for all gene loci. All the equations used here rely on genotypic, rather than allele numbers, and are identified as follows.

- **H_0**: observed proportion of heterozygotes.
- **H_S** : within sample gene diversity.
- **H_t** : overall gene diversity
- **D_{st}** : gene diversity among samples.

Thereafter, the multilocus Weir and Cockerham (1984) estimator of F_{ST} (theta, Φ) between all pairs of samples was calculated. Confidence intervals were based on bootstrapping, which randomly re-sampled (10 000 replicates) all loci analysed. Pairwise F_{STs} were calculated for each replicate and sorted in ascending order. The significance levels (Rice, 1989) were re-adjusted over the number of pair-wise comparisons made: hence the α levels were 0.005 and 0.001. The so-called "Nm", which in population genetic represents the number of migrant per generation (Heydrick, 1999), is normally output simply as a function of pairwise $F_{ST,}$ namely $1/(4\ F_{ST}) - 1/4$. In many cases, however, the assumptions necessary to transform F_{ST} into a number of migrants are not fulfilled (Whitlock and McCauley, 1999). This was particularly the case in this study that dealt with organisms whose dispersal and migration can occur at different life stages and, therefore, are subjected to many evolutionary variables. Ultimately, this can influence the outcome of the model under investigation (Kimura 1983; Wright, 1969). Consequently calculations of these values were not applied to the present research.

2.4 Results

Penaeus indicus populations along the East Coast of Africa had never been screened for polymorphism at allozyme loci. For this reason, a total of 26 loci (Table 2.2, page 27) were tested by starch gel electrophoresis in order to detect enzymatic activity of abdominal tissue from the 261 individuals collected.

After several attempts to integrate the best buffer and running conditions, 20 loci were found to be consistently active and well resolved. Three of these were expressions of two distinct gene products, MDH and IDH, whereas two loci expressed MPI. These loci were chosen for routine use in estimating protein polymorphism in the five populations of *P. indicus* studied (Table 2.2, page 27).

Because of excessive substrate activity in the gels, it was not possible to score the GPI locus, while variation on loci IDH2 and LP was not genetically interpretable, owing to inconsistency in banding patterns, even among replicates on the same gel. A measure of inconsistent activity was generally noticed also between samples loaded onto the same gel. Initially, each sample was applied more than once in the same place on the gel, allowing for absorption into the starch between successive sample applications. However, this method, which improved staining intensity, was only useful in cases of widely separated alleles.

Simple, single banding patterns were usually observed for homozygotes, and either two or three bands were observed for heterozygotes. However, SDH and PGM zones were often accompanied by secondary banding i.e. a triple banding pattern, typical of dimer molecules was observed. Only four loci (LDH, LGG, MPI2, MDH1) of the 20 loci screened (Fig 2.2), displayed polymorphism at a level of 0.99 percent level frequency of "most common allele" (Table 2.3).

Table 2.3 - Polymorphic loci in *P. indicus*: overall observed number of alleles (na) and effective number of alleles (ne) (Kimura and Crow, 1964). Shannon's information index (I) (Lewontin, 1972). Only polymorphic Loci are presented, but calculations include monomorphic loci. Sample size calculations are based on number of alleles while table presents number of genotypes.

Locus	Sample Size	na	ne	I
LDH	261	2	1.9895	0.6905
LGG	261	2	1.0713	0.1500
MDH	261	2	1.0351	0.0871
MPI	261	3	1.9647	0.7655
Mean	**261**	**1.2500**	**1.1038**	**0.0847**
St. Dev		0.5501	0.2994	0.2235

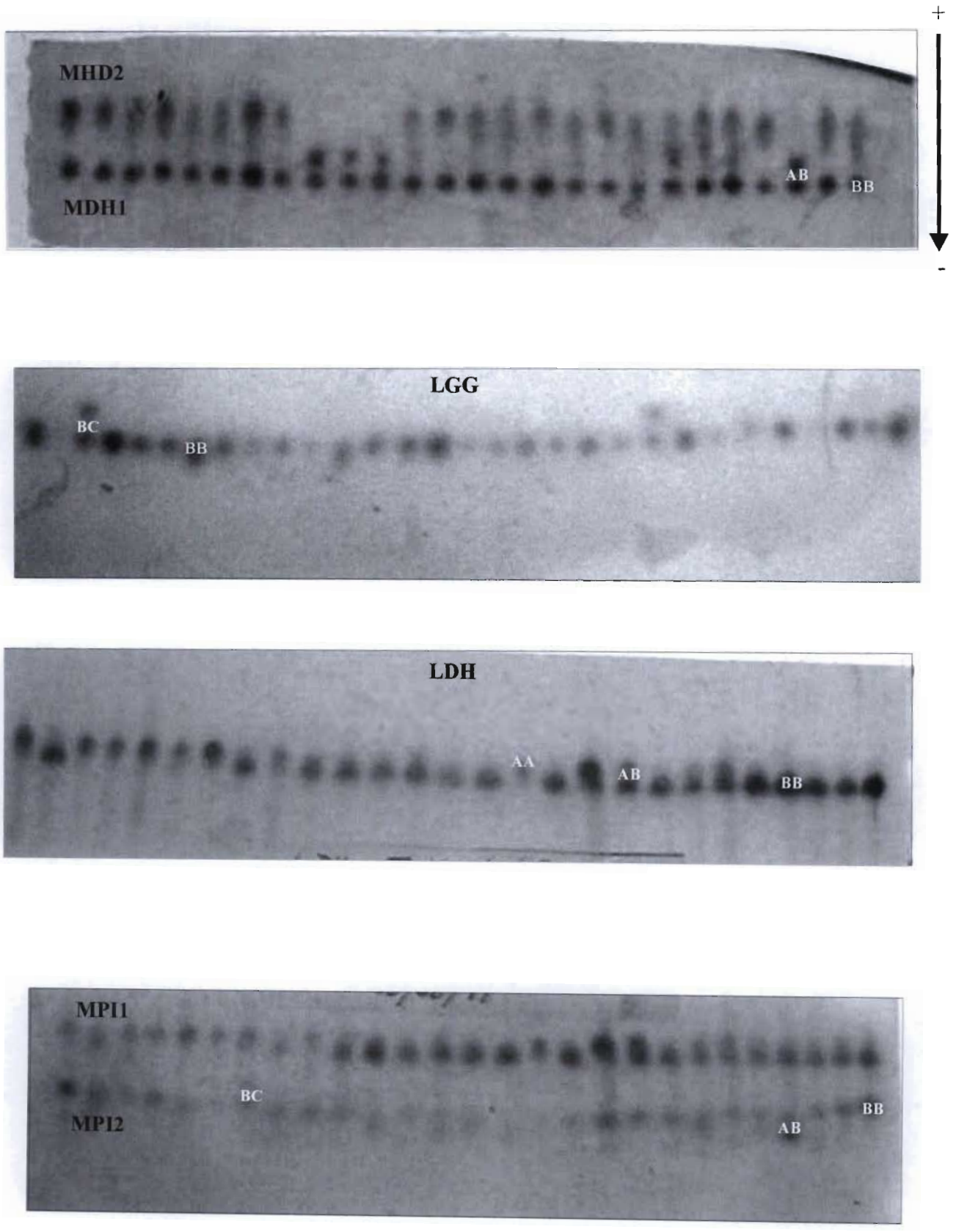


Figure 2.2 - Examples of allozyme migrations on starch gel of the four polymorphic loci and genotype interpretations (see table 2.2 for acronyms).

The remaining loci (ALD, HK, LP, LT, LA, LV, CK, AK, IDH1, PGD, PGM, FBP, ME, SDH, XDH, MPI1 and MDH2) were monomorphic with one or more alleles present at a frequency higher than 0.99 in all the individuals studied. Overall, the percentage of polymorphic loci varied from 15 % to 20 %. All four polymorphic loci had mean alleles frequencies per locus ranging between 1.27 and 1.57 (SE 0.20); observed mean heterozygosities per locus ranged from 0.048 to 0.058 (Table 2.4).

Table 2.4 - Overall observed and expected homozygosity and heterozygosity in *P. indicus* computed using Levene (1949) and Nei (1973) indices. Only values of polymorphic loci are reported.

Locus	Sample Size	Obs_Hom	Obs_Het	Exp_Hom	Exp_Het	Nei	Ave_Het
LDH	261	0.5326	0.4674	0.5017	0.4983	0.4974	0.4824
LGG	261	0.9310	0.0690	0.9333	0.0667	0.0666	0.0684
MDH	261	0.9655	0.0345	0.9660	0.0340	0.0339	0.0345
MPI	261	0.5096	0.4904	0.5080	0.4920	0.4910	0.4772
Mean	261	**0.9469**	**0.0531**	**0.9455**	**0.0545**	**0.0544**	**0.0531**
St. dev		0.1467	0.1467	0.1516	0.1516	0.1513	0.1469

These results are similar to those found for several other species of *Penaeus* and conform to the finding of Mulley and Latter (1980) who described an extremely low heterozygosity in 13 Penaeus species they analysed. Such low frequencies ranged from 0.06 to 0.03, although the analysis did not include *Penaeus indicus*. Forbes *et al.*, (1999) recorded very similar results, with mean heterozygosities ranging from 0.8 to 0.01 in five samples of ***Penaeus monodon*** collected from different locations along the east coast of Africa, although their analysis included only selected polymorphic loci.

Allele frequencies at the four polymorphic loci from the present study did not show a significant departure from Hardy-Weinberg equilibrium (Appendix IIC). Appendix IIC also lists data for the LGG locus, where the most common allele was allele B however in the samples listed above some individuals showed the presence of heterozygotes BC. It should be noted that this allele (C) was not found in the Oman population for which several additional individuals (approx. 20) were also screened for that locus (Fig 2.2). Appendix IIC also lists other details of allele frequencies, homozygosity and heterozygosity levels for all

polymorphic loci. No significant values of linkage disequilibrium were recorded among pairwise comparisons of loci (Table 2.5 – section A). The within sample gene diversity was low (Hs = 0.074), the overall gene diversity (Ht) and the amount of gene diversity among samples (D_{st}) were respectively 0.078 and 0.004. Finally, the estimator of F_{st}, the measure of heterozygote deficit in a population, or G_{st} (Wright 1970), computed to a value of 0.056 (Table 2.5- section B).

Table 2.5- Section A: *p*-value for genotypic disequilibrium based on 4620 permutations. Adjusted *p*-value for 5 % nominal level is 0.000216 (Rice, 1989). Section B: Nei's estimation of heterozygosity

Section A:

Loci tested	*p*-values
LDH X AAT	0.83117
LDH X LGG	0.60087
LDH X MDH	0.3744
LDH X MPI	0.29892
LGG X MDH	0.00022
LGG X MPI	0.75844
MDH X MPI	0.52749

Section B:

Locus	H_o	H_s	H_t	D_{st}	$D_{st'}$	$H_{t'}$	G_{st}	$G_{st'}$	G_{is}
LDH	0.468	0.487	0.497	0.010	0.013	0.500	0.020	0.025	0.040
LGG	0.078	0.076	0.075	0.000	0.000	0.075	-0.002	-0.003	-0.036
MDH	0.036	0.035	0.035	0.000	0.001	0.035	0.012	0.015	-0.028
MPI	0.489	0.482	0.489	0.007	0.009	0.491	0.014	0.018	-0.016
Overall	**0.073**	**0.074**	**0.078**	**0.004**	**0.005**	**0.079**	**0.056**	**0.069**	**0.005**

The index F_{is} , the measure of the deviation from Hardy-Weinberg equilibrium within sub-populations, although not statistically significant, was found to be negative at the loci LGG, MDH and MPI. F_{st}, defined from Nei (1987) as G_{st}, was 0.024 (Table 2.6- section A) and in accordance with Forbes *at al.* (1999).

Table 2.6- Summary of F-statistics for all polymorphic loci (Section A). Section B: *p*-values obtained from 1000 permutations of F_{st} pairwise values. Adjusted significance level (5%) for multiple comparisons: 0.005 (Rice, 1989)

Section A

Locus	Sample Size	F_{is}	F_{it}	F_{st}
LDH	261	0.0307	0.0580	0.0282
LGG	261	-0.0488	-0.0372	0.0111
MDH	261	-0.0380	-0.0182	0.0191
MPI	261	-0.0256	-0.0032	0.0219
Mean	**261**	**-0.0019**	**0.0222**	**0.0240**

Section B

Pop ID	MZ	TZ	MD	OM
ZA	**0.78800**	**0.22800**	**0.73400**	**0.00100****
MZ		0.61100	0.60100	0.00900
TZ			0.06500	**0.00200***
MD				0.00600

p-value of F_{st} pairwise comparison among populations (Weir and Cockerham, 1984) showed a significant level of genetic differentiation at the 0.005 level, between Oman and South Africa and between Oman and Tanzania (Table 2.6- section B). However, it is important to point out that the major weight in the output of the F_{st} analysis is given by the two loci MDH and LGG. These genes were in fact recorded as monomorphic in the Oman population while only MDH was found to be monomorphic in Tanzanian samples (see Appendix IIC).

Nei's original measure of genetic distance (Nei, 1978) ranged from 0.0001 to 0.0056 and are presented in table 2.7, while the phenogram (Fig. 2.3) shows the presence of the cluster combining South Africa, Mozambique and Tanzania, while the Madagascar and Oman samples separate from this group at higher clustering levels.

Table 2.7- Nei's original measures of genetic identity (Nei, 1978) and genetic distance. Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

Pop ID	ZA	MZ	TZ	MD	OM
ZA	*****	0.9995	0.9993	0.9996	0.9944
MZ	0.0005	*****	0.9997	0.9999	0.9973
TZ	0.0007	0.0003	*****	0.9996	0.9967
MD	0.0004	0.0001	0.0004	*****	0.9968
OM	0.0056	0.0027	0.0033	0.0032	*****

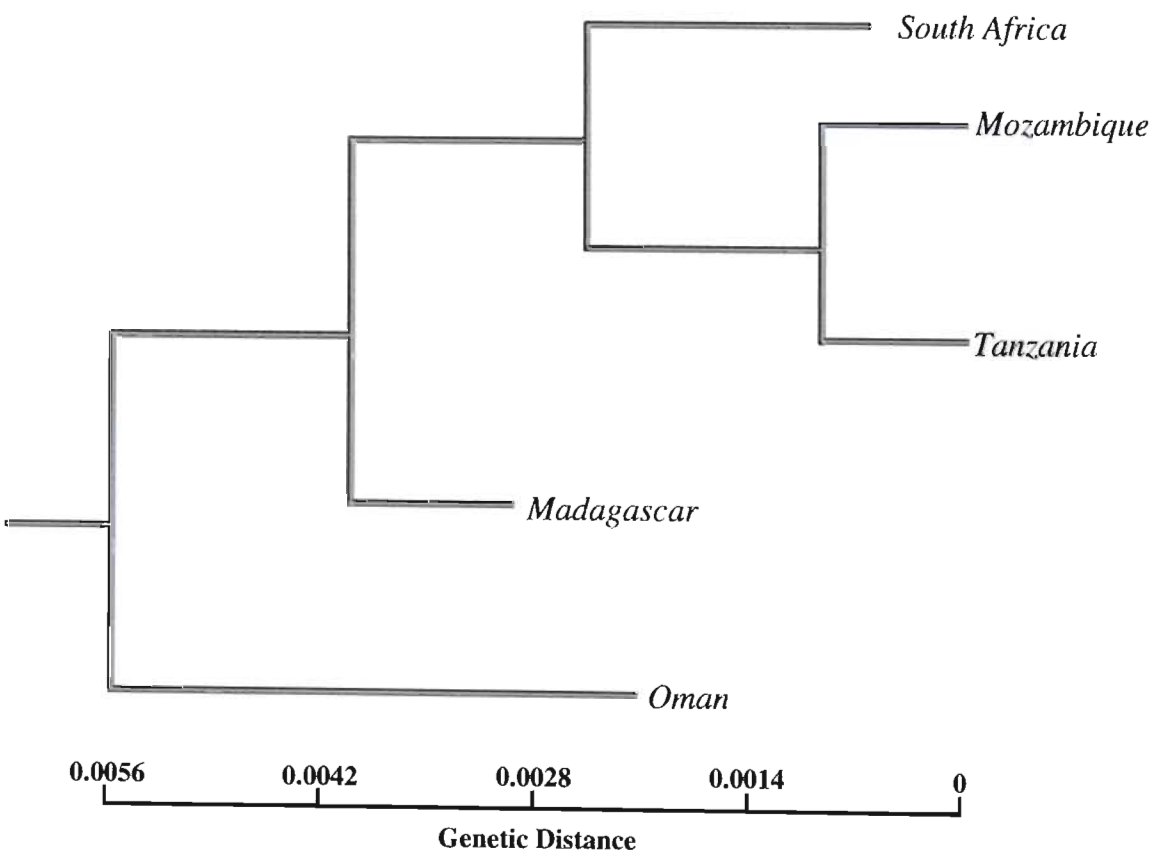


Figure 2.3 - UPGMA phenogram based on Nei's (1978) genetic distance.

In conclusion, although allozymes were an important source of information in detecting polymorphism levels and to test possible departure from Hardy-Weinberg equilibrium, they proved not to be variable enough to reliably detect any population structure of *P. indicus* along the East coast of Africa.

2.5 Discussion

The application of allozyme electrophoresis is no longer the first choice when testing the hypothesis of panmixia in natural populations, especially in cases where highly dispersing larval stages are present (Palumbi, 1999). However, in this research, allozymes have been helpful at least in detecting Hardy-Weinberg equilibrium. This technique has been by far the most utilised during the past 30 years or so and offers a reliable and relatively inexpensive approach to gaining genotypic information. The ease with which non-commercial starch can be prepared and the ability to test different enzymatic reactions in a single electrophoretic run, make this biochemical approach an important source of information when dealing with large sample numbers.

The most interesting feature revealed in the present study was the low level of polymorphism (20 %). Most enzymes showed the presence of a single allele fixed in all populations across all loci screened. However, the degree of polymorphism that has been recorded in this research is not uncommon in studies on penaeids (Mulley and Latter, 1980; Redfield *et al.*, 1980; Ko *et al.*, 1983; Benzie *et al.*, 1992; Forbes *et al.*, 1999). Allozyme polymorphism has been studied in different regions of the world (East coast of America, South East Asia, East coast of Africa, Australia). Several species, *P. monodon*, *P. japonicus*, *P. merguensis*, *P. indicus*, *P. semisulcatus*, to cite some which have an economical importance, have been the target of such studies. Although wide geographical studies of African *P. indicus* have not been published yet (Benzie, 2000), allozyme variation of a single population of *P. indicus* from Kenya has been reported by Ko *et al.* (1983). Their results based on a mean number of 20 individuals and 29 loci, showed a similar pattern to that detected by the present research. They estimated a low level of polymorphism of only four loci and mean heterozygosities of 0.026. Based on these reports one could hypothesise that the genus *Penaeus* has arisen fairly recently and speciation events happened to evolve quite rapidly compare to other marine organisms. Alternatively, a likely explanation of the degree of polymorphism encountered in penaeidae could reflect their generalist adaptation to the ocean environment. Species that

inhabit three different ecological niches during their life cycle, could be characterised by monomorphic loci due to selection pressure (Mulley and Latter, 1980). Such a pattern could also be achieved in the case that populations were subjected to periodic crashes, bottlenecks or founder events, resulting in reduction of population sizes (Heydrick, 1999). In the case of reduction of the effective population size, the neutral theory of Kimura (see chapter one, section 1.5) predicts a relatively fast fixation of alleles. These alleles, if present at high frequency in the genetic pool of the population, reach fixation in spite of the degree of selection acting on these loci. However, some species, such as *Penaeus plebejus*, complete their life cycle in oceanic waters so that the survival of breeding stocks would not have been endangered by population bottleneck if postlarvae were prevented from entering estuaries by climatic upheavals (Racek, 1970; Racek and Dall, 1965). In *P. plebejus*, a relatively low degree of polymorphism has also been recorded (Mulley and Latter, 1980). With this scenario Latter (1980) thought it unlikely that the genetic variation of penaeid prawns could be attributed to periodic population crashes.

A more parsimonious explanation for the extremely low levels of allozyme variation generally observed in Penaeidae is the hypothesis of selective elimination of mutational variation presented by Mulley and Latter (1980). This is applicable only in the case of relatively large effective population size, maintained for long periods, in accordance with the neutral theory (Kimura, 1983; see also chapter one, section 1.5.2).

The bottleneck hypothesis assumes that genetic drift is the major causal force of variability and, hence, a neutralist vision of evolution. On the other hand, selection is the major force implicitly advocated by Mulley and Latter (1980): a classic neo-Darwinian view. Both these factors are equally possible and present analytical methods do not have the power to discriminate between these two evolutionary forces. This issue is further discussed in chapter five.

For the limited number of species where data are available there is a strong positive correlation between duration of the larval stage and the realised dispersal distance (Jackson, 1986). It is generally believed that species with long larval periods disperse widely on ocean currents (Barber *et al.*, 2002). However there is increasingly evidence that larvae fail to fully achieve their dispersal capability (Briggs, 1984; Scheltema, 1971, 1986; Jackson, 1986; Burton and Lee, 1994; Swearer *et al.*, 1999) and local retention of larval stages may be high in conjunction with local currents and wind direction (Barber *et al.*, 2002). The combined

effects of these oceanographic mechanisms as well as a complex set of ecological and genetic processes, acting across a range of spatial and temporal time scale would determine the population structure within and among groups of individuals (Grosberg and Cunningha, 1999). For example, Forbes and Benzie (1999) analysed five populations of *P. monodon* collected from South Africa and Madagascar. Screening only polymorphic loci, they detected no significant differentiation among those samples ($F_{st} = 0.012$). These authors attributed this genetic pattern to significant reduction of effective population size during Pleistocene glacial periods. The marine current patterns were different from those of today and oceans receded from tidal lakes and river estuaries (Figure 2.4) thus reducing the availability of suitable habitat for the life cycle to be completed.

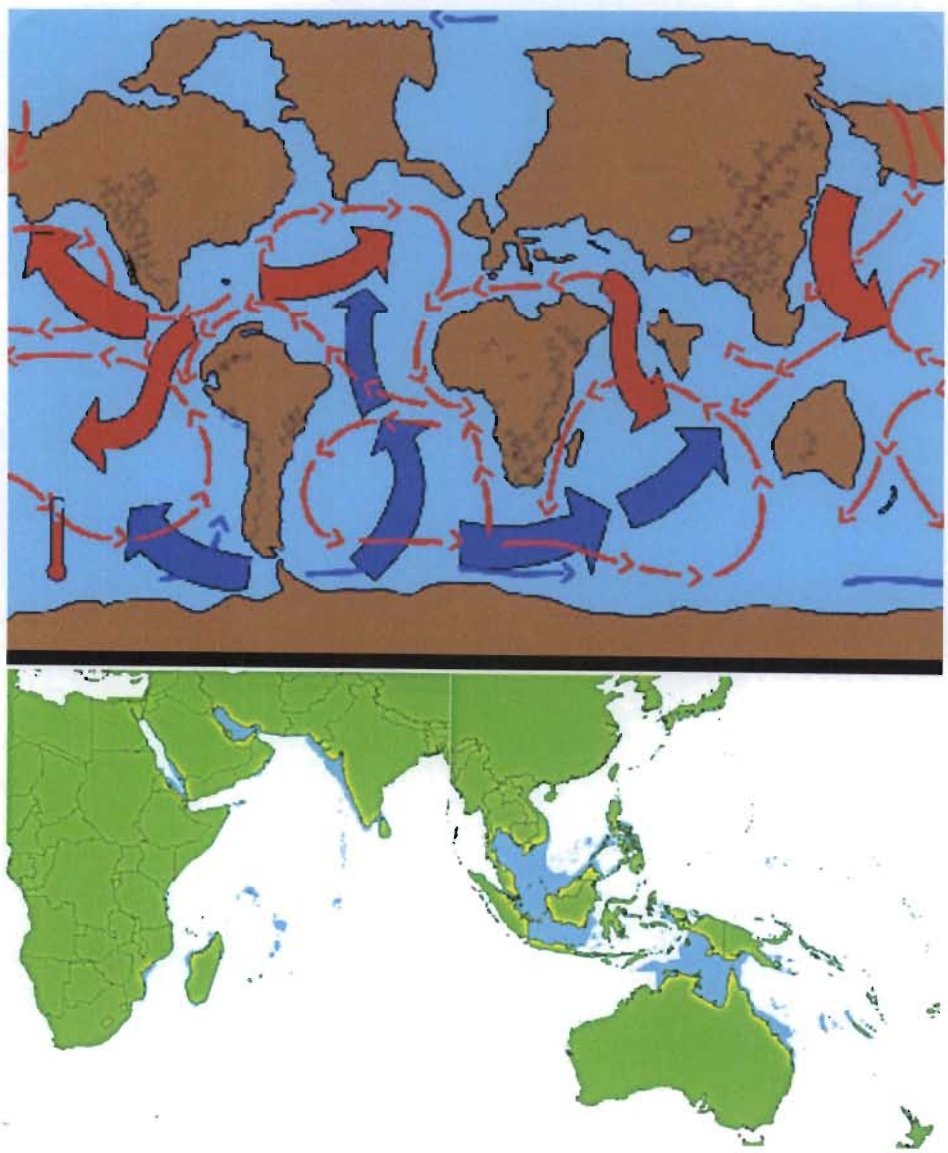


Figure 2.4 - Pleistocene global ocean currents (upper map). Red arrows: warm currents; blue arrows: cold currents. Average continental shelf (light blue) during Pleistocene glaciations in the IWP (lower map)

In this research, tests of population structure with F_{st} values in pairwise comparisons showed a low level of inter-population differentiation: overall F_{st} was 0.025. *P*-values of pairwise comparison were found to be significant in the case of Oman samples in comparison with Tanzania and South Africa, but were below 0.009 when compared with Madagascar and Mozambique, indicating perhaps a genetic differentiation of the sample from the South Arabian Peninsula from populations of East Africa. The Oman samples displayed monomorphic LGG and MDH loci, whereas all other specimens (except Tanzania at MDH locus) exhibited two more alleles at these loci, although these were detected at very low frequencies. This genetic pattern is suggestive of past gene flow restriction, although this does not necessarily apply to the present day gene exchange. Looking at figure 2.4, major warm ocean currents move southward from the Arabian Peninsula and thus would have not prevented dispersal of larvae towards the East coast of Africa. Hence this could be an indication that other evolutionary forces played a major role in determining this recorded scenario. Within the East African coast no significant population structure has been revealed among Mozambique, Tanzania and Madagascar, while a pattern conforming to an isolation by distance model is recognised in the significant statistical value of South African samples towards Tanzania.

Sometimes population sub-structuring is not obvious, and as a result, a sample may consist of heterogeneous sub-samples from populations. For example, sub-populations may be separated by subtle physical or ecological barriers that limit movement between groups or sampling activity might not be able to discern local isolated groups. If there are fairly large differences in allelic frequencies among these sub-samples, then when they are lumped together, there will be a net of deficiency of heterozygotes. In this research, the F_{is} index revealed a general excess of heterozygotes at most polymorphic loci (LGG, MDH, and MPI) as an indication at a sub-population level of deviation from Hardy-Weinberg equilibrium. Although this value was found to be not statistically significant, such a result is quite unexpected in this kind of organism where the hypothesis of panmixia is the normal rule. As a more likely explanation of this result, one is left with a lack of sampling effort per population, which might have influenced the outcome of the analysis. However, in the case of population bottlenecks the most characteristic effect experienced by the genome of the organism would be the reduction of the expected number of alleles compared with the expected heterozygosity (Nei *et al.*, 1975). Rare alleles do not contribute significantly to the heterozygosity of a population and are more likely to be lost through the action of selection and genetic drift. Therefore

conditional on the low number of alleles observed in this study, the recorded excess of heterozygosity could be the result of population, which did not reach yet a mutation-drift equilibrium state.

P. indicus is a widely dispersed species and its range is probably, together with that of *P. monodon*, the widest recorded for any *Penaeidae* species in the IWP region. If these samples are representative of the surveyed populations, the general lack of population structure along the African east coast may be an indication that all specimens derived from the same breeding population. If cladogenetic events in the Indo-West Pacific diversity triangle involved drastic reduction of effective breeding population sizes (founder effect) and *P. indicus* expanded towards the west by progressive colonisation, a lack of population structure would be expected. This is because a population may descend from a small number of individuals either through founder events or because of population crashes (bottlenecks). These events can lead to random genetic changes altering allelic frequencies of the ancestral population. The results of this would be lower heterozygosity and fewer alleles, as in the case of the Oman population. It is important to point out, however, that if the alleles are equal in frequency, the founder population size need not be very large in order to afford a high probability of inclusion of both alleles. If the frequencies of two alleles in the parental population are close to, say 0.95 and 0.05, then the founder number needs to be 30 or greater to give a 95 % chance of including both alleles (Heydrick, 1999). In general the founder population size has a large effect on the number of alleles in a population because founding events often eliminate alleles of low frequency. The average heterozygosity, on the other hand, is influenced primarily by the rate of population growth after the founding event, rather than by the number of founder individuals. The reduction of heterozygosity in each generation is dictated by the finite size of the population[2]. After a founder event through small founding numbers, if population growth is slow, heterozygosity is lost at each generation, until the population grows to a substantial size. It is possible that the rate of growth per generation in the east African samples could have been substantially high due to various ecological factors, such as lack of competition and/or availability of suitable habitats. Finally, it is also feasible that polymorphism at these two loci is maintained by balancing selection, as discussed earlier. Allozymes analysis did not succeed in revealing any substantial population structuring in East African samples. Therefore, it is possible to suggest that this technique was not sufficiently sensitive or sampling effort extensive enough to detect a hypothetical cryptic population structure.

Chapter three

Molecular analysis:

Random Amplification of Polymorphic DNA (RAPD)

3.1 Introduction

The suspected low sensitivity of allozyme analysis was highlighted in the previous chapter, based on the inconsistent population genetic structure demonstrated by this method. As a result molecular methods were utilised in subsequent studies. A lack of suitable molecular markers, such as primers from microsatellite genomic libraries for *P. indicus*, dictated the choice of random amplification of nuclear DNA, which had already been successful in providing useful information on population genetic structure of some penaeids (Meruane *et al.*, 1997; Tassanakajon *et al.*, 1997).

3.2 Materials

All populations analysed by allozyme electrophoreses (Table 2.1, page 26) were also investigated in this molecular study, DNA extraction was, however, carried out on a total of 50 specimens, 10 specimens per population.

3.3 Methods

3.3.1 DNA extraction

DNA extraction was carried out by the following two different protocols: the

Kessing method and the

CTAB method.

Quantitative comparison between the DNA yielded by the two techniques was made by means of spectrophotometric readings (Beckman DV 530) in order to establish the most efficient methodology.

Kessing's protocol:

Samples of muscle tissue were ground with small Teflon pestles in microcentrifuge tubes containing 100 mM EDTA, 10 mM TRIS (pH 7.5), 1 % SDS, and 1 μg/ml of Proteinase K. Samples were then incubated overnight at 55 °C and extracted twice with phenol/chloroform/isoamyl alcohol (24: 24: 1). The extracted DNA was precipitated twice with 100 % ethanol, once with 50 % ethanol and centrifuged at 4 °C in a Sepatech microcentrifuge (Heraeus, Switzerland) for 1.5 minutes at 13 000 g. The solution was then left at –78 °C for half an hour to aid precipitation of the DNA. The samples were then dried in a vacuum centrifuge and resuspended in TE buffer (10 ml 1 M Tris, at pH 7.69- 8.0 and 2 ml 0.5 M EDTA at pH 8.0-80 ml H_2O).

CTAB: modified extraction protocol

Abdominal muscle (50 mg) was placed in Eppendorf tubes with 500 μl of extraction buffer (100 mM TRIS, 50 mM EDTA, 1 % SDS at pH 8.0). These were heated at 65 °C for 30 minutes, 25 μl proteinase K (10 mg/ml) was then added and incubated at 55 °C overnight. Following this incubation, 80 μl of 5 M NaCl was then added, to a total volume of 605 μl, and gently swirled to mix the solution. Thereafter 75 μl of warmed (65 °C) C-Tab solution (10 %) were added and the mixture incubated at 65 °C for an hour. The DNA was extracted from this solution with an equal volume (700 μl) of chloroform: isoamyl alcohol (24: 20). The aqueous phase (500 μl) obtained from this extraction was transferred to a fresh tube and precipitated in 1 ml cold 100 % ethanol for 12 hr at –78 °C. This was then centrifuged (13 000 g in a Sepatech microfuge) for 30 sec and washed by replacing the absolute ethanol with 1 ml of 70 % ethanol. Tubes were then inverted on a paper towel for a few minutes; residual excess liquid was tapped off. Finally each tube was air dried overnight in a clean area at 37 °C. The extracted DNA was subsequently resuspended in 100 μl TE buffer or distilled water. Quantity and purity of the extracted products were examined spectrophotometrically and by electrophoretic separation on agarose gels.

3.3.2 Random Amplification of Polymorphic DNA (RAPD)

In contrast to isozymes, RAPD (Welsh and Mc Clelland, 1990; Williams *et al.*, 1990) provides by far a more extensive sampling of the genome and, ideally, can generate unlimited numbers of loci for use in genetic analyses. The RAPD procedure employed in this study was not substantially modified from that originally described by Williams *et al.* (1990). A total of 20 single-stranded oligonucleotide primers (decamers in Kit A from Operon Technologies, U.S.A.) were utilised in a polymerase chain reaction (Lynch and Milligan, 1994) using genomic DNA as a template. Typically, one primer at a time is used by this technique and only 10-25 ng of genomic DNA is necessary for the reaction. If the primer binds to sites on different strands that are within about 3 kilobases (Kb) of each other, the region between the ends of the priming sites is amplified. Because this can occur at any number of locations within the genome for any given primer, more than one DNA fragment may result from a single chain reaction governed by one primer. The decamer primers employed for RAPD have G-C rich content ranging from 50 % to 70 %. Thereafter, amplified fragments can be separated from the reaction mixture by poly-acrylamide or agarose gel electrophoreses.

The RAPD banding pattern is superficially similar to a multilocus minisatellite pattern in that it is generated by a number of loci. There is a problem in assigning bands to loci, co-migration cannot readily be disregarded, and two bands of differing mobility may be the two alleles from the same locus. Although heterozygotes may have more bands on average then homozygotes, the presence of a band is considered a dominant characteristic since its sequence is also present, albeit in single dose, in heterozygotes. Therefore, heterozygotes cannot be scored nor can allele frequencies be estimated directly.

3.3.3 Amplification

Amplification reactions were performed in volumes of 25 µl of reaction mix: containing 25 ng of DNA, 200 µM each $dNTP_s$, 3.0 mM $MgCl_2$, 0.2 µM 10-mer primer, and 1.0 unit of *Taq* DNA polymerase (Perkin Helmer, Stoffel fragment, USA) which is a truncated version of the *Taq* polymerase. Truncated polymerases, marketed under various names, have missing amino acids in their N-terminal region. They have higher thermostability than other versions of *Taq*, and no 3' to 5' or 5' to 3' exonuclease activity.

Two types of thermal cyclers (Hybaid) were programmed for one initial denaturing stage of 180 seconds, in order to separate the double stranded DNA. This was followed by 35 cycles of 45 seconds each at 94 °C, 45 seconds at 33° / 40°C, depending on the primer used, and 90 seconds at 72 °C. A final stage at 72 °C for 10 minutes ended the programme in order to allow the enzyme to complete amplification of all amplified fragments.

The reason for using two instruments lies in the fact that thermal cycler make use of different temperature transition speeds between different stages (fast, Sprint PCR and slow, Thermal cycler). This could influence the reproducibility of the RAPD patterns because the enzyme may have specific time requirements in order to complete amplification and transfer of heat to different Eppendorf tubes could be a further variable. Hence agarose gel band patterns produced by the two thermal cyclers were compared in terms of band intensity and numbers at each locus. All the amplification products were analysed by electrophoresis in 1.4 % agarose gels (Sigma, USA) and detected by ethidium bromide staining (0.5 μg/ml).

In order to score the banding pattern produced by the amplifications, as in the case of allozyme analysis discussed in chapter two, digital images were taken of each gel and used as a record for data analysis (Figure 3.1, page 48). As a measure of band intensity, the grey level (RGB) of each band was digitally recorded. The eyedropper tool of Adobe Photoshop (version 5.0) was placed across the band in order to record its grey value. If the reading across the all band was between 200 and 300 the band was considered as present for the analysis. However sometimes one band would yield a lower intensity of 200 but because of its presence in other individuals within the same population was still recorded. Furthermore each band on the gel was digitally enhanced to achieve more reliable reading after the bands with unmatched grey scale (< 200) value were discarded. Matrices of presence/absence of characters were then constructed. Band reproducibility was estimated as a proportion of individuals showing homologous bands in both replicates run on the two different PCR machines.

3.3.4 Statistical Analysis

For statistical analysis of RAPD data, the amount and distribution of variability of the RAPD profiles obtained from the 50 individuals studied (Table 2.1, page 26) using 20 primers (OPA1 – 20, Operon Technologies, USA) were used. Under the assumption that variation in

banding patterns represents allelic segregation at homologous and independent loci, each locus was treated as a two-allele system corresponding to the presence/absence of the amplified band. For each of the five populations, a matrix was constructed by coding presence of each band as 1 and absence as 0 (Appendix IIIA).

These presence-absence data were then analysed in order to obtain an estimate of the degree of polymorphism (P) for each band (POPGENE, version 3.2) and to estimate divergence values among all pairs of individuals in each population using the FORTRAN programme RAPDdist. The latter also performed a bootstrapping procedure that generated 1000 distance matrices that were used as input to the programme Neighbour from the suite PHYLIP 3.5 (Felsenstein, 1999) in order to construct a consensus tree (UPGMA clustering method). Most programmes that analyse matrices of RAPD patterns make the following assumptions regarding the inheritance and expression of RAPD polymorphisms among individuals in populations: RAPD bands are inherited as dominant alleles in a Mendelian fashion, individuals with a given amplified band are homozygous or heterozygous for a dominant allele whereas individuals without that band are homozygous recessive. Genotypes are assumed to be in Hardy-Weinberg equilibrium; the frequency q of the recessive allele at a locus is estimated as a square root of the frequency of homozygous recessive individuals and the frequency of p, the dominant allele consequently is estimated from $1 - q$.

The analysis was performed on a total of 50 individuals, under the assumption that a high number of RAPD loci may counterbalance the relatively low number of individuals and may, therefore reduce the sampling variation of divergence estimates (Nei, 1997). To estimate the variation attributable to intra and inter-population differences an AMOVA (Analysis of Molecular Variance) test (Excoffier *et al.*, 1992) was performed using the computer software ARLEQUIN version 2.0 (Schneider *et al.*, 2000). This utilises both frequency and sequence divergence between genotypes. The divergence between genotypes was estimated calculating the pairwise mean differences between haplotypes.

This type of analysis is based on the assumption that all populations surveyed have the same mating pattern. In such a case, the data can be treated as RFLP markers and AMOVA becomes an appropriate test (Excoffier, 2001). For each level, the sum of the squared deviation (SSD), the mean squared deviation (MSD) and the variance component were calculated, thus partitioning the genotypic variance, rather than the variance of allele

frequencies, as is the case for co-dominant markers. The variance component was then expressed as a percentage. The significance of the fixation index is tested using a non-parametric permutation approach described by Excoffier *et al.* (1992) consisting of a process of permuting haplotypes, individuals, and populations or groups of populations. After each permutation round, re-computation of all statistics is performed to get their null distribution. Under this procedure the normality assumption, usual in analysis of variance tests, is no longer necessary nor is it necessary to assume equality of variance among populations or groups. Nevertheless it is not possible to compare these F-statistics to those inferred from co-dominant markers, such as allozymes and microsatellites, because these are genotype correlations, whereas F-statistics refer to gene correlations for dominant markers (Excoffier *et al.*, 1992). Pairwise F_{st} values generated by the AMOVA were used to construct a matrix to test for the correlation between geographical distances and genetic distances using the programs IBD version 1.2, edited by Bohonak (2002).

As a form of comparison inference methodology RAPD band diversity was also analysed with the approach of Borowsky (2000), who showed that band diversity is a simple function of the nucleotide diversity (π) thus allowing for conversion of band survey data to equivalent measures of nucleotide diversity (Appendix IIIA).

A phenotypic heterogeneity index, Phi (ϕ_e) was defined as the proportion of unmatched RAPD bands between two individuals randomly chosen from the population according to the following equation:

$$\phi_e = 1/n \sum_{i=1 \text{ to } n} 2q_i^2 (1 - q_i^2)$$

It follows (for details see Borowsky, 2000) that the nucleotide diversity index (π) can be estimated as $\pi = 6/5 \ \phi_e (1 - \eta) / m$, where $\eta = 0.38$ is the estimated proportion of undetected loci for population sizes greater than 10^3 and $m = 20$ is the number of bases in the random primers used.

3.4 Results

The use of two thermal cyclers with the specific characteristic presented in chapter two, showed that a longer transition time between stages results in more consistent and reproducible banding pattern across all loci scored. The optimal program parameters for

amplification of *P. indicus* were 40 cycles of 5 seconds at 95 °C, 45 seconds at 34 °C to 40 °C, and 150 seconds at 72 °C. This program is considerably different from the standard program recommended by Williams *et. al.* (1990). The 20-decamer primers chosen from Kit A (Operon Technologies), consistently amplified a total of 64 scorable bands whose estimated sizes range from 200 to 1000 base pairs. The number of polymorphic sites ranged from 30 in the South African samples to 36 in the Mozambique samples, which had the highest percentage of polymorphisms (Table 3.1) (56 %). Each primer had between three and eight bands, with a mean of 5.2. Examples of RAPD profiles generated by primer OPA-2, OPA-6 and OPA-17 are shown in Fig 3.1. The Shannon Index (I) ranged from 0.39 to 0.42 showing very similar levels of intra-population diversity. Nei's analysis of gene diversity (G_{st}) in subdivided populations (Nei, 1987), averaged over all loci, was 0.12 (Table 3.2).

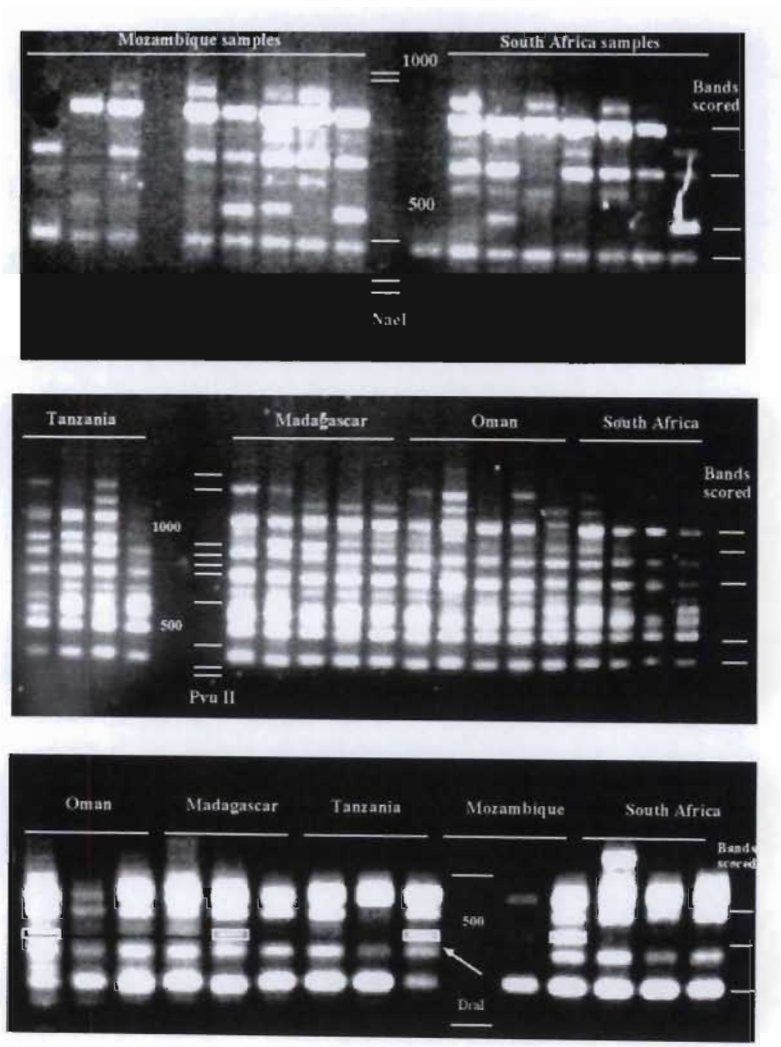


Fig 3.1- Example of RAPD patterns obtained during screening stages for optimal annealing temperature and DNA concentration. Respectively OPA2, OPA6, OPA17. Presented in the figure also the band scored after grey scale setting and computer generated restriction enzyme digests (pBR322 DNA). Arrow: OPA 17 bands present in the samples from Mozambique, Tanzania, Madagascar and Oman

Table 3.1 - Number and percentage of polymorphic loci in the samples from five populations of *P. indicus*.

Population ID	Number of Polymorphic Loci	Percentage Polymorphic Loci
South Africa	30	46.88
Mozambique	36	56.25
Tanzania	34	53.12
Madagascar	35	54.69
Oman	32	50.00

Table 3.2 - Nei's analysis of gene diversity in subdivided populations.

Locus	Sample Size	Ht	Hs	Gst
OPA02-1	40	0.4747	0.4747	0.0000
OPA02-2	40	0.1314	0.0914	0.3044
OPA02-3	40	0.2428	0.1828	0.2471
OPA02-4	40	0.2069	0.1657	0.1991
OPA02-5	40	0.2411	0.2196	0.0888
OPA03-1	40	0.3325	0.2789	0.1612
OPA03-2	40	0.4990	0.4949	0.0081
OPA03-3	40	0.4571	0.4571	0.0000
OPA04-4	40	0.0763	0.0706	0.0747
OPA05-1	40	0.4571	0.4571	0.0000
OPA05-3	40	0.4914	0.4692	0.0452
OPA07-1	40	0.3164	0.2060	0.3490
OPA08-1	40	0.4931	0.4828	0.0209
OPA08-2	40	0.4571	0.4571	0.0000
OPA09-1	40	0.2150	0.2050	0.0468
OPA09-2	40	0.0503	0.0483	0.0398
OPA09-3	40	0.4431	0.1126	0.7458
OPA10-1	40	0.3406	0.3146	0.0764
OPA10-2	40	0.1225	0.1189	0.0293
OPA10-3	40	0.4989	0.3521	0.2943
OPA11-1	40	0.3206	0.2491	0.2230
OPA11-2	40	0.4673	0.4556	0.0252
OPA11-3	40	0.5000	0.3041	0.3919
OPA11-4	40	0.4931	0.4828	0.0209
OPA12-1	40	0.3006	0.2815	0.0634
OPA12-2	40	0.4571	0.4571	0.0000
OPA12-3	40	0.4294	0.3743	0.1284
OPA13-1	40	0.4497	0.3828	0.1487
OPA14-1	40	0.4603	0.4319	0.0616
OPA14-2	40	0.4571	0.4571	0.0000
OPA14-3	40	0.4963	0.4642	0.0648
OPA15-2	40	0.4760	0.4191	0.1195
OPA16-1	40	0.2352	0.1778	0.2440
OPA16-2	40	0.4070	0.2813	0.3088
OPA16-3	40	0.4725	0.4657	0.0145
OPA17-2	40	0.3588	0.3314	0.0765
OPA18-1	40	0.4721	0.4363	0.0758
OPA18-2	40	0.5000	0.4491	0.1019
OPA18-4	40	0.4772	0.3864	0.1903
OPA19-2	40	0.3949	0.2914	0.2619
OPA20-1	40	0.4999	0.4864	0.0271
OPA20-3	40	0.0522	0.0464	0.1101
Mean	**40**	**0.2457**	**0.2152**	**0.1242**
St. Dev		**0.0447**	**0.0378**	

Nei's original measures of genetic identity and genetic distance (Nei, 1978) are shown in table 3.3. The majority of the loci identified, displayed shared bands among all populations and the similarity were striking even between geographically well-separated populations. Exceptions were represented by loci generated by OPA 03, 05, 10, 11 and 17 where recorded bands unique to South African, Tanzania and Oman samples were identified.

Genetic distances ranged from 0.018 to 0.097 the former value identifying the distance between Madagascar and Tanzania and the latter between Oman and South Africa.

Table 3.3 - Nei's original measures of genetic identity and genetic distance. Nei's genetic identity (above diagonal) and genetic distance (below diagonal).

Pop ID	ZA	MZ	TZ	MD	OM
ZA	****	**0.9631**	**0.9479**	**0.9255**	**0.9072**
MZ	**0.0376**	****	**0.9640**	**0.9476**	**0.9456**
TZ	**0.0535**	**0.0367**	****	**0.9767**	**0.9580**
MD	**0.0774**	**0.0539**	**0.0236**	****	**0.9815**
OM	**0.0974**	**0.0559**	**0.0429**	**0.0187**	****

In order to test the statistical significance of this variation, intra- and inter-population variances, as determined by the Analysis of Molecular Variance, was carried out. AMOVA confirmed a general lack of genetic differentiation but significantly isolated South Africa from Tanzania and Oman as well as Oman from Tanzania (Table 3.4). Genetic diversity among populations (d.f.= 4) was 6.34 %, whereas the percentage of intrapopulation variation (d.f. = 20) was 93.66.

Table 3.4 - Analysis of molecular variance of five populations of *P. indicus* (ZA: South Africa; MZ: Mozambique; TZ: Tanzania; MD: Madagascar; OM: Oman) Distance method: Pairwise difference; Significance tests (5 %) based on 1000 permutations of population pairwise F_{ST}

POP ID	ZA	MZ	TZ	MD	OM
ZA	0.00000				
MZ	0.01929	0.00000			
TZ	**0.12052****	0.00307	0.00000		
MD	0.09780	0.05322	0.01017	0.00000	
OM	**0.18803****	0.00585	**0.10819***	0.04605	0.00000

F_{ST} value was 0.063, however distance method of F_{ST}s' pairwise comparisons among population was found to be significant (p=0.008 at α = 0.05) in the case of Oman and Tanzania towards South Africa as well as Tanzania towards Oman. This separation was confirmed by UPGMA clustering of Nei's (1978) genetic distances. This clustering is graphically represented by the dendrogram in figure 3.2. In this phenogram, a cluster of northern populations (Madagascar, Oman, Tanzania) appeared relatively well differentiated from the southern populations of Mozambique and South Africa.

Increased genetic distances were significantly correlated with geographical distances in the case of South Africa towards the other populations (Z= 3.15, r = 0.64, p< 0.02; with 1000 permutation, Mantel test) (Fig 3.3).

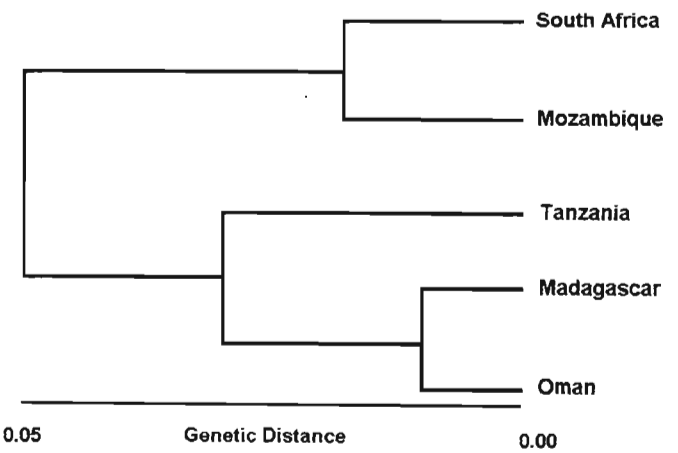


Figure 3.2 - UPGMA dendrogram based on Nei's (1978) genetic distances (NEIGHBOR procedure of PHYLIP Version 3.5).

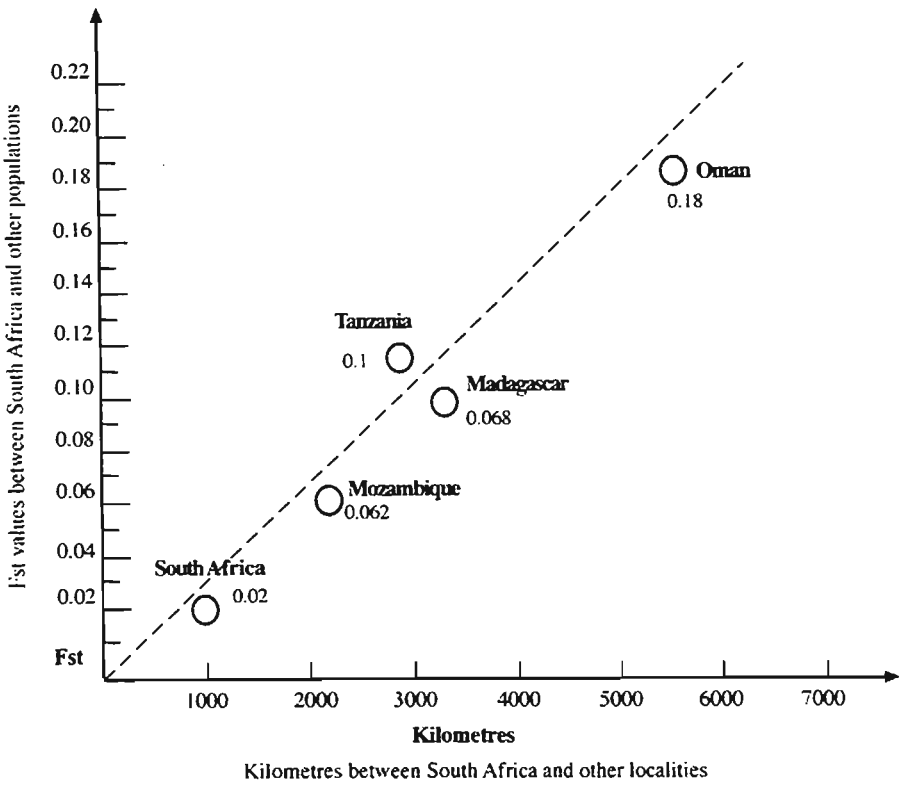


Figure 3.3- Correlation between geographical distance in kilometres (X axis) and genetic distance (pairwise F_{st}) between South African samples of *P. indicus* and Mozambique, Tanzania, Madagascar and Oman populations. Significance values were estimated with Mantel test implemented in the IBD software (Bohonak, 2002).

Table 3.5 present the values of the phenotypic diversity index ($Ø_e$) and nucleotide diversity (estimated from $Ø_e$), which ranged from 0.008 to 0.004 in Oman and South African samples respectively. It is interesting to notice the relation of the phenotypic diversity with geographical distance. Figure 3.4 depicts the variability in the five populations of *P. indicus* using the phenotypic index and nucleotide diversity.

Table 3.5 - Average phenotypic diversity, $Ø_e$ and nucleotide diversity π, from the 64 loci screened.

Pop ID	$Ø_e$	π
South Africa	0.072	0.0045
Mozambique	0.092	0.0057
Tanzania	0.087	0.0054
Madagascar	0.102	0.0063
Oman	0.134	0.0083
Total	**0.107**	**0.0066**
Stand. Er	**0.045**	**0.0012**

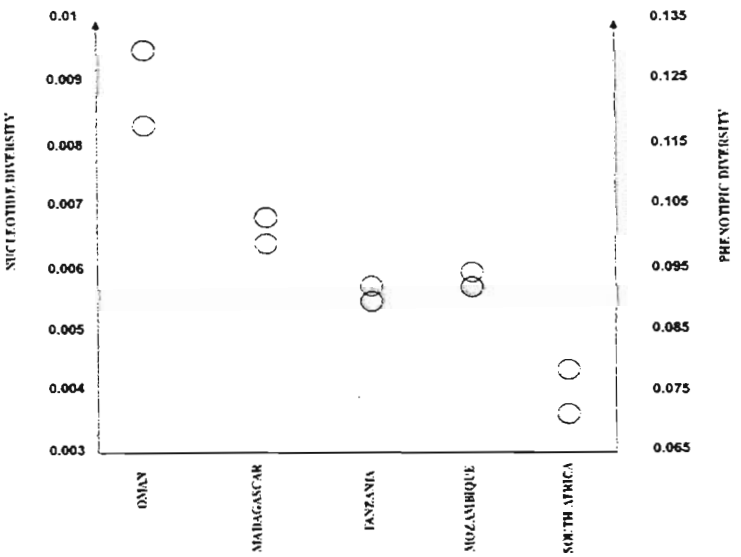


Figure 3.4 - The relation of nucleotide diversity (right axis) and phenotypic diversity (left axis) among the five populations (X axis) of *P. indicus*. Total standard error indicated in table 3.5

3.5 Discussion

RAPD is a feasible technical approach in cases where little is known about the genome of the species being investigated and allozyme variation is not informative. Highly polymorphic markers such as microsatellite loci available for *P. monodon* (Pmo 09, 27, 25) were tested for *P. indicus* in this research, but proved uninformative in spite of the efforts in attempting

various experimental approaches. Microsatellites are isolated from penaeid prawns with great difficulty and are normally characterised by large size (100 repeats) and have degenerate ends, which make designing effective primers difficult (Benzie, 1999).

In this study, RAPD profiles showed a low diversity index (Nei, 1987) among all populations surveyed with most of the bands (64) shared by most specimens, although some primers displayed enough polymorphism to indicate significant population structure. The reliability of the technique was tested with gel replicates for each primer, and 95 % of the bands were reproducible. The technique was however found to be extremely sensitive to different DNA concentrations among individuals and also dependent on the amplification efficiency of the PCR machine employed. Nonetheless, once the DNA concentrations were adjusted and an appropriate annealing temperature selected, this procedure allowed the quick identification of a number of population specific bands.

Although G_{st} (0.12) is a good measure of the relative degree of gene differentiation among sub-populations it worth noticing that it is highly dependent on the value of H_T (0.24) (Nei, 1987). The relatively high value of H_T depends on both H_S (0.21), the within population gene diversity, and D_{st} the gene diversity among population. In this research probably the low number of individuals analysed per population and the high variance of band variation within each sample inflated the values of gene diversity. However the high number of loci screened (polymorphic and monomorphic) should have been able to counterbalance this phenomenon (Nei, 1987). Although the F_{st} values, calculated with pairwise distance methods, could not be compared with those obtained from allozyme analysis (see chapter three, section 3.1.4), they were more informative revealing a significant population structure within the east coast of Africa. South African, Tanzania and Oman samples were found to be significantly different among each other ($p< 0.05$). In agreement with these results, South Africa formed a cluster that also included Madagascar, Mozambique and Oman (Fig 3.2) but excluded Tanzania. What seems inconsistent, however, is the fact that the Mozambique and Madagascar populations did not appear to be significantly different from the other populations. However such an outcome could be the result of a founder effect centred in the Zambesi basin. As previously pointed out (Forbes *et al.*, 1999), Pleistocene glaciations could have reduced not only the effective population sizes of penaeid species along the East Coast of Africa but also the availability of suitable habitats. In the Southern Hemisphere water temperatures were much lower and probably the range of distribution of *Penaeus sp.*

along the coast would have been reduced to a narrow band of latitude around the equator. Eventually the delta of the Zambesi may have represented a refuge for remaining populations of penaeid species, which subsequently spread southward and northward as the water temperature began to rise again during inter-glacial periods. Not enough time would have elapsed for mutations to accumulate and population divergence to occur even under restricted gene flow.

Another characteristic observed was the high value of the intra-population variance component (97 %). Although this is not an uncommon result of the RAPD approach (Moya *et al.*, 2001; Sebastiani *et al.*, 2001; Pearson *et al.*, 2002), this variability could be the result of different statistical as well as evolutionary components. For example, same authors have pointed out that this variation could be artifactual, irrespective of annealing temperature and even after excluding the faintest bands and individuals with rarer profiles (Perez *et al.*, 1998). However, because these markers sample along the whole genome, they may be more likely to reflect temporal genetic drift caused by differential survival of early life stages. Additionally, the large fluctuations in recruitment commonly found in marine invertebrates, may occasionally result in populations expanding their boundaries, causing considerable inter-population mixing, rather than a constant migration between demes. This intermittent level of migration is unlikely to result in temporally stable population units. Secondly, a population structure conducive to high variation, because it is composed of numerous demes, could be a plausible hypothesis for high intra-population variability levels (Wahlund effect). This phenomenon may not necessarily have been created by a precise biological factor, but also by sampling since most of the specimens in this investigation were obtained from fishing trawlers. Oceanic trawlers collect specimens from different locations that can be separated by as many as hundreds of metres and, probably, these organisms are mixed in the subsequent packaging processes.

Ultimately the estimates of π for the five populations of *P. indicus* using the averaged RAPD data was 0.0066 (± 0.0012) and seemed to be correlated with distance (Table 3.5) in agreement with the pairwise F_{st} (Fig 3.3). This result proved that the heterogeneity index can be an useful parameter of the degree of differentiation of populations especially when a low number of individuals is taken into account. A simple correspondence between geographical distance and genetic distance, such as the one recorded in this investigation from RAPD data would not necessarily mean that the populations are at equilibrium. When a

population consists of an aggregate of more or less contiguous local populations that are internally panmictic, but that do not exchange migrants with the other sub populations, an analysis of genetic structure that includes both intra- and inter population comparisons can yield a significant relationship between genetic and geographic distances, despite the absence of gene flow among populations.

Chapter four

Sequencing of Nuclear and mtDNA loci

4.1 Materials

The DNA collected from specimens of *P. indicus* from the east coast of Africa that was used for the RAPD analyses was also used for the sequencing of the Pi06 nuclear locus and for the COI mtDNA locus. A total of 25 individuals were sequenced at locus Pi06, comprising five samples from each population, while only 14 sequences were obtained for the mitochondrial survey. In addition, since it proved impossible to obtain *P. indicus* specimens from Thailand (Amornat, pers. comm.), it was decided to carry out an interspecific comparison of sister species and, therefore, 26 specimens of *P. merguensis* from Thailand were obtained. These were kindly supplied by the Marine Biology Department of the University of Songala (Thailand) in the guise of 30 sequences of COI each 558 bp long. These samples had been collected from different locations in the Gulf of Thailand and the Andaman Sea (Fig. 4.4) and four individuals of *Penaeus silasi* were also included as an out-group for phylogenetic comparison.

4.2 Methods

4.2.1 Technical Hurdles of Nuclear Sequencing

In diploid organisms, each locus of the nuclear genome can carry two different alleles on the homologous chromosomes and the major technical challenge for a researcher, is to be able to identify each allele of a particular nuclear gene. In the polymerase chain reaction (PCR), usually, both alleles of heterozygous individuals are amplified from a target locus, such that subsequent assays fail to distinguish between alternative genetic configurations. Further complications can arise if heterogeneous sequencing products are present as a result of the target locus belonging to paralogous genes with similar PCR priming sequences. It can happen, although rarely, that the PCR reaction itself generates recombinant DNA products from two amplifying alleles of a single gene in heterozygous individuals (Bradley and Hillis, 1997).

Various approaches can be employed to circumvent these difficulties (Avise, 1994). One method has broad taxonomic applicability and involves cloning of PCR products through biological vectors and designing new primers for direct sequencing of the PCR products. When analysing such products homogeneity of sequences may be taken as an indication of homozygosity, whereas lack of congruence can indicate that a heterozygous sample was cloned and amplified (Avise, 2000). Furthermore, the possibility of nucleotide misincorporation by *Taq* polymerase in the preceding PCR step is of some concern, hence several clones from each individual should be sequenced and compared in order to distinguish true allelic variants from PCR cloning artefacts (Bernardi *et al.*, 1993). Consequently, this allows one to ignore possible mis-incorporations in phylogenetic analyses (Palumbi and Baker, 1994). However, as stated by Avise (1999), this is a costly and laborious methodology and a cost/benefit assessment indicated that this would not have been a worthwhile approach for this study.

Therefore, a "poor-man" approach to the resolution of this hurdle was specifically developed for this study. This was based on the assumption that misincorporation is a rare event (Beltran *et al.*, submitted) occurring at random at a frequency of 0.001 or 0.002 (Palumbi and Baker, 1994). Therefore, the lowest frequency of incorporation errors in a short sequence of 500 bp can be estimated at one base for every sequence analysed. Because *Taq* polymerase errors are random occurrences, any two erroneous singletons have a low probability of being observed in the same position. Hence, elimination of possibly misincorporated bases from the analyses may be achieved by a more or less arbitrary elimination of unique substitutions. Therefore, in this study of five individuals per population, any single-base substitution that occurred in only one individual was interpreted as a misincorporation and eliminated from the analysis. Furthermore, a sampled specific index was calculated as a measure of diversity among the population. The assumption made in order to calculate this index was that sequences were representing both alleles in a random proportion. In this manner a private nucleotide index was defined as a measure of the frequency of the private nucleotides screened at the locus, belonging to a single population. Intuitively, a higher value of this arbitrary index would be indicative of an isolated population.

4.2.2 The Pi06 Autosomal Locus

Using DNA sequence data to study the effect of evolutionary forces on marine organisms is useful but there are several obstacles, both practical and conceptual, that will limit its applicability within the short term (Ford, 2002). The primary practical limitation for most molecular ecologists studying non-model organisms is to find appropriate genes to study. Mitochondrial genes are extensively used for sequencing because of their mutation rate and haploid manner of maternal inheritance, but in some cases the rate of evolution of these loci is not high enough to discern genetic structure of relatively close species (Baldwin *et al.*, 1998). An innovative approach as a potential source of informative genes could be that of randomly, or systematically, sampling a species genome for loci that show evidence of positive selection or genetic drift (Pogson, 1995). This author surveyed genetic variation in Atlantic cod, within and among populations, using a group of randomly cloned cDNA fragments, and found that random clones marked greater levels of inter-population diversity than allozyme loci, suggesting that major evolutionary forces acted on such sequences. Although this approach can fail to identify a particular gene functions it can, however, provide nuclear markers capable of differentiating populations (Pogson, 2001). Combined nuclear and mtDNA genealogies can help to distinguish different models of speciation and identify both, intra- and inter-specific gene flow (Wang *et al.*, 2002; Hey, 1999).

Also, RAPD markers have been successfully employed (Amornrat *et al.*, 1999) to develop sequence characterisation of amplified regions (SCARs). SCARs have been developed to increase the usefulness of RAPD markers, thus avoiding uncertainties of interpretation due to co-migrating fragments (Paran and Michelmore, 1993). In this approach, RAPD bands are excised from the gel and the fragment cloned to develop sequence specific primers that will amplify only the targeted region. SCARs have been used successfully in many studies on, *inter alia*, wheat (Hernandez *et al.*, 1999; Boora *et al.*, 1999) but never to elucidate the population genetics of marine organisms.

A rapid Kit-C (Operon Technologies, U.S.A.) was utilised, in a RAPD analysis carried out according to the method described in Section 3.3, to isolate a monomorphic band in order to create a genetic marker for identification of *P. indicus* from *P. merguensis*. Following a

protocol of Amornrat, 1999[1], a band of 500 bp was identified amongst those generated by the random primer OPAC 13. The choice of this band was dictated also by the fact that its sequence contains a small microsatellite (TA), as shown in *P. indicus* specimens from The Thai peninsula (Amornrat, 1999). Because length polymorphism might be expected in a sequence extracted from randomly sampled specimens, the result of the sequence generated from the same locus (Pi06) from African *P. indicus* could be compared with the already available sequence of Pi06 for *P. indicus* sampled from the Thailand coasts.

This methodology consists of performing a standard RAPD procedure followed by the identification of a single monomorphic band of an arbitrary molecular weight. Thereafter, elution of this band from the gel is carried out and, in turn, followed by cloning as a means to obtain the full sequence of the band (SCAR). From this sequence, band-specific primers are designed and amplification and sequencing of the SCAR can be performed (Amornrat *et al.*, 1999).

The amplification reaction was performed using the following primers: clone Pi06 forward 5' – ACT CTC TCC TTC GAC TCT and clone Pi06 reverse 5'- TGC GTA AAT AAT CAT CAG GA – 3'. The PCR reaction was performed in a 12.5 µl reaction volume containing: 50 ng of genomic DNA, 200 µM of each dNTPs, 2.5 mM $MgCl_2$, 0.5 µM of forward and reverse primers and 0.05 U of *Taq* (Bioline, U.K.). The amplification reaction was performed in a Hybaid thermal cycler under the following condition: denaturation at 94 °C for two minutes, followed by 35 cycles of 94 °C for 60 seconds, 57 °C for 30 seconds and 72 °C for one minute. A final extension step of 5 minutes was allowed at the end of these 35 cycles.

PCR products were separated on 1.4 % agarose gel electrophoresis. The target band of 500 bp was excised from the agarose gel using a sterile scalpel blade and purified, using a QuiaGen Quick agarose gel DNA purification kit (Quiagen). The filtrate was tested again in a one % agarose gel in order to determine the purity and the concentration of the DNA amplification products. Sequences were obtained using an ABI automated 370A sequencer (see Appendix VIA for an example of an ABI chromatogram). Both, forward and reverse strands were sequenced in order to confirm the matching of double-stranded sequences although, in some cases, results from either forward or reverse strand sequencing

[1] A rapid Kit-C (Operon technologies, U.S.A.), was employed in order to isolate a monomorphic band of a minimum length of 500 bp to create a genetic marker for identification of *Penaeus indicus* versus *P. merguensis*.

only was used. In order to compare the validity of the assumption made in the statistical analysis of the locus, three randomly chosen *P. indicus* specimens from Tanzania, South Africa and Oman respectively were cloned using PGem-T easy vector kit (Promega). The 500 bp purified band was cloned in competent *E. coli* cells (protocol one, Appendix VIB), the plasmid containing the insert was recovered from a total number of six clones and sequenced with an ABI PRISM 310 automated sequencer (Molecular Biology Department University of Cape Town). A GENBANK search using BLAST was carried out for Pi06 sequences of African *P. indicus* in order to test for possible relationships with coding regions of known mitochondrial and autosomal genes. As may be expected from this kind of random approaches, no matches were found, as also reported by Amornarat (1999). Sequences obtained were aligned and edited with the programme BIOEDIT 4.7.0 (Hall, 2001) removing terminal regions that displayed either repeats or insertions that could hinder alignment. Subsequently, nucleotide diversity using a Jukes and Cantor (1969) approach, D_{xy} (average pairwise number of nucleotide differences) were calculated using the software DNAsp version 3.5.3 (Rozas and Rozas, 2001) and graphically presented in figure 4.1 (page, 64). A homology index was calculated, after re-alignment with a fast algorithm (Higgins and Sharp, 1988) using DNAman 4.13 (Lynnox Biosoft, 1999). Nucleotide diversity index π and average number of nucleotide substitutions per site between populations were plotted as a pairwise comparison between populations. Furthermore a private nucleotide index was arbitrarily defined (see 4.2.1). This index represented the number of nucleotides along the sequence of Pi06 that were unique to any single population. Subsequently, a graph was plotted representing the frequencies of private and shared nucleotides among the samples. This empirical methodology allowed the identification of populations with a higher percentage of private nucleotides, a finding that may be expected under conditions of limited gene flow. Ultimately values of nucleotide diversity extrapolated from each population were compared with data obtained from mtDNA markers to ascertain the reliability of the nuclear data set and eventually to confirm the potential of this methodology as a low cost-benefit ratio approach for the inference of population genetic structure.

4.2.3 The Cytochrome Oxidase I mtDNA Locus

It is not always understood that a failure to reject a null hypothesis does not mean that hypothesis is true. In the case of penaeids, there could be a single panmictic population or there could be a multitude of separate meta-populations exchanging just sufficient individuals

to prevent the detection of population differentiation. In marine invertebrates with relatively high effective population sizes, a few hundreds animals each generation will prevent such differentiation unless the sampling effort is intense or specific molecular markers are employed. Mitochondrial DNA sequences have been widely used to reconstruct intraspecific phylogenies and determine relationships between closely related species (Carr and Marshall, 1991; Park *et al.*, 1993; Baldwin *et al.*, 1998). The mtDNA genome typically shows a rapid rate of evolution and short coalescence times, which means that phylogenies are often well resolved, even between recently separated populations and species complexes (Avise, 1994). For these reasons, and in the light of the previous results, it was decided to apply a mitochondrial gene sequencing approach to the present study so as to be able to combine nuclear and mtDNA genealogies. Identification of intra-specific gene flow or finer scale population structure within *P. indicus* was tested with the mitochondrial locus Cytochrome Oxidase I (Wang *et al.*, 1997)

The cytochrome oxidase subunit I (COI) was sequenced from 15 specimens of *P. indicus*. An attempt to sample *Penaeus indicus* from Thailand was made, but no individuals of this species could be found, although its range is considered to extend from the African coasts to Australia. However, *P. indicus* is typically difficult to separate from *P. merguensis* and other *Penaeus* species based on morphological evidence because of the presence of several sibling and cryptic species in Thai waters, (Amornrat P., pers. comm.).

The primers were especially designed after alignment of COI sequences from *P. monodon*, *P. merguensis* and African *P. indicus*. PCR primers were: 5' CAA CAT TTA TCT TGA TCT TTT GG 3' and 5' TCC AAT GCA CTA ATC TGC CAT ATT A 3' . PCR conditions were as follows: one denaturing step at 95 ^{0}C for 4 min. This was followed by 35 cycles of these three steps: 94 ^{0}C for 30 sec., 45-50 ^{0}C for 30 sec. and 72 ^{0}C for 1 min. which were followed by one cycle of 72 ^{0}C 10 min. This resulted in the amplification of a 464 bp region of the COI gene. PCR reactions were carried out in 25 µl volumes each containing 20 µM primer, 1.25 mM dNTPs, 10x buffer, *Taq* polymerase (Bioline), 50 ng of template DNA, and purified water (Milli-Q, Millipore). As in the case of Pi06 locus, PCR reaction products were separated on 1.3 % agarose gels (Sigma). The band of interest was eluted from the gel, filtered through a 20 µl yellow disposable tip inserted in a 1.5 ml Eppendorff tube and then centrifuged for 1 minute at 12 000 RPM. The filtrate was then resuspended in 20 µl of TE buffer. Previous to sequencing 3 µl of solution were analysed on 1.4 % agarose gel and the

concentration determined by a densitometric measurement of band intensity and comparison with a standard DNA concentration. Sequences were obtained using an ABI Prism 310 automated sequencer. Both, forward and reverse strand sequencing were carried out for confirmation of correct sequences.

Sequence alignments were performed manually as described above using BIOEDIT (version 5.0.6) and DNAman (version 4.13) for subsequent analyses. A dendrogram was constructed using Neighbour Joining method (Saitau and Nei, 1987). Pairwise exact tests between population were performed using haplotype frequency data following AMOVA (Excoffier, 2001). An analysis of molecular variance (AMOVA) was carried out using the package ARLEQUIN (Schneider *et al.*, 2000) to test for population genetic structure. The AMOVA (Excoffier *et al.*, 1992) programme in ARLEQUIN was used to calculate the θ_{st} (analogous of $F_{st)}$ and to perform a hierarchical analysis on two arbitrarily assigned, different groups. The first group (Southern group) comprised sequence data from South Africa and Mozambique, while the northern group included sequences from Oman, Tanzania and Madagascar samples. Tajima's test of selective neutrality, Tajima's D (Tajima, 1989a), was calculated to assess evidence for population expansion (Tajima, 1989b). The distribution of the observed pairwise difference and mismatch distribution (Rogers and Harpending, 1992) were plotted using the program DNAsp (version 3.53) in order to infer the occurrence of past demographic expansion in finite populations. In the case of COI sequences from *P. merguensis* from Thailand, the AMOVA was performed on a west-coast group comprising the Ranong, Puket and Satun localities while to the east coast group belonged the sample populations of Trad, Songala and Surat.

4.3 Results

4.3.1 Pi06 autosomal locus

The location of the Pi06 sequence on the genome of *Penaeus indicus* is unknown. Sequence comparison with GENBANK stored data, failed to recognise any similarity with already cloned chromosomal genes or mitochondrial genes, which is not unexpected.

Not all sequences resulted in an interpretable chromatogram, probably due to the amplification of paralogous genes with similar annealing primer sites. A total of 25 sequences displayed a sufficiently clear pattern for interpretation. Five of these sequences came from individuals from each of the sampling locations: South Africa, Mozambique, Tanzania, Madagascar and Oman. The sequences of the four clones obtained from two specimens from South Africa and Tanzania are presented in Appendix IVC. In all the sequences obtained from each clone the Pi06 insert showed only one segregating site at position 138 bp, probably due to misincorporation of *Taq* polymerase. This is consistent with the previous assumption of an average of two mis-incorporations every 1000 bases reported by other authors (Section 4.1.1). Variation in the length of sequenced products was not observed in the short microsatellite $(TA)_8$ discovered from position 330, except in one case, for one Tanzanian sample. Insertion of sequences, such as CATAG in one sample from South Africa in position 72 and TCTTA in one Oman specimen, were observed. In all Tanzanian specimens a region of insertion was found at position 25. These regions caused alignment difficulties, especially because of length variation, and were eliminated in order to maximise group alignment. All sequences were aligned to form a 400 bp data matrix (Appendix IVD). Nucleotide composition (relative value) was 11.65 % for C, 45.30 % for T, 30.15 % for A and 12.90 % for G. The average number of polymorphic sites was 24.6 and there was a bias towards transversions, with 14 transversions found in South African samples and 11, 23, 16, and 16 respectively, for Mozambique, Tanzania, Madagascar and Oman. Although length variation was observed in some specimens, species-specific differences were determined by fixed polymorphisms. The nucleotide diversity index (π) ranged from 0.026 between South Africa and Mozambique to a maximum value of 0.048 between Oman and Madagascar. Comparison of sequence alignment between spatially related samples indicated a high degree of identity between South Africa and Mozambique (98.89 %), with a highly conserved region between 60 bp and 180 bp (Appendix IVD). When comparing specimens from Tanzania with those from Mozambique and South Africa, the degree of identity was lower (97.57 %). Comparison between Madagascar specimens and samples from South Africa and Mozambique indicated again a value of 97.75 %. In the case of Oman, comparison with Madagascar and Tanzania populations was of the same magnitude as that of other comparisons (97.52 %). When testing an isolation by distance model, an increased similarity index of almost one percent was noted when comparing Oman samples with those from South Africa and Mozambique. When this data set was compared with a published sequence of *P. indicus* collected in Thailand (Amornrat, 1999) one could identify a region of deletion (between base 90 to 102) and region

of insertion (between base 140 to 150). It is noteworthy that all sequences reached a value of 97 % homology from 150 to 460 bp. The pairwise comparison with the east Africa data set found a total of 25 segregating sites. Furthermore, the pairwise nucleotide diversity index was plotted together with the average number of nucleotide differences (D_{xy}), for each population pair and this is graphically presented in figure 4.1.

As reflected in the graphs, nucleotide diversity indexes (π) plotted with nucleotide position are very similar in the case of the pairwise comparison of South Africa versus Mozambique

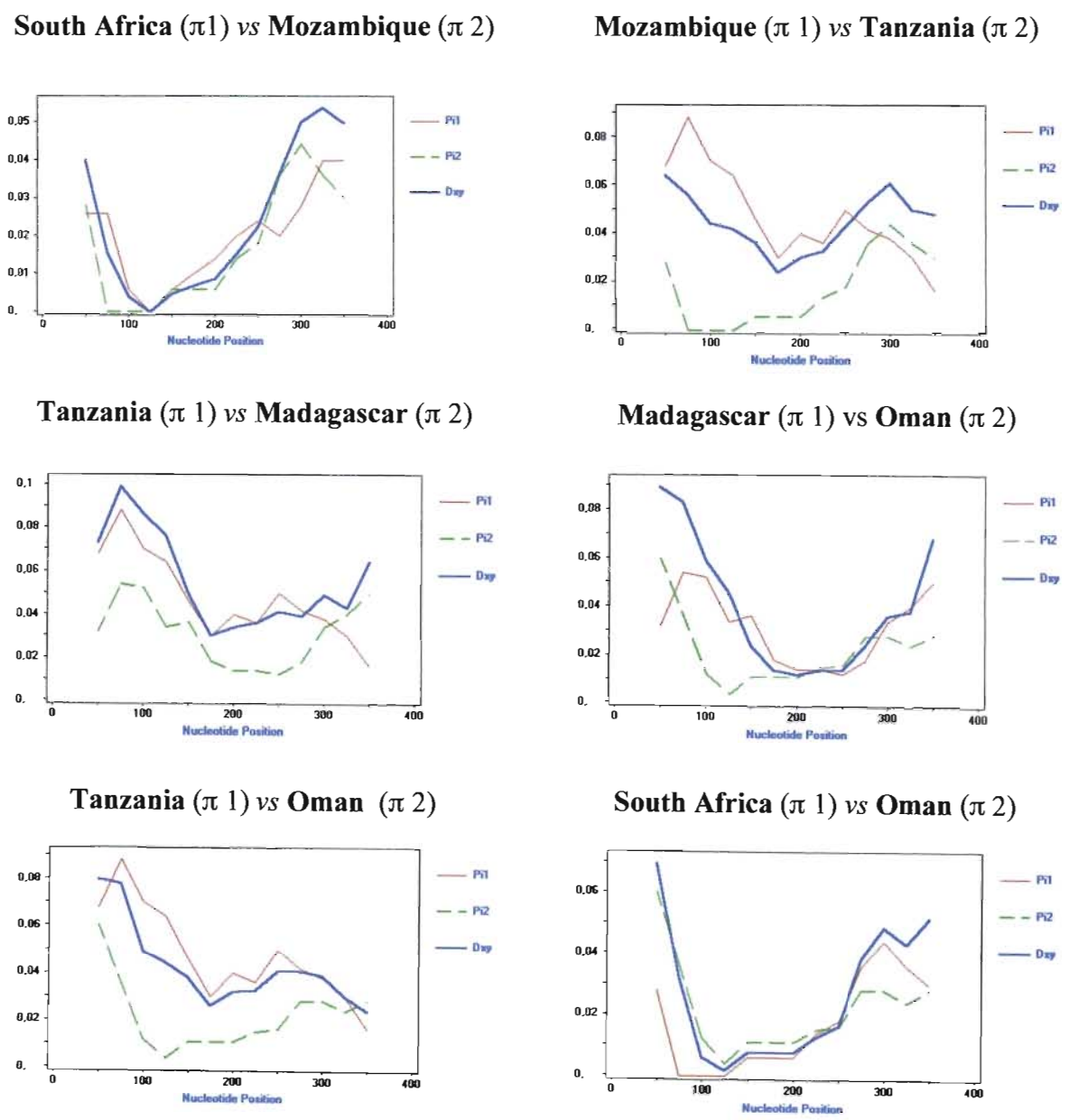


Figure 4.1- Nucleotide diversity index (π) (Y-axis) and nucleotide position (X-axis) in pairwise comparisons of the five populations of *P. indicus*. Blue line: D_{xy} along the Pi06 locus; Red line: P1 (π) nucleotide diversity of Population 1; Green broken line: P2 (π) nucleotide diversity of Population 2.

The private nucleotide index represented graphically in figure 4.2 indicated a higher frequency of private alleles for Tanzanian (12) and Madagascar samples (9).

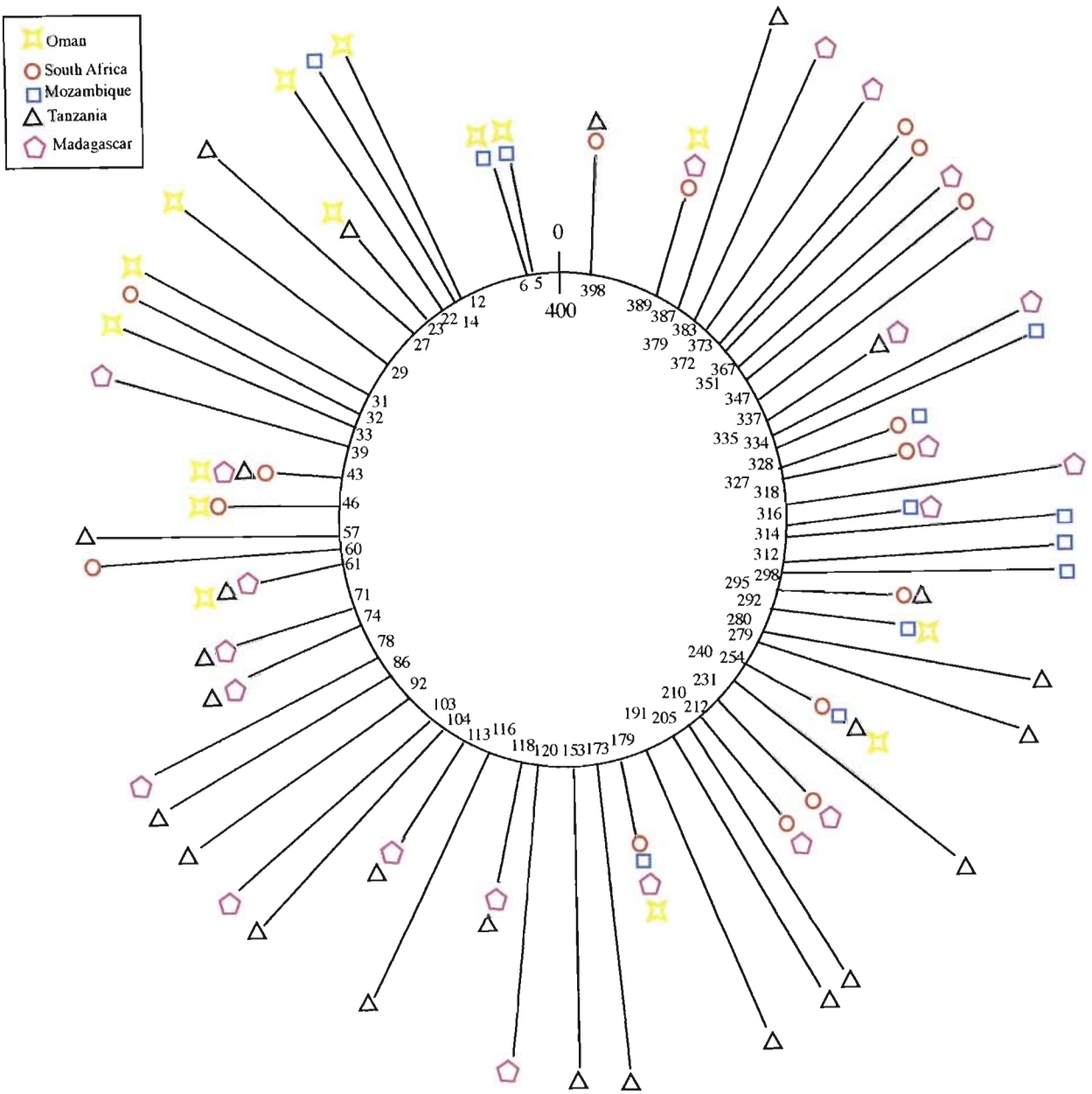


Figure 4.2 - Private nucleotide index along the 400 bp sequence of Pi06 locus from the five African populations of *P. indicus*. The circle represents the sequence of the locus and the lines show the position of private and shared nucleotides along the sequence.

4.3.2 Cytochrome Oxidase I

4.3.2.1 *Penaeus indicus* from the African coast

A total of 15 sequences of the COI mtDNA locus were analysed and their total alignment for these two species of *Penaeus* is presented in Appendix IVE. As previously mentioned, the difficulty in obtaining a reliable morphological identification of *P. indicus* from Thailand suggested that it may be profitable to collect penaeid samples from this area although *P. indicus* is thought to be totally absent from Thai waters (Amornrat P., pers. comm.).

In east African samples (South Africa, Mozambique, Tanzania, Oman) the 464 bp portion of this gene showed an overall lack of insertions or deletions and contained four variable sites included in the region 270 - 462 bp (Appendix IVE). There was no sequence variation at the second codon position. Average nucleotide composition was 17 % for C, 25 % T, 17 % for G and 39 % for A. A high AT content (average 184-A and 117-T) was consistent with descriptions of other arthropod mtDNA sequences (Spicer, 1995) as well as other *Penaeus* mtDNA sequences (Garcia *et al.*, 1996). There were two polymorphic sites for transition in South African samples and three and one in Mozambique and Oman respectively.

AMOVA analysis performed on pairwise F_{st} (Jukes and Cantor distance method, 1969) using single population sequences failed to reveal any significant structure. Thus a second analysis was performed. The AMOVA was used to analyse two different groups: South Africa and Mozambique specimens were pooled together in a single group arbitrarily named South samples. Tanzania, Madagascar and Oman specimens were pooled to define a second group named North samples. This type of pooling technique was performed to enhance the power of the analysis since in some cases (Madagascar) only two sequences were used. The average pairwise population F_{st} value (Jukes and Cantor, 1969) was 0.082 and was found not to be significant, after testing 1000 permutations (MCM), $p = 0.092$; $p < 0.05$. AMOVA estimates of within group diversity were very high as previously detected by RAPDs and reached a value of 91 %. Nucleotide diversity π calculated over all sites, was 0.0028 and the nucleotide sequence divergence ranged from 0 % within population to 0.26 % among population (identity index). The total divergence between populations was very low since it ranged from

a minimum of 0.0022 when Oman was compared with the samples from South Africa, Mozambique and Tanzania, to 0.0043 among the latter populations. The total nucleotide diversity index (π) from East Africa and Oman was 0.0032. It is noteworthy that although the number of sequences per sample was low and haplotype frequency does depend on samples size (Nei and Kumar, 2001), there was again a particular pattern in the shared haplotypes. The South group samples shared one haplotype (H7) with the north group comprising Oman, Madagascar and Tanzania (H11, H12). The values of the average number of nucleotide substitutions per site pairwise (D_{xy}) and between groups (D_a) were 0.0046 and 0.00044 respectively. The mismatch distributions of the pairwise comparison between the two groups are presented in figure 4.3 and show a distinct unimodal pattern of distribution.

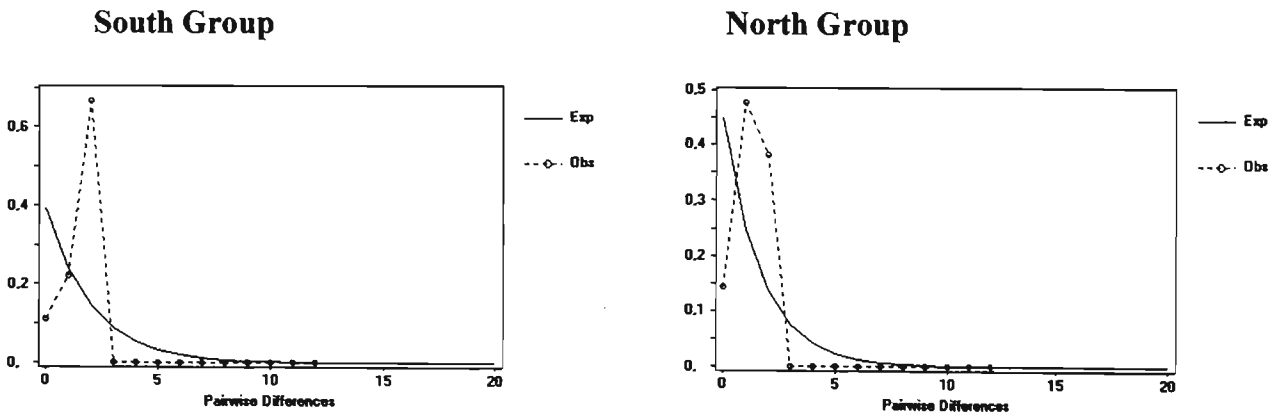


Figure 4.3 - Mismatch distribution (dotted line) for each of the two African groups of *P. indicus*. Expected curve under a constant size coalescent model of population size (solid line). Pairwise differences on the abscissa and relative frequencies on the ordinates.

The value of Tajima's D (Tajima, 1989a) was – 0.65 and was not significant in the south group, whereas in the north group this parameter had a value of 0.05. However, in both cases no statistical significance was found ($p > 0.01$). When comparing haplotype diversity (h) and nucleotide diversity (π), in both groups there were high values of h (0.83 south group; 0.72 north group) and extremely low values of π (0.0033 south group; 0.0026 north group).

A comparison between the nucleotide diversity of the Pi06 locus with that of the COI locus was also carried out and the values are shown in table 4.1.

Pop. ID	h (COI)	π (COI)	π (Pi06)
South Africa	0.83	0.0033	0.026
Mozambique	0.83	0.0033	0.026
Tanzania	0.72	0.0026	0.046
Madagascar	0.72	0.0026	0.046
Oman	0.72	0.0026	0.046
Total	0.79 +/- 0.64		

The values of nucleotide diversity are both very small for COI and Pi06 although the latter is an order of magnitude higher.

4.4.3.2 *P. merguensis* from Thailand

Sequences from this species were analysed according to two groups: an east coast group, from the Gulf of Thailand, and a west coast group, from the Andaman Sea (Fig 4.4).

Figure 4.4 – Map showing the collecting sites of *P. merguensis* samples in Thailand.

The number of polymorphic sites ranged from 10 to 36, (see Appendix IVE). The number of transitions was higher than the number of transversions in all samples (6/3, Ranong; 4/1

The number of polymorphic sites ranged from 10 to 36, (see Appendix IVE). The number of transitions was higher than the number of transversions in all samples (6/3, Ranong; 4/1 Puket; 12/2 Satune; 6/4 Songala; 23/6 Trad and 5/1 in Surat), and the nucleotide composition was consistent with the *P. indicus* sequence data from the present study. Thai specimens showed a percent of AT higher than that of GC (C: 19.61 %; T: 35.36 %; A: 27.80 %; G: 17,23 %). Population structure using the Jukes and Cantor (1969) distance method of F_{ST} estimation with AMOVA was calculated on sequences from the Gulf of Thailand and the Andaman Sea group, and its value was 0.27. The total within samples variation was 72 %, whereas the remaining 28 % was due to among populations structuring (Table 4.3). Furthermore, F_{ST} pairwise comparison (1000 permutations, MCM) failed to recognise a consistent significant difference, at a 0.05 *p*-value, between the two groups. However, the west coast group was significantly different from Songala. On a small geographical scale, significant structure was recorded also between Surat and Songala (Table 4.2).

Table 4.2 - F_{st} pairwise values and significance of AMOVA among samples of the Gulf of Thailand and the Andaman Sea (west coast comprising sequences of Puket and Ranong). Distance method: Jukes and Cantor Matrix of significant F_{st}. ** Significant *p* values at a 0.05 significance level.

Pop ID	Westcoast	Songala	Trad	Surat
West coast	0.00000			
Songala	**0.37225****	0.00000		
Trad	0.24412	0.06382	0.00000	
Surat	0.10023	**0.50702****	0.38444	0.00000

Nucleotide diversity was 0.030 and 0.037 for west and east group respectively and again haplotype diversity was high 0.76 and 0.81 respectively. Furthermore, Tajima's D was not found to be significant at $p<0.01$ or 0.05 and had a negative value (-0.35) in the east group, while it was positive (1.16) in the west group. Contrary to the analysis of east African samples of *P. indicus*, the mismatch distribution in *P. merguensis* displayed an uneven distribution of pairwise differences compared to the expected model graph (Fig 4.5).

West Group: Andaman sea **East Group: Gulf of Thailand**

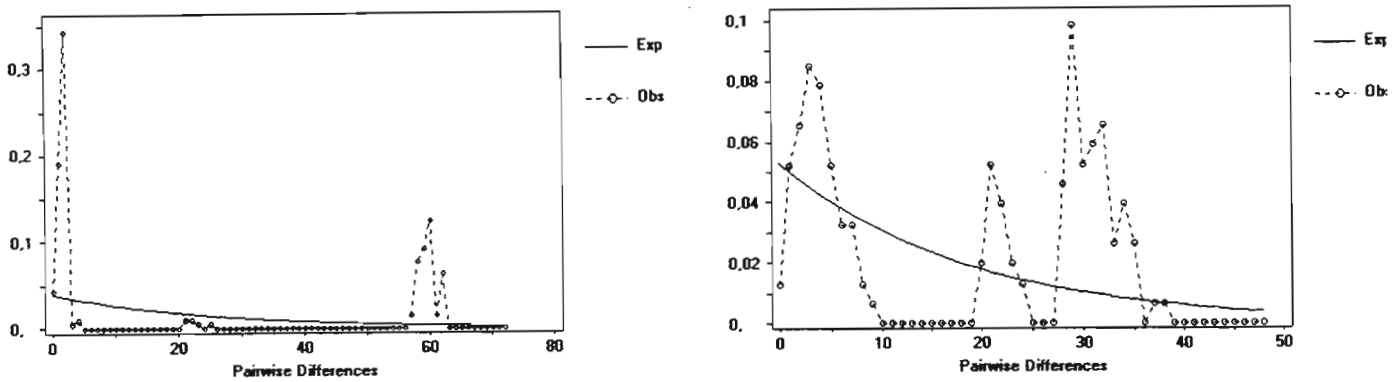


Figure 4.5 - Mismatch distribution for each of the two groups of *Penaeus merguensis* in the Gulf of Thailand and the Andaman Sea.

In this case the values of nucleotide diversity (π) for the COI mtDNA locus were an order of magnitude higher than the those recorded for *P. indicus* African samples, whereas the haplotype diversity (h) was similar (Table 4.3).

Table 4.3 - Haplotype diversity (h) and nucleotide diversity (π) recorded from COI mtDNA locus for the *Penaeus merguensis* groups (West and East coast of Thailand).

Pop ID	h (COI)	π (COI)
West Coast	0.76	0.030
East Coast	0.81	0.037
Total (SE)	0.78 ±0.58	

4.5 Comparison of results between *P. indicus* and *P. merguensis*

The data set comprising the COI sequences of *Penaeus indicus* was pooled into a single group and comparisons were made with the West and East coast groups of *P. merguensis*. Nucleotide diversity in the all-African samples was 0.0032: an order of magnitude lower than that of pooled Thailand sequences (0.052). Haplotype diversity was very similar, being 0.77 in *P. indicus* and 0.78 in *P. merguensis*. The value of Tajima's D was negative (-1.32) in the African sample, although this was not statistically significant, and the mismatch distribution followed a pattern similar to that of the African pooled group. The tree obtained using the neighbour-joining tree-construction method, included sequences downloaded from GENEBANK for the species *P. monodon* (outgroup) and *P. silasi*. This dendrogram indicated a strong differentiation of *P. indicus* from *P. merguensis* with 100 % bootstrap support and suggested a closer genetic relationship between *P. merguensis* and *P. silasi* (Fig 4.6). There was no bootstrapping support for strong differentiation between African populations. A stronger diversification is found in the *P. merguensis* samples but the branching was not consistent with the sampling locations. The haplotypes from Ranong (RNW), from the west coast of Thailand, are grouped with samples collected from Suratthani (SRE) in the Gulf of Thailand. The same pattern is evident in the first group (A) where samples from Trad (Gulf of Thailand) are grouped together with Satun (STW). The most striking feature is the higher level of population structuring among the samples of *P. merguiensis* compared to that of African *P. indicus*.

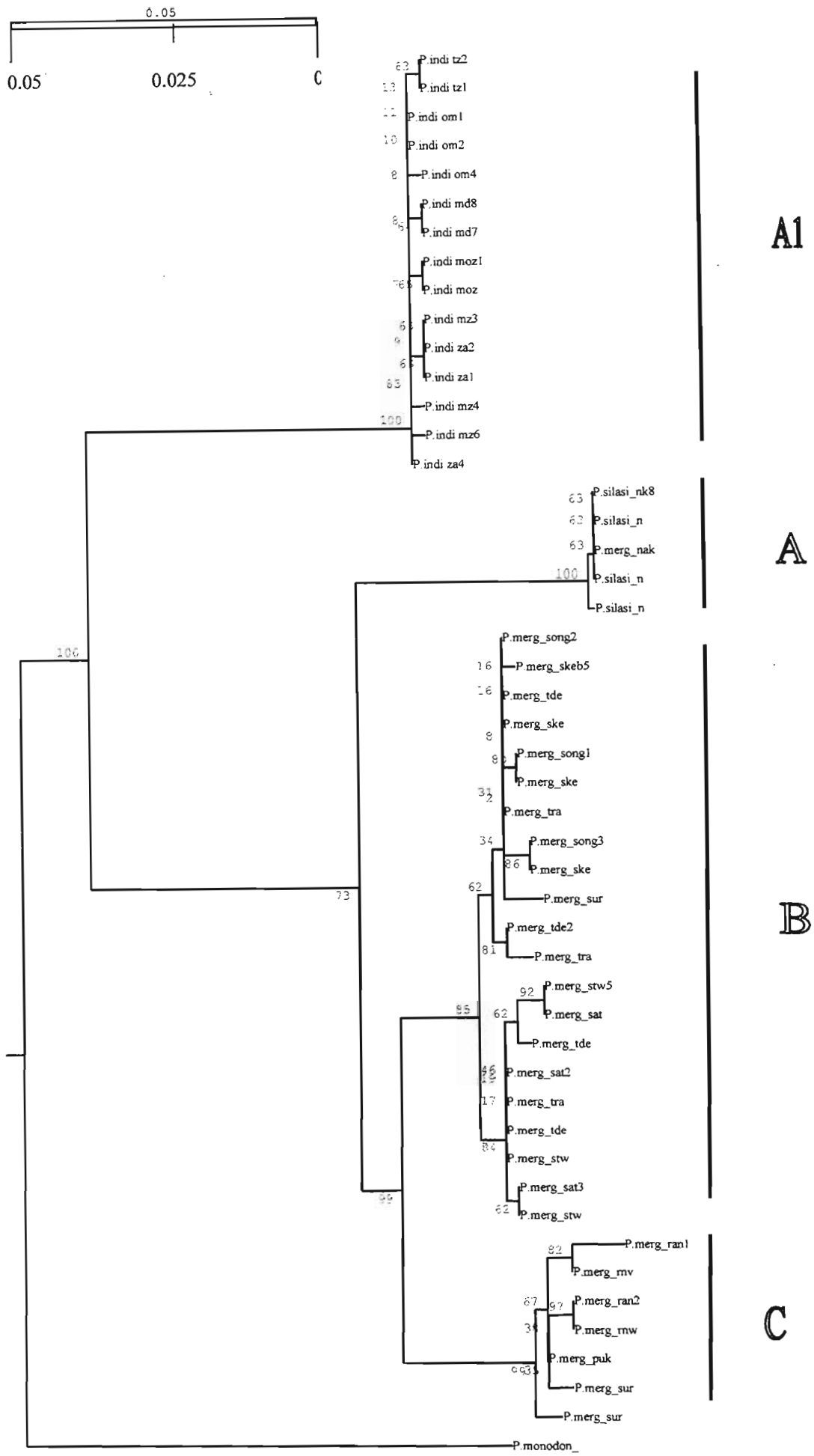


Figure 4.6- Neighbour joining tree (Jukes and Cantor distance) of *P. indicus* populations from East Africa and *P. merguensis* and *P. silasi* from Thailand with *P. monodon* as out-group. Values at the nodes represent bootstrapping estimates.

4.6 Discussion

The nuclear locus (Pi06) revealed a discrete pattern of genetic differentiation among populations of *Penaeus indicus*. Sequence alignment showed that interpopulation nucleotide diversity indices were lower than the diversity between the group composed of South Africa and Mozambique, on the one hand, and Tanzania and Madagascar group on the other. Nevertheless, all sequences were highly conserved within the region 150-460 even after including the only published sequence (Amornrat *et al.*, 1999) of Pi06 from *P. indicus*. This result is quite interesting in the light of the fact that downstream and upstream of this conserved region, differences among populations were high, especially between African and Thailand samples. Selection could be a possible explanation, but no particular function has been attributed to this locus at present. The private nucleotide index as a measure of population differentiation has no statistical properties, but is a simple frequency measure that can empirically ascertain if there is gene flow restriction between populations at any given locus. The values recorded for this parameter and presented in Fig 4.2 revealed a higher frequency of private nucleotides for the populations from Tanzania and Madagascar. However, it is difficult to explain such a pattern from an ecological point of view. Although oceanographic data are not readily available, one might suppose that Tanzania has been isolated due to ocean-current patterns that retain larvae in that area. It may be suggested further, that similar current patterns do not occur in Mozambique and South Africa and it is probable that the distance separating these two localities is not sufficient to prevent gene exchange.

Although analyses of nuclear genes are influenced by different technical constraints (see introduction), this study made an attempt to reduce the cost of procedures normally employed to obtain reliable results from nuclear genes. It is important to note that this approach has been developed to test the possibility to deploy SCARs as a practical and novel technique to discern variability among populations of the same species. Noteworthy is the results of the pairwise comparison analysis that underlined the similarity of the Oman population with those from South Africa and Mozambique. This result is discordant from previous analyses based on allozymes and the RAPDs profile (see chapter 3). Both markers showed a significant degree of differentiation between South Africa and Oman, at least at some loci. However, a polyphyletic relationship among recently isolated populations is

often caused by the persistence of ancestral polymorphisms, but similar genealogical patterns can be caused by moderate gene flow over longer divergence times (Wakeley, 1995). Furthermore, when population genetic analyses are performed on a wide range of loci, it is possible that unexpected results may appear, especially when dealing with organisms with high dispersal rates. As discussed by Slatkin (1985a), if the average number of generations needed by a local population to become extinct is less than the effective number of breeding adults, extinction and re-colonisation would prevent the genetic differentiation of this local population by means of evolutionary forces such as genetic drift. Screening more than one locus for polymorphism, affords a greater probability to highlight possible restrictions of gene flow having occurred in more recent evolutionary times since the last re-colonisation event.

COI variation of *Penaeus indicus* was low (mean nucleotide diversity 0.0045) and only three haplotypes were recorded. No significant difference was present when analysis was performed by population haplotypes or when haplotypes were pooled into North and South groups (see page 63). The high value of the within population variance component might be an artefact of the low number of sequences utilised in this investigation, despite the pooling procedure adopted. However as previously discussed in chapter tree (page 55), this could also be a result of dealing with structured populations collected within the same sampling area. Nevertheless, the lack of spatial variation evidenced does differ from the results obtained from the analysis of the Pi06 locus.

Another intriguing feature revealed by the analysis of mtDNA sequences, was the mismatch distribution pattern. This methodology showed graphically the presence of an unimodal distribution of haplotype diversity for east African *P. indicus*. Looking a the graph (Fig. 4.3) and following the mutation model of Roger and Harpending (1992), the drop of the curve occurs because, following a population expansion, the mean pairwise difference increases much more rapidly than its standard deviation. On the other hand, Aris-Brosou and Excoffier (1996) have shown that an uneven mutation rate model, a diametrically opposed hypothesis, could also lead to an unimodal distribution of pairwise differences. This latter hypothesis was also tested by comparing the values of π and h. Nucleotide diversity (π) is a weighted sequence divergence between individuals in a population. Haplotype diversity (h) condenses information on the numbers of different alleles at a locus, regardless of their sequence relationships (Grant and Bowen, 1998).

Intuitively, loss of nucleotide diversity compared with high values of haplotype diversity suggests rapid population growth from an ancestral population with small effective population size, provided the time was sufficient for recovery of haplotype variation via mutation, yet too short for an accumulation of large sequence differences (Avise, 2000). Nucleotide diversity was very low when compared with haplotype diversity, an indication of rapid growth in East African populations starting from a very low N_e. Such kind of population history is not unique for this part of the Indian Ocean. For example, Benzie *et al.* (2002) calculated divergence times of several populations of *Penaeus monodon* in the IWP region based on RLFP mtDNA haplotypes and suggested that two population expansions might have happened along the east coast of Africa around 22 000 and 43 000 years ago. As will be discussed in chapter five, this could have drastic effects on the effective population size and, hence, on the effect of genetic drift towards the gene frequency of these penaeid populations.

The COI sequences of *Penaeus merguensis* collected from the Gulf of Thailand were very different from those of African *P. indicus*, although AMOVA failed to recognise a consistent difference between samples collected in the Andaman Sea and those from the Gulf of Thailand (Table 4.3). The west coast sequences were significantly different from those from Songala and within the Gulf of Thailand and significant population differentiation was recorded between Surat and Songala. This pattern of spatial variation was not consistent in all samples, probably due to the small number of sequences utilised for the analysis, although hybridization among sympatric species among regions could have resulted in an homogenisation of gene genealogy. The haplotype diversity and nucleotide diversity (Table 4.4) were more consistent with a hypothesis of stable population sizes with large and long-term N_e. In this case too, the mismatch distribution was congruent with that recorded from the above-mentioned index (h and π) and, following Roger and Harpending (1992), showed an uneven distribution consistent with a lack of population expansion. An alternative hypothesis to explain these findings, would be the possible admixture of samples, derived from historically sundered populations.

It is noteworthy that the average surface of the continental shelf in the Gulf of Thailand is very different from the rest of the western IWP and the east coast of Africa. During the Pleistocene, lowered sea levels connected much of SE Asia, New Guinea and Australia,

almost closing the sea connection between the Indian and the Pacific Oceans. These tectonic changes might have restricted gene flow and created a sundering barrier which, ultimately, lead to genetic differentiation among several marine species (Williams and Benzie, 1996, 1997, 1998; Nishida and Lucas, 1988; Mc Millan and Palumbi, 1995). These results, although preliminary because based on a relatively low number of sequences, are consistent with other reports that surveyed the extremes of other species ranges (Klinbunga *et al.*, 1998; Benzie *et al.* 2000, 2002; Forbes *et al.*, 1999). It is possible that, historically, large numbers of semi isolated populations in the highly geographically structured IWP region generated, or at least maintained, the diversity in Southeast Asia. As mentioned in the previous paragraph, surveys of mitochondrial DNA variation in the tiger prawn, *Penaeus monodon*, using restriction fragments length polymorphisms (Benzie *et al.*, 2002), have provided clear evidence that the Indo West Pacific region (IWP) is a site of accumulation of genetic diversity rather than a site of origin of diversity as stated by Briggs (1999). In the report by Benzie *et al.* (2002) no shared haplotypes were found in common between group of south east African, Australian and south east Asian populations. Therefore, these authors suggested that southeast African samples could have gone through a relatively recent population expansion and further analysis allowed them to conclude that an increase in population size occurred approximately 43 000 years ago. Their observations provide evidence for the evolution of genetic variants in peripheral populations in one of the oceans and their subsequent migration into the East Asia region. An important factor common to the present investigation is the fact that the least diverse populations were found at the periphery of the species range. Possibly the bimodal shapes for both groups in the Gulf of Thailand and the Andaman sea are consistent with two evolutionary scenarios: the first one suggests a lack of population expansion in these populations and the second one indicates a possible population mixing.

The results presented in the phylogenetic tree (Fig 4.6), show a different population structuring between the two penaeids species. Following the genealogical concordance rules stated by Avise (2000), if co-distributed species with comparable natural histories or habitat requirements are phylogenetically structured in different fashions, this is not be a reflection of shared historical elements in the evolutionary or ecological factors. *Penaeus indicus* and *P. merguensis* are closed related species that happen to share similar environments and life cycles (Dall *et al.*, 1990). This issue therefore, draws attention to different evolutionary forces acting on these two species. The genetic variation of *Penaeus*

sp appears to be lost along the way to East Africa. Presently the only sampling gap is the Indian peninsula which moved away from Africa five million years before present and efforts are under way to obtain samples of *Penaeus indicus* from the east and west coasts of India. If the genetic break of diversity is within the range of SE Asia, we expect to see a gradual lowering of genetic diversity between the two Indian coasts, with more recent haplotypes along the east coast of India. Two alternative hypotheses can be proposed to explain the genetic diversity registered at the COI locus for the two species of *Penaeus*:

(i) The COI locus could be under selection by environmental pressure.

(ii) The lower genetic diversity in East Africa may reflect bottlenecks sometimes in the last few millions years or founder effects following re invasion after extinction of peripheral populations during the last ice age (Benzie, 2000).

This explanation contrasts with the previous one of Mulley and Latter (1980), although also in this case a bottleneck effect is hypothesised, the process that generated it is considered to be strictly biological (speciation) and not environmentally caused (geological). It will be necessary to develop specific predictions of the spatial patterns of genetic variation expected under different biogeographic hypotheses so that these can be explicitly tested. Some approaches have been suggested (Palumbi, 1998) which predict that older genotypes and consequently, higher diversity at neutral loci are likely to be found where species originate. However, this pattern might be masked by differences in population size and extinction rates among regions (e.g.: higher population sizes and lower extinction rates or lower population sizes and higher extinction rates). A phylogeographic analysis of *P. monodon* (Benzie *et al.*, 2002) has shown evidence for the evolution of haplotypes in an oceanic region and of their subsequent migration into the Indonesian region, contributing to the increased level of genetic diversity in these areas. The present study of *P. indicus, P. merguensis* is in accordance with this hypothesis. Nevertheless, further application of molecular genetic tools to examine the structure of penaeid species within these regions, will provide an important means of advancing our understanding of the origin and maintenance of biodiversity in the IWP. In this case the higher values of the nucleotide diversity could be an expected signature of a stable population with large long term N_e or, due to admixture of samples of individuals from historically sundered populations.

Chapter five

Synthesis

5.1 General considerations

As can be gathered from much of the literature cited in this thesis, the latter part of the XX Century has seen a flourishing of diverse molecular techniques, aimed at eliciting and analysing genetic information in order to interpret it in terms of evolutionary history of biological forms (Beaumont, 1994). The present study is no exception to this trend since several bio-chemical and molecular methods were utilised to gain an understanding of *P. indicus* genetic structure and the phylogenetic relationships between its populations. Each of these sources of genetic information, however, tends to be better suited to a given taxonomic level (Avise, 2000; Hillis and Moritz, 1996) and, hence, a given evolutionary time frame. Thus, an overall understanding of biological diversity can only be gained with great difficulty from any single method. Therefore, as pointed out earlier, comparing and integrating the information derived from more than one method best assesses genetic polymorphism of natural populations and this is the purpose of this chapter.

For the study of genetic variation in natural populations, DNA sequences are much more informative than protein sequences or electrophoretic variation of proteins. A large part of the DNA sequences such as pseudo-genes, introns and microsatellites, is not translated into proteins and is thus not detected by gene product analyses. The degeneracy of the genetic code further contributes to this "hidden" genome variation. Genetic variation in these particular regions, therefore, can be studied only through the analysis of DNA sequence variation. As stated by Avise (2000), when gathering molecular information about different populations or species, one must take into consideration the agreement of genetic results across independent loci. The occurrence of such concordance demonstrates almost conclusively that particular partitions in the gene trees can accurately register fundamental phylogenetic subdivisions at a population or species level (Avise, 2000). For these reasons collection of molecular data should be normally conducted not only at varying degrees of polymorphism, but also from both nuclear and mitochondrial genomes. Mitochondrial DNA is the most utilised source of molecular data mostly due to the small size (16 000 bp) of its circular genome (Wilson *et al.*, 1985; Avise, 1986). Because there is little or no paternal

contribution of mitochondria, and no known recombination event between mitochondrial genomes has been recorded (Avise, 1994), mtDNA is considered to be clonally inherited. All these factors combine to reduce the mtDNA effective population size to one-fourth of that of the nuclear genes of the same organisms (Nei and Tajima, 1981). A smaller effective population size means that genetic drift can cause frequency differences between isolated gene pools more readily in mtDNA than in nuclear DNA. In many organisms, the mtDNA also seems to accumulate mutations more rapidly than do single-copy nuclear genes (Lynch and Jarrell, 1993). In other words it provides markers with greater variability and sensitivity to drift, and is, therefore, more likely to show differences between populations or species. Because different regions of the mitochondrial genome evolve at different rates, certain regions of the mtDNA have been targeted for certain types of studies. The Cytochrome b locus and the D-loop region have been examined in a number of species (Carr and Marshall, 1991; Park *et al.*, 1993) and different evolutionary hypotheses have been successfully tested by means of these genetic markers. Furthermore the COI locus had been successfully used in other investigations of prawns (Baldwin *et al.*, 1998) and other marine organisms with long dispersal larval stages (Barber *et al.*, 2002).

Most natural populations of organisms carry large amounts of genetic variation. In sexually reproducing or out-breeding species, for example, any pair of individuals is genetically different and most loci often contain two or more alleles so that populations are genetically polymorphic. Polymorphism at a locus is generated by mutations such as nucleotide substitution, insertion or deletion, gene conversion or inter-allelic recombination. However, most of these new mutations are eliminated from the population by genetic drift or purifying selection and only a minority of them are incorporated in the population by chance, positive selection, or over-dominant selection (selective advantage of heterozygotes). The neutral theory proposed by Kimura (1968) plays an important role in the study of molecular variation, since it proposes that genetic variation at a molecular level (see chapter one) is largely neutral and that the extent of variation is determined primarily by the mutation rate and the effective population size (Kimura and Crow, 1964; Nei, 1987). Therefore due to the fact that, normally, mutation rates are considered to be quite slow in most organisms, the effective population size would remain the major evolutionary force driving the genetic sculpturing of out-breeding populations. Thus neutrality predicts that an increase in effective population size and/or mutation rate results in increased heterozygosity. Nevertheless, as mentioned in section 1.6 (page 18), in recent years, many genes have been shown to be subjected to positive selection

so that assumptions of strict neutrality may not always be appropriate to an evolutionary interpretation of genetic data, although examples of this kind would be more the exception than the rule. Following Nei (1987), the present genetic variability of natural populations is simply a product of evolutionary changes accumulated in the past. In the present survey *Penaeus indicus* populations were almost completely depleted of protein and molecular polymorphisms despite the fact that these organisms were subjected to different, as well as quite variable, environmental conditions. In addition to this, phylogenetically closer species, such as *P. merguensis*, showed different levels of genetic diversity at the same locus (COI). This thesis, in turn, should be interpreted as a significant indication that neutral theory should be the null hypothesis tested in most evolutionary frameworks. Hence the data presented in this thesis allows the formulation of the following two major predictions in the light of the neutral hypothesis:

1. The equilibrium heterozygosity in the population, following Kimura (1964) is proportional to the effective populations size and mutation rate. Because the production of new mutants is generally quite slow and the population size in most species is relatively large, making the effect of genetic drift small, a long time is required for this equilibrium to be reached. Of course, low population sizes experienced at some time in the past, as in the case of *P. indicus,* would reduce the heterozygosity to below the level of equilibrium expected. Hence, the neutrality theory predicts that increases in effective population size and/or mutation rate result in an increase in heterozygosity. On the other hand low levels of heterozygosity can be related to low effective population size that, in the case of *P. indicus,* could only be due to population bottlenecks. The concordance of the data obtained at different molecular levels in identifying extremely low polymorphism or number of segregating sites in all population of *P. indicus,* is consistent with extremely low N_e (effective population size) assuming a mutation rate μ of 10^{-6}.
2. Low values of genetic distances and/or nucleotide divergence are a consequence of the low polymorphism, referring again to population size depression; otherwise these values are more commonly interpreted as an indication of panmixia and, hence, of large effective population sizes maintained at least since the last Pleistocene glaciations.

5.2 *Penaeus indicus*: comparisons of Allozyme and RAPD variation

Despite the availability of recently developed molecular techniques, analysis of allozyme variation will continue to have a major impact on studies of population genetics as well as phylogenetic analyses (Hoelzen, 1998). Cost-benefit ratios of allozymes electrophoresis are probably lower than any molecular approach, a great amount of published data are available for different species and, usually, the co-dominant expression of most loci screened allows an unbiased and prompt identification of genotypes and alleles. Furthermore, the loci usually chosen for this type of approach often include enzymes that play important roles in biochemical pathways. Therefore, they may be more informative than sequences of genes of unknown function, because they can show the effects of selection and this in turn, can be related to important ecological factors influencing the species under survey. Nevertheless, the following drawbacks may apply to methods, such as allozyme electrophoresis, that rely on the detection of gene product variation. For example, nucleotide differences can be masked by the redundancy of the genetic code or by synonymous substitution at the third position in the codon. In addition, nucleotide substitutions can lead to insertion of different amino acids with the same charge, which would then not interfere with the electrophoretic mobility of the protein and consequently, would remain undetected. For these reasons allozyme electrophoresis can be prone to underestimate the degree of genetic polymorphism and consequently, erroneous conclusions could be drawn if expectations of rejecting the null hypothesis of no differentiation between population is entertained.

The RAPD technique (Welsh and McClelland 1990; Williams *et al.*, 1990) has been described as a valid methodology to overcome these limitations in that it is more sensitive to polymorphism in out-breeding organisms. This technique is useful for population genetic analysis (Haig, 1998), taxonomy (Chapco *et al.*, 1992) and mapping (Michelmore *et al.*, 1991), because it allows the rapid acquisition of a lot of genetic information (Helzel and Green, 1998; Bruford *et al.*, 1998). It is suitable to genomes for which microsatellite markers are not available (Hadrys *et al.*, 1992) and has the potential to detect polymorphism anywhere in a genome. However, although RAPD is a fast and inexpensive molecular technique with applications in different fields, it also has practical and statistical limitations (see chapter 3.3.2 page 44).

References

Abbot R.T., 1960. The genus *Strombus* in the Indo-Pacific. *Indo-Pacific Molluscs* 1: 33-146.

Allen G. R., 1975. Damselfishes of the South Seas. T.F.H. publications, Neptune City, NJ.

Amornat, P. C. University of Songala, Thailand. Personal Communication.

Amornrat P., C. Wilaiwan, C. Ausa, T. Chanpa and D. Piyanan, 1999. Identification of *Penaeus merguensis* and *Penaeus indicus* by RAPD-PCR derived DNA marker. *Science Asia* 25: 143-151.

Amornrat P., H. Watcharasuda, J. Querci, S. Pairoj and W. Chotige, 2003. Molecular Phylogenetic analysis of the white prawns species and the existence of two clades in *Penaeus merguensis*. *Journal of Experimental Molecular Biology* 11: 456-465

Angel M.V., 1984. Deep-water biological processes in the Northwest region of the Indian Ocean. *Deep-sea Research* 31: 935-950.

Aris-Brosou S. and L. Excoffier, 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution* 13: 494-504

Avice J.C., 2000. *Phylogeography: The History and Formation of Species*. Harward University Press, London: 447 pp

Avice J.C., 1994. Molecular Markers, Natural History and Evolution. Chapman & Hall, New York: 511 pp

Avise J.C., 1986. Mitochondrial DNA and the evolutionary genetics of higher animals. *Philosophical Transactions of the Royal Society*, London, 312B: 325-342.

Baldwin J. D., L. Anna Bass, W. Brian, L. Bowen and W.H. Clark, 1998. Molecular Phylogeny and Biogeography of the Marine Shrimp *Penaeus*. *Molecular Phylogenetic and Evolution* 10: 399-407.

Ball A.O., S. Leonard, and R.W. Chapman, 1998. Characterization of $(GT)_n$ microsatellites from native white shrimp (*Penaeus setiferus*). *Molecular Ecology* 7: 1251-1253.

Barber P.H., S.R. Palumbi, M.V. Erdmann and M.K. Moosa, 2002. Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns causes and consequences. *Molecular Ecology* 11: 659-674

Bauer, S., G.L. Hitchcock and D.B. Olson, 1991. Influence of monsoonally forced Ekman dynamics upon surface layer depth and plankton biomass distribution in the Arabian Sea. *Deep-sea Research* 38: 531-553.

Beaumont A.R., 1994. Genetics and Evolution of Aquatic Organisms. Chapman & Hall, London UK: 509 pp

Beaumont M.A. and R.A. Nichols, 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceeding of the Royal Society*, London, 263 B: 1619-1626.

Beerli P. and J. Felsenstein, 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152: 763-773.

Beltran M, Jiggins CD, Bull V, Linares M, Mallet J, McMillan WO, Bermingham E. Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating Heliconius butterflies. *Molecular Biology and Evolution*, submitted.

Begun D.J. and C.F. Aquadro, 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* 356: 519-520.

Benzie J. A. H. and S. T. Williams, 1997. Genetic structure of giant clam (*Tridacna maxima*) populations in the West Pacific is not consistent with dispersal by present day ocean currents. *Evolution* 51: 768-783.

Benzie J. A. H., 1999. Genetic improvement of prawns. *Report,* Australian Institute of marine science and cooperative centre for aquaculture, Townsville, Australia.

Benzie J. A. H., E. Ballment, A. T. Forbes, N. T. Demetriades, K. Sugama, Haryanti and S. Moria, 2002. Mitochondrial DNA variation in the Indo-Pacific populations of the giant tiger prawn, *Penaeus monodon. Molecular Ecology* 11: 2553-2569.

Benzie J.A.H. and S.T. Williams, 1997. Phylogenetic relationships among giant clam species (Mollusca: Tridacnidae) determined by protein electrophoresis. *Marine Biology* 132: 123.133

Benzie J.A.H., 2000. Population genetic structure in penaeid prawns. *Aquaculture Research* 31: 95-119.

Benzie J.A.H., S. Frusher and E. Ballment, 1992. Geographical variation in allozyme frequencies of *Penaeus monodon* (Crustacea: Decapoda) populations in Australia. *Australian Journal of Marine and Freshwater Research*, 43: 715-725.

Benzie J.A.H., 1993. Pattern of gene flow in the Great Barrier Reef and Coral Sea. Genetic and evolution of aquatic organisms, *A.R. Beamont*: 67-79.

Bernardi G., Sordino P. and D.A. Powers, 1993. Concordant mitochondrial and nuclear phylogenies for populations of the teleost fish *Fundulus heteroclitus. Proceedings of the National Academy of Science,* USA, 90: 9271- 9274.

Beurlen K., 1930. Vergleichende Stammesgeschichte Grundlagen, Methoden, Probleme unter besonderer Berucksichtigung der hoheren Krebse. *Fortsschritte der Geologie und Palaeontologie* 8: 317-586.

Bohonak A. J., 2002. IBD (Isolation By Distance): a program for analyses of isolation by distance. *Journal of Heredity* in press.

Bohonak A.J., 1999. Dispersal, gene flow and population structure. *Quarterly Review of Biology* 74: 21-45.

Boora K.S., Frederiksen R.A. and C.W. Magill, 1999. A molecular marker that segregates with sorghum leaf blight resistance in one cross is maternally inherited in another. *Molecular and General Genetic* 261 (2): 317-322

Borowsky L.R., 2000. Estimating Nucleotide Diversity from Random Amplified Polymorphic DNA and Amplified Length Polymorphism Data. *Molecular Phylogenetic and Evolution* 10: 1-5.

Borowsky, L.R. and C. Vidthayanon, 2001. Nucleotide diversity in populations of balitorid cave fishes from Thailand. *Molecular Ecology* 10: 2799-2805.

Bowen B. W. and W. S. Grant, 1997. Phylogeography of the sardines (*Sardinops spp.*): Assessing biogeographic models and population histories in temperature upwellings zones. *Evolution* 51: 1601-1610.

Bowen B. W., A. M. Clark, Abreu-Grobois, F. A. Chavez, A. Reichart and R. J. Ferl, 1998. Global phylogeography of the ridley sea turtles (*Lepidochelys spp.*) inferred from mtDNA sequences. *Genetics* 101: 179-189.

Bradley R.D. and D.M. Hillis, 1997. Recombinant DNA sequences generated by PCR amplification. *Molecular Biology and Evolution* 14: 592-593.

Briggs J.C., 1984. Marine Zoogeography. Mc.Graw Hill, New York: 158-193

Briggs J.C., 1995. Global Biogeography. Elsevier, Amsterdam: 320

Briggs J.C., 1999. Coincident Biogeographic Patterns: Indo-West Pacific Ocean. *Evolution* 53 (2): 326-335

Brinton E. and K. Gopalakrishnam, 1973. The distribution of Indian Ocean euphausiids. *The Biology of the Indian Ocean, Springer-Verlag, New York,* Zeitzschel, B. (Ed): 357-382

Brooker A.L., J.A.H. Benzie, D. Blair, and J.J. Versini, 1999. Population structure of the giant tiger prawn, *Penaeus monodon*, in Australian waters using microsatellite markers. *Marine Biology* 136: 149-157

Brown B.E., Darcy, G.H. and W. Overholtz, 1987. Stock assessment/stock identification: an interactive process. *Proceeding of the stock identification workshop,* Washington, Kumpf, H.E., Vaught, R.N. and Grimes, C.B. editions. US. Department Commerce: 1-24.

Brown O.B., J.G. Bruce and R.H. Evans, 1980. Evolution of the sea surface temperature in the Somali basin during the Southwest monsoon of 1979. *Science* 209: 595-597.

Brusher H.A., 1974. The magnitude, distribution and availability of prawn (Penaeidae) resources in coastal and estuarine waters of Kenya. *Journal of Marine Biology Association India* 16: 335-348.

Brutford M.W., O. Hannotte, J.F.Y. Brookfield and T. Burke, 1998. Multilocus and single locus DNA fingerprinting. *Molecular Genetics of Populations*. Oxford University Press, UK: 287-336

Burton R.S. and B.N. Lee, 1994. Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the marine copepod *Tigriopus californicus*. *Proceeding National Academy of Sciences*, USA, 91: 5197-5201

Calman W.T., 1909. Crustacea. A treatise on zoology, Adam and Black, London.part 7 3rd fasc: 1-346.

Carr S.M. and H.D. Marshall, 1991. Detection of intraspecific DNA sequence variation in the mitochondrial genome cytochrome b gene of Atlantic cod (*Gadus murphua*) by the polymerase chain reaction. *Canadian Journal of Fisheries and Aquaculture Sciences* 48: 48-52.

Carvalho G.R. and L. Hauser, 1995. Molecular genetics and stock concept in fisheries. *Molecular genetics in Fisheries*. Edited by Gary R. and R. Cavalho and Tony J. Pitcher. Published by Chapman & Hall, London.

Champion H.F.B., 1970. Aspects of the biology of *Penaeus indicus* (Milne-Edwards) with notes on associated Penaeidae occurring off Natal on the East coast of South Africa. *Oceanography in South Africa, Durban*, August 1970.

Chan T. Y. and K. H. Chu, 1996. On the different forms of *Panulirus longipes femoristriga* (Von Martens, 1872) (Crustacea: Decapoda: Palinuridae), with a description of a new species. *Journal of Natural History* 30: 367-387.

Chapco W., N.W. Ashton, R.K. Martel, N. Antonishishyn, W.L. Crosby, 1992. A feasibility study of the use of random amplified polymorphic DNA in the population genetics and systematics of Grasshoppers. *Genome* 35: 569-574.

Chalesworth D., Mc Vean, 2001. Genome sequences and evolutionary biology, a two-way interaction. *Trends in Ecology and Evolution* 16: 235-242.

Cockerham C.C., 1969. Variance of gene frequencies. *Evolution* 23: 72-83.

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

Contrafatto G., D. David and V. Gossens, 1994. Genetic variation in the African rodent sub-family Otomynae (Muridae): phylogeny inferred by immuno-electro-transfer analysis. *Durban Museum Novitates* 19: 1-7.

Crow J.F. and K. Aoky, 1984. Group selection of a polygenic behavioural trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences*, USA, 81: 234-241.

Currie R.I., A.E. Fisher and P.M. Hargreaves, 1973. Arabian sea upwellings. *The Biology of the Indian Ocean,* in Zeitzschel, B. (Ed.), Springer-Verlag, New York: 57-63.

Dakin, W.J., 1973. The habits and life history of penaeid prawn (*Penaeus plebejus*). *Proceeding of Zoological Society London*, (A) 108(2): 781-815.

Dall W., BJ. Hill, P.C. Rothlisberg and D.J. Staples, 1990. The biology of the Penaeidae. *Advance Marine Biology*, Academic Press New York, 27: 489

Davies D.H. and L. Joubert, 1964 (unpublished report). Preliminary report on the second survey of the penaeid prawns of the St. Lucia lake system carried out in October/November 1964. Durban, *Oceanographic Research Institute*: 1-20.

Duncan C.P., 1970. The Agulhas Current. *Unpublished Ph.D thesis*, University of Hawaii: 1-76.

Eanes WF, 1999. Analysis of selection on enzyme polymorphisms. *Annual Review of Ecology and Systematics* 30: 301-326.

Ekman S., 1953. Zoogeography of the Sea. Sidgwick and Jackson editions, London.

Emerson W.K., 1991. First records for *Cymanthium mundum* (Gould) in the eastern Pacific Ocean, with comments on the zoogeography of the tropical trans-Pacific tonnacean and non-tonnacean prosobranch gastropods with Indo-Pacific faunal affinities in western American waters. *The Nautilus* 105: 62-80.

Excoffier L., 2001. Analysis of Population Subdivision. *Handbook of Statistical Genetics*, D. J. Balding, Bishop M., Cannings C., John Wiley and Sons Ldt. editions: 271-307

Excoffier L., P. Smouse and J. Quattro, 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Fairbridge R.W., B.C. Heezen, T. Ichme and M. Tharp, 1966. Indian Ocean. *Encyclopedia of Oceanography*, Fairbridge, R.W. editions, New York: 370-402.

FAO, 1997a. Fisheries Statistics Capture Production. FAO Fisheries Series n°52. Rome, Food and Agricolture of United Nations

FAO, 1990. Catches and Landings. *Fisheries Statististics* 52: 1-357.

Fast A. W. and L.J. Lester, 1992. Marine shrimp culture: principles and practises. *Development in Aquaculture and Fisheries Science* 23: 1-862.

Fay J.C., G.J. Wyckoff, C-I Wu, 2001. Positive and negative selection on the human genome. *Genetics* 158: 1227-1234.

Feldman, R.M. and W.J. Zinsmeister, 1984a. First occurrence of fossil decapods crustaceans (Callianassidae) from Mc Murdo sound region, Antarctica. *Journal of Palaeontology* 58: 1041-1045.

Feldman, R.M. and W.J. Zinsmeister, 1984b. New fossils crabs (Decapoda: Brachyura) from the la Meseta Formation (Eocene) of Antartica: Palaeographic and biogeographic implications. *Journal of Palaeontology* 58: 1046-1061.

Felsenstein J., 1999. PHYLIP (Phylogeny Inference Package) Version 3.5c. Distributed by the author, University of Washington, Seattle.

Forbes A. T., 1994. Review of possible penaeid culture in Natal republic of South Africa. *Aquaculture Magazine*: 46-55.

Forbes, A.T. and D.P. Cyrus, 1991. Recruitment and origin of penaeid prawn postlarvae in two southeast African estuaries. *Estuarine Coastal and Shelf Science* 33(3): 281-289.

Forbes, A.T., N.T. Demetriades, J.A.H. Benzie and E. Ballment, 1999. Allozyme frequencies indicate little geographic variation among stocks of giant tiger prawn *Penaeus monodon* in the southwest Indian Ocean. *South African Journal of Marine Science* 21: 271-277.

Ford J. M., 2002. Applications of selective neutrality tests to molecular ecology. *Molecular Ecology* 11: 1245-1262.

Fu Y.X. and W.H. Li, 1999. Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theoretical Populations Biology* 56: 1-10.

Fu Y.X., 1997. Statistical tests of neutrality mutations against population growth, hitchhiking and background selection. *Genetics* 147: 917-927

Futuyama, D. J., 1998. *Evolutionary Biology*. Sunderland, Mass. 3rd Ed., Sinauer.

Garcia S., 1977. Biologie et dynamique des populations de crevettes roses, *Penaeus duodarum notialis* en Cote D'Ivoire. *Travaux et documents*. ORSTOM 79: 1-271.

Garcia S., 1985. Reproduction, stock assessment models and population parameters in exploited penaeid shrimp populations. *Second Australian National Prawn Seminar*: 139-158.

Garcia, S., 1988. Tropical penaeid prawns. *Fish Population Dynamics* (J.A. Gulland ed.) 2nd edn., John Wiley, Chichester: 219-249

Garcia-Machado E., N. Dennebouy, M.O. Suarez, J.C. Mounolou and M. Monnerot, 1996. Partial sequence of the shrimp *Penaeus notialis* mitochondrial genome. *Academy of Science Paris* 319: 473-486.

Gaudet J., 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity* 86: 485-486.

Glaessner M.F., 1969. Decapoda. Treatise on Invertebrate Paleontology, *Geological Society of America,* Boulder, Colorado, and University of Kansas, R.C. Moore editions, Lowrence: 399-533.

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

Grosberg R.K. and C.W. Cunningham, 2001. Genetic structure in the sea: from populations to communities. *Marine Community Ecology*, Sinauer, Sunderland, Massachusetts.: 61-84

Hadrys H., M. Balick and B. Schierwater, 1992. Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Molecular Ecology* 1: 55-63.

Haig S.M., 1998. Molecular contributions to conservation. *Ecology* 79: 413-425.

Hall T., 2001. BIOEDIT (version 5.0.6): Biological Sequence Editor, available on request at tahall@unity.ncsu.edu.

Hall D.N.F., 1962. Observation on the taxonomy and biology of some Indo West Pacific Penaeids (Crustacea: Decapoda). *Fishery Publications*, Colonial office, Her Majesty's stationary office, London, 17: 1-229.

Hare, P.M., 2001. Prospects for nuclear gene phylogeography. *Trends in Ecology and Evolution* 16: 700-706.

Harris M. and D.A. Hopkinson, 1976. Handbook of enzyme electrophoresis in human genetics. Oxford: *North Holland Publishing Co*. (Looseleaf).

Hartl D.L. and A.G. Clark, 1997. Principles of Population Genetics, 3rd Edition. Sinauer, Sunderland, Massachusetts: 682 pp

Hedgecock D. and S.R. Malecha, 1991. Prospects for the application of biotechnology to the development and improvement of shrimp and prawns. Shrimp Culture in North America and the Caribbean. *Advances in World Aquaculture* 4, Aquaculture society, Baton Rouge, LA, USA, P.A. Sandifer editions: 161-200.

Hedgecock D., M.L. Tracey and K. Nelson, 1982. *The biology of Crustacea*, L.G. Abele editions, Academic press New York: 284-403.

Hedgecock, D., 1986. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Sciences* 39: 550-564.

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

Hey J., 1999. The neutralist, the fly and the selectionist. *Trends in Ecology and Evolution* 14: 35-38.

Heydrick P.W., 1999. Genetics of populations. 2nd edition, Jones and Bartlett publishers, Sudbury, Massachusetts.

Hernández P., Martín A. and Dorado G., 1999. Development of SCARs by direct sequencing of RAPD products: a practical tool for the introgression and marker-assisted selection of wheat. *Molecular Breeding* 5: 233-244

Higgins D.G. and P.M. Sharp, 1988. Clustal: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73: 237-244.

Hillis C J and P. M. Morritz, 1996. Molecular Systematics. 2nd edition. Sinauer Associates, Sunderland, Massachusetts

Hoelzen A.R., 1998. Molecular Genetic Analysis of Populations. A practical approach. 2nd edition. Oxford University Press, New York: 445 pp

Hoelzen A.R. and A. Green, 1998. PCR protocols and population analysis by direct DNA sequencing and PCR- based DNA fingerprinting. Molecular Genetic Analysis of Populations. A practical approach. Edited by A.R. Hoelzel, Oxford University Press, 6: 201-235.

Hoelzer G.A., 1997. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear gene trees revisited. *Evolution* 51: 622-626.

Hoguane A., 1997. Shrimp abundance and river runoff in Sofala Bank: the role of the Zambesi. *Workshop on the sustainable use of Cahora Bassa dam and the Zambesi valley*, Songo, Mozambique.

Holmes A. and D.L. Holmes, 1964. Holmes' Principles of Physical Geology. Thomas Nelson, Middlesex.

Hudinaga M., 1942. Reproduction, development and rearing of *Penaeus japonicus* Bate. *Journal of Zoology* 10 (2): 305-394.

Hudson R.R., 1983. Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology* 23: 183-201.

Hudson RR., 1990. Gene genealogies and the coalescent approach. *Evolutionary Biology*, Futuyma D. edition, Oxford University press, New York: 1-44

Hughes A.L., 1999. Adaptive Evolution of Gene and Genome. Oxford University press, New York.

Ihssen P.E., H.E. Brooke, J.M. Casselman, J.M. Mc Glade, N.R. Payne and F.M. Utter, 1981. Stock identification: materials and methods. *Canadian Journal of Fisheries and Aquaculture Science* 38: 1838-1855.

Jackson J.B.C., 1986. Modes of dispersal of clonal benthic invertebrates: Consequences for species distribution and genetic structure of local populations. *Bulletin of Marine Science* 39: 588-606.

Jensen T., 1991. Modeling the seasonal undercurrents in the Somali Current system. *Journal of Geophysical Research* 96: 22151-22168.

Johnson, M.C. and J.R. Fielding, 1956. Propagation of the white shrimp *Penaeus setiferus* (Linn.) in captivity. *Tulane Studies of Zoology* 4(6) : 175-190.

Jukes J. H. and C.R. Cantor, 1969. Evolution of protein molecules. Mammalian protein metabolism, H.N. Munro, edition, Academic, New York: 21-132

Karl, S.A. and J.C. Avise, 1993. PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. *Molecular Biology and Evolution* 10: 342-361.

Kensley B., H.A. Tranter and D.J.C. Griffin, 1987. Deepwater decapods crustacean from eastern Australia (Penaeidae and Caridae). *Records of the Australian Museum* 39: 263-331.

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

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

Kimura M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, New York.

Kingman J.F.C., 1982a. The coalescent. *Stochastic Processes and their Applications* 13: 235-248.

Kingman J.F.C., 1982b. In Exchangeability in Probability and Statistics. G. Koch and F. Spizzichino eds., North-Holland, Amsterdam: 97-112.

Kingman J.F.C., 1982c. Essays in Statistical Science. *Journal of Applied Probability* special Papers in honour of P.A.P. Moran, J. Gani and E.J. Hannan, editions, 19(A): 27-43

Kirkegaard I., 1975. Observations of penaeids larvae around Australia. *National Prawn Seminar of Australian Government Publications Service*, P. C. Young editions, Canberra: 54-59.

Klinbunga S., D.J. Penman, B.J. McAndrew, A. Tassanakajon and P. Jarayabhand, 1998. Genetic variation, population differentiation, and gene flow of the giant tiger shrimp (*Penaeus*

monodon) inferred from mtDNA-RLFP data. *Advances in Shrimp Biotechnology*, National Centre for Genetic Engineering and Biotechnology, Bangkok: 51-59

Knowlton N., 1993. Sibling species in the sea. *Annual Review of Ecology and Systematic*. 24: 189-216.

Ko G. L., N. Pasteur, Bonhomme F., Aquacop, Seafdec and D. Liao, 1983. Variabilite genetique de qualques especes de crevettes peneides d'interet aquicole. Bases biologiques de l'aquaculture, Montpellier, *Actes de colloques* 1: 174-180.

Kraft T., T. Sall, I.M. Magnusson-Rading, N-O Nilsson and C. Hallden, 1998. Positive correlation between recombination rates and levels of genetic variation in the natural populations of sea beet (*Beta vulgaris*). *Genetics* 150: 1239-1244.

Kreitman M. and H. Akashi, 1995. Molecular evidence for natural selection. *Annual Review of Ecology and Systematic* 26: 403-422.

Kreitman M., 1996. The neutral theory is dead. Long live the neutral theory. *Bioessays* 18: 678-683.

Kuang H., T.E. Richardson, S.D. Carson and B.C. Bongarten, 1998. An allele responsible for seedling death in *Pinus radiata* D. Don. *Theoretical and Applied Genetics* 96: 640-644.

Kutkuhn J.F., 1966a. The role of estuaries in the development and perpetuation of commercial shrimp resources. *American Fisheries Society*, special publication 3: 16-36.

Ladd H. S., 1960. Origin of the Pacific Island molluscan fauna. *American Journal of Science* 258: 137-150

Leis J.J., 1984. Larval fish dispersal and the East Pacific barrier. *Oceanographique Tropicale* 19: 181-192.

Lester L.J., 1979. Population genetics of penaeid shrimp from Gulf of Mexico. *Journal of Heredity* 70: 175-180.

Levene H., 1949. On a Matching Problem in Genetics. *Annual Mathematic Statistic* 20: 91-94.

Lewontin R.C., 1972. Apportionment of human diversity. *Evolutionary Biology* 6: 381-398

Li G.D. and R. Hedgecock, 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 Aquaculture Sciences*, 55: 1025-1033

Li W-H, and D. Graur, 1991. Fundamentals of Molecular Evolution. Sinauer, Sunderland Massachusetts.

Linnox Biosoft, 1999. DNAman software program: licensed at Department of Microbiology University of Cape Town, South Africa.

Lynch M and P.E. Jarrell, 1993. A method for calibrating molecular clocks and its application to animal mitochondrial DNA. *Genetics* 135: 1197-1208.

Lynch M. and B. Milligan, 1994. Analysis of population genetic structure with RAPD markers. *Molecular Ecology* 3: 91-99.

Mc Creary J.P. and P.K. Kundu, 1988. A numerical investigation of the Somali current during the Southwest monsoon. *Journal of Marine Research* 46: 25-58.

McMillan W.O. and S.R. Palumbi, 1995. Concordant evolutionary patterns among Indo-West Pacific butterfly fishes. *Proceeding of the Royal Society of London* 260: 229-236.

Meanche, M., 1961. Decouvertè d'ùn phenomene dé rémontee d'eaux profundes de canal de Mozambique. *Memorial Institute Scientific*, Madagascar, (F) 4: 166-171.

Meruane J., M. Takagi and N. Taniguchi, 1997. Species identification and polymorphisms using RAPD-PCR in penaeid prawns *Penaeus japonicus* and *Metapenaeus ensis. Fisheries Science* 63: 149-150.

Milchemore R.W., I. Paran, R.V. Kesseli, 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences*, USA, 88: 9828-9832.

Miquel, J.C., 1982. Le Genre *Metapenaeus* (Crustacea, Penaeidae): *Taxonomie, Biologie et Peches Mondiales.* E.J. Brill, Leiden.

Moosa, J., 2002. Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes and consequences. *Molecular Ecology* 11: 659-674.

Moya A., P. Guirao, D. Cifuentes, F. Beitias and L.J. Cenis, 2001. Genetic diversity of Iberian populations of *Bemisia tabaci* (Hemiptera: Aleyrodidae) based on random amplified polymorphic DNA-polymerase chain reaction. *Molecular Ecology* 10: 891-897.

Mulley J.C. and B.D.H. Latter, 1980. Genetic variation and evolutionary relationships within a group of thirteen species of penaeid prawns. *Evolution* 34: 904-916.

Munroe I. S. R., 1975. Biology of the Banana prawn (*Penaeus merguensis*) in the south east corner of the Gulf of Carpenteria. *National Prawn Seminar Australian Government Publication Service*, Canberra P. C., Young edition: 60-78.

Neal R. A., 1969. Methods of marking shrimp. *FAO Fish Report* 57: 1149-1165.

Nei M. and F. Tajima, 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 105: 207-217.

Nei M. and S. Kumar, 2000. Molecular Evolution and Phylogenetics. Oxford University Press, London: 333 pp

Nei M., 1972. Genetic distance between populations. *American Nature* 106: 283-292.

Nei M., 1973. Analysis of gene diversity in subdivided populations. *Proceeding of the National Academy of Science*, USA 70: 3321-3323.

Nei M., 1977. F-Statistics and analysis of gene diversity in subdivided populations. *Annual Human Genetics* 41: 225-233.

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

Nei M., 1987. Molecular evolutionary genetics. Columbia University Press, New York

Nei M., T. Maruyama and R. Chakraborty, 1975. The Bottleneck Effect and Genetic Variability in Populations. *Evolution* 29: 1-10

Nelson G.J. and D.E. Rosen, 1981. Vicariance Biogeography: A critique. Columbia University Press New York

Nevo E., 1978. Genetic variation in natural populations: patterns and theory. *Theoretical Population Biology* 13: 121-177.

Nichida M. and J.S. Lucas, 1988. Genetic differences between geographic populations of crown-of-thorns starfish throughout the pacific region. *Marine Biology* 98: 359-368.

Nordborg M., 2001. Coalescent Theory. *Handbook of Statistical Genetics*. Edited by Baldwin D.J., Bishop M., Cannings C. Wiley.: 179-212.

Palumbi S. R., 1997. Molecular biogeography of the pacific. *Coral Reefs* (Suppl) 16: 547-552.

Palumbi S.R. and C.S. Baker, 1994. Contrasting population structure from nuclear intron sequences and mtDNA of Humpback whales. *Molecular Biolology and Evolution* 11: 426-435.

Palumbi S.R. and J.A.H. Benzie, 1991. Large mitochondrial DNA differences between morphologically similar Penaeid shrimp. *Molecular Marine Biology and Biotechnology* 1: 27-34.

Paran I. and R.W. Michelmore, 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lattuce. *Theretical Application Genetics* 85: 985-993

Park L.K., M.A. Brainard and D.A. Dightman, 1993. Low levels of intraspecific variation in the mitochondrial DNA of chum salmon (*Oncorynchus keta*). *Molecular Marine Biology and Biotechnology* 2: 362-370.

Paulay G., 1990. Effects of late Cenozoic sea-level fluctuations on the bivalve faunas of tropical oceanic islands. *Paleobiology* 16: 415-434.

Paulay G., 1997. Diversity and distribution of reef organisms. C.E. Birkeland edition, Chapman *et* Hall, New York: 298-353.

Pearson C.V.M., A.D. Rogers and M. Sheader, 2002. The genetic structure of rare lagoonal sea anemone, *Nematostella vectenis* (Cnidaria; Anthozoa) in the United Kingdom, based on RAPD analysis. *Molecular Ecology* 11: 2285-2293.

Perez T., J. Albornoz and A. Dominguèz, 1998. An evaluation of RAPD fragment reproducibility and nature. *Molecular Ecology* 7: 1347-1357.

Piccoli G., S. Sartori and A. Franchino, 1987. Benthic molluscs of shallow Tethys Sea and their destiny. *Proceedings International Symposium on Shallow Tethys,* K. Mc Kensie, edition. 2: 333-373.

Pogson G.H., 1995. Genetic population structure and gene flow in the Atlantic cod, *Gadus morhua*: a comparison of allozyme and nuclear RLFP loci. *Genetics* 139: 375-385.

Pogson G.H., 2001. Nucleotide polymorphism and natural selection at the panthophysin (Pan I) locus in the Atlantic cod, *Gadus morhua* (L.). *Genetics* 157: 317-330.

Racek A. A. and W. Dall, 1965. Littoral Penaeidae (Crustacea: Decapoda) from northern Australia, New Guinea and adjacent waters. Verh. Akad. Wet. Amst. 56:1-119.

Racek A.A., 1959. Prawns investigations in eastern Australia. *Research Bulletin St. Fisheries* N.S.W. 6: 209-241.

Racek A.A., 1970. Indo-west Pacific penaeid prawn species of commercial importance. *Indo-Pacific Fishing Coun.* 14th Session Symposium, 3: 1-33.

Redfield J.A., D. Hedgecock, K. Nelson and J.P. Salini, 1980. Low heterozygosity in tropical marine crustaceans of Australia and the trophic stability hypothesis. *Marine Biology Letters* 1: 303-313.

Reynolds J., B.S. Weir and C.C. Cockerham, 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105: 767-779.

Rice W.R., 1989. Analysing tables of statistical tests. *Evolution*, 43: 223-225.

Roger A. R. and H.C. Harpending, 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552-569.

Rogers J.S., 1972. Measures of genetic similarity and genetic distance. *University of Texas Publication* 7213: 145-153.

Roger A.R. and H.C. Harpending, 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552-569

Rosemblatt R.H. and R.S. Waples, 1986. A genetic comparison of allopatric populations of shore fish species from the eastern and central pacific ocean: dispersal or vicariance? *Copeia* 2: 275-284.

Rozas J. and R. Rozas, 2001. DNA sequence polymorphism (Version 3.5.3): available at www.bioub.es/-julio/dnasp.html.

Saitau N.and Nei M., 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4: 406-425

Sbordoni V., E. De Matthaeis, M.C. Sbordoni, G. La Rosa and M. Mattoccia, 1986. Bottleneck effects and the depression of genetic variability in hatchery stocks *of Penaeus japonicus* (Crustacea: Decapoda). *Aquaculture* 57: 239-251.

Scheltema R.S., 1971. Larval dispersal as a means of genetic exchange between geographically separated populations of shallow-water marine gastropods. *Biology Bulletin* 140: 284-322.

Scheltema R.S., 1988. Initial evidence for the transport of teleplanic larvae of benthic invertebrates across the East Pacific barrier. *Biology Bulletin* 174: 1445-1452.

Scheltema R.S., 1986. On dispersal and planktonic larvae of benthic invertebrates: an eclectic overview and summary of problems. *Bulletin of Marine Science* 39: 290-322

Schnider S., D. Roessli and L. Excoffier, 2000. ARLEQUIN ver 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland,

Schott P.H.A. and R.R. Sokal, 1973. Numerical taxonomy, Freeman, San Francisco.

Sebastiani F., R. Meiswinkel, L.M. Gomulski, C.R. Guglielmino, P.S. Mellor, A.R. Malacrida and G. Gasperi, 2001. Molecular differentiation of Old World *Culicoides imicola* species complex (Diptera: Ceratopogonidae), inferred using random amplified polymorphic DNA markers. *Molecular Ecology* 10: 1773-1786.

Shackleton N.J. and J.P. Kennett, 1975. Paleo-temperature history of the Cenozoic and the initiation of Antartic glaciation: oxygen and carbon isotope analyses in DSDP Sites 277, 279 and 281. *Initial Reports of the Deep Sea Drilling Project* 29: 743-756.

Shaw K.L., 2002. Conflict between mitochondrial and nuclear DNA phylogenies of a recent species radiation: what mitochondrial reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences* USA 99: 16122-16127

Sinclair M., 1988. Marine populations. An Essay on Population Regulation and Speciation. Seattle and London: University Washington Press, pp 53.

Slatkin M., 1987. Gene flow and geographic structure of natural populations. *Science* 236: 787-792.

Slatkin M., 1985a. Gene flow in natural populations. Annual Review of Ecology and Systematics 16: 393-430

Sodsuk S., B.J. Mc Anrew and D.J. Penman, 1992. Genetic population structure of the giant tiger prawn (*Penaeus monodon*: Fabricius, 1798) in the Gulf of Thailand and the Andaman sea. *Proceedings of the ASEAN-EEC Aquaculture Development and Coordination Program* D. Penman, N. Roograti and B. McAndrew edition Stirling, UK.:161-165.

Sokal R.R. and F.J. Rohlf, 1981. *Biometry*, Freeman, San Francisco, California.

Spicer G.S., 1995. Phylogenetic utility of the mitochondrial cytochrome oxidase gene. Molecular evolution of the *Drosophila buzzatii* species complex. *Journal of Molecular Evolution* 41: 749-759.

Springer V. G. and J.T. Williams, 1990. Widely distributed Pacific Plate endemics and lowered sea-levels. *Bulletin of Marine Science* 47: 631-640.

Staples D.J. and D.J. Vance, 1985. Short-term and long term influences on the immigration of postlarval banana prawns *Penaeus merguensis*, into a mangrove estuary of the Gulf of Carpenteria, Australia. *Marine Ecology Progress Series* 23: 15-29.

Staples D.J., 1980a. Ecology of juvenile and adolescent banana prawns, *Penaeus merguensis*, in a mangrove estuary and adjacent offshore area of the Gulf of Carpentaria. Immigration and settlement of Postlarvae. *Australian Journal of Marine and Freshwater Research* 31: 635-652.

Staples, D.J., D.J. Vance and D.S. Heales, 1985. Habitat requirements of juvenile penaeid prawns and their relationship to offshore fisheries. *Second Australian National Prawn Seminar*. Rothlisberg, P.C., Hill, B.J. and Staples, D.J. edition, Cleveland, Australia: 47-54.

Stephens M., 2001. Inference under the Coalescent. *Handbook of Statistical Genetics*, Baldwin D.J., Bishop M., Cunnings C.. John Wiley and Sons edition, 8: 213-238.

Swearer S.E., J.E. Caselle, D.W. Lea and R.R. Warner, 1999. Larval retention and recruitment in an island population of the coral reef fish. *Nature* 402: 799-802

Swofford D.L. and R.B. Selander, 1981. BYOSYS 1: a FORTRAM program for comprehensive analysis of electrophoretic data in population genetics and systematic. *Journal of Heredity* 72: 281-283.

Tajima F., 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Tajima F., 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tajima F., 1989b. The effect of change in population size on DNA polymorphism. *Genetics* 123: 597-601.

Tam Y.K. and K.H. Chu, 1993. Electrophoretic study of the phylogenetic relationships of some species of *Penaeus* and *Metapenaeus* (Decapoda: Penaeidae) from the South China Sea. *Journal of Crustacean Biology* 13: 697-705.

Taniguchi N. and H. Han, 1989. Population genetic analysis of the fish and shellfish by isozymes. *Japanese Fisheries Resource Conservation Association Report*: 275-281.

Tassanakajon A., A. Tiptawonnukul, P. Supungul and V. Rimphanitchayakit, 1998a. Isolation and characterization of microsatellite markers in the black tiger prawn *Penaeus monodon*. *Molecular Marine Biology and Biotechnology* 7: 55-61.

Tassanakajon A., P. Supangul, S. Klinbunga, P. Jarayabhand and V. Boosaeng, 1998b. Microsatellite variation in wild populations of black tiger prawn, *Penaeus monodon*, in Thailand. *Proceeding of the 4th Asia-Pacific Conference on Agricultural Biotechnology*, CPN Publications, Canberra: 255-257.

Tassanakajon A., S. Pongsomboom, P. Jarayabhnd, S. Klinbunga and V. Boonsaeng, 1998. Genetic structure in the wild populations of black tiger shrimp (*Penaeus monodon*) using randomly amplified polymorphic DNA analysis. *Journal of Marine Biotechnology* 6: 249-254.

Tassanakajon A., S. Pongsomboon, V. Rimphanitchayakit, P. Jarayabhand and V. Boosaeng, 1997. Random amplified polymorphic DNA (RAPD) markers for determination of genetic variation in wild populations of black tiger prawn (*Penaeus monodon*) in Thailand. *Molecular Marine Biology and Biotechnology* 6: 110-115.

Templeton A.R., 1998. Nested clade analyses of phylogeographic data: Testing hypotheses about gene flow and population history. *Molecular Ecology* 7: 381-397.

Thorpe J.P. and A. M. Solè-Cava, 1994. The use of allozyme electrophoresis in invertebrate systematics. *Zoologica Scripta* 23: 3-18.

Thorpe J.P., A.M. Solè-Cava and P.C. Watts, 2000. Exploited marine invertebrates: genetics and fisheries. *Hydrobiologia* 420: 165-184.

Tom M., M. Goren and M. Ovadia, 1988. The benthic phase of the life cycle of *Parapenaeus longirostris* (Crustacea: Decapoda, Penaeidae) along the Mediterranean coast of Israel. *Hydrobiologia* 169: 339-352.

UNEP (United Nations Environment Programme), 1985a. Ocean Energy potential of the West African Region. UNEP *Regional Seas Reports and Studies*: 30.

Van der Elst R.P. and S.T. Fennessy, 1990. An investigation into the discard catch of the Tugela bank prawn trawlers. *Unpublished Research Report*, Oceanographic Research Institute, South Africa 66.

Veron J. E. N., 1995. Corals in Space and Time. University of South Wales, Sydney.

Wahlund S., 1928. Zusammensetzung von populationen und Korrelation-serscheinungen von Standpunkt der Verebeungslehre aus betrachtet. *Hereditas* 11: 65-106.

Wakeley J., 1995. Distinguishing migration from isolation using the variance of pairwise differences. *Theoretical Population Biology* 49: 369-386.

Wang R.L., J. Wakeley and J. Hey, 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. *Genetics* 147: 1091-1106.

Wang RL., A. Stec, J. Hey, L. Lukens and J. Doebley, 1999. The limits of selection during maize domestication. *Nature* 398: 236-239.

Wang W., K. Thornton, A. Berry and M. Long, 2002. Nucleotide variation along the *Drosophlila melanogaster* fourth chromosome. *Science* 295: 134-137.

Weir B.S. and C.C. Cockerham, 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

Weir B.S., 1990. Genetic data analysis, Sinauer Associates, Sunderland, Massachusetts.

Welsh J. and M. Mc Clelland, 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* 18: 7213-7218.

Whatley R.C., 1987. The Southern end of Tethys: an important locus for the origin of both deep and shallow water Ostracoda. *Proceeding of the International Symposium on Shallow Tethys* 2, The Netherlands: 461-474.

Williams G.K., A.R. Kubelik, K.L. Livak, J.A. Rafalski and S.V. Tingey, 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetics markers. *Nucleic Acids Research* 18: 6531-6535.

Williams S.T. and J.A.H. Benzie, 1996. genetic uniformity of widely separated populations of the coral reef starfish *Linkia laevigata* from the west Pacific and east Indian Oceans, revealed by allozyme electrophoresis. *Marine Biology* 126: 99-108

Williams S.T. and J.A.H. Benzie, 1997. Indo-West Pacific patterns of differentiation in the high-dispersal starfish *Linkia laevigata. Molecular Ecology* 6: 559-573

Williams S.T. and J.A.H. Benzie, 1998. Evidence of biogeographic break between populations of high dispersal starfish, congruent regions within the Indo-west Pacific defined by colour morphs, mtDNA and allozyme data. *Evolution* 52: 87-99

Wilson A.C., R.L. Cann, , S.M. Carr, M. George, U.B. Gyllensten, K.M. Helm-Bychowsky, R.G. Higuchi, S.R. Palumbi, E.M. Prager, R.D. Sage and M. Stoneking,1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biology Journal of Linnean Society* 26: 375-400.

Wilson K., V. Cahill, E. Ballment and J.H.A. Benzie, 2000. The complete sequence of the mithichindrial genome of the crustacean *Penaeus monodon*: Are Malacostracan Crustacea more closely related to Insects than to Branchiopods? *Molecular Biology and Evolution* 17(6): 863-874.

Withlock M.C. and D.E. Mc. Cauley, 1990. Some population genetic consequences of colony formation and extinction, genetic correlation within founding groups. *Evolution*: 1717-1724

Williams S.T. and J.A.H. Benzie, 1996. Genetic uniformity of widely separated populations of the coral reef starfish *Linckia laevigata* from the West Pacific and East Indian Oceans, revealed by allozyme electrophoresis. *Marine Biology* 126: 99-108

Williams S.T. and J.A.H. Benzie, 1997. Indo-West Pacific patterns of differentiation in the high-dispersal starfish *Linckia laevigata*. *Molecular Ecology* 6: 559-573

Williams S.T. and J.A.H. Benzie, 1998. Evidence of the biogeographic break between populations of the high-dispersal starfish: congruent regions within the Indo-West Pacific defined by colour morphs, mtDNA and allozyme data. *Evolution* 52: 87-99.

Woods H., 1925. A monograph of the fossil Macrurous (Crustacea) of England. Palaeontographical society, London.

Workman P.Z. and J.D. Niswander, 1970. Population structure of the South-western Indian tribes. II. Local genetic differentiation in papago. *American Journal of Human Genetics* 1: 24-29.

Wright S., 1943. Isolation by distance. *Genetics* 28: 114-138.

Wright S., 1969. Evolution and the Genetics of Populations II. *Theory of Gene Frequencies*. University of Chicago Press, Chicago.

Wright, S. 1978. Variability within and among natural populations. *Evolution and the genetics of populations*, University of Chicago Press, Chicago 4.

Wright S., 1986. Evolution: Selected Papers. University of Chicago Press, Chicago.

Yang Z., JP. Bielawski, 2000. Statystical methods for detecting molecular adaptation. *Trends in Ecology and Evolution* 15: 496-503.

Yeatman J. and J. A. H. Benzie, 1994. Genetic structure and distribution of *Photololigo spp. Australian Marine Biology* 118: 79-87.

Yeh F.C. and R. Yang, 1999. Popgene version 1.31. Microsoft Window-based freeware for population genetic analysis. Available for free at www.ualberta.ca/-fyeh

Zhivotovsky L.A., 1999. Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* 8: 907-913.

Zinsmeister, W.J., 1982. Late cretaceous-early tertiary molluscan biogeography of the southern circum-Pacific. *Journal of Paleontology* 56: 84-102.

APPENDIX

Map of the Indo West Pacific region showing continental shelves. Modified from "Atlant of the Oceans", Geographic Institute De Agostini 1983.

Red circles delimited enlarged sections of sampling areas of speciments from *Penaeus indicus* and *Penaeus merguensis* respectively along the East African coast and the Thailand peninsula.

Oman

Tanzania

Madagascar

Mozambique

South Africa

Gulf of Thailand

Andaman Sea

APPENDIX IIA

Buffers used for starch gel electrophoresis of *Penaeus indicus*

1. **TEB**:

Gel buffer:	48mM Tris, 1mM EDTA, 37mM Boric acid, pH 8.4
Electrode buffer:	150mM Tris, 3 mM EDTA, 117 mM Boric acid, pH 8.4
Stock solution:	Dissolve the following in distilled water and make up to 2 litres 181,67 g Tris 12,42 g EDTA (Na2 salt) 72,57 g Boric acid
Running buffers:	
Gel buffer:	16ml stock solution diluted to 250 ml with distilled water
Electrode buffer	1 part stock solution plus 4 parts distilled water
Running conditions:	30-35 mA/gel at 200-400 V (mosto often 400V) for 5 hours

---

2. **TEC 7.9**:

Gel buffer:	8,5 mM Tris, 2 mM Citric acid, 0,27 mM Na2EDTA, pH 7,87
Electrode buffer:	135mM Tris, 32 mM Citric acid, 4 mM Na2EDTA, pH 7,87
Stock solution:	Dissolve the following in distilled water and make up to 2 litres 163.5 g Tris 67.25 g Citric acid H20 15.2 g Na2EDTA
Running buffers:	
Gel buffer:	3.15ml stock solution diluted to 250 ml with distilled water
Electrode buffer	1 part stock solution plus 4 parts distilled water
Running conditions:	30-35 mA/gel at 200V for 5 hours

---

3. **TC7**

Gel buffer:	9.6 mM Tris, 3mM Citric acid, pH7.0
Electrode buffer:	135 mM Tris, 43mM Citric acid, pH 7.0
Stock solution:	Dissolve the following in distilled water and make up to 2 litres 163.5 g Tris 90.4 g Citric acid H20

Running buffers:

Gel buffer:	3.5ml stock solution diluted to 250 ml with distilled water
Electrode buffer	1 part stock solution plus 4 parts distilled water
Running conditions:	30-35 mA/gel at 175-200V for 5 hours

---

4. **LI:**

Gel buffer:	48.6 mM Tris, 7.8 mM Citric acid, 3.2 mM LiOH, 20.7 mM Boric acid pH8.4
Electrode buffer:	192mM Boric acid, 30 mM LiOH, pH 8.15
Stock solution: (Electrode stock solution	Dissolve the following in distilled water and make up to 2 litres 12,6 g LiOH 118.9 g Boric acid
Stock solution: (Gel stock solution	Dissolve the following in distilled water and make up to 2 litres 54.5 g Tris 15.1 g Citric acid H2O 200ml Electrode Stock solution

Running buffers:

Gel buffer:	27ml stock solution diluted to 250 ml with distilled water
Electrode buffer	1 part stock solution plus 4 parts distilled water
Running conditions:	40-45 mA/gel at 300-400 V (mosto often 400V) for 5 hours (It is normally necessary to cool down the gel with ice)

APPENDIX IIB

Enzymes' staining recepies used in the this thesis (modified from Harris and Hopkinson, 1976)

ASPARTATE AMINO TRASNFERASI — EC 2.6.1.1

TRIS –A BUFFER	40ml
L-ASPARTIC acid	200mg
KETOGLUTARIC acid	100mg
PIRIDOXAL-5-PHOSPHATE	10mg

Dissolve the above substances in a beaker and check that pH is at least 7.4 (is not adjust with TRIS 1M

Incubate with the gel for 30 minutes at room temperature, than add:
FAST BLUE BB

ACONITASE — EC 4.2.1.3

TRIS – B BUFFER	18ml
CIS-ACONITIC acid	75mg
Mg Cl2 (0.5M)	5ml
NADP (1% IN H2O)	0.5ml

Before staining add:

MTT (1% IN H2O)	18ml
PMS (1% IN H2O)	0.2ml
AGAROSE (1% IN H2O)	25ml

ALCOHOL DEHYDROGENASE — EC 1.1.1.1

TRIS-A buffer	40ml
MgCl2 (0.5M)	0.2ml
ETHANOL (95°)	3ml
NAD (1% IN H2O)	2ml

Before use add:

NBT (1%H2O)	1ml
MTT (1% in H2O)	0.3ml
PMS (1% in H2O)	0.5ml

ADENYLATE KINASE — EC 2.7.4.3

TRIS-A buffer	5ml
MgCl2 (0.5M)	1.2ml
H2O	45ml
D+ - GLUCOSE	450ml
ADP	10mg
NAD (1% in H2O)	1ml
Before use add:	
HEXOKINASE (1.7U)	6µl
GLUCOSE-6-PHOSPHATE DEHYDROGENASE (17U)	6µl
Incubate for 60 minutes than add:	
PMS (1% in H2O)	0.5ml

ALDOLASE **EC 4.1.2.13**

TRIS-A buffer	25ml
FRUCTOSE-1,6-DIPHOSPHATE	100mg
NAD (1% in H2O)	2ml
SODIUM ARSENATE	60mg
Before staining add:	
G-3PDH (800 U/ML)	50µl
MTT (1% in H2O)	0.74ml
PMS (1% in H2O)	0.25ml
AGAROSE (1.5% in H2O)	

CREATINE KINASE **EC 2.7.3.2**

TRIS-A buffer	15ml
D+ GLUCOSE	100mg
MgCl2 (0.5M)	1ml
ADP	30ml
PHOSPHOCREATINE	15mg
NAD (1% in H2O)	1ml
NADP (1% in H2O)	0.5ml
Before use add:	
GLUCOSE-6-PHOSPHATE DEHYDROGENASE (7U)	6µl
HEXOKINASE (1.7U)	6µl
PMS (1% in H2O)	0.5ml
MTT (1% in H2O)	0.5ml
AGAROSE (1.5% in H2O)	10ml

FUMARASE — EC 4.2.1.2

TRIS-A buffer	40ml
FUMARIC acid	100mg
PYRUVATE	30mg
NAD (1% in H2O)	2ml

Before staining add:

MALATE DEHYDROGENASE	100µl
NBT (1% in H2O)	1ml
PMS (1% in H2O)	0.5ml

GLYCERALDEHYDEPHOSPHATE DEHYDROGENASE — EC 1.2.12

TRIS-A buffer	10ml
NAD (1% in H2O)	2ml
PYRUVATE	50mg
SODIUM ARSENATE	50mg

60 minutes before staining, make up:

FRUCTOSE-1.6 DIPHOSPHATE	50mg
ALDOLASE (27U)	30µl
TRIS A buffer	1ml
H2O	2ml

At staining mix the above solutions and add:

MTT (1% in H2O)	0.4ml
PMS (1% in H2O)	0.2ml
AGAROSE (1.5% in H2O)	15ml

ALFA – GLYCEROPHOSPHATE DEHYDROGENASE — EC 1.1.1.8

TRIS-A buffer	40ml
D.L – GLYCEROPHOSPHATE	250mg
MgCl2 (0.5M)	0.2ml
NAD (1% in H2O)	2ml

Before staining add:

NBT (1% in H2O)	1ml
MTT (1% in H2O)	0.3ml
PMS (1% in H2O)	0.5ml

GLUCOSEPHOSPHATE ISOMERASE — EC 5.3.1.9

TRIS A buffer	10ml
MgCl2	1ml
FREUCTOSE 6 PHOSPHATE	10mg
NAD (1% in H2O)	1ml
NADP (1% in H2O)	0.5mg

Before staining add:

GLUCOSE 6 PHOSPHATE ISOMERASE (17U)	6µl
PMS (1% in H2O)	0.5ml
NBT (1% in H2O)	0.5ml
MTT (1% in H2O)	0.5ml
AGAROSE (1.5% IN WARM H2O)	10ml

HEXOKINASE — EC 2.7.1.1

TRIS-A buffer	10ml
ATP	20mg
D-GLUCOSE	500mg
MgCl2 (0.5M)	1ml
NAD (1% in H2O)	1ml
NADP (1% in H2O)	0.5ml

Before staining add:

NBT (1% in H2O)	1ml
GLUCOSE-6-PHOSPHATE DEHYDROGENASE (17U)	6µl
PMS (1% in H2O)	0.5ml
AGAROSE (1.5% in H2O)	10ml

ISOCITRATE DEHYDROGENASE — EC 1.1.1.42

TRIS-A buffer	10ml
MgCl2 (0.5M)	1ml
NADP (1% in H2O)	0.3ml

Before staining add:

NBT (1% in H2O)	0.3ml
PMS (1% in H2O)	0.3ml
MTT (1% in H2O)	0.3ml

Incubate withgel for 30minutesnthan add:

D.L- ISOCITRIC ACID	0.1M

Preparation of 0.1m citric acid na-d-l- isocotric acid 2.9g
H2O make up to 100ml

LACTATE DEHYDROGENASE	**EC 1.1.1.27**
TRIS-A buffer	35ml
D.L- LACTIC acid (0.5M)	6ml
NAD (1% in H2O)	1ml
Before use add:	
NBT (1% in H2O)	0.3ml
PMS (1% in H2O)	0.5ml
Preparation of lactic acid:	
lithium D.L-LACTATE	9.6g
H20, make up to (adjust pH with LiOH)	200ml

MALIC ENZYME	**EC 1.1.1.40**
TRIS-A buffer	10ml
MgCl2 (0.5M)	1.5ml
MALIC acid 2M (pH7)	1ml
NADP (1% in H2O)	0.1ml
Before use add:	
PMS (1% in H2O)	0.1ml
NBT (1% in H2O)	0.2ml
MTT (1% in H2O)	0.2ml

MANNOSE PHOSPHATE ISOMERASE	
Mannose – 6 – phosphate	35mg
Glucose – 6 – phosphate dehydrogenase	60U
Glucose phosphate isomerase	100U
NAPD	10mg
Before use add	
MTT (1% in H2O)	0.5ml
PMS (1% in H2O)	0.5ml

PHOSPHOGLUCOMUTASE

Glucose – 1 – phosphate	100mg
Glucose –6 – phosphate dehydrogenase	60U
NAPD	10mg

Before use add:

MTT (1% n H2O)	1ml
PMS (1% in H2O)	0.5ml

PHOSPHOGLUCONATE DEHYDROGENASE

6 – phosphogluconate	80mg
NAPD	10mg
MTT (1% in H2O)	1ml

SORBITOL DEHYDROGENASE

Sorbitol	250mg
NAD (1% in H2O)	20mg
MTT (1% in H2O)	0.5ml
PMS (1% in H2O)	0.5ml

SUPEROXIDE DISMUTASE

NBT	20mg
PMS	6mg
NAD	20mg
(leave exposed to ight)	

XANTHINE DEHYDROGENASE

Hypoxantine	60mg
NAD	20mg
MTT (1% in H2O)	1ml
PMS (1% in H2O)	1ml

Heat substrate until it dissolves (do not bring to boil) and allow to cool down to room temperature before adding other ingredients.

APPENDIX IIC

1. Hardy-Weinberg equilibrium of observed and expected genotypes. : Chi square (χ^2) test for significant departure form Hardy-Weinberg equilibrium are also presented. d.f. = degree of freedom.

Population : South Africa

LDH

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	20	19.1010	0.0423
(B, A)	22	23.7980	0.1358
(B, B)	8	7.1010	0.1138

χ^2 test for Hardy-Weinberg equilibrium :

χ^2: 0.291964
Df : 1
p-Value : 0.588965

LGG

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	0	0.0000	0.0000
(B, A)	0	0.0000	0.0000
(B, B)	46	46.0606	0.0001
(C, A)	0	0.0000	0.0000
(C, B)	4	3.8788	0.0038
(C, C)	0	0.0606	0.0606

χ^2 test for Hardy-Weinberg equilibrium :

χ^2: 0.064474
Df: 1
p-Value : 0.799560

MDH

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	0	0.0101	0.0101
(B, A)	2	1.9798	0.0002
(B, B)	48	48.0101	0.0000

χ^2 test for Hardy-Weinberg equilibrium :

χ^2: 0.010309
Df: 1
p-Value : 0.919126

MPI

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	3	3.2828	0.0244
(B, A)	20	18.9091	0.0629
(B, B)	25	25.8182	0.0259
(C, A)	0	0.5253	0.5253
(C, B)	2	1.4545	0.2045
(C, C)	0	0.0101	0.0101

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 0.853131
Df: 3
p-Value : 0.836721

Population: Mozambique

LDH

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	17	15.2475	0.2014
(B, A)	22	25.5050	0.4817
(B, B)	12	10.2475	0.2997

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 0.982778
Df: 1
p-Value : 0.321514

LGG

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	0	0.0000	0.0000
(B, A)	0	0.0000	0.0000
(B, B)	47	47.0594	0.0001
(C, A)	0	0.0000	0.0000
(C, B)	4	3.8812	0.0036
(C, C)	0	0.0594	0.0594

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 0.063118
Df: 1
p-Value : 0.801634

MDH

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	0	0.0099	0.0099
(B, A)	2	1.9802	0.0002
(B, B)	49	49.0099	0.0000

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 0.010101
Df: 1
p-Value : 0.919944

MPI

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	6	5.2277	0.1141
(B, A)	21	21.5644	0.0148
(B, B)	21	21.2376	0.0027
(C, A)	0	0.9802	0.9802
(C, B)	3	1.9604	0.5513
(C, C)	0	0.0297	0.0297

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 1.692721
Df: 3
p-Value : 0.638553

Population: Tanzania

LDH

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	13	15.5556	0.4198
(B, A)	30	24.8889	1.0496
(B, B)	7	9.5556	0.6835

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 2.152907
Df: 1
p-Value : 0.142300

LGG

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	0	0.0000	0.0000
(B, A)	0	0.0000	0.0000
(B, B)	44	44.1515	0.0005
(C, A)	0	0.0000	0.0000
(C, B)	6	5.6970	0.0161
(C, C)	0	0.1515	0.1515

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 0.168154
Df: 1
p-Value : 0.681758

MPI

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	6	4.6970	0.3615
(B, A)	19	20.6667	0.1344
(B, B)	22	21.6667	0.0051
(C, A)	0	0.9394	0.9394
(C, B)	3	2.0000	0.5000
(C, C)	0	0.0303	0.0303

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 1.970720
Df: 3
p-Value : 0.578506

Population : Madagascar

LDH

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	20	17.8788	0.2517
(B, A)	20	24.2424	0.7424
(B, B)	10	7.8788	0.5711

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 1.565189
Df: 1
p-Value : 0.210907

LGG

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	0	0.0000	0.0000
(B, A)	0	0.0000	0.0000
(B, B)	46	46.0606	0.0001
(C, A)	0	0.0000	0.0000
(C, B)	4	3.8788	0.0038
(C, C)	0	0.0606	0.0606

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 0.064474
Df: 1
p-Value : 0.799560

MDH

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	0	0.1010	0.1010
(B, A)	5	4.7980	0.0085
(B, B)	45	45.1010	0.0002

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 0.109742
Df: 1
p-Value : 0.740438

MPI

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	5	6.7273	0.4435
(B, A)	27	22.7980	0.7745
(B, B)	16	18.4848	0.3340
(C, A)	0	0.7475	0.7475
(C, B)	2	1.2323	0.4782
(C, C)	0	0.0101	0.0101

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 2.787816
Df: 3
p-Value : 0.425510

Population: Oman

LDH

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	9	8.6975	0.0105
(B, A)	28	28.6050	0.0128
(B, B)	23	22.6975	0.0040

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 0.027352
Df: 1
p-Value : 0.868641

MPI

Genotypes	Obs. (O)	Exp. (E)	$(O-E)^2/E$
(A, A)	15	12.9412	0.3275
(B, A)	26	27.7647	0.1122
(B, B)	14	14.3782	0.0099
(C, A)	0	2.3529	2.3529
(C, B)	5	2.4790	2.5637
(C, C)	0	0.0840	0.0840

χ^2 test for Hardy-Weinberg equilibrium :
χ^2: 5.450361
Df: 3
p-Value : 0.141638

2. Allele frequencies of polymorphic loci

South Africa

Allele \ Locus	**LDH**	**LGG**	**MDH**	**MPI**
Allele A	0.6200		0.0200	0.2600
Allele B	0.3800	0.9600	0.9800	0.7200
Allele C		0.0400		0.0200

Mozambique

Allele \ Locus	LDH	LGG	MDH	MPI
Allele A	0.5490		0.0196	0.3235
Allele B	0.4510	0.9608	0.9804	0.6471
Allele C		0.0392		0.0294

Madagascar

Allele \ Locus	LDH	LGG	MDH	MPI
Allele A	0.6000		0.0500	0.3700
Allele B	0.4000	0.9600	0.9500	0.6100
Allele C		0.0400		0.0200

Tanzania

Allele \ Locus	LDH	LGG	MDH	MPI
Allele A	0.5600			0.3100
Allele B	0.4400	0.9400	1.0000	0.6600
Allele C		0.0600		0.0300

Oman

Allele \ Locus	LDH	LGG	MDH	MPI
Allele A	0.3833			0.4667
Allele B	0.6167	1.0000	1.0000	0.4917
Allele C				0.0417

APPENDIX III A
RAPDS

Frequency						
p	0,625	0,975	0,95	0,3	0,675	0,725
q	0,375	0,025	0,05	0,7	0,325	0,275
Tot Sample	0,241699219	0,001249219	0,0049875	0,4998	0,188936719	0,139811719
p	0,625	1	0,875	0,375	0,875	0,75
q	0,375	0	0,125	0,625	0,125	0,25
SAfr	0,241699219	0	0,030762	0,476074219	0,030761719	0,1171875
p	0,625	0,875	0,875	0,5	0,625	0,75
q	0,375	0,125	0,125	0,5	0,375	0,25
Moza	0,241699219	0,030761719	0,030761719	0,375	0,241699219	0,1171875
p	0,625	1	1	0,125	0,625	0,625
q	0,375	0	0	0,875	0,375	0,375
Tanz	0,241699219	0	0	0,358886719	0,241699219	0,241699219
p	0,625	1	1	0,125	0,625	0,75
q	0,375	0	0	0,875	0,375	0,25
Mada	0,241699219	0	0	0,358886719	0,241699219	0,1171875
p	0,625	1	1	0,375	0,625	0,75
q	0,375	0	0	0,625	0,375	0,25
Oman	0,241699219	0	0	0,476074219	0,241699219	0,1171875

Frequency						
p	0,875	1	1	1	0,075	0,875
q	0,125	0	0	0	0,925	0,125
Tot Sample	0,030761719	0	0	0	0,247061719	0,030761719
p	0,875	1	1	1	0	0,875
q	0,125	0	0	0	1	0,125
SAfr	0,030761719	0	0	0	0	0,030761719
p	0,875	1	1	1	0	0,875
q	0,125	0	0	0	1	0,125
Moza	0,030761719	0	0	0	0	0,030761719
p	0,875	1	1	1	0,125	0,875
q	0,125	0	0	0	0,875	0,125
Tanz	0,030761719	0	0	0	0,358886719	0,030761719
p	0,875	1	1	1	0,25	0,875
q	0,125	0	0	0	0,75	0,125
Mada	0,030761719	0	0	0	0,4921875	0,030761719
p	0,875	1	1	1	0	0,875
q	0,125	0	0	0	1	0,125
Oman	0,030761719	0	0	0	0	0,030761719

Frequency						
p	1	0,8	1	1	1	1
q	0	0,2	0	0	0	0
Tot Sample	0	0,0768	0	0	0	0
p	1	0,625	1	1	1	1
q	0	0,375	0	0	0	0
SAfr	0	0,241699219	0	0	0	0
p	1	0,75	1	1	1	1
q	0	0,25	0	0	0	0
Moza	0	0,1171875	0	0	0	0
p	1	0,875	1	1	1	1
q	0	0,125	0	0	0	0
Tanz	0	0,030761719	0	0	0	0
p	1	0,875	1	1	1	1
q	0	0,125	0	0	0	0
Mada	0	0,030761719	0	0	0	0
p	1	0,875	1	1	1	1
q	0	0,125	0	0	0	0
Oman	0	0,030761719	0	0	0	0

Frequency						
p	1	0,8	0,875	1	1	0,25
q	0	0,2	0,125	0	0	0,75
Tot Sample	0	0,0768	0,030761719	0	0	0,4921875
p	1	0,875	0,875	1	1	0,125
q	0	0,125	0,125	0	0	0,875
SAfr	0	0,030761719	0,030761719	0	0	0,358886719
p	1	0,75	0,875	1	1	0,125
q	0	0,25	0,125	0	0	0,875
Moza	0	0,1171875	0,030761719	0	0	0,358886719
p	1	0,875	0,875	1	1	0,375
q	0	0,125	0,125	0	0	0,625
Tanz	0	0,030761719	0,030761719	0	0	0,476074219
p	1	0,75	0,875	1	1	0,375
q	0	0,25	0,125	0	0	0,625
Mada	0	0,1171875	0,030761719	0	0	0,476074219
p	1	0,75	0,875	1	1	0,25
q	0	0,25	0,125	0	0	0,75
Oman	0	0,1171875	0,030761719	0	0	0,4921875

Frequency						
p	0,075	0,725	1	0,375	1	0,7
q	0,925	0,275	0	0,625	0	0,3
Tot Sample	0,247061719	0,139811719	0	0,476074219	0	0,1638
p	0	0	0	0,5	1	0,875
q	1	1	1	0,5	0	0,125
SAfr	0	0	0	0,375	0	0,030761719
p	0,125	1	1	0,375	1	1
q	0,875	0	0	0,625	0	0
Moza	0,358886719	0	0	0,476074219	0	0
p	0,125	1	1	0,625	1	0,5
q	0,875	0	0	0,375	0	0,5
Tanz	0,358886719	0	0	0,241699219	0	0,375
p	0	0,375	1	0,25	1	0,5
q	1	0,625	0	0,75	0	0,5
Mada	0	0,476074219	0	0,4921875	0	0,375
p	0,125	0,25	1	0,125	1	0,625
q	0,875	0,75	0	0,875	0	0,375
Oman	0,358886719	0,4921875	0	0,358886719	0	0,241699219

Frequency						
p	0,375	0,6	0,65	0,8	0,6	0,875
q	0,625	0,4	0,35	0,2	0,4	0,125
Tot Sample	0,476074219	0,2688	0,2149875	0,0768	0,2688	0,030761719
p	0,75	0,75	1	0,875	0,625	0,875
q	0,25	0,25	0	0,125	0,375	0,125
SAfr	0,1171875	0,1171875	0	0,030761719	0,241699219	0,030761719
p	0,5	0,5	0,875	0,75	0,75	0,875
q	0,5	0,5	0,125	0,25	0,25	0,125
Moza	0,375	0,375	0,030761719	0,1171875	0,1171875	0,030761719
p	0,5	0,625	0,75	0,875	0,75	0,875
q	0,5	0,375	0,25	0,125	0,25	0,125
Tanz	0,375	0,241699219	0,1171875	0,030761719	0,1171875	0,030761719
p	0	0,625	0,375	0,75	0,625	0,875
q	1	0,375	0,625	0,25	0,375	0,125
Mada	0	0,241699219	0,476074219	0,1171875	0,241699219	0,030761719
p	0,125	0,5	0,25	0,75	0,25	0,875
q	0,875	0,5	0,75	0,25	0,75	0,125
Oman	0,358886719	0,375	0,4921875	0,1171875	0,4921875	0,030761719

Frequency						
p	0,875	0,85	1	1	0,575	0,875
q	0,125	0,15	0	0	0,425	0,125
Tot Sample	0,030761719	0,0439875	0	0	0,295999219	0,030761719
p	1	1	1	1	0,75	0,875
q	0	0	0	0	0,25	0,125
SAfr	0	0	0	0	0,1171875	0,030761719
p	0,875	0,875	1	1	0,5	0,875
q	0,125	0,125	0	0	0,5	0,125
Moza	0,030761719	0,030761719	0	0	0,375	0,030761719
p	0,875	0,875	1	1	0,5	0,875
q	0,125	0,125	0	0	0,5	0,125
Tanz	0,030761719	0,030761719	0	0	0,375	0,030761719
p	0,75	0,75	1	1	0,75	0,875
q	0,25	0,25	0	0	0,25	0,125
Mada	0,1171875	0,1171875	0	0	0,1171875	0,030761719
p	0,875	0,75	1	1	0,375	0,875
q	0,125	0,25	0	0	0,625	0,125
Oman	0,030761719	0,1171875	0	0	0,476074219	0,030761719

Frequency						
p	0,775	1	1	0,6	1	0,475
q	0,225	0	0	0,4	0	0,525
Tot Sample	0,096124219	0	0	0,2688	0	0,399311719
p	0,875	1	1	0,625	1	0,875
q	0,125	0	0	0,375	0	0,125
SAfr	0,030761719	0	0	0,241699219	0	0,030761719
p	0,875	1	1	0,75	1	0,125
q	0,125	0	0	0,25	0	0,875
Moza	0,030761719	0	0	0,1171875	0	0,358886719
p	0,875	1	1	0,75	1	0,625
q	0,125	0	0	0,25	0	0,375
Tanz	0,030761719	0	0	0,1171875	0	0,241699219
p	0,625	1	1	0,75	1	0,625
q	0,375	0	0	0,25	0	0,375
Mada	0,241699219	0	0	0,1171875	0	0,241699219
p	0,625	1	1	0,125	1	0,125
q	0,375	0	0	0,875	0	0,875
Oman	0,241699219	0	0	0,358886719	0	0,358886719

Frequency						
p	0,85	1	1	0,55	1	1
q	0,15	0	0	0,45	0	0
Tot Sample	0,0439875	0	0	0,3229875	0	0
p	0,875	1	1	1	1	1
q	0,125	0	0	0	0	0
SAfr	0,030761719	0	0	0	0	0
p	0,75	1	1	0,75	1	1
q	0,25	0	0	0,25	0	0
Moza	0,1171875	0	0	0,1171875	0	0
p	0,875	1	1	0,75	1	1
q	0,125	0	0	0,25	0	0
Tanz	0,030761719	0	0	0,1171875	0	0
p	0,875	1	1	0,625	1	1
q	0,125	0	0	0,375	0	0
Mada	0,030761719	0	0	0,241699219	0	0
p	0,875	1	1	0,625	1	1
q	0,125	0	0	0,375	0	0
Oman	0,030761719	0	0	0,241699219	0	0

Frequency						
p	0,6	0,725	1	0,8	1	0,875
q	0,4	0,275	0	0,2	0	0,125
Tot Sample	0,2688	0,139811719	0	0,0768	0	0,030761719
p	0,625	0,875	1	1	1	1
q	0,375	0,125	0	0	0	0
SAfr	0,241699219	0,030761719	0	0	0	0
p	0,625	0,75	1	0,875	1	1
q	0,375	0,25	0	0,125	0	0
Moza	0,241699219	0,1171875	0	0,030761719	0	0
p	0,75	0,875	1	0,75	1	0,875
q	0,25	0,125	0	0,25	0	0,125
Tanz	0,1171875	0,030761719	0	0,1171875	0	0,030761719
p	0,75	0,75	1	0,75	1	0,75
q	0,25	0,25	0	0,25	0	0,25
Mada	0,1171875	0,1171875	0	0,1171875	0	0,1171875
p	0,25	0,375	1	0,625	1	0,75
q	0,75	0,625	0	0,375	0	0,25
Oman	0,4921875	0,476074219	0	0,241699219	0	0,1171875

Frequency					
p	0,75	1	0,05		
q	0,25	0	0,95		
Tot Sample	0,1171875	0	0,1759875	ϕ_e =	0,107407292
				π =	0,006689326
p	0,875	1	0		
q	0,125	0	1		
SAfr	0,030761719	0	0	ϕ_e =	0,053152902
				π =	0,003310363
p	0,75	1	0,25		
q	0,25	0	0,75		
Moza	0,1171875	0	0,4921875	ϕ_e =	0,092238653
				π =	0,005744623
p	0,625	1	0		
q	0,375	0	1		
Tanz	0,241699219	0	0	ϕ_e =	0,087355841
				π =	0,005440522
p	0,75	1	0		
q	0,25	0	1		
Mada	0,1171875	0	0	ϕ_e =	0,102539063
				π =	0,006386133
p	0,75	1	0		
q	0,25	0	1		
Oman	0,1171875	0	0	ϕ_e =	0,134254092
				π =	0,008361345

APPENDIX IVA

APPENDIX IVB

APPENDIX IV

Cloning protocol procedures performed in this research for Pi06 locus:

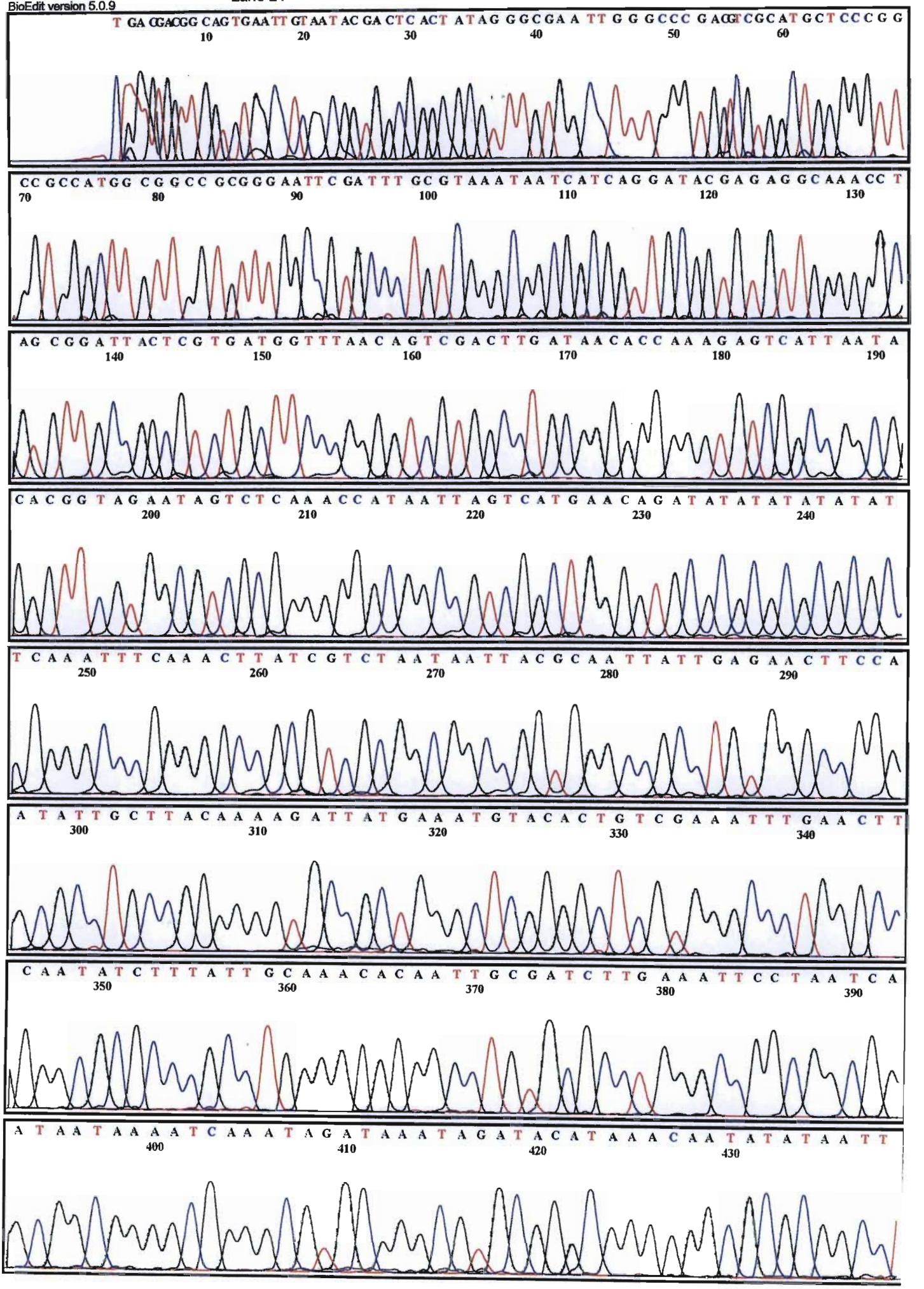
1. Preparation of the plateswith ampicillin (100µl in 50 ml of media)
2. Set ligation reaction as described in Pgem – Easy cloning kit pag 11 (use only half reaction mix)
3. Store ligation reaction overnight at 4°C
4. Take two µl of ligation reaction mix and add to 1.5ml eppendorf
5. Add 50 µl of competent cells (E.Coli: DH5α lacZΔM13)
6. Insert the eppendorf with the solution (Ligation reaction and competent cells) on ice for two minutes
7. Then place it on a heating block at 42 °C for two minutes.
8. Under the fume cupboard add 950µl of SOC media to the eppendorf.
9. Place the eppendorf at 36°C for 1.5 hr.
10. Take two plates out of the 4°C and add to each one 100µl of IPTG and 20 µl of X-GAL
11. Place then the plates upside down in the 36°C
12. Take eppendorf from the 36°C and centrifuge it at 10000 g for 10 minutes
13. Remove 700µl of supernatant and re-suspend the cells in the media left
14. Under the fume cupboards place 125µl of solution in each plates and place it in the 36°C overnight.
15. If colony grew, pick up one white colony (containing the insert) and place it in a 5ml eppendorf containing 4 ml of LB media with ampicillin 1%
16. Store in the 36°C overnight

Miniprep for 5ml culture in order to sequence: all steps are performed on ice except as stated otherwise. Centrifuge steps are performed at 4°C.

1. Pipette 1.5 ml of overnight culture into eppendorf
2. Centrifuge cells (12 000 g) for 30 seconds
3. Discard supernatant and repeat step 2 and 3
4. Re-suspend the pellet in 100 µl of ice-cold Solution 1 by vortexin
5. Incubate for a maximum of 5 minutes at room temperature (RT)
6. Add 200 µl of fresh solution 2 (0.1 mL of 2M NaOH, 0.1mL of 10 % SDS, 0.8mL of distilled water) and mix by inverting the tube rapidly 5 times.

7. Incubate on ice for 5 minutes
8. Transfer 400 µl of supernatant to a fresh tube and add Rnase (heat inactivated) at a concentration of 50µg/mL (stock 10mg/mL = 2µl) and incubate for 30 minutes at 37° C
9. Add an equal volume of phenol:chloroform:isoamyl (25:24:1) and vortex Centrifuge for 2 minutes.
10. Transfer 400 µl of iso-propanol at room temperature for 2 minutes
11. Centrifuge for 5 minutes
12. Discard supernatant and allow pellet to dry air-dry
13. Add 1 ml of ice-cold 70 % ethanol
14. Centrifuge for 5 minutes
15. Discard supernatant and dry pellet at 70°C for not more than 5 minutes and then air-dry pellet for 5 minutes
16. Re-dessolve pellet in 50 µl of milliQ water
17. Run on 1 % agarose in 1 time TBE buffer to check purity state
18. Clean with sephadex- G50 from excess salts
19. Store at –20°C

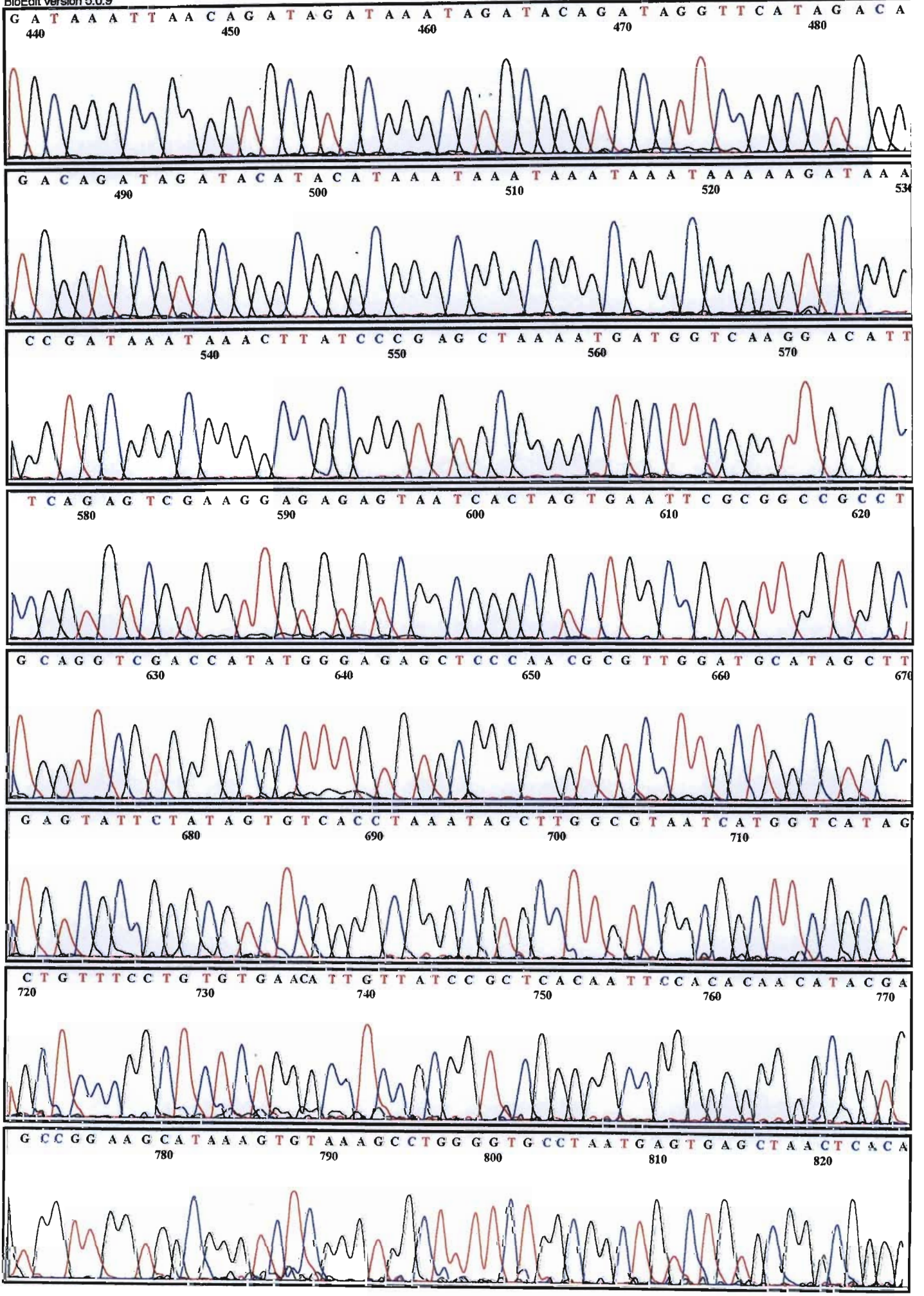
APPENDIX IVC

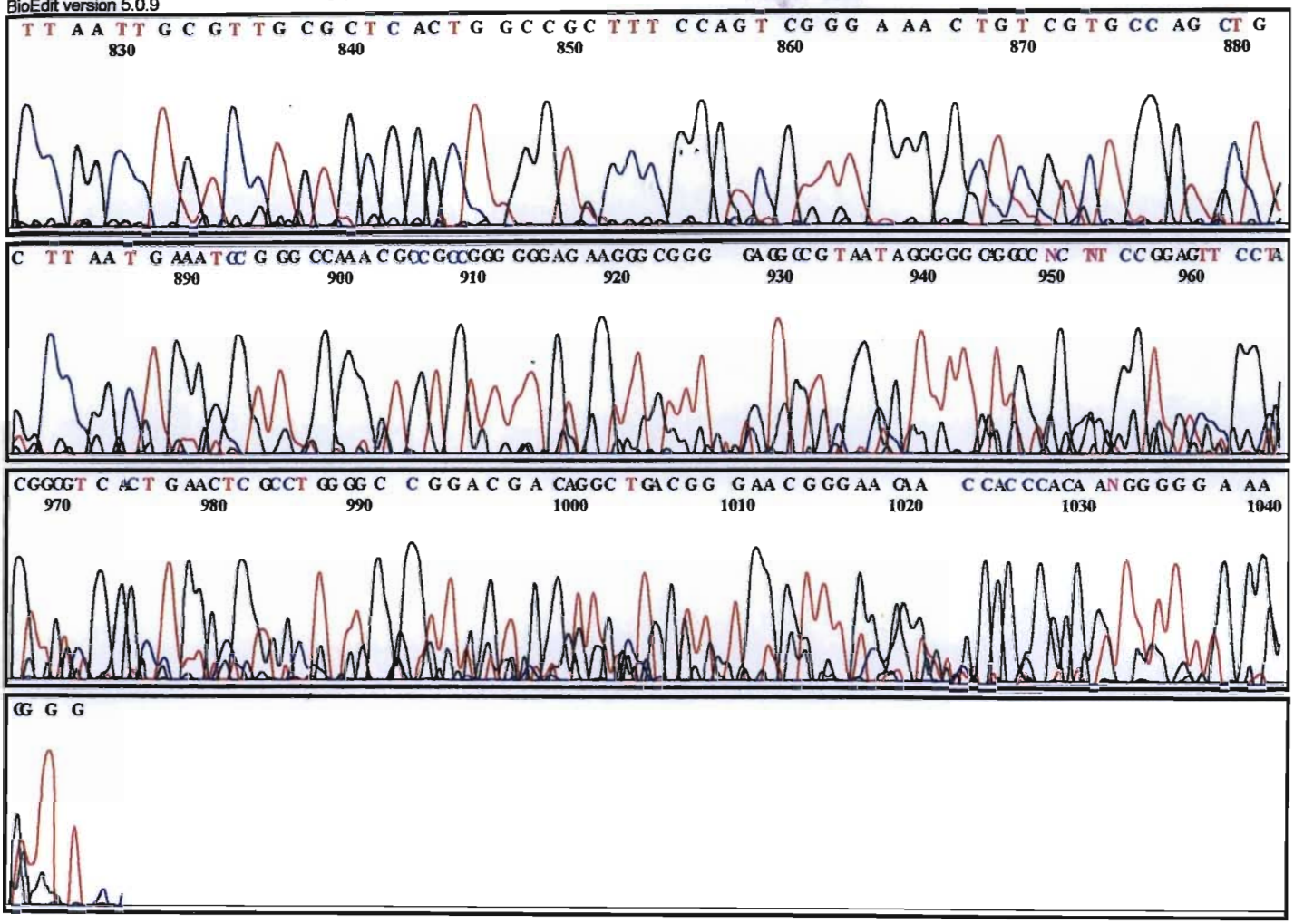

T GA CGACGG CAG TGAATT GT AAT AC GA CTC ACT ATAG GGCGA A TT G G GCCC G ACGTCGCAT GCTCCC G G
10 20 30 40 50 60
CC GCCATGG CGGCC GCGGGAATTC GATTT GC GTAAATAATCATCAGGATACGAGAGGCAAACC T
70 80 90 100 110 120 130
AG CGGATTACTCGTGATGGTTTAACAGTCGACTTGATAACACCAAAGAGTCATTAATA
140 150 160 170 180 190
CACGGTAGAATAGTCTCAAACCATAATTAGTCATGAACAGATATATATATATATAT
200 210 220 230 240
TCAAATTTCAAACTTATCGTCTAATAATTACGCAATTATTGAGAACTTCCA
250 260 270 280 290
ATATTGCTTACAAAAGATTATGAAATGTACACTGTCGAAATTTGAACTT
300 310 320 330 340
CAATATCTTTATTGCAAACACAATTGCGATCTTGAAATTCCTAATCA
350 360 370 380 390
ATAATAAAATCAAATAGATAAAATAGATACATAAACAATATATAATT
400 410 420 430

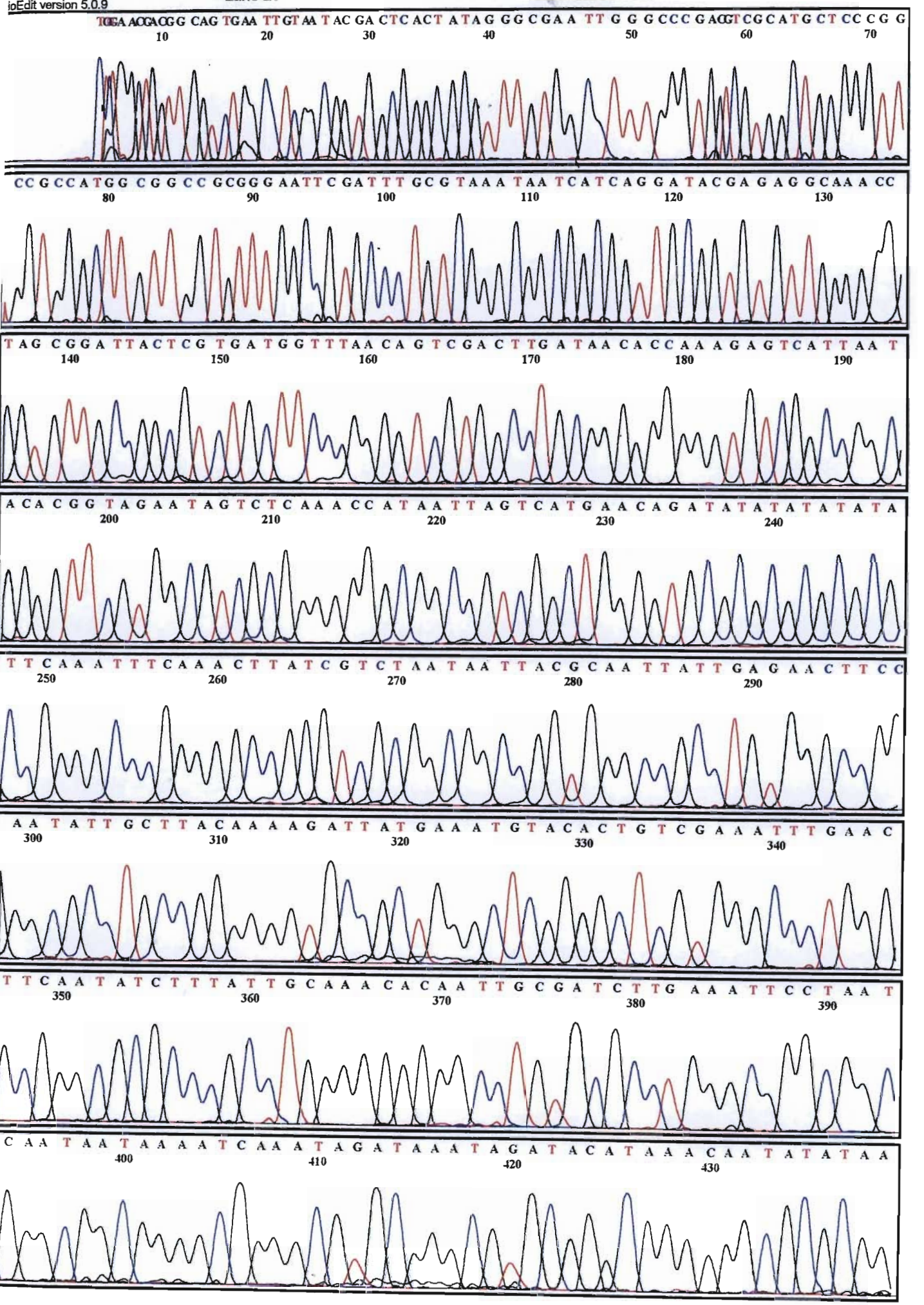
Model 1000 File: M13- Cllone01- TZ.abd
HÙ8 ETOH
Cimarron 3.12Jack_1-_M13_F
Lane 21

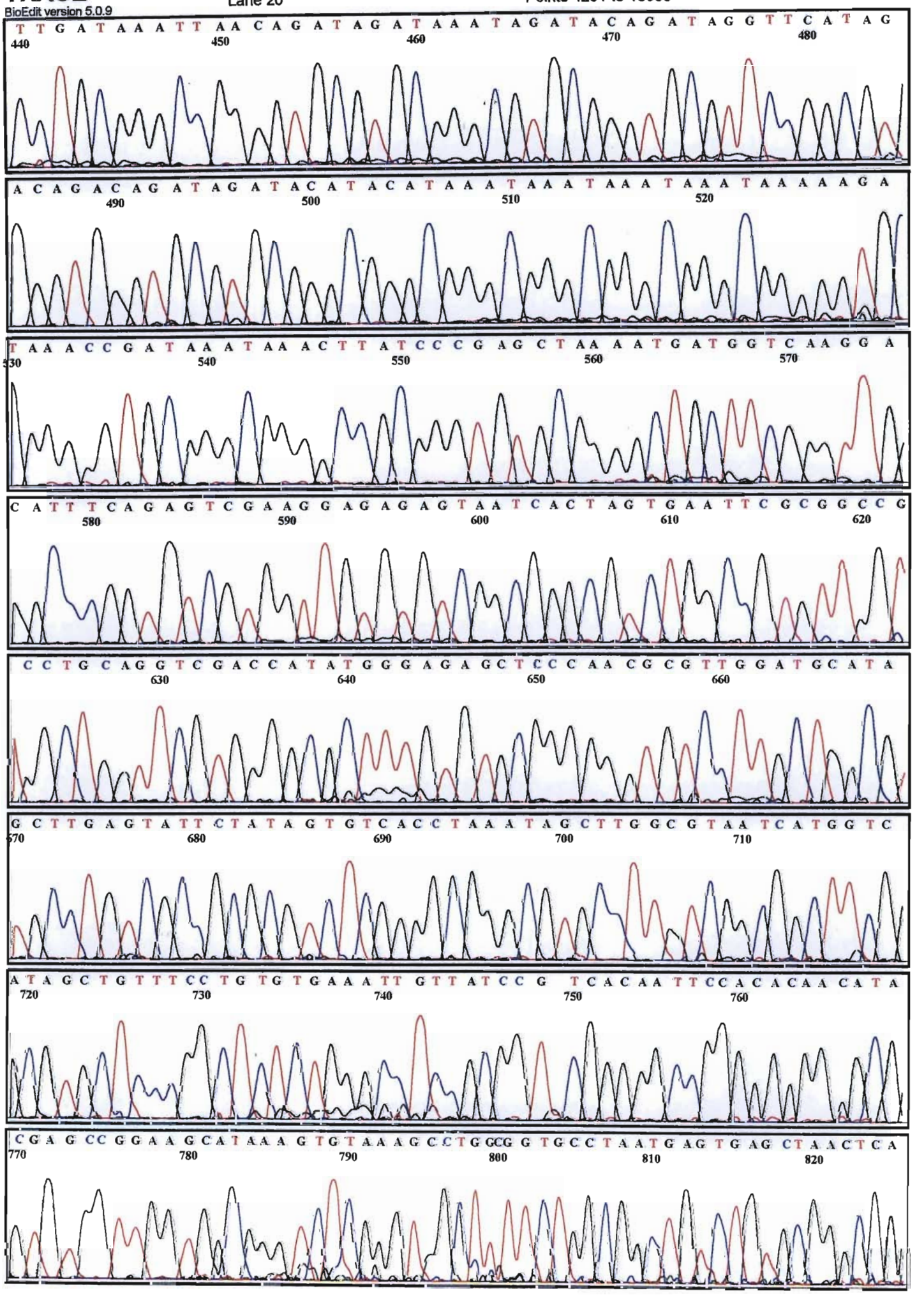
Signal C:0 A:0 G:0 T:0
ET Terminators
?? no 'MTXF' field
Points 4201 to 18956

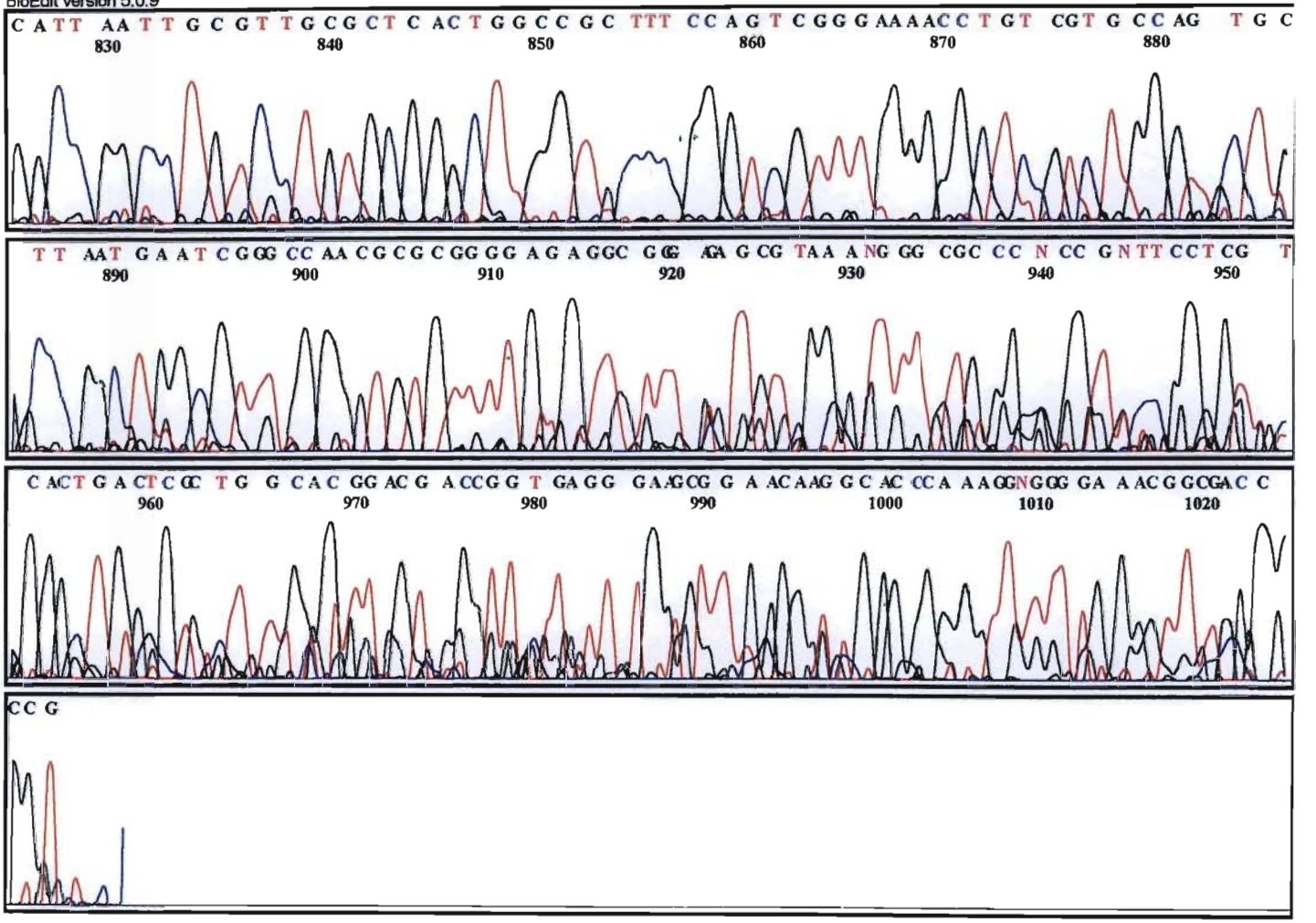
5/2/2003
Spacing: 0




T T A A T T G C G T T G C G C T C A C T G G C C G C T T T C C A G T C G G G A A A C T G T C G T G C C A G C T G
830 840 850 860 870 880
C T T A A T G A A A T C C G G C C A A A C G C C G C C G G G G G A G A A G G C G G G G A G G C C G T A A T A G G G G G C G G C C N C N T C C G G A G T T C C T A
890 900 910 920 930 940 950 960
C G G C G T C A C T G A A C T C G C C T G G G C C G G A C G A C A G G C T G A C G G G A A C G G G A A C A A C C A C C C A C A A N G G G G G A A A
970 980 990 1000 1010 1020 1030 1040
C G G G


TGGAACGACGG CAG TGAA TTGTAA T AC GA CTC ACT ATAG GGCGAA TT G G GCCC GACGTCGCATGCTCCCGG
10 20 30 40 50 60 70
CC G CC ATGG CGGCC GCGGG AATTC GATTT GC G TAAA TAA TCATCAG GA TACGA GAG GCAAA CC
80 90 100 110 120 130
T AG CGGA TTACTCG TGATGGTTTAA CA G TC GA CTTG A TAA CA CC AAA GAG TCATTAA T
140 150 160 170 180 190
A CA C GG TAGAA TAG TCTC AAA CC AT AA TTAG TCATG AA CAG A TATATATATATA
200 210 220 230 240
T TC AAA TTTC AAA A CTTATC G TCTAA T AA TTACG CAA TTATT GA G AA CTTCC
250 260 270 280 290
AA TATT GCTTACAAAAAG ATTATG AAA T G TACACTG TCGAAA TTTGAAC
300 310 320 330 340
T TCAA T ATCTTTATTG CAAA A CACAA TTG C GA TCTTG AAA TTCCTAA T
350 360 370 380 390
C AA T AA T AAAA TCAAA TAGA TAAA TAGA TACATAAA CAA TATATAA
400 410 420 430

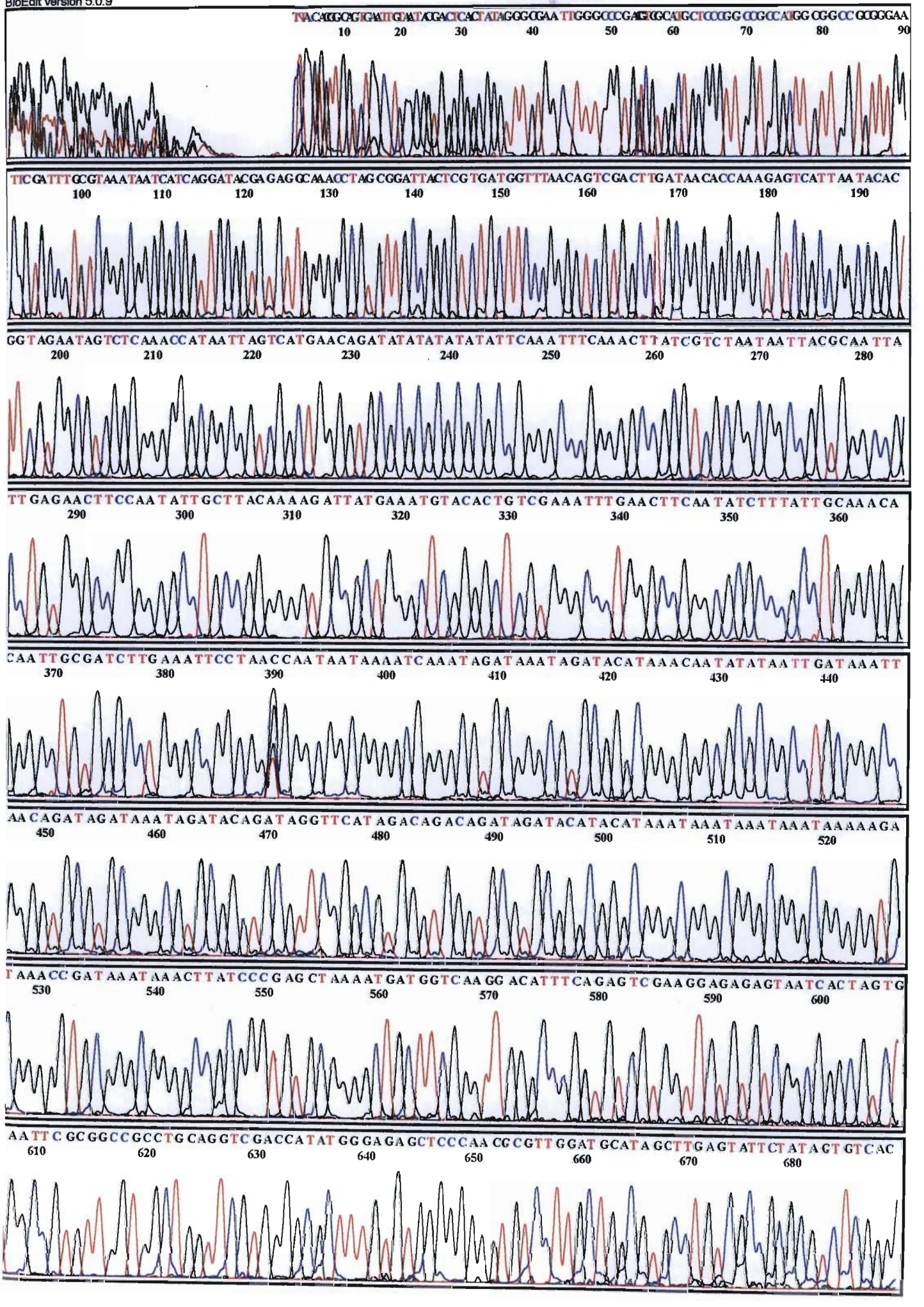

T T G A T A A A T T A A C A G A T A G A T A A A T A G A T A C A G A T A G G T T C A T A G
440 450 460 470 480
A C A G A C A G A T A G A T A C A T A C A T A A A T A A A T A A A T A A A T A A A A A A G A
490 500 510 520
T A A A C C G A T A A A T A A A C T T A T C C C C G A G C T A A A A T G A T G G T C A A G G A
530 540 550 560 570
C A T T T C A G A G T C G A A G G A G A G A G T A A T C A C T A G T G A A T T C G C G G C C G
580 590 600 610 620
C C T G C A G G T C G A C C A T A T G G G A G A G C T C C C A A C G C G T T G G A T G C A T A
630 640 650 660
G C T T G A G T A T T C T A T A G T G T C A C C T A A A T A G C T T G G C G T A A T C A T G G T C
670 680 690 700 710
A T A G C T G T T T C C T G T G T G A A A T T G T T A T C C G T C A C A A T T C C A C A C A A C A T A
720 730 740 750 760
C G A G C C G G A A G C A T A A A G T G T A A A G C C T G G G G T G C C T A A T G A G T G A G C T A A C T C A
770 780 790 800 810 820

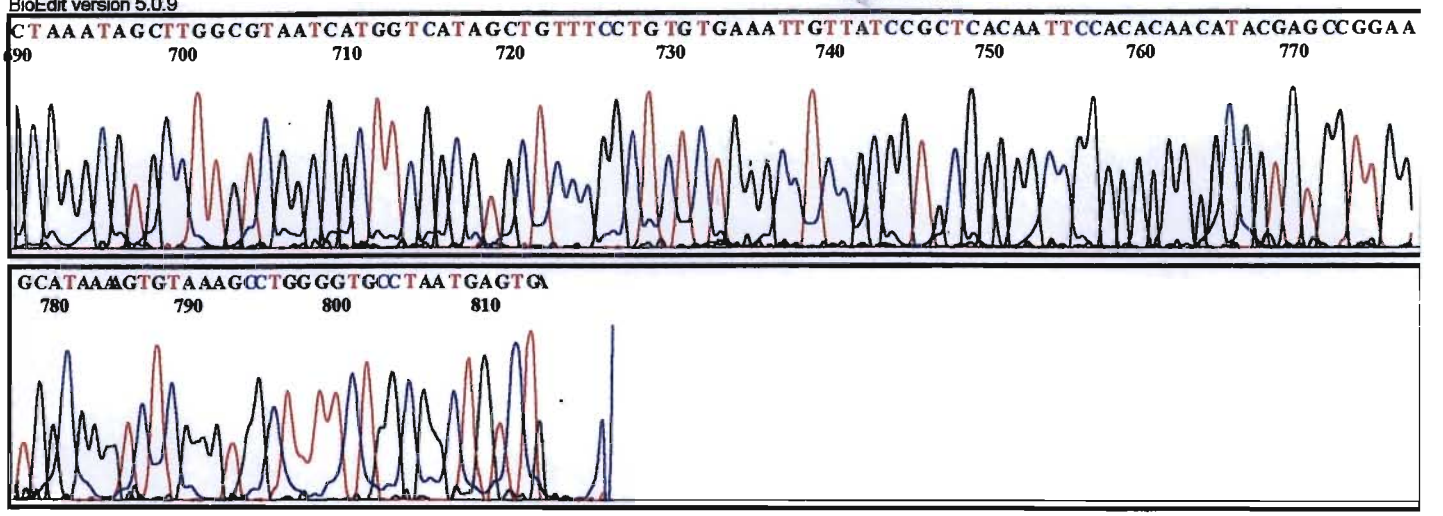

C A TT AA T T G C G T T G C G C T C AC T G G C C G C TTT CC A G T C G G G AA A AC C T G T CG T G C C AG T G C
830
840
850
860
870
880
T T AAT G A A T C G GG CC AA C G C G C GG G G A G A G GC G GG AA G C G TA A A NG GG CG C CC N CC G N T T C C T CG T
890
900
910
920
930
940
950
C AC T G A C T C CC T G G C A C GG AC G A CC G G T GA G G GA AG CG G A A C AA G G C AC CC A A A GG NG GG GA A A C G G C GA C C
960
970
980
990
1000
1010
1020
C C G

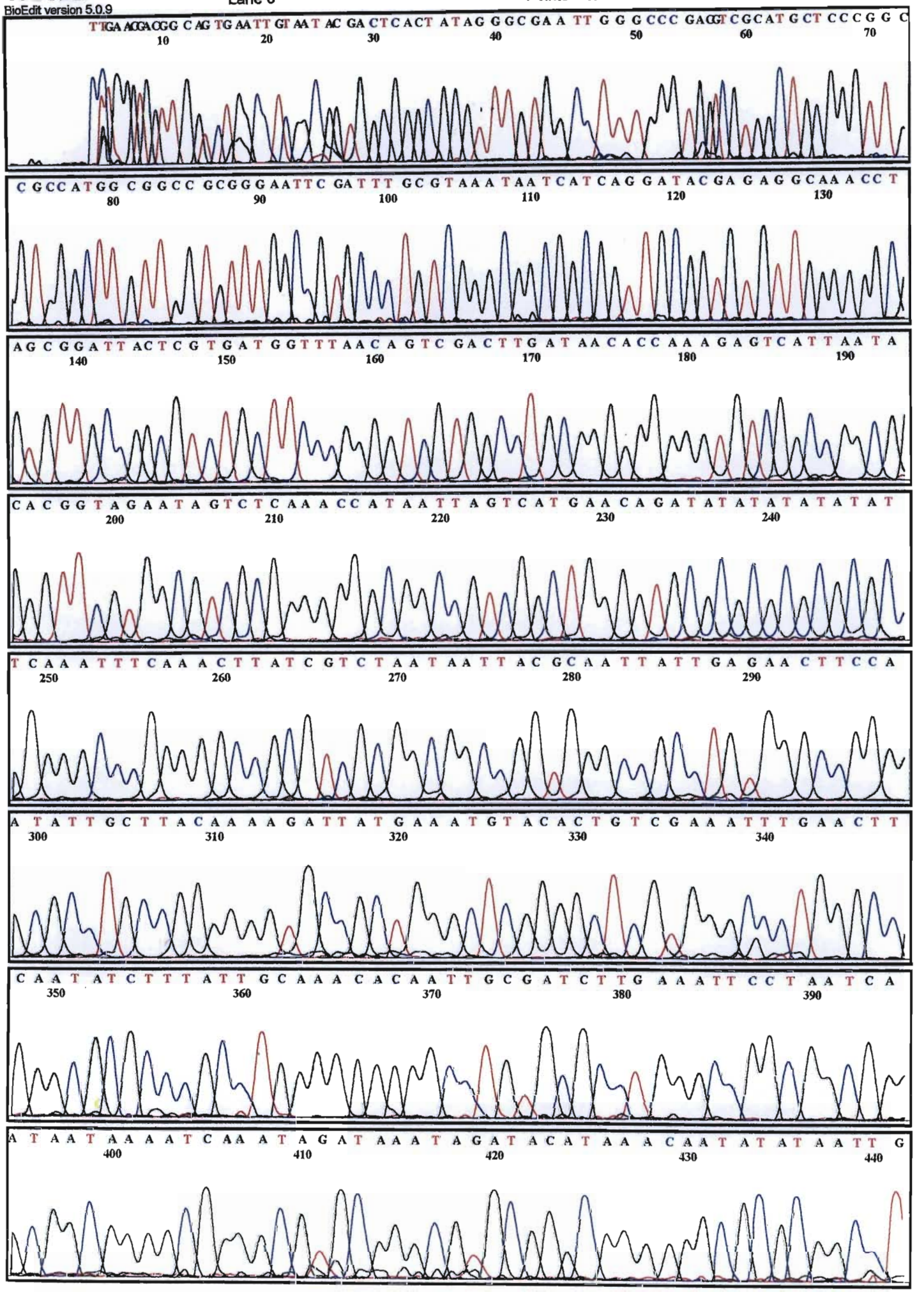
Model 1000 File: Jack_3-_M13_F.abd
HÙ8 ETOH
Cimarron 3.12Jack_3-_M13_F
Lane 19

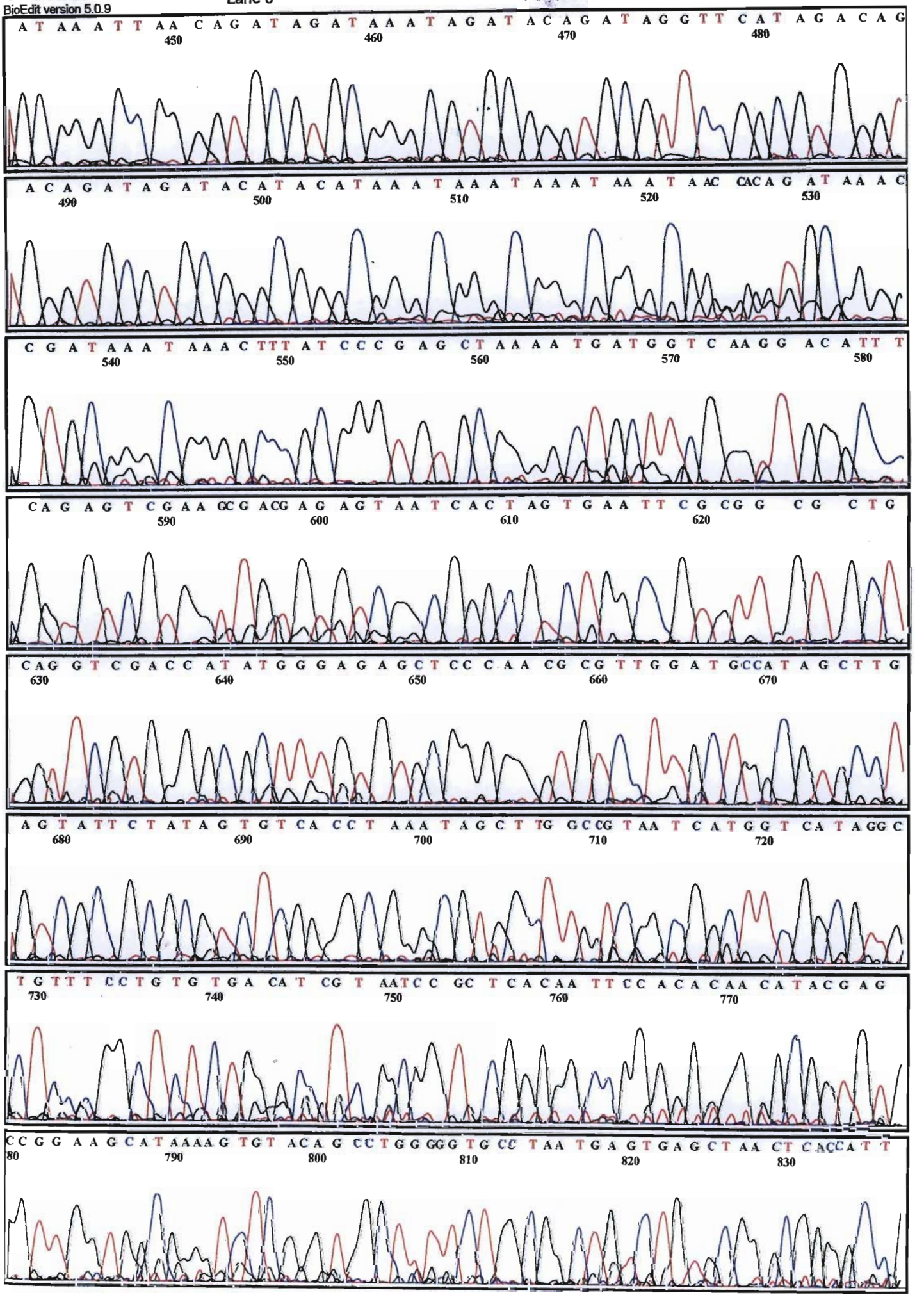
Signal C:0 A:0 G:0 T:0
ET Terminators
?? no 'MTXF' field
Points 4002 to 9363

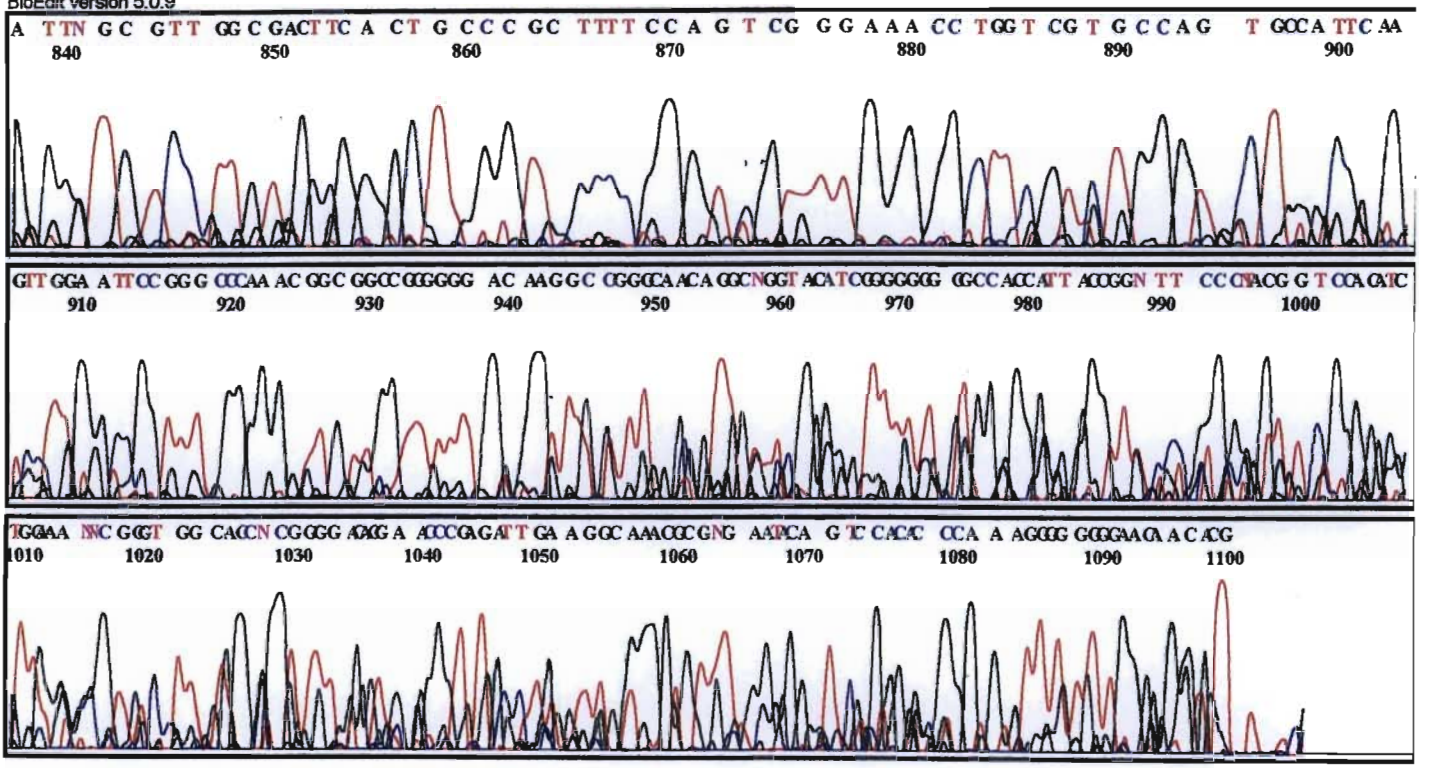
5/2/2003
Spacing: 0



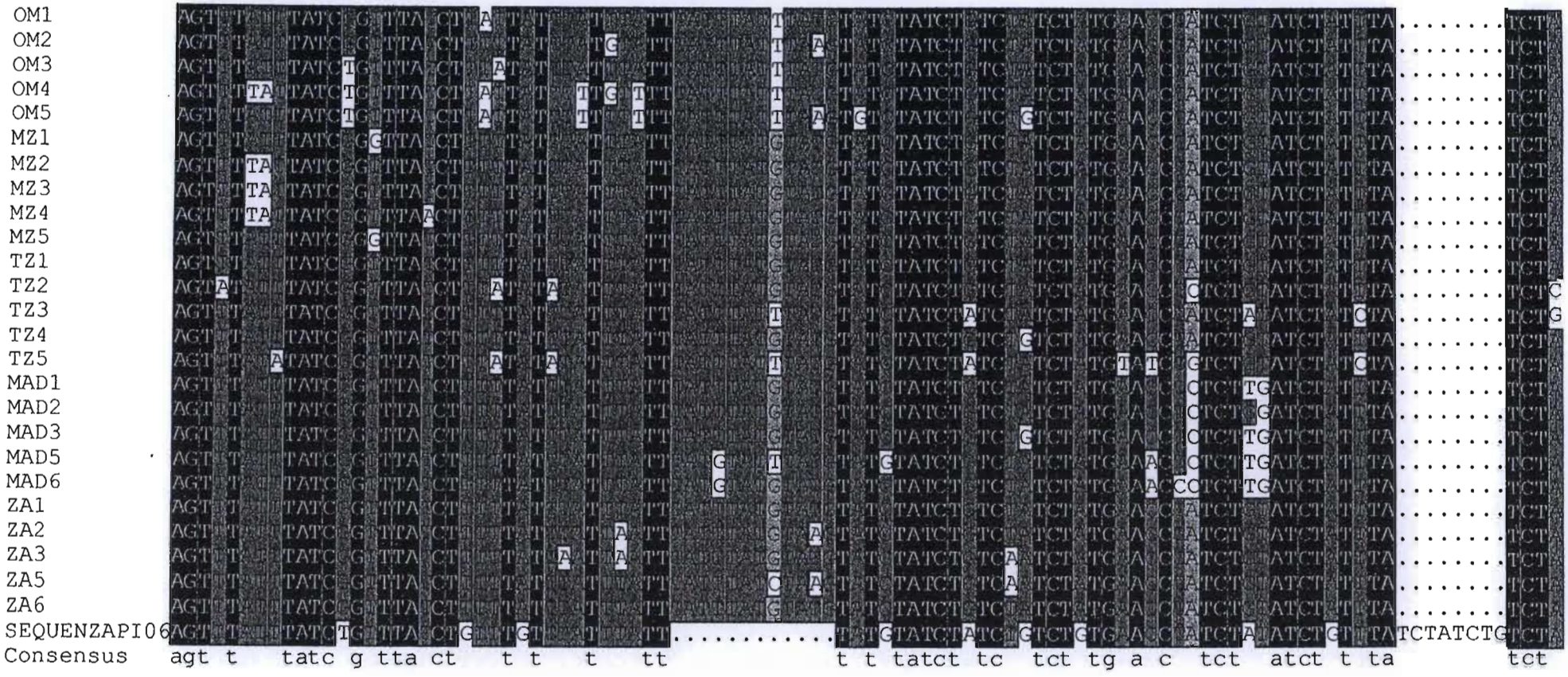

CTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAA
690
700
710
720
730
740
750
760
770
GCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGA
780
790
800
810

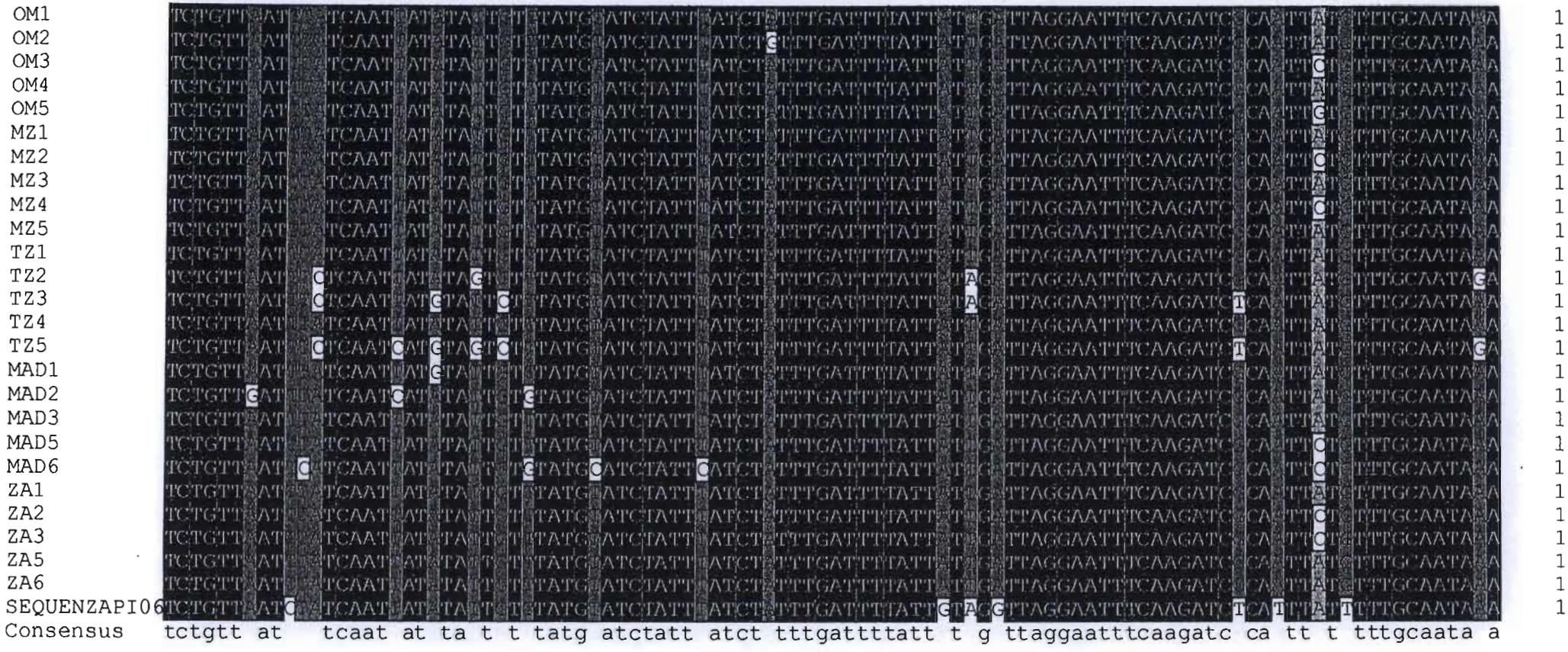

TTGAACGACGG CAG TG AATT GT AAT AC GA CTCACT ATAG GGC GAA TT G G G CCC GACGTCGCATG CT CCCGG C
10 20 30 40 50 60 70
C GCC ATGG C GGCC GCGGG AATTC GA TTT GC GT AAA T AA T CAT CAG GA T ACGA G AG GCAAA CC T
80 90 100 110 120 130
AG C GGA TT AC T C G T GA T GG TTT AA C A G T C GA C TT G A T AA C AC C AA A G AG T C A TT AA T A
140 150 160 170 180 190
CA C GG T A G AA T A G T C T C AA A C C A T AA TT A G T C A T G AA C A G A T A T A T A T A T A T
200 210 220 230 240
T C AA A T T T C AA A C T T A T C G T C T AA T AA T T AC G C AA T T A T T G A G AA C T T C C A
250 260 270 280 290
A T A T T G C T T A C AA A A G A T T A T G AA A T G T A C A C T G T C G A A A T T T G A A C T T
300 310 320 330 340
C A A T A T C T T T A T T G C A A A C A C A A T T G C G A T C T T G A A A T T C C T A A T C A
350 360 370 380 390
A T A A T A A A A T C A A A T A G A T A A A T A G A T A C A T A A A C A A T A T A T A A T T G
400 410 420 430 440

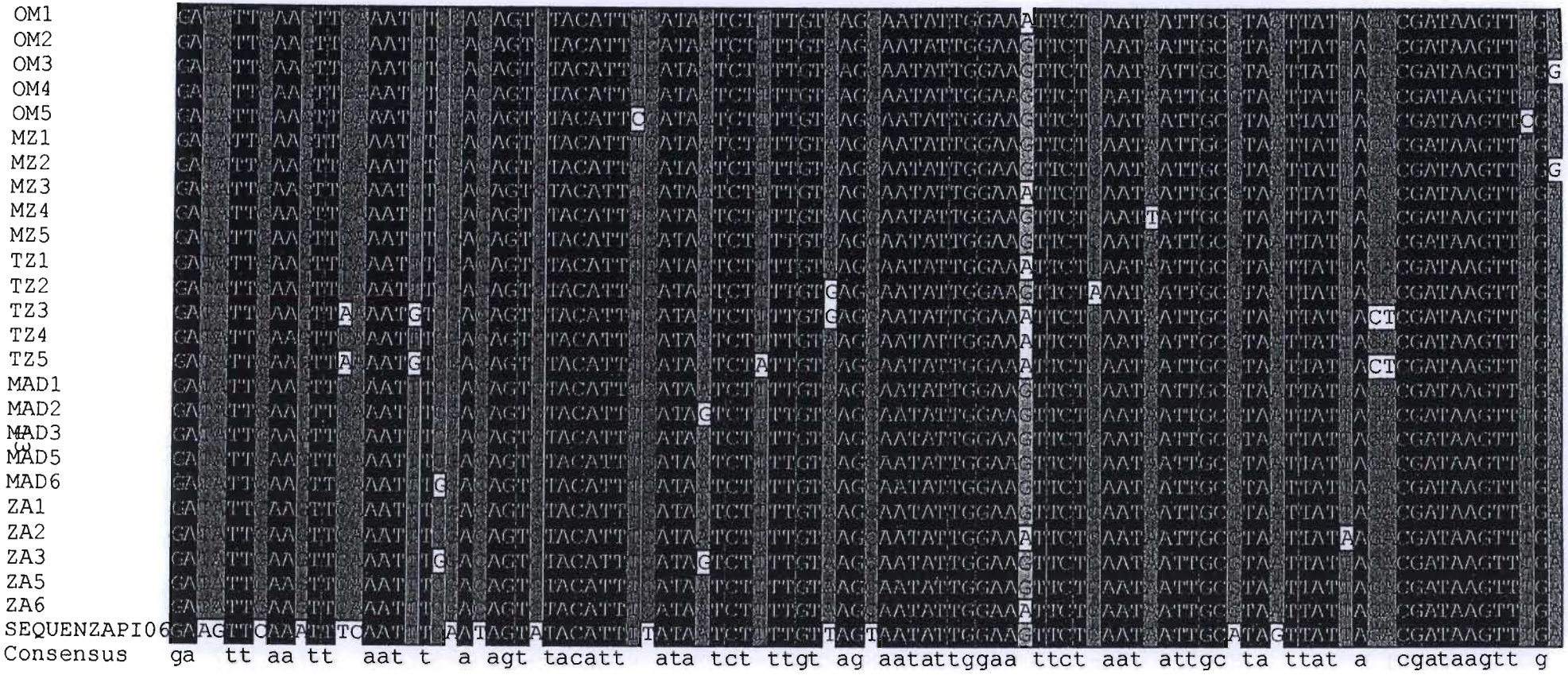

A T AA A T T AA C A G A T A G A T A A A T A G A T A C A G A T A G G T T C A T A G A C A G
450
460
470
480
A C A G A T A G A T A C A T A C A T A A A T A A A T A A A T AA A T A AC CAC A G A T AA A C
490
500
510
520
530
C G A T A A A T A A A C T T T A T C C C G A G C T A A A A T G A T G G T C A A G G A C A T T T
540
550
560
570
580
C A G A G T C G A A GC G A CG A G A G T A A T C A C T A G T G A A T T C G C G G C G C T G
590
600
610
620
C A G G T C G A C C A T A T G G G A G A G C T C C C A A C G C G T T G G A T G C C A T A G C T T G
630
640
650
660
670
A G T A T T C T A T A G T G T C A C C T A A A T A G C T T G G C C G T A A T C A T G G T C A T A G G C
680
690
700
710
720
T G T T T C C T G T G T G A C A T C G T AA T C C GC T C A C A A T T C C A C A C A A C A T A C G A G
730
740
750
760
770
C C G G A A G C A T A A A A G T G T A C A G C C T G G G G G T G C C T A A T G A G T G A G C T A A C T C A C C A T T
80
790
800
810
820
830

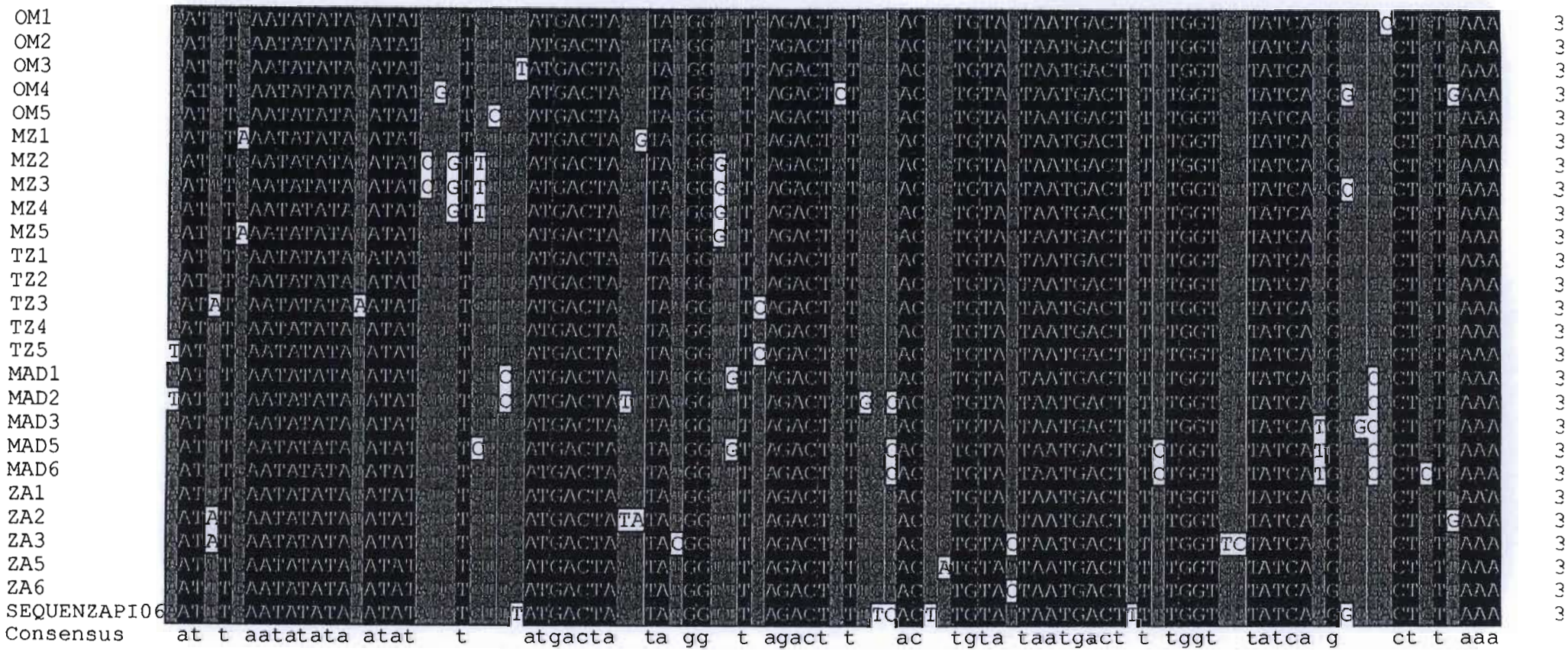

A TTN G C GTT GG C GACT TC A CT G C C C GC TTTT C C A G T CG G G A A A CC TGG T CG T G C C A G T GCC A TTC AA
840 850 860 870 880 890 900
GTT GGA A TTCC GG G CCCAA AC GGC GGCC GGGGGG AC AAGGC GGGCCA ACA GGCNGGT ACATCGGGGGGG GGCC ACCATT ACCGGN TT CC CNACG G T CCA CATC
910 920 930 940 950 960 970 980 990 1000
TGGAA NNC GGGT GG CACCN CGGGG ACAG A ACCCGAGATT GA A GGC AAACGC GNG AATACA G TC CACAC CCA A AGGGG GGGAA CA A C ACG
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100

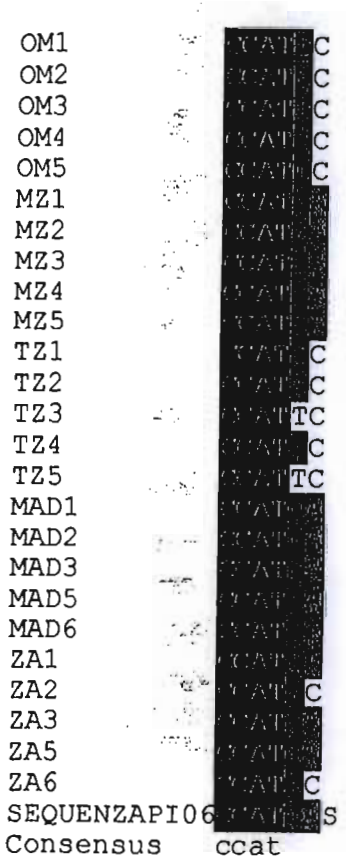
APPENDIX IVD


OM1
OM2
OM3
OM4
OM5
MZ1
MZ2
MZ3
MZ4
MZ5
TZ1
TZ2
TZ3
TZ4
TZ5
MAD1
MAD2
MAD3
MAD5
MAD6
ZA1
ZA2
ZA3
ZA5
ZA6
SEQUENZAPI06
Consensus
agt t tatc g tta ct t t t tt t t tatct tc tct tg a c tct atct t ta tct

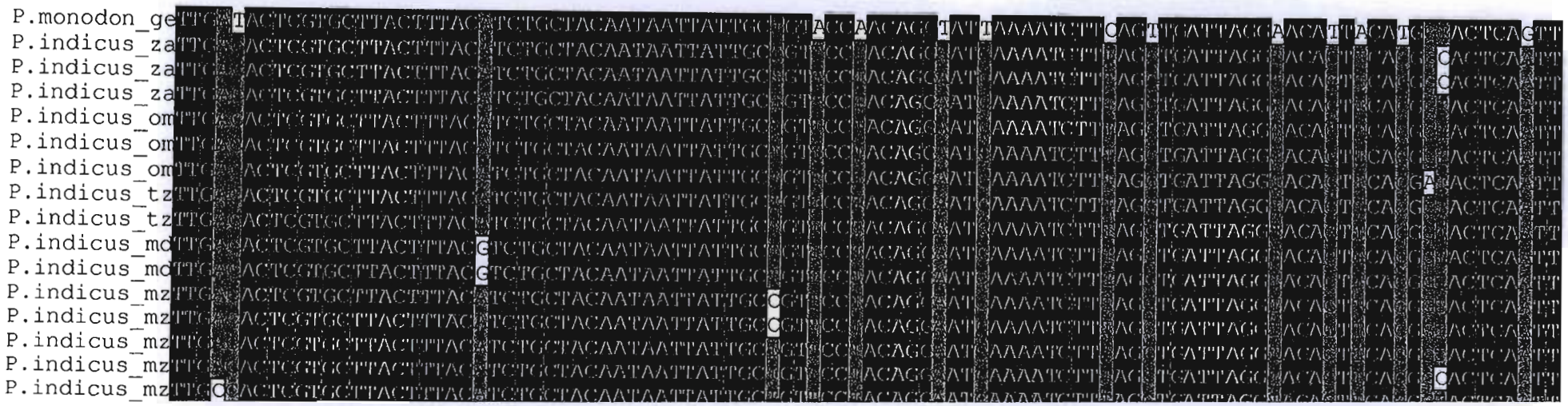

OM1 1
OM2 1
OM3 1
OM4 1
OM5 1
MZ1 1
MZ2 1
MZ3 1
MZ4 1
MZ5 1
TZ1 1
TZ2 1
TZ3 1
TZ4 1
TZ5 1
MAD1 1
MAD2 1
MAD3 1
MAD5 1
MAD6 1
ZA1 1
ZA2 1
ZA3 1
ZA5 1
ZA6 1
SEQUENZAPI06 1
Consensus tctgtt at tcaat at ta t t tatg atctatt atct tttgattttatt t g ttaggaatttcaagatc ca tt t tttgcaata a




OM1
OM2
OM3
OM4
OM5
MZ1
MZ2
MZ3
MZ4
MZ5
TZ1
TZ2
TZ3
TZ4
TZ5
MAD1
MAD2
MAD3
MAD5
MAD6
ZA1
ZA2
ZA3
ZA5
ZA6
SEQUENZAPI06
Consensus at t aatatata atat t atgacta ta gg t agact t ac tgta taatgact t tggt tatca g ct t aaa


OM1
OM2
OM3
OM4
OM5
MZ1
MZ2
MZ3
MZ4
MZ5
TZ1
TZ2
TZ3
TZ4
TZ5
MAD1
MAD2
MAD3
MAD5
MAD6
ZA1
ZA2
ZA3
ZA5
ZA6
SEQUENZAPI06
Consensus ccat

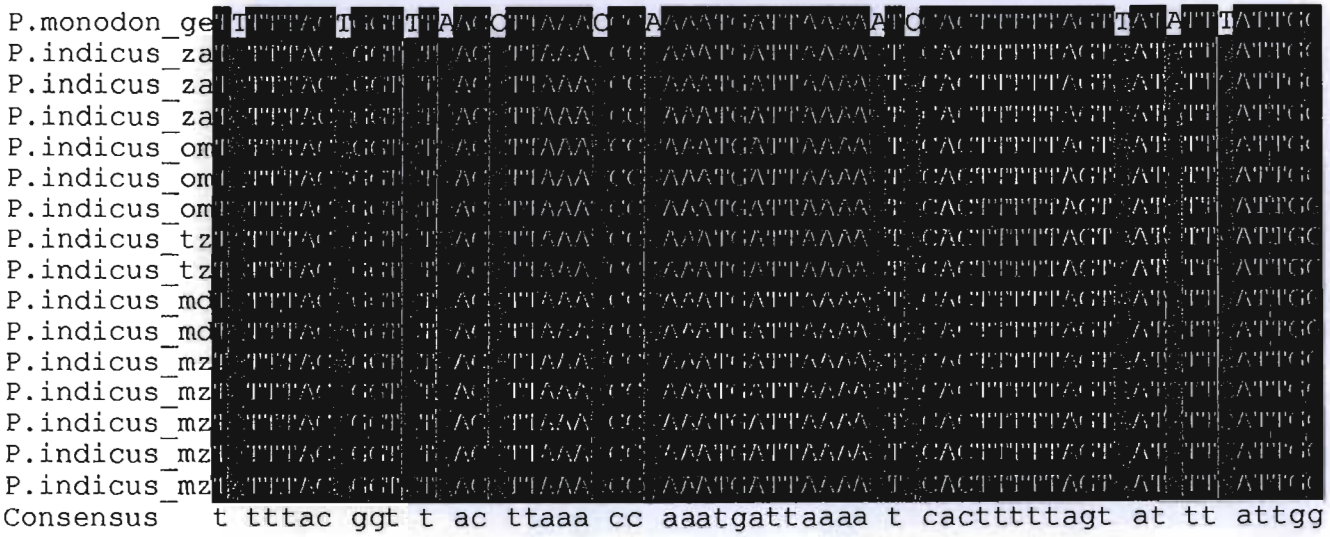
APPENDIX IVE



DNAMAN File: C:\WINDOWS\Desktop\Stampa tesi\Multipleallineamento_indicus_merguensis_tot_tesi.MSD

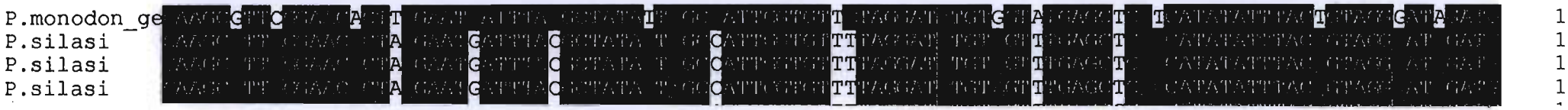

P.monodon_ge
P.indicus_za
P.indicus_za
P.indicus_za
P.indicus_om
P.indicus_om
P.indicus_om
P.indicus_tz
P.indicus_tz
P.indicus_md
P.indicus_md
P.indicus_mz
P.indicus_mz
P.indicus_mz
P.indicus_mz
P.indicus_mz
Consensus gaattatagtcc tctttaatttgagc tagg tttgt tt ttatt ac gtaggaggt t acaggagt gt t gctaa tcatc at ga at

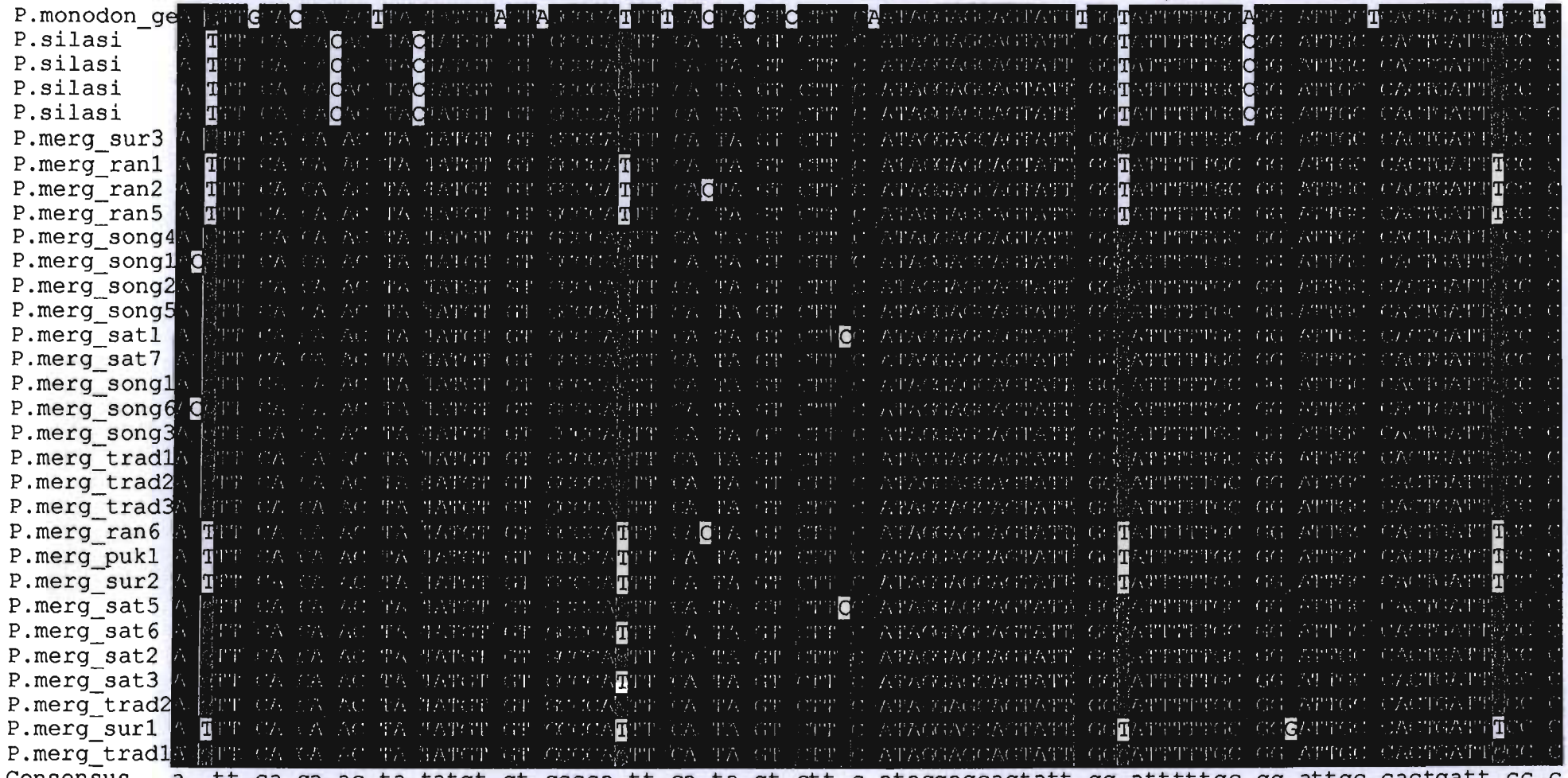

P.monodon_ge
P.indicus_za
P.indicus_za
P.indicus_za
P.indicus_om
P.indicus_om
P.indicus_om
P.indicus_tz
P.indicus_tz
P.indicus_md
P.indicus_md
P.indicus_mz
P.indicus_mz
P.indicus_mz
P.indicus_mz
P.indicus_mz
Consensus atctt cacga acttattatgt gtagc ca tt cact cgt ctttc at ggagc gtatttggtatttttgc ggtattgctcactgatttcc

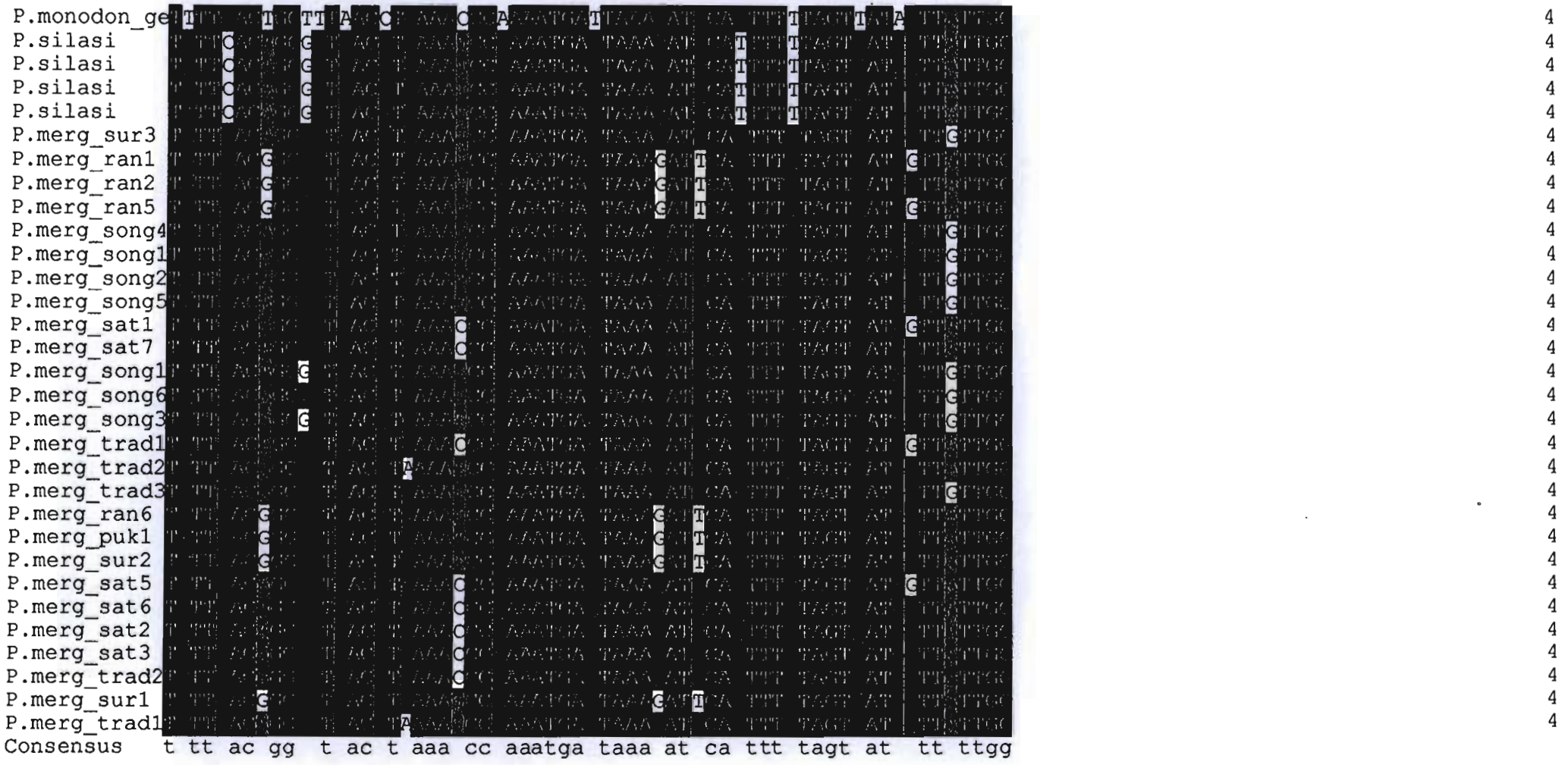

P.monodon_ge
P.indicus_za
P.indicus_za
P.indicus_za
P.indicus_om
P.indicus_om
P.indicus_om
P.indicus_tz
P.indicus_tz
P.indicus_mo
P.indicus_mo
P.indicus_mz
P.indicus_mz
P.indicus_mz
P.indicus_mz
P.indicus_mz
Consensus t tttac ggt t ac ttaaa cc aaatgattaaaa t cacttttagt at tt attgg

DNAMAN File: C:\WINDOWS\Desktop\Stampa tesi\Multipleallineamento_indicus_merguensis_tot_tesi.MSD

DNAMAN File: C:\WINDOWS\Desktop\Stampa tesi\Multipleallineamento_indicus_merguensis_tot_tesi.MSD




P.monodon_ge
P.silasi
P.silasi
P.silasi
P.silasi
P.merg_sur3
P.merg_ran1
P.merg_ran2
P.merg_ran5
P.merg_song4
P.merg_song1
P.merg_song2
P.merg_song5
P.merg_sat1
P.merg_sat7
P.merg_song1
P.merg_song6
P.merg_song3
P.merg_trad1
P.merg_trad2
P.merg_trad3
P.merg_ran6
P.merg_puk1
P.merg_sur2
P.merg_sat5
P.merg_sat6
P.merg_sat2
P.merg_sat3
P.merg_trad2
P.merg_sur1
P.merg_trad1
Consensus a tt ca ga ac ta tatgt gt gccca tt ca ta gt ctt c ataggagcagtatt gg atttttgc gg attgc cactgatt cc c


P.monodon_ge
P.silasi
P.silasi
P.silasi
P.silasi
P.merg_sur3
P.merg_ran1
P.merg_ran2
P.merg_ran5
P.merg_song4
P.merg_song1
P.merg_song2
P.merg_song5
P.merg_sat1
P.merg_sat7
P.merg_song1
P.merg_song6
P.merg_song3
P.merg_trad1
P.merg_trad2
P.merg_trad3
P.merg_ran6
P.merg_puk1
P.merg_sur2
P.merg_sat5
P.merg_sat6
P.merg_sat2
P.merg_sat3
P.merg_trad2
P.merg_sur1
P.merg_trad1
Consensus t tt ac gg t ac t aaa cc aaatga taaa at ca ttt tagt at tt ttgg