POPULATION GENETICS AND PHYLOGENETIC PLACEMENT OF THE ENDANGERED KNYSNA SEAHORSE, *HIPPOCAMPUS CAPENSIS*

PETER R. TESKE

PROMOTERS:
PROF. MICHAEL I. CHERRY
DR. CONRAD A. MATTHEE

Dissertation presented for the Degree of Doctor of Philosophy
at the University of Stellenbosch

December 2003

DECLARATION

I, the undersigned, hereby declare that the work contained in this dissertation is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature:

Date:

ABSTRACT

The aims of this study were to investigate genetic issues pertaining to the conservation of the Knysna seahorse, *Hippocampus capensis*, and to determine the phylogenetic placement of this endangered estuarine species among marine seahorses. This was accomplished by focusing on three aspects of the taxonomy: the inter-specific level; the inter-population level; and the intra-population level. To determine which species are closely related to *H. capensis*, and how the evolutionary history of this lineage relates to that of other seahorses, sequence data derived from four gene fragments (the nuclear RP1 and Aldolase and the mitochondrial 16S rRNA and cytochrome *b* genes) were used to determine the phylogenetic relationships among 30 species belonging to the genus *Hippocampus*. There were marked differences in the rate of evolution among these gene fragments, with Aldolase evolving the slowest and the mtDNA cytochrome *b* gene the fastest. Among individual partitions, the RP1 gene recovered the highest number of nodes supported by >70% bootstrap values from parsimony analysis, and >95% posterior probabilities from Bayesian inference. The combined analysis based on 2317 nucleotides resulted in the most robust phylogeny. A distinct phylogenetic split was identified between the pygmy seahorse, *H. bargibanti*, and a clade including all other species. Three species from the western Pacific Ocean included in this study, namely *H. bargibanti*, *H. breviceps*, and *H. abdominalis*, occupy basal positions in the phylogeny. This and the high species richness in the region suggest that the genus probably originated in this region. There is also fairly strong molecular support for the remaining species being subdivided into three main evolutionary lineages: two West Pacific clades and a clade of species present in both the Indo-Pacific and the Atlantic Ocean, which includes *H. capensis*. The phylogeny obtained herein suggests that seahorses belonging to the latter clade colonised the Atlantic Ocean at least twice, once before the closure of the Tethyan Seaway, and once afterwards. Phylogenies reconstructed using mitochondrial DNA gene fragments (16S rRNA, cytochrome *b* and 382 bp of the rapidly evolving control region) indicate that *H. capensis* is closely related to an Indian Ocean lineage of *H. kuda* and a Red Sea lineage of *H. fuscus*. Other lineages closely associated with these taxa include *H. kuda* from the West Pacific, the East Atlantic species *H.*

algiricus, the West Atlantic species *H. reidi*, the East Pacific species *H. ingens*, and the Hawaiian species *H. fisheri*. No control region alleles were shared among *H. capensis* and any of the marine seahorses, suggesting that the Knysna seahorse is phylogenetically distinct. The evolutionary history of *H. capensis*, and the extent of gene flow between its three known populations, were investigated using control region sequences from 138 specimens. Most samples were obtained by taking fin clips; this method was studied on captive seahorses and no negative effects were found. Similarly high levels of genetic diversity were found in two of the wild populations (Knysna and Keurbooms Estuaries), whereas diversity in the third population (Swartvlei Estuary) was lower. Although most haplotypes are shared among at least two populations, based on the haplotype frequency distributions the three assemblages constitute distinct management units. The extant population structure of *H. capensis* suggests that the Knysna seahorse originated in the large Knysna Estuary. The presence of seahorses in the two smaller estuaries is either the result of a vicariance event at the beginning of the present interglacial period, or colonisation of the estuaries via the sea, or a combination of the two. Population genetic parameters of the Knysna population and those of two populations of closely related marine seahorses (*H. kuda* from the Philippines and *H. fuscus* from the Red Sea) were similar, suggesting that the Knysna population is not genetically impoverished, despite its comparatively small area of occupancy.

OPSOMMING

Die doelwitte van hierdie studie was om die Knysna seeperdjie, *Hippocampus capensis*, te ondersoek relatief tot die spesie se bewaring asook om die filogenetiese posisie van hierdie bedreigte estuariene spesie binne mariene seeperdjies te bepaal. Drie aspekte van die taksonomie word ondersoek: interspesie verwantskappe, inter-bevolking verwantskappe en intra-bevolking verwantskappe. Om te bepaal watter spesies na verwant is aan *H. capensis*, asook om die evolusionêre geskiedenis van hierdie groep met die van ander groepe te vergelyk, word nukleotieddata van vier DNS fragmente (die nukleêre RP1 intron en Aldolase, en die mitochondriale 16S rRNA en sitokroom *b* fragmente) van 30 spesies van die genus *Hippocampus* gebruik. Aansienlike verskille in die tempo van evolusionêre verandering tussen hierdie DNS fragmente word gevind: Aldolase was die stadigste en die mitochondriale sitokroom *b* die vinnigste. Die RP1 intron het die meeste knoesteringe gehad wat ondersteun word deur hoë stewelvasgordnommers (>70%) van parsimoniese analises en hoë agter-waarskynlikheide (>95%) van Bayesiese gevolgtrekkinge. Die kombineerde analise wat 2317 nukleotiede ingesluit het, het die beste filogenie geproduseer. 'n Besliste filogenetise verdeling was gevind tussen die pigmee seeperdjie, *H. bargibanti*, en 'n groep wat al die ander spesies ingesluit het. Drie spesies van die westelike Stille Oseaan wat in hierdie studie ingesluit is, *H. bargibanti*, *H. breviceps* en *H. abdominalis*, neem primitiewe posisies in die filogenie in. Dit, en die hoë spesiesrykdom in daardie gebied dui aan dat dit moontlik is dat die genus in die westelike Stille Oseaan ontstaan het. Daar is ook taamlike goeie molekulêre ondersteuning dat al die ander spesies in drie evolusionêre hoofgroepe verdeel kan word: twee groepe wat hoofsaaklik in die westelike Stille Oseaan voorkom, en 'n groep van spesies wat in die Stille Oseaan, die Indiese Oseaan en in die Atlantiese Oseaan voorkom, wat *H. capensis* insluit. Die filogenie wat hier gevind is dui aan dat seeperdjies van hierdie laas genoemde groep die Atlantiese Oseaan minste twee keer gekoloniseer het, een keer voor die sluiting van die Tetiese Seepad, en een keer daarna. Filogenies wat met mitochondriale DNS fragmente gerekonstrueer is (16S rRNA, sitokroom *b* en 382 nukleotide van die vinnig evolveerende kontrolestreek) dui aan dat *H. capensis* na verwant is aan 'n groep van *H. kuda* wat in die Indiese Oseaan voorkom en *H. fuscus* van die Rooi See. Ander groepe wat na verwant is aan hierdie

takson is *H. kuda* van die westelike Stille Oseaan, *H. algiricus* van die Oos Atlantiese Oseaan, *H. reidi* van die Wes Atlantiese Oseaan, en die Hawaiise spesie *H. fisheri*. Geen kontrolestreek allele was gedeel tussen *H. capensis* en enige mariene seeperdjie spesies; dit dui aan dat die Knysna seeperdjie filogeneties verskillend is. Die evolusionêre geskiedenis van *H. capensis*, en die omvang van die genetiese interaksies tussen sy drie bekende bevolkings, word ondersoek met kontrolestreek nukleotieddata van 138 monsters. Die meeste van hierdie monsters was verkry deur vinknipsels; hierdie metode was getoets op seeperdjies in gevangenskap en geen negatiewe gevolge was gevind nie. Genetiese diversiteit was omtrent dieselfde in twee van die natuurlike bevolkings (Knysna en Keurbooms Estuariums), maar diversiteit in die derde bevolking (Swartvlei Estuarium) was laër. Alhoewel die meeste allele gedeel was tussen ten minste twee bevolkings, dui die verspreiding van allelfrekwensies aan dat die drie bevolkings aparte bestuurseenhede is. Die ekstante bevolkingsstruktuur van *H. capensis* dui aan dat die Knysna seeperdjie in die groot Knysna Estuarium ontstaan het. Die teenwordigheid van seeperdjies in die twee kleiner estuariums is óf die resultaat van 'n vikariansie voorval aan die begin van hierdie interglasiale tydperk, óf kolonisasie van die estuariums deur die see, óf 'n kombinasie van albei. Bevolkingsgenetiese parameters van die Knysna bevolking en van twee bevolkings van na verwante seeperdjie spesies (*H. kuda* van die Filippyne en *H. fuscus* van die Rooi See) was soortgelyk, wat aandui dat die Knysna bevolking nie geneties verarm is nie, alhoewel dit 'n betreklik kleiner streek bewoon.

ZUSAMMENFASSUNG

Die hier präsentierte wissenschaftliche Studie beschäftigte sich mit genetischen Themen relevant für den Artenschutz des Knysna Seepferds, *Hippocampus capensis*, und den phylogenetischen Beziehungen dieser ausschliesslich in Estuaren (Flussmündungen) vorkommenden gefährdeten Art mit den im Meer lebenden Seepferden. Die folgenden taxonomischen Einheiten wurden verglichen: Arten, Populationen und Sub-Populationen. Um festzustellen, welche Arten nah mit *H. capensis* verwand sind, und wie die Evolution dieser Gruppe sich von der anderer Seepferdgruppen unterscheidet, wurden genetische Sequenzen von vier Genen (den nuklearen RP1 und Aldolase und den mitochondrischen 16S rRNA und Cytochrom *b* Genen) von 30 Seepferdarten verwendet und phylogenetische Beziehungen rekonstruiert. Beträchtliche Unterschiede wurden festgestellt hinsichtlich der Geschwindigkeit in der Mutationen stattgefunden haben: Aldolase mutierte am langsamsten und Cytochrom *b* am schnellsten. Eine auf RP1 Sequenzen basierende Phylogenie hatte die höchste Anzahl von Gabelungspunkten, die sowohl von parsimonischen Analysen, als auch von bayesischer Inferenz unterstützt wurden. Die robusteste Phylogenie wurde jedoch gefunden, wenn Sequenzen von allen vier Genen kombiniert wurden (im ganzen 2317 Nukleotide). Eine beträchtliche genetische Distanz wurde zwischen dem Pygmäen-Seepferd, *H. bargibanti*, und einer Gruppe, die aus allen anderen Arten bestand, gefunden. Drei Arten vom westlichen Pazifik, nämlich *H. bargibanti*, *H. breviceps* und *H. abdominalis*, hatten basale Positionen in der Phylogenie. Das, und der Artenreichtum dieser Region, sind Anzeichen dafür, dass Seepferde möglicherweise ursprünglich aus dem westlichen Pazifik stammen. Es wurde weiterhin gefunden, dass alle übrigen Seepferdarten in drei Hauptgruppen unterteilt werden können: die Verbreitungsgebiete zweier dieser Gruppen beschränken sich hauptsächlich auf den westlichen Pazifik, aber die dritte Gruppe kommt sowohl im Indo-Pazifik, also auch im Atlantik vor (*H. capensis* ist mit dieser letzteren Gruppe assoziiert). Es gibt gute Anzeichen dafür, dass die Seepferde der letztgenannten Gruppe den Atlantik mindestens zweimal kolonisiert haben, einmal vor der Schliessung der tethyschen Seeverbindung, und einmal danach. Phylogenien, die ausschliesslich mit mitochondrischen Genen rekonstruiert wurden (16S rRNA,

Cytochrom *b* und 382 Nukleotide der schnell-mutierenden Kontollregion), zeigen, dass *H. capensis* sehr nah verwandt mit *H. kuda* aus dem Indischen Ozean und *H. fuscus* aus dem Roten Meer ist. Andere nah verwandte Arten sind *H. kuda* from westlichen Pazifik, *H. algiricus* vom östlichen Atlantik, *H. reidi* vom westlichen Atlantik, *H. ingens* vom östlichen Pazifik, sowie die in Hawaii vorkommende Art *H. fisheri*. Keine der Kontrollregionallele, die in *H. capensis* gefunden wurden, kamen in anderen Arten vor. Dies zeigt, dass das Knysna Seepferd eine eigenständige Art ist, und Paarungen mit anderen Arten nicht vorkommen. Die Evolutionsgeschichte von *H. capensis*, und das Ausmass von genetischem Austausch zwischen den drei Populationen dieser Art, wurden untersucht, indem Kontrollregionsequenzen von 138 Individuen analysiert wurden. Die meisten Proben stammten von Flossenschnitten; diese Methode wurde zuvor an in Gefangenschaft lebenden Seepferden ausprobiert, und es wurden keine negativen Folgeerscheinungen beobachtet. Genetische Diversität war ungefähr gleich hoch in zwei der Populationen (Knysna und Keurbooms Estuare), aber eine deutlich niedrigere Diversität wurde in der dritten Population gefunden (Swartvlei Estuar). Obwohl die meisten Allele in mindestens zwei Populationen gefunden wurden, sind die drei Populationen unterschiedliche genetische Einheiten, eine Schlussfolgerung, die hauptsächlich auf Unterschiede in der relativen Häufigkeit der Allele beruht. Die Populationsstruktur von *H. capensis* deutet darauf hin, dass diese Art ihren Ursprung im Knysna Estuar hat. Die Präsenz von Seepferden in den beiden anderen Estuaren ist entweder das Resultat von Vikarianz (eine Spaltung der ursprünglichen Population) zu Beginn der jetzigen Interglazialzeit, oder Kolonisierung der Estuare durchs Meer, oder eine Kombination beider Szenarios. Populationsgenetische Parameter der Knysna Population und die zweier Populationen von nah verwandten Arten (*H. kuda* aus den Philippinen und *H. fuscus* aus dem Roten Meer) zeigten keine grossen Unterschiede. Dies deutet darauf hin, dass das Knysna Seepferd trotz seines vergleichbar kleinen Verbreitungsgebietes nicht unter geringer genetischer Diversität leidet.

ACKNOWLEDGEMENTS

I am grateful for the various intellectual and financial contributions that made this study possible. First and foremost, I wish to thank my supervisors, Prof. Michael I. Cherry and Dr. Conrad A. Matthee, for their intellectual input, encouragement, support and advice. I also wish to thank other individuals affiliated with the Evolutionary Genomics Group at Stellenbosch University for various useful suggestions and fruitful discussions, including Dr. Deryn Alpers, Dr. Krystal Tolley, Dr. Bettine Jansen van Vuuren, Dr. Claudine Montgelard, Dr. Michael Cunningham, Dr. Savel Daniels, Gavin Gouws, Sandi Willows-Munro and James Rhodes.

A prerequisite for the success of this work was the contribution of seahorse specimens, without which many sections of this study would have remained largely anecdotal. In this regard, I particularly wish to thank Jackie Lockyear (Knysna Seahorse Project, WWF Green Trust) for providing expertise and equipment, and for helping with the sampling. Her contributions ensured that a sufficiently large number of Knysna seahorses could be collected to discover new insights about the demographic history and population structure of this most endangered of the world's seahorses. I also wish to thank Ian Russell (Rondevlei Research Station) for providing most of the samples from the Swartvlei Estuary, and Peet Joubert and Rob Milne (National Parks Board, Knysna) for providing expertise and logistical support during the early phases of the research. Sara Lourie (Project Seahorse, University of British Columbia, Canada) is gratefully acknowledged for providing specimens from her organisation's extensive collection of dried syngnathids. These samples formed the backbone of the phylogenetic sections of this study. Various individuals and institutions are thanked for contributing additional samples that helped clarify details and fill gaps, including Healy Hamilton (California Academy of Sciences, San Francisco, USA), Adam Jones (Georgia Institute of Technology, Atlanta, USA), Janelle Curtis (Project Seahorse, University of British Columbia, Canada), Takahito Mukai (Ocean Research Institute, University of Tokyo, Japan), Howaida Gabr (Suez Canal University, Ismaila, Egypt), Melchor Santos (Manila, Philippines), Sesh

Serebiah (Jeghovah Shumma, Tamil Nadu, India), Anantha Sreepada (National Institute of Oceanography, Goa, India), Zacharie Sohou (University of Benin, Benin City, Benin), Mark Erdmann (Manado, Sulawesi, Indonesia), Paul Southgate (James Cook University, Queensland, Australia) and Kerryn Parkinson (Australian Museum, Sydney, Australia). Marlene Gunter (Port Nolloth Sea Farms, South Africa) is thanked for allowing me to take fin clips from her pet seahorses.

For financial assistance, I thank the National Research Foundation (NRF) and the Harry Crossley Foundation. Supplementary financial support for conference participation was provided by a grant to C. A. Matthee, the University of Stellenbosch research office and the South African Network for Coastal and Oceanic Research (SANCOR).

Lastly, I would like to thank my father, Dr. Eberhard Teske, for his encouragement throughout my studies, his hands-on contributions (hunting for seahorses in Mozambique), and his appreciation of the work I was doing. This dissertation is dedicated to him.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

APPENDIX I

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

The following publications have emanated from this thesis so far:

Teske, P.R., Cherry, M.I. & Matthee, C.A. 2003. Population genetics of the endangered Knynsa seahorse, *Hippocampus capensis*. *Molecular Ecology* **12:** 1703-1715.

Teske, P.R., Cherry, M.I. & Matthee, C.A. (in press). The evolutionary history of seahorses (Syngnathidae: *Hippocampus*) – molecular data suggest an Indo-Pacific origin and two invasions of the Atlantic Ocean. *Molecular Phylogenetics and Evolution.*

Both papers are included in APPENDIX II

GENERAL INTRODUCTION

The Knysna seahorse, *Hippocampus capensis* Boulenger, 1900, was the first seahorse species to be listed as endangered on the IUCN Red Data List (Hilton-Taylor 2000). This conservation status derives from its small area of occupancy and, consequently, its low abundance and vulnerability. Being the world's only seahorse known to exclusively inhabit estuaries, *H. capensis* has the smallest distributional range of any seahorse and has been recorded from only three estuaries on the south coast of South Africa (Knysna, Swartvlei and Keurbooms). Despite the high conservation status of *H. capensis*, little scientific information is available on population sizes, population structure and habitat requirements (Skelton 1987; Day 1997). Due to the elusive nature of this species, most of the studies on *H. capensis* have been conducted under captive conditions (Fourie 1997; Lockyear *et al.* 1997; Tops 1999; Le Cheminant 2000), and *in situ* biological and ecological research is limited (Bell *et al.* 2003). Human settlement, and associated industrial, domestic and recreational activities are increasing along the species' habitats. As pollution events or other disturbances affecting the submerged plant beds of these estuaries will have a direct and indirect impact on seahorse populations (Skelton 1987), more detailed scientific research on *H. capensis* is urgently required in order to aid management decisions.

Phylogenetic relationships among the different species within the genus *Hippocampus* Rafinesque, 1810, and the position of the Knysna seahorse among them, are currently poorly understood. Although morphological data and cytochrome *b* sequence data suggest that Knysna seahorses are part of a widely distributed clade of seahorses, which includes the Indo-Pacific *H. kuda* complex, the West African *H. algiricus* Kaup, 1856, the East American *H. reidi* Ginsburg, 1933 and the West American *H. ingens* Girard, 1859 (Lourie *et al.* 1999; Casey 1999), the exact associations with these marine species, and the position of this clade within the phylogeny of the genus, are as yet unresolved.

The different components of this thesis were aimed at addressing each of the above issues by sequencing molecular markers that are likely to provide phylogenetic signal at different taxonomic levels (i.e. interspecific level, inter-population level and intra-population level). The placement of *H. capensis* and its closest relatives within the phylogeny of the genus *Hippocampus* was addressed by sequencing two nuclear and two mitochondrial genes of a limited number of individuals from 28 of the 32 species considered to be valid by Lourie *et al.* (1999), as well as two recently described species. Species that were found to be closely related to *H. capensis* were then focused on by sequencing as many specimens of each species as could

be obtained. To ensure sufficient resolution, sequences of the rapidly evolving mitochondrial control region were primarily used. Population structure within and among the different estuaries inhabited by *H. capensis* was investigated by obtaining mitochondrial control region sequences from a total of 138 specimens. In order to be able to include such a large sample without risking negative impacts on the populations, samples were obtained either from specimens that were already dead, or by taking fin clips. The effects of fin clipping on growth and survival of seahorses were investigated in captive seahorses. Lastly, in order to understand how population parameters of *H. capensis* compare with those of marine seahorses, control region sequences of *H. capensis* were compared with those of two populations of the closely related species *H. fuscus* and *H. kuda*.

CHAPTER 1: THE EVOLUTION OF THE GENUS *HIPPOCAMPUS*

1.1 Introduction

1.1.1 The taxonomy of the Sygnathidae

Seahorses belong to the family Syngnathidae, a family of fishes whose oldest fossils date back to the Eocene (Lutetian: 52 mya; Patterson 1993). The family also includes the pygmy pipehorses (grouped with seahorses in the subfamily Hippocampinae), pipehorses and seadragons (Solegnathinae), flag-tail pipefishes (Doryrhamphinae) and pipefishes (Syngnathinae; Kuiter 2000). The monophyly of seahorses is not questioned, as it is based on a number of synapomorphic morphological characters distinguishing them from other syngnathids. These characters include a prehensile tail, the absence of a caudal fin, the head at a right angle to the trunk, a brood pouch sealed along the midline (except for a small anterior opening), and a raised dorsal fin base (Fritzsche 1980). Herald (1959) suggested that the possession of a fused brood-pouch placed seahorses as highly derived members of this family. Seahorses (genus *Hippocampus*) and possibly also pygmy pipehorses (genera *Amphelikturus* Parr, 1930, *Acentronura* Kaup, 1853 and *Idiotropiscis* Whitley, 1947), are phylogenetically most closely associated with pipefishes of the genus *Syngnathus* Linnaeus, 1758 (Wilson *et al.* 2001, 2003).

1.1.2 The evolution of seahorses

The world's tropical marine faunas can be divided into those associated with the Atlantic Ocean (including the Caribbean and Mediterranean), and those associated with the Indo-Pacific,

a pattern that probably resulted from the closure of the Tethyan seaway between Africa and Eurasia (Rosen 1988). The exact date of Tethys closure is not agreed upon (Adams *et al.* 1983; Rögl & Steininger 1983, 1984; Steininger & Rögl 1984), and it seems that there were several closures and re-openings throughout the Oligocene and mid-Miocene (approx. 10-35 mya, Rosen 1988).

Seahorses are found throughout the tropical and temperate regions of both the Atlantic and Indo-Pacific biomes, and most species are restricted to shallow, coastal waters (Lourie *et al.* 1999). Because of this circumglobal distribution of the extant members of the genus, Fritzsche (1980) suggested that the genus *Hippocampus* is pre-Tethyan in origin. In a study based on cytochrome *b* sequences, Casey (1999) concluded that the genus evolved in the Atlantic biome. An Atlantic origin is also supported by the fact that most species of the closely related pipefish genus *Syngnathus* are associated with the Atlantic biome (Kuiter 2000), as well as the fact that the only known seahorse fossils have been found in Italy (Sorbini 1988). On the other hand, it is interesting to note that the majority of seahorse species are found in the Indo-Pacific region (>27 species, Lourie *et al.* 1999). This pattern is not unique to seahorses; it has been shown that the majority of tropical marine families have their greatest concentration of species within the East Indies triangle formed by the Philippines, the Malay Peninsula, and New Guinea (Briggs 1999).

The Indo-West Pacific region has an interesting history of faunal convergences and vicariance events resulting from tectonic movements and changes in sea level. Collision of Australia with south-east Asia during the late Oligocene (Pigram & Davis 1987) to early Miocene (Woodland 1986) resulted in faunal convergences among Australian and Asian marine faunas and has been attributed to the evolution of a number of new genera of marine organisms. Glacial-interglacial cycles during the late Pliocene and particularly during the Pleistocene (Shackleton & Opdyke 1977) caused drastic changes in the sea level, which resulted in a series of vicariance events in the East Indies (Greenfield 1968; Woodland 1983; Winterbottom *et al.* 1984). These coincided with tectonic movements in this region, most notably the closing of the Indonesian seaway between New Guinea and eastern Indonesia 3-4 mya (Cane & Molnar 2001). Divergences between Indian Ocean and western Pacific sibling species have been attributed to the isolation of formerly continuously distributed species (McManus 1985; McMillan & Palumbi 1995; Palumbi 1996). The present-day marine fauna in the Indo-West Pacific is characterised by comparatively young genera (Stehli & Wells 1971; Newman *et al.* 1976; Roux 1987) many of which are apomorphic species (Menon 1977; Fricke 1988; Ricklefs & Latham 1993; Specht 1981). Although these characteristics may suggest that the high species richness

in the Indo-West Pacific is a result of recent speciation or colonisation, several authors suggested that the region is a centre of origination and radiation of various marine taxa (Rosen 1984; Briggs 1999; Lessios *et al.* 2001), as the more advanced species tend to be characteristic of the center of radiation, whereas their more primitive relatives are displaced towards the periphery (Matthew 1915; Darlington 1957, 1959). It is thus not unlikely that seahorses could have rather originated in the Indo-West Pacific.

Irrespective of the origin of the genus, the circumglobal distribution of seahorses reflects long distance dispersal events. It is possible that some tropical shore species have been able to migrate around the Cape of Good Hope to establish themselves in the Atlantic Ocean, but there is no evidence for such dispersal events in the opposite direction (Briggs 1995). Migration events from the Atlantic Ocean towards the Indo-West Pacific via the Central American Seaway prior to its final closure 2.7 mya (Marshall 1988) are theoretically possible, but the expanse of the Pacific Ocean has been shown to be a formidable barrier to dispersal of shallow-water fishes and invertebrates (Ekman 1953; Briggs 1961, 1999; Rohde & Hayward 2000).

1.1.3 The taxonomy of the genus *Hippocampus*

The occurrence of uncertain species boundaries and species complexes are common problems associated with the systematics of marine organisms (Knowlton 1993; Avise 1994; Gosling 1994; Powers *et al.* 1991; Bernardi *et al.* 1993; Miya & Nishida 1997). This phenomenon is no different in seahorses, where morphology-based taxonomic methods have shown to be problematic. More than 100 species have been described (Eschmeyer 1998), but recent attempts at revising the genus accept between 33 (Lourie *et al.* 1999; Lourie & Randall 2003) and 56 (Kuiter 2000, 2001) valid species names (Table 1.1). These controversies are mainly due to differences in the interpretation of morphological characteristics and it is possible that these are under strong selection pressure due to the unique life history of the group. Seahorses avoid predators by relying on camouflage, and as external appearance is one of the main survival strategies employed by members of this genus, it seems reasonable to assume that many morphological characters are convergent. It is possible that because of this, no robust hypothesis regarding the phylogeny of the genus *Hippocampus* exists to date.

Genetic methods have great potential to both resolve disputed taxonomic issues and to infer phylogenetic relationships among different marine species (McMillan & Palumbi 1995; Arnaud *et al.* 1999; Burridge & White 2000; Bowen *et al.* 2001; Colborn *et al.* 2001; Grant & Leslie

2001; Muss *et al.* 2001). Two nuclear DNA gene fragments (the first intron of the S7 ribosomal protein and a section of the Aldolase gene) and two mitochondrial gene fragments (16S rRNA and cytochrome *b*) were used to investigate phylogenetic relationships among the species within the genus *Hippocampus.* By using three independent evolutionary markers that presumably evolve under different phylogenetic constraints and four gene fragments, we were hoping to obtain a robust phylogenetic tree addressing both the recent and older evolutionary events.

Table 1.1 Seahorse species considered valid by Kuiter (2000, 2001), and those considered valid by Lourie *et al.* (1999) and Lourie & Randall (2003). In the case of some species names listed in Kuiter (2000, 2001) that were not considered valid by Lourie *et al.* (1999), alternative interpretations of type specimens are shown in brackets as listed in Lourie *et al.* (1999).

Kuiter 2000	Lourie *et al.* 1999	Authority
H. abdominalis	*H. abdominalis*	Lesson 1827
H. alatus		Kuiter 2001
	H. algiricus	Kaup 1856
H. angustus	*H. angustus*	Günther 1870
H. arnei		Roule 1916
H. barbouri	*H. barbouri*	Jordan & Richardson 1908
H. bargibanti	*H. bargibanti*	Whitley 1970
H. biocellatus		Kuiter 2001
H. bleekeri	*H. abdominalis*	Fowler 1908
H. breviceps	*H. breviceps*	Peters 1870
H. borboniensis	*H. borboniensis*	Dumeril 1870
H. camelopardalis	*H. camelopardalis*	Bianconi 1855
H. capensis	*H. capensis*	Boulenger 1900
H. comes	*H. comes*	Cantor 1850
H. coronatus	*H. coronatus*	Temminck & Schlegel 1850
H. dahli	(synonym of *H. trimaculatus*)	Ogilby 1908
H. deanei	(synonym of *H. algiricus*)	Dumeril 1857
	H. denise	Lourie & Randall 2003
H. erectus	*H. erecuts*	Perry 1810
H. elongatus	(synonym of *H. subelongatus*)	Castelnau 1873
H. fuscus	*H. fuscus*	Ruppell 1838
H. grandiceps		Kuiter 2001
H. guttulatus	*H. guttulatus*	Cuvier 1829
H. hendriki		Kuiter 2001
H. hippocampus	*H. hippocampus*	Linnaeus 1758
H. histrix	*H. histrix*	Kaup 1856
H. ingens	*H. ingens*	Girard 1858
H. japonicus	(synonym of *H. mohnikei*)	Kaup 1856
H. jayakari	*H. jayakari*	Boulenger 1900
H. jugumus		Kuiter 2001
H. kampylotrachelos	(synonym of *H. trimaculatus*)	Bleeker 1854
H. kelloggi	*H. kelloggi*	Jordan & Snyder 1902
H. kuda	*H. kuda*	Bleeker 1852

Table 1.1 (continued)

Kuiter 2000	Lourie *et al.* 1999	Authority
H. lichtensteinii	*H. lichtensteinii*	Kaup 1853
H. manadensis	(synonym of *H. trimaculatus*)	Bleeker 1856
H. minotaur	*H. minotaur*	Gomon 1977
H. mohnikei	*H. mohnikei*	Bleeker 1851
H. moluccensis	(synonym of *H. kuda*)	Bleeker 1852
H. montebelloensis		Kuiter 2001
H. multispinus		Kuiter 2001
H. planifrons	(synonym of *H. trimaculatus*)	Peters 1877
H. polytaenia	(synonym of *H. kuda*)	Bleeker 1852
H. procerus		Kuiter 2001
H. queenslandicus		Horne 2001
H. reidi	*H. reidi*	Ginsburg 1933
H. semispinosus		Kuiter 2001
H. sindonis	*H. sindonis*	Jordan & Snyder 1901
H. spinosissimus	*H. spinosissimus*	Weber 1913
	H. subelongatus	Castelnau 1973
H. suezensis	(synonym of *H. kelloggi*)	Duncker 1940
H. taeniopterus	(synonym of *H. kuda*)	Bleeker 1852
H. takakurai		Tanaka 1916
H. trimaculatus	*H. trimaculatus*	Leach 1814
H. tuberculatus	(synonym of *H. breviceps*)	Calstelnau 1875
H. tristis	(synonym of *H. kuda*)	Castelnau 1872
H. villosus	(synonym of *H. erectus*)	Gunther 1880
H. whitei	*H. whitei*	Bleeker 1855
H. zebra	*H. zebra*	Whitley 1964
H. zosterae	*H. zosterae*	Jordan & Gilbert 1882

1.1.4 DNA markers used to construct a phylogeny

The two groups of markers used in the present study (i.e. nuclear and mitochondrial DNA) both have unique advantages to study evolution: Although some nDNA regions develop rapidly and are thus suitable to address lower-level phylogenetic questions (e.g. ITS), nDNA has been shown to be powerful to address higher-level phylogenetic questions (e.g. Gatesy *et al.* 1996; Matthee *et al.* 2001; Murphy *et al.* 2001; Lavoué *et al.*, submitted). However, Zhang & Hewitt (2003) list five concerns when using nDNA that makes their application difficult: the unavailability of markers suitable for carrying out such analyses, technical laboratory hurdles for resolving haplotypes (several copies of a particular marker may exist, and these will amplify

simultaneously), difficulty in data analysis because of recombination, low divergence levels and intraspecific multifurcation evolution.

Phylogenetic research using mtDNA sequences of animals has been used most frequently at the lower taxonomic level (i.e. population, species, and genera; Moritz *et al.* 1987; Avise *et al.* 1988; Meyer 1994; Simon *et al.* 1994), although the slowly mutating 12S and 16S rRNA genes are primarily used at higher taxonomic levels (Mindell & Honeycutt 1990; Hillis & Dixon 1991). As mtDNA represents only a single locus, it can reflect only the matrilineal history of a population or species. Moreover, as the effective population size of mtDNA is only a fourth of that of nuclear autosomal sequences, mtDNA lineages have a much faster lineage sorting rate and higher allele extinction rate. This may result in evolutionary relationships being oversimplified (Zhang & Hewitt 2003).

Properties of the four gene fragments used are discussed below:

a) First intron of the S7 ribosomal protein (RP1)

Chow & Hazama (1998) detected significant intraspecific variation in the first and second introns of the S7 ribosomal protein in yellowfin tuna (*Thunnus albacares*) and suggested that this fragment may be useful to investigate intraspecific genetic structure and phylogenetic relationships among closely related species. Lavoué *et al.* (submitted) considered the introns useful to reconstruct the phylogeny of a highly divergent group of African electric fishes (Mormyroidea) and stated that the homoplasy content of this marker is very low at this level of resolution.

b) Aldolase

In order to find a highly conserved marker with which we aimed to address older evolutionary events within the genus *Hippocampus*, we attempted to amplify seahorse DNA using primers originally designed for frogs. Amplification was successful in the case of Aldolase.

c) 16S ribosomal RNA

Ribosomal RNA (rRNA) is involved in the process of translating messenger RNA (mRNA) into polypeptides. Although the regions coding for these RNAs should strictly speaking be referred to as rDNAs, we are following convention (e.g. Alves-Gomes *et al.* 1995; Abouheif *et al.* 1998; Van de Peer *et al.* 2000; Zardoya & Meyer 2001). The sequences of the ribosomal subunits can be divided into two types characterised by different evolutionary rates: those that are transcribed into the rRNAs' stem regions, and those that are transcribed into loop regions. Because paired nucleotides forming stem regions appear to undergo compensatory mutations that maintain secondary structure, their evolutionary rate is lower than that of loop regions (Wheeler & Honeycutt 1988; Ortí *et al.* 1996). It has been argued that because mitochondrial ribosomal genes are the most conserved genes in the mitochondrial genome, they are suitable to elucidate relationships among taxa that diverged as long as 300 (Mindell & Honeycutt 1990) and 65 (Hillis & Dixon 1991) million years ago. Simon *et al.* (1994) suggested that the 12S and 16S rRNA are likely to be useful at the population level, where highly variable sites are not saturated (i.e. characterised by multiple substitutions) as well as at more ancient levels of divergence, where the more conserved regions contain useful phylogenetic information. Ortí *et al.* (1996) found that 12S and 16S rRNA provided phylogenetic signal to resolve phylogenetic relationships among the subfamily Serrasalminae, which is believed to have diverged as recently as the mid- to late Tertiary (30-13 mya).

d) Cytochrome *b*

The cytochrome *b* gene codes for a transmembrane protein important in the respiratory chain of cellular metabolism (Esposti *et al.* 1993) and is one of the most frequently used genetic markers to study the evolution of fishes (Kocher *et al.* 1989; Meyer *et al.* 1990; Carr & Marshall 1991; Block *et al.* 1993; Zhu *et al.* 1994; Carr *et al.* 1995; Stepien & Kocher 1997). As in all protein coding sequences, most variation is found in third codon positions and least is found in second codon positions (Lydeard & Roe 1997). Cytochrome *b* has been used to assess questions at different taxonomic levels, ranging from population genetics (e.g. Bermingham *et al.* 1997) to the family level (Bernardi 1997). However, several authors have questioned the ability of this marker to resolve deep nodes in the phylogenies (Martin *et al.* 1990; Hillis & Huelsenbeck 1992; Graybeal 1993; Meyer 1994; Matthee *et al.* 2001).

1.2 Materials and methods

1.2.1 Sample acquisition, DNA extraction, amplification and sequencing

Most of samples used in this study were provided by Sara Lourie[1] and had been preserved by drying. Ethanol preserved samples, comprising fin clips, muscle, operculum or internal organs were obtained from various other sources (Table 1.1). The total sample size consisted of 51 individuals from 30 species, following the classification system of Lourie *et al.* (1999). Two newly described species not yet included in Lourie *et al.* (1999), namely *Hippocampus procerus* (Kuiter 2001) and *H. queenslandicus* (Horne 2001), and a specimen of the South African pipefish *Syngnathus temminckii* Kaup 1856 (used as outgroup taxon), were also included.

[1] Sara Lourie, Project Seahorse, UBC Fisheries Centre, 2204 Main Mall, Vancouver, British Columbia, V6T 1Z4 Canada

Table 1.2 List of specimens used in this study, including the code as indicated on all phylogenetic trees, region of origin, and the individual or institution who provided the samples. Alternative species names (based on Kuiter 2000) are *H. bleekeri* for Australian *H. abdominalis*, *H. elongates* for *H. subelongatus*, and *H.* cf. *reidi* for Brazilian *H. reidi*. Specimens not seen by the author or Sara Lourie (tissue samples only) are marked with asterics. Specimens used to amplify a particular gene fragment are marked with a dot. All cytochrome *b* sequences used in this study, as well as some additional 16S rRNA sequences, were downloaded from GenBank (16S rRNA accession numbers: AF355013 [*Hippocampus abdominalis*], AF354999 [*H. barbouri*], AF355007 [*H. erectus*], AF354991 [*Syngnathus acus*], Wilson *et al.* 2001; cytochrome *b* accession numbers AF192679-AF192686, Casey, 1999; AF356040 [*S. acus*], Wilson *et al.*, 2001)

Species name	Code	Collection locality	Collector/Source	Gene fragments sequenced		
				Ald	RP1	16S rRNA
Hippocampus abdominalis	abdoAU	SE Australia	(Project Seahorse)	•	•	
H. abdominalis	abdoNZ	New Zealand	(Project Seahorse)	•	•	
H. algiricus	algiBN	Benin	Z. Sohou	•	•	•
H. angustus	anguAU	W Australia	(Project Seahorse)	•		
H. barbouri	barbID1	Indonesia	(Project Seahorse)	•	•	
H. barbouri	barbID2	Indonesia	A. Tuwo		•	
H. bargibanti	bargID1	Indonesia	Sara Lourie		•	
H. bargibanti	bargID2	Indonesia	M. Erdmann			•
H. borboniensis	borbTZ	Tanzania	J. Schulz	•		
H. borboniensis	borbMG	Madagascar	(Project Seahorse)	•		
H. breviceps	brevAU	SE Australia	(Project Seahorse)	•	•	
*H. breviceps**	brevAU2	SE Australia	(Australian Museum)	•		•
H. camelopardalis	cameTZ	Mozambique	M. Cherry	•	•	•
H. capensis	capeZA	S Africa	P. Teske	•	•	•
H. capensis	capeZA2	S Africa	P. Teske		•	
H. comes	comeVN	Vietnam	A. Vincent	•		•
H. comes	comePH	Philippines	N. Perante	•	•	•
H. comes	come??	(captive)	M. Gunter			
H. coronatus	coroJP	Japan	C. Kawamura	•	•	•
H. erectus	erecUS	Florida	J. Campsen	•	•	
H. fisheri	fishUS	Hawai'i	(Project Seahorse)	•	•	
H. fuscus	fuscEG	Egypt	H. Gabr		•	•
*H. fuscus?**	fuscSL	Sri Lanka	Leo Smith		•	
H. fuscus	fusc??	(captive)	(Project Seahorse)	•	•	
H. guttulatus	guttIT	Italy	(Project Seahorse)	•	•	
*H. guttulatus**	guttPG	Portugal	J. Curtis	•		•
*H. hippocampus**	hippPG	Portugal	J. Curtis	•	•	•
H. histrix	histVN	Vietnam	Hoang		•	
H. ingens	ingePE	Peru	(Project Seahorse)	•	•	•
H. ingens	ingeMX	Mexico, Pacific coast	J. Baum		•	
H. kelloggi	kellIN	India?	(Project Seahorse)	•	•	
H. kelloggi	kellVN	Vietnam	(Project Seahorse)	•	•	•
H. kuda	kudaZA	S Africa	(Sea World Durban)	•	•	•
H. kuda	kudaPH	Philippines	M. Santos	•	•	•
H. kuda?	kudaIN	India?	(Project Seahorse)		•	
H. mohnikei	mohnVN	Vietnam	L.-S. Feng		•	
H. mohnikei	mohnJP1	Japan	C. Kawamura		•	•
H. mohnikei	mohnJP2	Japan	T. Mukai	•	•	
*H. procerus**	procAU	Australia? (captive)	L. Smith		•	
H. queenslandicus	queeAU	NE Australia	P. Southgate	•	•	•
H. reidi	reidMX	Mexico, Caribbean coast	J. Baum	•	•	•
*H. reidi**	reidBR	Brazil (captive)	L. Smith		•	
*H. sindonis**	sindJP	Japan	T. Mukai	•	•	•
H. spinosissimus	spinPH	Philippines	(Project Seahorse)	•	•	•
H. subelongatus	subeAU	SW Australia	A. Jones	•	•	•
H. trimaculatus	trimID	Indonesia	(Project Seahorse)	•	•	
H. trimaculatus	trimVN	Vietnam	(Project Seahorse)	•		
H. trimaculatus	trimHK	Hong Kong	(Project Seahorse)			•
*H. whitei**	whitAU1	SE Australia	(Australian Museum)	•	•	•
*H. whitei**	whitAU2	SE Australia	(Australian Museum)		•	
H. zosterae	zostUS	USA, Gulf of Mexico	(Project Seahorse)	•		
Syngnathus temminckii	temmZA	S Africa	P. Teske	•	•	

Tissue samples were cut into small pieces in order to improve digestion. Samples were then placed into 1.5 ml Eppendorf tubes containing 200 µl of amniocytic buffer (50mM Tris-Cl [pH 7.6], 100mM NaCl, 1mM EDTA [pH 8.0], 0.5% SDS) and 20 µl of 10mg/ml proteinase K. Tubes were incubated at 37°C for at least 24 h and occasionally agitated in order to improve digestion. Problematic samples were left to digest over several days. Once samples were sufficiently dissolved, the mixture was subjected to a standard phenol-chloroform extraction procedure (Sambrook *et al.* 1989). The remaining supernatant was precipitated with 50 µl of lithium chloride and 1 ml of cold (-20 °C) 100% ethanol, and stored overnight at -70 °C. Samples were then centrifuged for 15 min at 13,000 rpm. The liquid portion containing the ethanol was removed by inverting the tubes. In cases where DNA quantities were too low for pellets to be visible, most of the liquid portion was removed using a micropipette, and approximately 20 µl of fluid was left in the tube to evaporate. Pellets were subsequently resuspended in 20-50 µl of TE buffer (10mM Tris-Cl [pH 7.6], 0.1 mM EDTA) and these DNA samples were stored at -20°C.

The polymerase chain reaction (PCR) was used to amplify the first intron (RP1) of the S7 ribosomal protein (Chow & Hazama 1998), a portion of the Aldolase gene containing both coding and non-coding regions (forward primer: 5'-T G T G C C C A G T A T A A G A A G G A T G G–3'; reverse primer: 5'-C C C A T C A G G G A G A A T T T C A G G C T C C A C A A-3') and mitochondrial 16S rRNA using universal primers (Palumbi 1996). In some species, a tandem repetitive series in the RP1 intron resulted in sequencing complications. In order to sequence the remaining portions of the fragment, two internal primers were designed, forward primer S7RPEX3F (5'-T G G T G G A G T W G C A G T G A-3') and reverse primer S7RPEX4R (5'-A C A A A C A A C A G A C Y R G T A A-3').

Each 50 µL PCR reaction contained approximately 200 ng/ml of DNA, 0.2 µM of each dNTP, 5 µL of reaction buffer containing 100 mM NaCl, 0.1 mM EDTA and 20 mM Tris-HCl (pH 8.0), variable $MgCl_2$ (2 mM for RP1 and 16S rRNA and 2.5 mM for Aldolase), 0.4 µM of each primer respectively, and 1 unit of thermostable polymerase (Southern Cross Biotechnology). The PCR profile consisted of an initial denaturation step (5 min at 94 °C), followed by 35-40 cycles of denaturation (30 s at 94 °C), annealing for 1 min (at 60 °C for RP1 and Aldolase and 50 °C for 16S rRNA), and extension (1 min at 72 °C), and a final extension step (10 min at 72 °C). PCR products were then purified using a QIAquick PCR purification kit (Qiagen Ltd.), cycle-sequenced using BigDye sequencing kit (Applied Biosystems) and analysed on a 3100 AB automated sequencer.

Due to the poor quality of the DNA of many of the samples, PCR conditions were optimised in individual cases using one or several of the following methods: simplified "hot start" PCR, "touchdown PCR" and the addition of stabilizing PCR reagents. A "hot start" is achieved by withholding at least one reagent from the reaction mixture until the tubes containing the PCR reaction mixture can be placed in a thermal cycler approaching the denaturation temperature (Chou *et al.* 1992). By performing a "hot start" procedure, the frequency of misprimed products as well as the creation of primer oligomers is reduced (Chou *et al.* 1992). A "hot start" can also be achieved by keeping the reaction tubes on ice until the heat block of the thermal cycler had reached denaturing temperature (94°C), upon which the reactions are placed into the machine. This simplified method was used in this study. In "touchdown" PCR, the initial annealing temperature is set 10°C above the desired temperature and reduced by 1°C every second cycle until the desired annealing temperature is reached, at which temperature the remaining cycles are performed (Don *et al.* 1992). This procedure reduces the mispriming of the amplimers and hence the creation of short, spurious amplification products, and is particularly useful when working on longer target sequences. Lastly, the addition of bovine serum albumin (BSA) to the PCR reaction mixture improves amplification by binding to a large number of compounds that are potential polymerase inhibitors, including heme and fatty acids (Sensabaugh 1994).

1.2.2 Sequence alignment

When sequences were characterised by considerable length variation (which was the case in RP1 and 16S rRNA sequences), POY software (Gladstein & Wheeler 1997) was used to establish character homologies. The programme uses parsimony as an optimality criterion (Wheeler 1996): the preferred alignment is the one that produces the shortest tree. In the case of 16S rRNA sequences, a considerable difference was found between the number of transitions and the number of transversions in each of the implied alignments (see Results). To incorporate this information in the alignment procedure, the following cost scheme was applied: gap openings = 3, gap extensions = 1, transversions =2 and transitions =1. No such difference in the frequency of occurrence of transitions and transversions was apparent in the case of RP1 sequences, and the cost parameters gap openings = 2, gap extensions = 1, transversions = 1 and transitions = 1 were thus specified. No distinctions were made between conserved and variable regions. In each case, one of the implied alignments (computed *a posteriori* for each of the equally parsimonious trees inferred) was used as an input matrix for further analyses. Length

differences in the Aldolase sequences were rare and confined to just a few nucleotides (see Results). Hence, it was considered sufficient to align these sequences using default parameters in the alignment programme CLUSTALX (Thompson *et al.* 1997). Manual optimisations were performed where necessary. All alignments are shown in Appendix I, Tables A1-A4.

1.2.3 Phylogenetic analyses

To ensure authenticity, we attempted to obtain tissue material from several specimens for each species and amplify Aldolase and the RP1 intron of these at different times. However, many of the samples used in this study were of poor quality and did not amplify readily during PCR. For this reason, some specimens included in the RP1 phylogeny are not represented in the Aldolase phylogeny, and vice versa. Mitochondrial 16S rRNA was sequenced only in specimens for which both Aldolase and RP1 sequences were available. Alignment gaps in Aldolase, RP1 and 16S rRNA sequences were treated as missing characters. However, the phylogenetic information of indels that consisted of two or more consecutive base pairs, had clearly defined alignment borders, and were present in at least two different species was incorporated into parsimony analyses by coding these indels as additional characters. Properties of the four gene fragments used in this study were compared by determining nucleotide frequencies, maximum uncorrected p-distances among specimens, and the relative proportion of informative sites and empirical transition:transversion ratios using maximum likelihood (ML) estimation in PAUP* version 4.0 beta 10 (Swofford 2002). The data-base used consisted of selected ingroup taxa for which sequences were available for all four gene fragments. This reduced data-base was also used to construct strict consensus MP trees for each of the partitions and to compare their topologies as a means of comparing their phylogenetic signal. Alternative tree topologies were compared with the Shimodaira-Hasegawa test (SH test, Shimodaira & Hasegawa 1999) as implemented in PAUP*. The SH test is conservative, as it minimises the type I error rate (Buckley 2002). To assess the degree of saturation of transitions and transversions of the four partitions, these two types of mutations were plotted against proportional sequence divergence in pair-wise comparisons. In this case, all of the sequences available for each gene fragment were included.

As phylogenetic signal emerges better through the interaction of all data (Baker & DeSalle 1997; Cognato & Vogler 2001; Murrell *et al.* 2001; Buckley *et al.* 2002), topologies recovered from combined data are generally better resolved than those based on individual partitions

(Baker & DeSalle 1997; Matthee & Davis 2001; Murphy *et al.* 2001; Buckley *et al.* 2002; Gatesy *et al.* 2002). In order to maximise the descriptive and explanatory power of the evidence, the four partitions used in this study were thus initially combined into a supermatrix. Although this data-base contained a fair amount of missing data, Wiens (1998) suggested that unless the proportion of missing data is very large, addition of incomplete data sets is more likely to improve phylogenetic accuracy than reduce it. In order to determine the effect of missing data and incomplete taxon sampling on the phylogenies, we used several different approaches of combining the data. Firstly, we compiled a combined data set that included a single representative from each species for which sequence data was available for at least one of the four data partitions (referred to as data-base 'combined I'). Our second combined analysis included all the specimens sequenced for this study (Table 1) in one large combined data set (data-base 'combined II'). Lastly, we analysed six data partitions to screen for topological congruence: 1) combined nuclear data; 2) combined mitochondrial data; 3-6) each of the four data partitions separately. Sequences of 16S rRNA from Wilson *et al.* (2001) and cytochrome *b* data from Casey (1999) were included if sequence data was available for a particular species, despite the fact that these originated from different specimens. As most of the samples used in this study and in the study by Casey (1999) had been identified by Lourie (pers. comm.), or were identified based on the criteria in Lourie *et al.* (1999), there was no reason to assume inconsistencies with regard to identification. Wrong place here moved down

Phylogenetic relationships among taxa were investigated using maximum parsimony (MP) and Bayesian inference (BI). In order to deptict how the partitions differed with regard to evolutionary rate, phylogenies from individual partitions were presented as parsimony phylograms, which were drawn using default parameters in PAUP*. Bootstrap support from MP and posterior probability values from BI were then added to these trees. Phylogenetic congruence among data sets was tested using the partition homogeneity test (Farris *et al.* 1994, 1995) as implemented in PAUP*. In the case of combined data sets, consensus cladograms from MP and BI analyses were presented.

Parsimony tree searches were performed in PAUP* using the heuristic search option with 10 random-addition sequences and tree-bisection-reconnection branch swapping. The reliability of nodes was assessed using 1000 nonparametric bootstrap replicates (Felsenstein 1985) and in each case the heuristic search was limited to a maximum of 10 000 saved trees. In order to verify that no shorter trees are likely to be found using an alternative method, trees were also searched using the 'parsimony ratchet' in NONA version 2.0 (Goloboff 1998) in combination with WINCLADA (Nixon 1999-2002). This method is highly efficient at finding different islands

of trees and finds shorter trees more quickly by avoiding time spent searching on new starting trees that are much less optimal than the last tree swapped (Nixon 1998). Strict consensus topologies of MP trees of individual partitions and combinations thereof were compared to each other in a pair-wise fashion using the SH test. For consistency, the data-base used included 20 ingroup species for which sequence data were available from all four partitions, as well as pipefish sequences.

The programme MODELTEST version 3.06 (Posada & Crandall 1998) was used to conduct hierarchical likelihood ratio tests as a method to determine the most appropriate substitution model for each data partition based on the Akaike Information Criterion (AIC; Akaike 1973). Bayesian analyses were performed using MRBAYES version 3.0B (Huelsenbeck & Ronquist 2001) available at:

http://brahms.biology.rochester.edu/software.html.

The Markov chain Monte Carlo process was set for 4 chains to run simultaneously for 1000 000 generations, with trees being sampled every 50 generations for a total of 20 000 trees in the initial sample. Maximum likelihood parameters in MRBAYES were specified for each partition according to the most appropriate evolutionary model identified by MODELTEST. Several methods were used to ensure that analyses were not trapped on local optima. Firstly, analyses were run independently five times, and their apparent stationarity levels were compared for convergence (Huelsenbeck & Bollback 2001). Secondly, Metropolis-coupled Markov chain Monte Carlo was used to enhance the tree-climbing ability of the Markov chains. This method generates incrementally heated Markov chains, which enables a more thorough exploration of parameter space (Marinari & Parisi 1992; Geyer & Thompson 1995; Huelsenbeck & Ronquist 2001). We used four incrementally heated Markov chains. Thirdly, the posterior probabilities for individual clades obtained from five separate analyses were compared for congruence (Huelsenbeck & Imennow 2002). Graphic examination of variation in maximum likelihood scores identified that in all cases, 'burnin' was complete prior to the 20 000th generation. Hence, the first 400 trees from each run were discarded, and the posterior probability of the phylogeny and its branches was determined from the remaining 19 600 trees. For each gene fragment, a single 50% majority rule consensus tree was then drawn in PAUP*. All trees were rooted using outgroups (see Results) and tree topologies obtained with MP and BI were compared using the SH test. In order to determine the effects of using different outgroup species on tree topologies and nodal support, several different data sets were analysed using MP only. Topologies derived from data sets utilising different outgroups were also compared using SH tests.

1.2.4 Molecular clock calibration

The concept of a molecular clock was introduced by Zuckerkandl & Pauling (1962) and proposes that DNA sequences evolve at rates constant across lineages and time. As the rate of substitution is constant across lineages, all tips on a phylogenetic tree should be at an equal distance to the root of the tree, and the mean and variance of the number of substitutions that occur in different lineages in the same amount of time should be equal (Zuckerkandl & Pauling 1965). The application of a molecular clock was considered useful when investigating the relationship between evolutionary patterns and their causes, particularly in the absence of adequate fossil evidence (Arbogast *et al.* 2002). However, although some genes may exhibit a clock-like behaviour, if only for a certain time period of relevance to a particular study (e.g. Ochman & Wilson 1987; Moran *et al.* 1993; Oleksiewicz *et al.* 2000; Lourie, in press), numerous studies have since indicated that the assumption of rate homogeneity among different lineages is often not met (e.g. Ayala 1986; Britten 1986; Jorgensen & Cluster 1988; Kwon *et al.* 1991; Gaut *et al.* 1992; Campbell *et al.* 1994; Clegg *et al.* 1994; Goremykin *et al.* 1997).

Despite the fact that the molecular clock hypothesis does not perfectly explain the substitution process, the concept remains an important tool in evolutionary bioloy: the molecular clock does provide a time scale, albeit an imperfect one (Huelsenbeck *et al.* 2000). Several methods have been proposed to deal with rate heterogeneity among lineages. For example, relative rate tests were developed to identify individual taxa which showed significant rate differences relative to other taxa under investigation (Tajima 1993; Robinson 1998). These problematic taxa were then exluded when estimating divergence dates. Recent work has suggested the possibility of estimating dates without assuming a global molecular clock (Sanderson 1997; Rambaut & Bromham 1998; Thorne *et al.* 1998; Huelsenbeck *et al.* 2000; Yoder & Yang 2000; Kishino *et al.* 2001). Sanderson (1997) developed a nonparametric method for smoothing the rate differences across speciation events on a phylogenetic tree, thus allowing rates to vary among branches. The method of Huelsenbeck *et al.* (2000) allows rates to vary across lineages according to a compound Poisson process. When an event of substitution rate change occurs, the current rate of substitution is modified by a gamma-distributed random variable. Rambaut & Bromham (1998) developed the 'quartet-dating method', a maximum likelihood-based approach which implements a specific rate model on a tree of four species. Yoder & Yang (2000) extended this model to the general case, with an arbitrary assignment of rate classes to branches. In the method developed by Thorne *et al.* (1998) and Kishino et al. (2001), a stochastic model of evolutionary change is used describe the

rate of change over evolutionary time, and the Bayes theorem is then used to derive posterior distributions of rates and times. Compared to likelihood-based methods, the Bayesian approach has been shown to perform better when rate changes affect many branches scattered across the tree, and when uncertainties exist regarding the exact dates of divergence events used to calibrate the clock (Yang & Yoder 2003).

We estimated ages of divergence events among different seahorse lineages by means of the Bayesian relaxed clock method for multiple genes implemented in the programmes ESTBRANCHES and MULTIDIVTIME (Thorne & Kishino 2002). Likelihood space is searched using Markov chain Monte Carlo, and provided that suitable data for divergence events are available, upper and lower calibration points can be specified for a particular node. To determine whether different combinations of partitions and taxa would yield similar results, two different combinations of these were used: one included the pygmy seahorse, *Hippocampus bargibanti* and two partitions for which sequence data from this species were available, and the other excluded this species and included sequence data from all four partitions. A pipefish (genus *Syngnathus*) was used as an outgroup taxon in both cases. MACCLADE version 4.06 (Maddison & Maddison 2003) was used to construct a topology congruent with the most likely topology derived from MP and BI using different combinations of taxa and partitions (section 1.2.3). The programme BASEML in the PAML package (Yang 1997) was used to estimate model parameters under the F84 model (Felsenstein & Churchill 1996). These were then used to estimate ML branch lengths for the rooted ingroup tree and their variance-covariance matrix using the programme ESTBRANCHES. The programme MULTIDIVTIME was then used to estimate the posterior ages of divergence events, as well as their standard deviations and 95% credibility intervals. In each case, the Markov chain was sampled 10 000 times with 100 cycles between each sample and burn-in after 10 000 cycles. 100 million years (a date which precedes the age of the oldest Syngnathid fossil by nearly 50 million years; Patterson 1993) was specified as the highest possible number of time units between tip and root. The application of even older dates (200, 300 and 500 million years) did not affect the results. The prior expected number of time units between tip and root (*rrtm*) was set to 30 million years. This value was chosen rather arbitrarily, as no hypothesis regarding the time of origin seahorses has been proposed in the literature. It was considered suitable as it falls into the Oligocene (Harland *et al.* 1989), i.e. the epoch between the Eocene (from which the oldest Syngnathid fossils are known) and the Miocene (from which the only fossils of 'true seahorses' are known; Sorbini 1988). The corresponding prior standard deviation was set to 15 million years. To determine whether this prior affected age estimates, the programme was also run at rrtm values of 20 and 40 million

years, with prior standard deviations of 10 and 20 million years, respectively. The mean of the prior distribution for the rate at the root node (*rtrate*) was estimated by dividing the median of all branch lengths from root to ingroup tips by *rrtm*., following the procedure described in the MULTIDIVTIME manual. It must be noted that relating *rtrate* to *rrtm* may not be appropriate, as priors should ideally be random, but if the programme is run for sufficiently long, the impact of the priors on the results should be minimal. This was confirmed by the fact that doubling or tripling *rtrate* did not significantly change the results. To check for convergence of MCMC analyses, two independent runs were performed in each case, which differed only with regard to their starting points.

1.3 Results

1.3.1 Characterisation of the four gene fragments

Primers for Aldolase amplified a single fragment 267-268 bp in length and this region included 127 exon and 140-141 intron characters. The RP1 primers amplified intron sequence that varied from 658 bp in *Hippocampus kelloggi* to 678 bp in *H. camelopardalis*. The fragment also included 30 bp of flanking sequence from exon 1 and 65 bp from exon 2. In some cases, two peaks equal in intensity were present at a single position on the chromatograph, indicating heterozygosity. All the heterogeneous sites were coded with IUB ambiguity codes. Total length of the aligned RP1 intron sequences was 706 bp (including alignment gaps). Sequencing of the complete RP1 intron was problematic in the three species *H. comes*, *H. subelongatus* and *H. sindonis* due to the presence of several AT-rich arrays. This problem was partly resolved by designing internal primers but unfortunately, the final data sets lack approximately 62 (*H. comes*), 184 (*H. subelongatus*) and 176 (*H. sindonis*) nucleotides respectively. Mitochondrial 16S rRNA sequences used were nearly complete (12 nucleotides were removed at the 5' end, and these were not variable among different species in cases where nucleotides could be readily interpreted on the chromatographs). The length of 16S rRNA sequences ranged from 473 bp (*H. bargibanti*) to 486 bp (*H. spinosissimus* and *H. hippocampus*). A total of 38 indels was found among aligned ingroup sequences in RP1, the largest one being 22 bp in length. However, as most of these indels overlapped, were present in only a single species, or were a single base-pair in length, only five indels were coded as characters. Among the Aldolase sequences, a single non-overlapping indel one base-pair in length was found in *H. capensis* and *H. breviceps*, respectively, whereas the pipefish sequence contained two indels. The aligned 16S rRNA sequences contained 16 indels. As all of the indels in Aldolase and 16S rRNA sequences either overlapped or were only a single base-pair in length, they were not coded as discrete characters. We pruned the edges of all sequences by removing characters close to the primers to reduce ambiguity. The remaining lengths of each of the aligned gene fragments analysed were 640 characters (RP1 intron; exon portions of the S7

ribosomal protein gene were removed entirely as they were invariable), 188 characters (Aldolase) and 464 characters (16S rRNA). Cytochrome *b* sequences downloaded from GenBank were reduced to a total length of 1020 nucleotides. All sequences generated in this study have been deposited in GenBank (accession nos. AY277286-AY277374).

Based on the number of parsimony informative characters, Aldolase is the most conserved gene fragment, and the cytochrome *b* gene evolves most rapidly (Table 1.2). The number of transitions and transversions was approximately equal in the two nuclear gene fragments, whereas mitochondrial gene fragments were characterised by high ti/tv ratios. Saturation plots show that there is evidence for saturation in case of the two mitochondrial partitions only (Fig. 1.1).

Table 1.3 Comparisons of partitions and combinations thereof using 20 ingroup species for which sequences are available for all four molecular markers.

Gene fragments	Type	Total sites	Informative sites	Nucleotide frequencies				ti/tv	Maximum p-distance
				%A	%C	%G	%T		
Aldolase	nDNA	188	11 (6%)	23.8	23.6	22.1	30.5	1.08	0.04
RP1	nDNA	645	58 (9%)	24.5	18.3	24.3	32.9	0.92	0.09
16S rRNA	mtDNA	464	62 (13%)	29.5	24.4	19.9	26.2	2.55	0.09
Cytochrome *b*	mtDNA	1020	308 (30%)	23.7	27.4	15.5	33.5	6.70	0.19
nDNA combined	nDNA	833	69 (8%)	24.4	21.8	23.2	30.6	0.94	0.09
mtDNA combined	mtDNA	1484	362 (24%)	30.6	25.5	13.6	30.4	5.60	0.16
Combined	Both	2317	431 (19%)	27.5	24.2	16.6	31.7	3.79	0.15

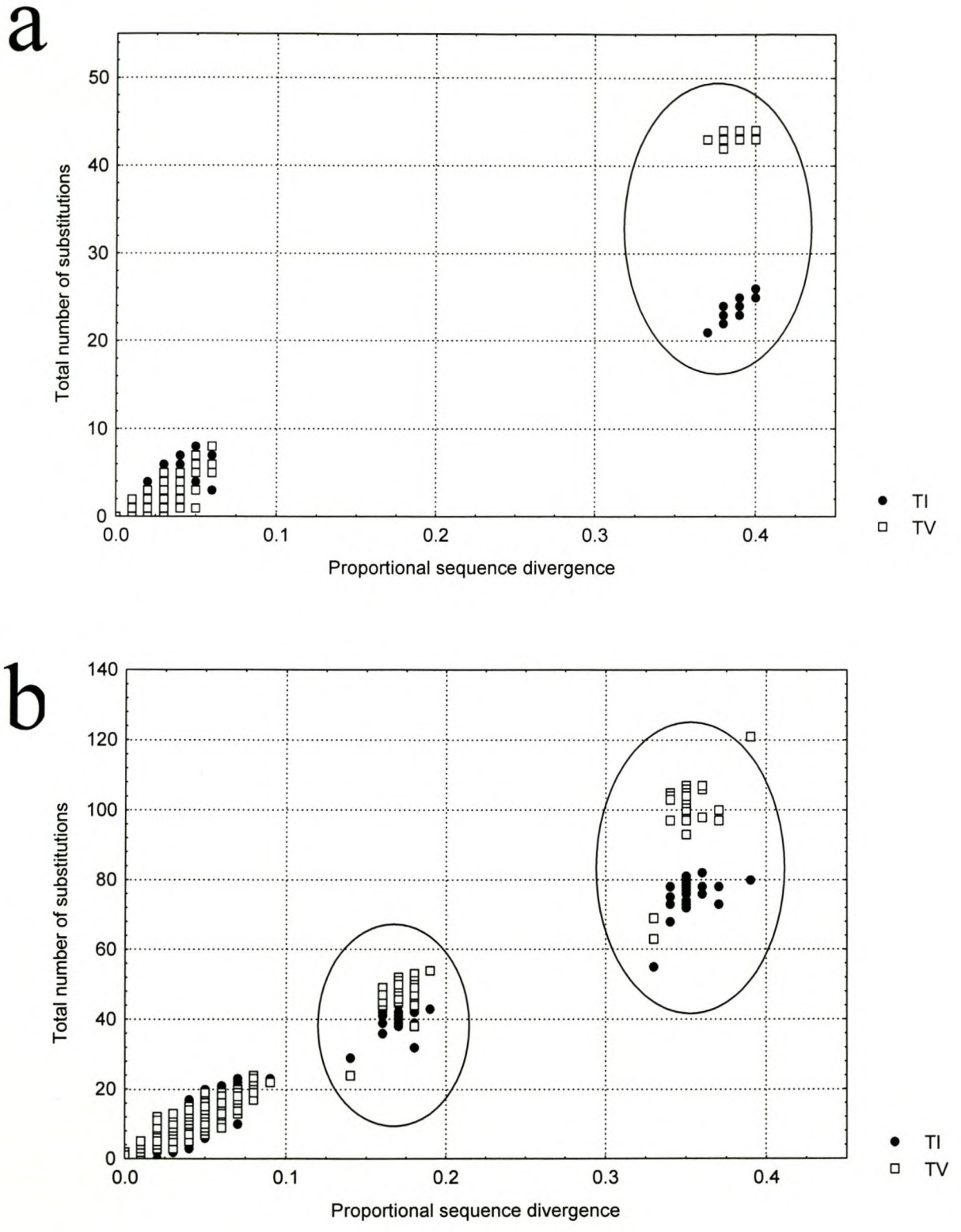

Fig. 1.1 Saturation plots of individual partitions: a) Aldolase; b) RP1. Encircled clusters represent pairwise sequence divergence vs. number of substitutions between ingroup and outgroup taxa.

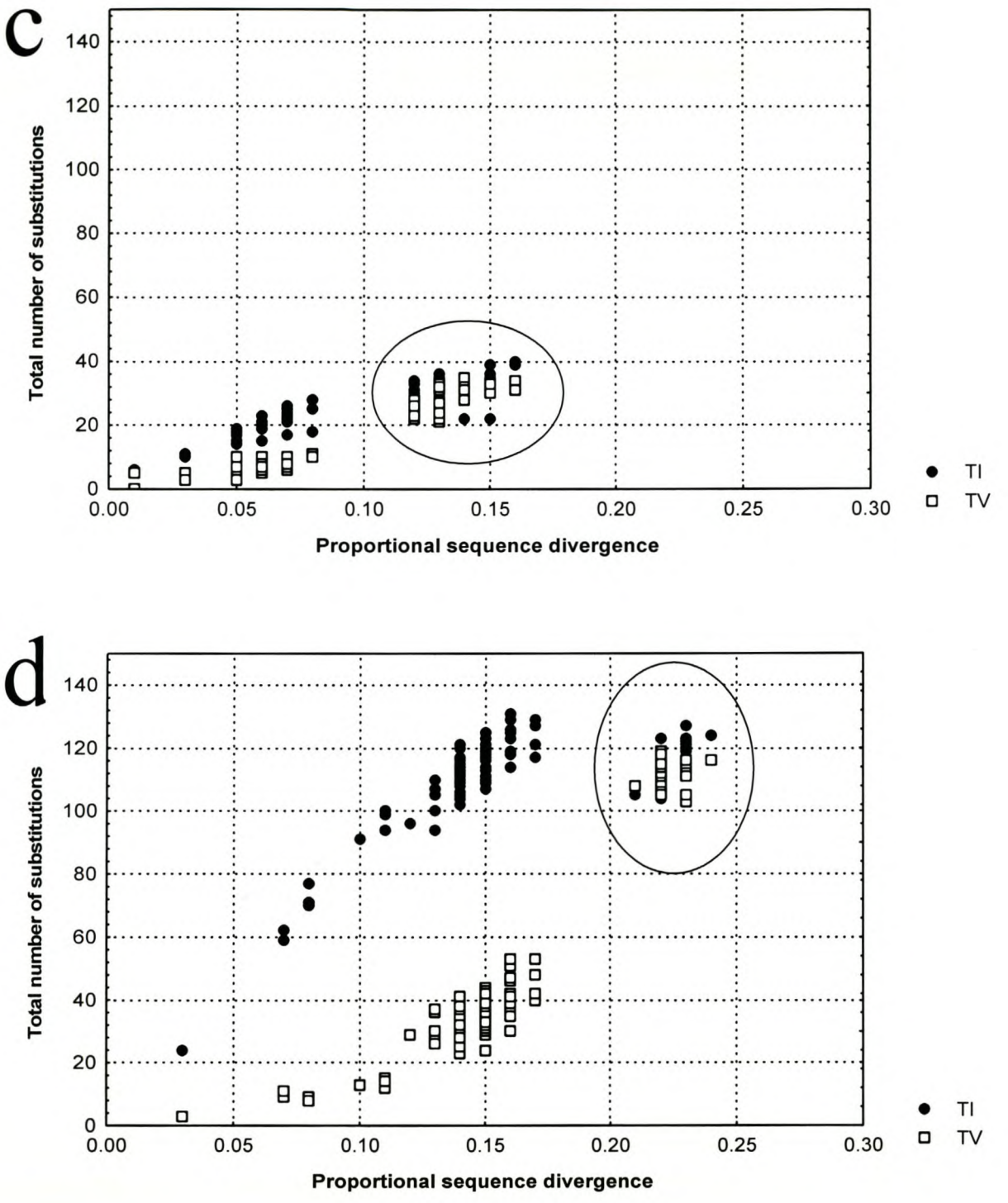

Fig. 1.1 (continued) Saturation plots of individual partitions: c) 16S rRNA; d) cytochrome *b*. Encircled clusters represent pair-wise sequence divergence vs. number of substitutions between ingroup and outgroup taxa.

Uncorrected sequence divergence values among pipefish, pygmy seahorse (*Hippocampus bargibanti*), and all other seahorse species were considerably larger in the case of nuclear markers than in the case of the two mitochondrial gene fragments (Table 1.3). For example, divergence values between the pipefish and the most distant ingroup species ranged from 0.16 in 16S rRNA (1.3 times the maximum value found in the ingroup) to 0.39 in Aldolase (6.5 times the maximum value found in the ingroup). This indicates that most sites in the nuclear gene fragments (mostly intron sequences) are free to vary, whereas real distances for the mtDNA data are probably underestimated due to saturation (see also Creer *et al.* 2003). As most indels were coded as missing data, their effect on p-distances was negligible (Table 1.3).

Table 1.4 Uncorrected p-distances among the two outgroup taxa *Syngnathus temminckii* and *Hippocampus bargibanti*, among outgroup and ingroup taxa, and maximum values found among ingroup taxa. Note that sequences of *H. bargibanti* were not available for Aldolase and cytochrome *b*. To determine the impact of indels on genetic distances, Aldolase, RP1 and 16S rRNA sequences were analysed both with and without sections containing indels (A and B, respectively).

Gene fragment	Type	*Syngnathus* vs. *H. bargibanti*		*Syngnathus* vs. ingroup		*H. bargibanti* vs. ingroup		Max. value within ingroup	
		A	B	A	B	A	B	A	B
Aldolase	nDNA	-	-	0.37-0.39	0.37-0.39	-	-	0.06	0.05
RP1	nDNA	0.37	0.38	0.31-0.35	0.31-0.35	0.14-0.20	0.17-0.21	0.10	0.09
16S rRNA	mtDNA	0.15	0.14	0.13-0.16	0.12-0.16	0.11-0.14	0.12-0.14	0.09	0.09
Cytochrome *b*	mtDNA	-		0.21-0.24		-		0.19	

1.3.2 Phylogenetic analyses

Initially, phylogenetic trees constructed in PAUP* using maximum parsimony were rooted using a pipefish (genus *Syngnathus*) as an outgroup taxon. However, due to the considerable sequence divergences between most seahorses and pipefishes (for example up to 35% in RP1 and up to 39% in Aldolase, Table 1.3) it was considered necessary to explore the utility of a basal ingroup species as an additional outgroup taxon. Apart from this, the inclusion of additional outgroup species may reduce the introduction of errors in the polarisation of characters (Milinkovitch *et al.* 1996; Milinkovitch & Lyons-Weiler 1998) and is also useful to test the assumption of ingroup monophyly (Baverstock & Moritz 1996). Evidence from the RP1, 16S rRNA and combined sequences (using both outgroup rooting with a pipefish sequence and midpoint rooting) indicated a strongly supported basal position of the pygmy seahorses (*Hippocampus bargibanti*) within the phylogeny. Hence, the pygmy seahorse was used as an additional outgroup taxon whenever sequence data for this species were available.

Partition homogeneity tests (Farris *et al.* 1994, 1995) only indicated significant conflict in the phylogenetic signal when RP1 sequence data were compared with either Aldolase or 16S rRNA data (Table 1.4). However, the reliability of the partition homogeneity test as a statistical method of deciding whether to combine independent data sets has been questioned (Yoder *et al.* 2001). Sanderson & Donoghue (1989) suggested that homoplasy may not be correlated with phylogenetic accuracy, Downtown & Austin (2002) argued that the test is unlikely to be an effective measure of congruence when two data sets differ markedly in size, and Gatesy *et al.* (1999) suggested that data sets which are in conflict over some parts of a tree may be congruent over other parts.

Table 1.5 Results of partition homogeneity tests comparing phylogenetic information from the four genetic markers. Above diagonal: number of steps; below diagonal: p-values. P-values below 0.05 are shown in italics.

	Aldolase	RP1	16S rRNA	Cytochrome *b*
Aldolase		6	8	23
RP1	***0.04***		17	20
16S rRNA	0.57	***0.01***		21
Cytochrome *b*	0.99	0.70	0.82	

Maximum parsimony analyses of the four separate data sets yielded from 3 (Aldolase and cytochrome *b*) to 706 (RP1) equally parsimonious trees (Table 1.5). In the case of combined analyses, the highest number of trees was found with the 'combined II' data-base. The shortest trees found using the parsimony ratchet in NONA were of equal length in all cases. SH tests revealed that most of the MP tree topologies constructed for each partition and combinations thereof using consensus data-bases that included 20 ingroup species and pipefish sequences were not significantly different from each other (Table 1.6). An exception was the topology of the Aldolase tree, which differed significantly (at $\alpha = 0.05$) from the RP1, 16S rRNA, cytochrome *b*, combined mtDNA and combined nDNA tree topologies. Nevertheless, despite differences in topology, resolution and the amount of homoplasy, no well-supported nodes were found that were in conflict among different partitions. Topologies based on BI were not significantly different from their MP counterparts in three of the individual partitions (Aldolase, RP1 and 16S rRNA), but they differed significantly in the case of cytochrome *b*, as well as in all analyses based on combined data (Table 1.7).

Table 1.6 Tree statistics from parsimony analyses of individual and combined data partitions.

	Number of				Tree length	RI
	ingroup specimens	characters	informative characters	equally parsimonious trees		
Aldolase[1]	34	188	21	3	97	0.95
RP1[2]	40	645	148	706	501	0.87
16S rRNA[2]	25	464	79	8	293	0.67
Cytochrome *b*[1]	23	1020	338	3	1471	0.54
Combined nDNA[2]	21	833	131	1	597	0.75
Combined mtDNA[2]	28	1484	415	3	1864	0.46
Combined I[2]	28	2317	536	218	2460	0.52
Combined II[2]	49	1297	242	2588	883	0.80

[1] *Syngnathus* used as outgroup species
[2] *Syngnathus* and *Hippocampus bargibanti* used as outgroup species

Table 1.7 Pair-wise SH tests of the topologies of parsimony trees constructed using a consensus set of 20 ingroup species for which sequence data are available from all four partitions. A pipefish (*Syngnathus*) was used as an outgroup species. As trees constructed using combined I (including cytochrome *b* sequences) and combined II (excluding cytochrome *b*) data sets resulted in trees with identical topologies (likelihood = 656), only one column is shown that compares topologies of trees constructed using data sets including more than two partitions. Above diagonal: difference in -ln likelihood values of trees; below diagonal: p-values. P-values below 0.05 are shown in italics.

	Aldolase	RP1	16S	Cyt *b*	nDNA	mtDNA	Combined
Aldolase		26.60	20.89	16.54	14.88	17.55	13.15
RP1	***0.018***		5.70	10.06	11.72	9.06	13.45
16S	***0.023***	0.173		4.35	6.02	3.35	7.74
Cyt *b*	***0.043***	0.052	0.227		1.66	1.00	3.39
nDNA	***0.039***	0.072	0.109	0.314		2.66	1.72
mtDNA	***0.034***	0.067	0.285	0.153	0.233		4.39
Combined	0.055	0.054	0.069	0.220	0.082	0.155	

Table 1.8 Results of SH tests for pair-wise comparisons of strict consensus tree topologies constructed with data-bases from four genes and combinations thereof using maximum parsimony (MP) and Bayesian inference (BI). Pipefish (*Syngnathus temminckii*) and pygmy seahorse (*Hippocampus bargibanti*, if available) sequences were included. P-values below 0.05 are shown in italics.

Data-base	Method	-ln L	Δ -ln L	p
Aldolase	MP	674	16	0.68
	BI	690		
RP1	MP	3371	15	0.06
	BI	3355		
16S rRNA	MP	2277	20	0.15
	BI	2298		
Cytochrome *b*	MP	8214	109	***<0.01***
	BI	8323		
nDNA combined	MP	3969	28	***0.02***
	BI	3941		
mtDNA combined	MP	11200	111	***<0.01***
	BI	11088		
Combined I	MP	14327	181	***<0.01***
	BI	14508		
Combined II	MP	6364	79	***<0.01***
	BI	6444		

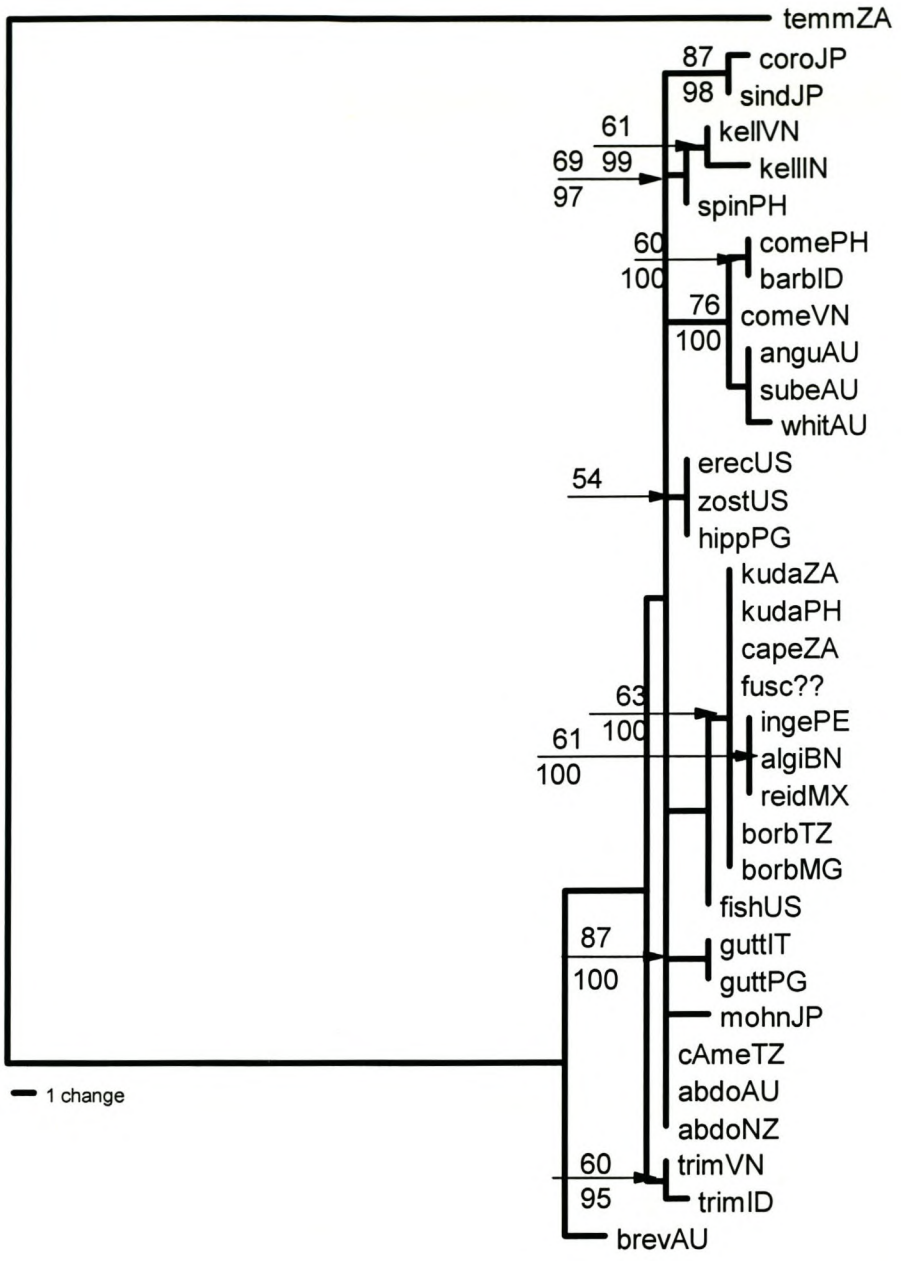

Fig. 1.2a Phylogram of one of three equally most parsimonious trees constructed from Aldolase sequences. Values above branches/arrows are bootstrap values from parsimony analysis, and values below branches are posterior probabilities from Bayesian inference.

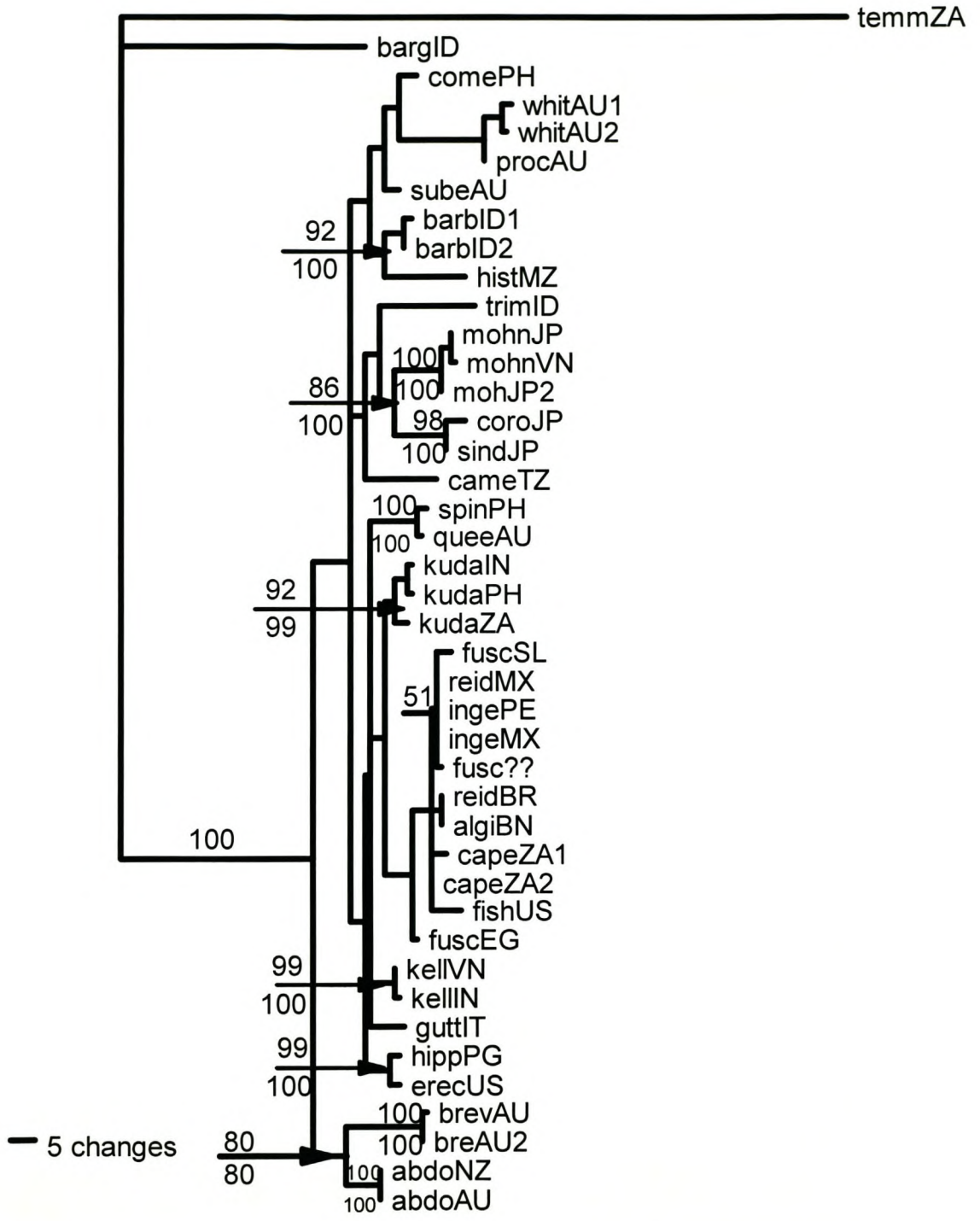

Fig. 1.2b Phylogram of one of 706 equally most parsimonious trees constructed from RP1 sequences. Values above branches/arrows are bootstrap values from parsimony analysis, and values below branches are posterior probabilities from Bayesian inference.

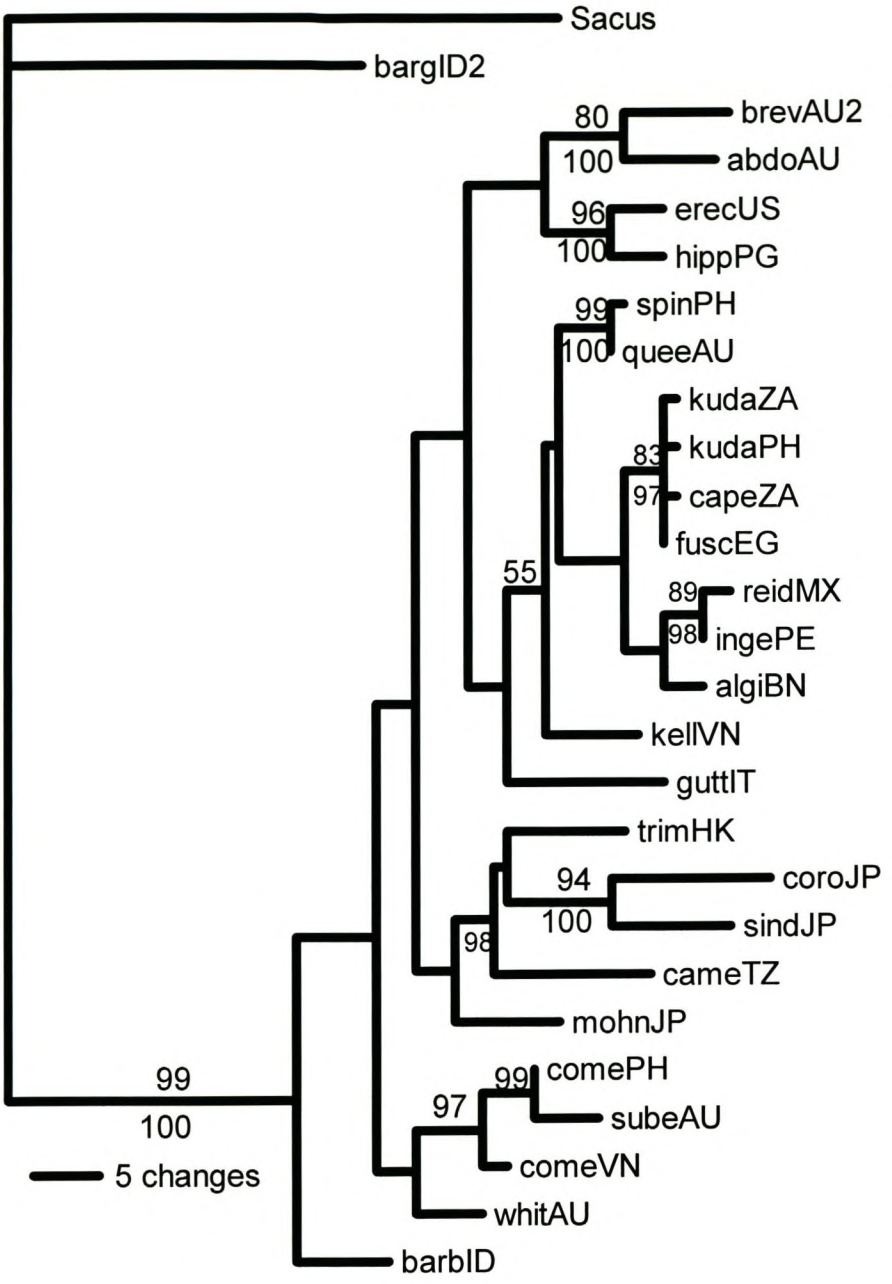

Fig. 1.2c Phylogram of one of eight equally most parsimonious trees constructed from 16S sequences. Values above branches are bootstrap values from parsimony analysis, and values below branches are posterior probabilities from Bayesian inference.

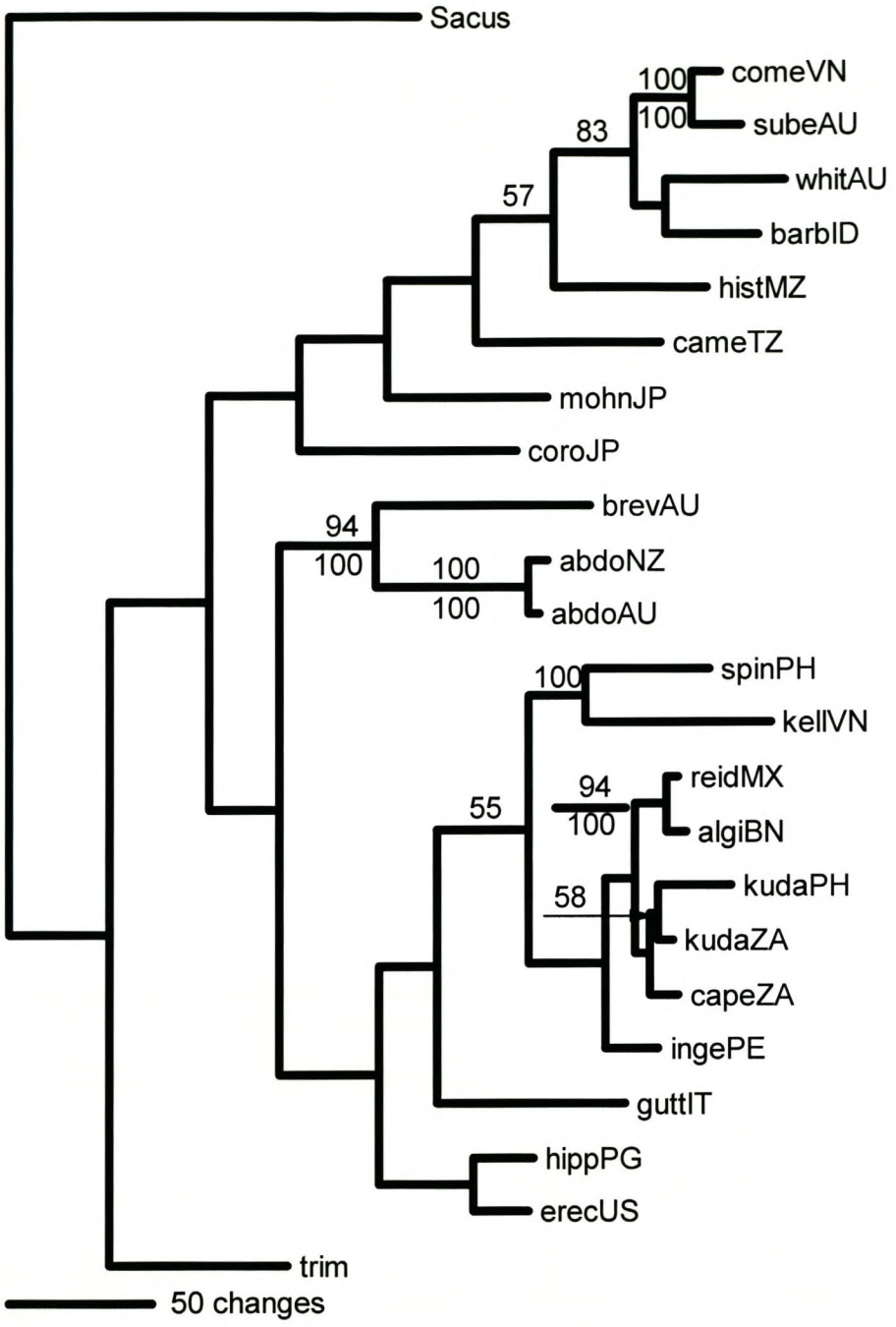

Fig. 1.2d Phylogram of one of three equally most parsimonious trees constructed from cytochrome *b* sequences. Values above branches are bootstrap values from parsimony analysis, and values below branches are posterior probabilities from Bayesian inference.

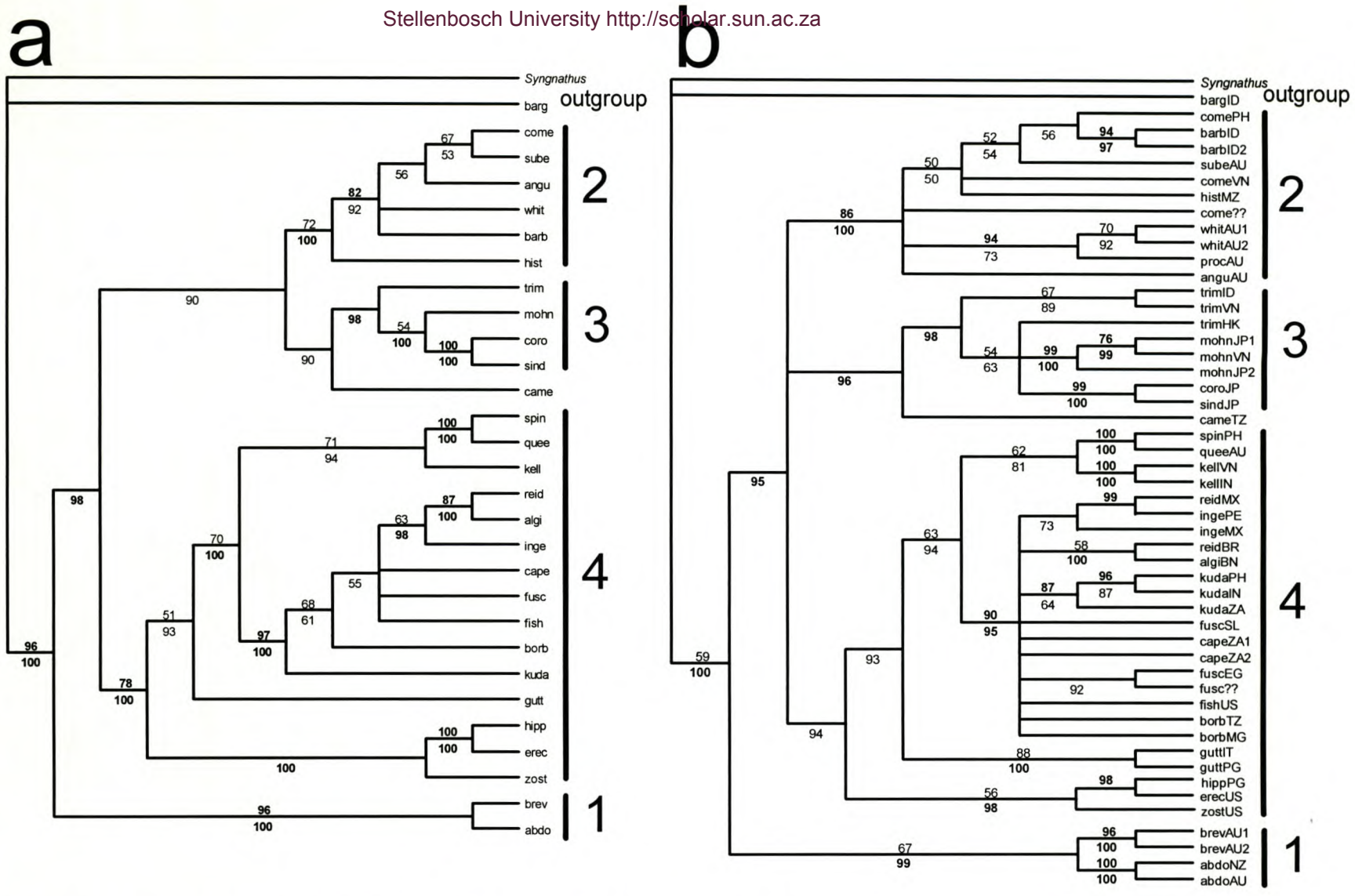

Fig. 1.3 Cladograms of combined analyses; data-bases used were (a) a combination of sequences from four partitions (Aldolase, RP1, 16S rRNA and cytochrome *b*) including a single representative from each species and (b) a combination of sequences from all the specimens sequenced in this study comprising three partitions (Aldolase, RP1 and 16S rRNA). Nodal support is indicated by bootstrap values from parsimony analyses and posterior probabilities from Bayesian inference (above and below branches, respectively). High nodal support (bootstrap values ≥75% and posterior probabilities ≥95%) is shown in boldface. Nodes not labeled had < 50% support.

The highest number of nodes was recovered using the data set including all four gene fragments (combined I). At least four monophyletic clades were retrieved irrespective whether cytochrome *b* was included or not, or whether several specimens were included from each species (Figs. 1.2 and 1.3). Apart from clade 1, which was always placed basal in the phylogeny, the exact branching patterns among clades differed when taxa/data were included and excluded. In order to determine whether the utility of a more closely related outgroup species resulted in significantly different phylogenies, the seahorses of clade 1 were used as an alternative outgroup. Pair-wise SH tests on MP trees constructed using different outgroups (i.e. *Syngnathus*/*H. bargibanti* vs. *H. breviceps*/*H. abdominalis*) revealed a significant topological difference in the case of combined mitochondrial data only; in contrast, trees constructed exclusively with nuclear data (both individual partitions and combined data) had identical topologies (i.e. p = 1.0, Table 1.7).

Because the inclusion of additional taxa and cytochrome *b* data did not alter the results of this study, we based our conclusions on a consensus phylogeny (Fig. 1.4), which is based on congruence among the results obtained from all the partitioned and combined analyses and also congruence among the different phylogenetic methods (Table 1.8). In this consensus approach we argue that nodes A, B, D, F, G, H, and L were generally well supported by at least one of the markers/combined data sets. Overall nodal support was weakest for nodes C, E and I.

Table 1.9 Results of SH tests for pair-wise comparisons of strict MP consensus tree topologies constructed using alternative outgroups: a) *Syngnathus/H. bargibanti* outgroup; b) *H. abdominalis/H. breviceps* outgroup.

Data-base	Outgroup	-ln L	Δ -ln L	p
Aldolase	a	446	0.00	1.00
	b	446		
RP1	a	2453	0.00	1.00
	b	2453		
16S rRNA	a	1870	0.00	1.00
	b	1870		
Cytochrome *b*	a	6688	4.05	0.37
	b	6692		
nDNA combined	a	2882	0.00	1.00
	b	2882		
mtDNA combined	a	9972	106.64	<0.01
	b	9865		
Combined I	a	12390	20.63	0.10
	b	12370		
Combined II	a	5269	0.089	0.49
	b	5269		

Table 1.10 Summary of branch support for nodes A-N in Fig. 1.4 using four gene fragments and combinations thereof. First row (MP1): bootstrap support from parsimony analysis using a pipefish (genus *Syngnathus*) and, if available, a pygmy seahorse (*Hippocampus bargibanti*) as outgroup species. Second row (MP2): Bootstrap support from parsimony analysis using *H. abdominalis* and *H. breviceps* as outgroup species. Third row (BI): posterior probabilities from Bayesian inference using the complete data sets. Boostrap and posterior probability values below 50% are indicated as "-". Strongly supported nodes (bootstrap values ≥75% and posterior probabilities ≥95%) are shown in boldface. Roman numerals for combined sequences refer to: I) supermatrix contained a single individual from each species and a maximum of four molecular markers each; II) supermatrix contained all specimens sequenced for this study and a maximum of three gene fragments each (cytochrome *b* excluded). If a clade could not be recovered because a particular sample was not represented in the data matrix (e.g. *H. zosterae* in clade N), this was indicated with a question mark.

		Nodes													
Data set		A	B	C	D	E	F	G	H	I	J	K	L	M	N
Aldolase	MP1	-	-	-	**76**	-	-	-	-	-	-	69	**87**	61	54
	MP2	?	-	-	**85**	-	-	-	-	-	-	67	**86**	-	56
	BI	-	-	-	**100**	-	-	-	-	-	-	**97**	**98**	**100**	-
RP1	MP1	**100**	**91**	-	**80**	72	**86**	**89**	**85**	-	-	-	**77**	-	?
	MP2	?	**100**	-	**83**	65	**81**	**88**	**88**	-	-	-	**81**	-	?
	BI	89	89	-	**100**	54	**100**	**100**	89	-	-	-	89	-	?
16S rRNA	MP1	-	-	-	-	-	-	-	-	-	55	-	-	**76**	?
	MP2	?	**82**	-	-	-	-	-	-	56	**75**	-	**90**	**81**	?
	BI	-	-	-	-	-	-	-	-	-	-	-	-	**94**	?
Cytochrome *b*	MP1	-	-	51	**77**	-	-	-	**88**	-	55	63	**100**	**92**	?
	MP2	?	**100**	-	**72**	-	-	51	**97**	-	-	58	**100**	-	?
	BI	-	-	-	-	-	-	-	**100**	-	-	-	**100**	-	?
Combined nDNA	MP1	**98**	**85**	-	**92**	69	**82**	**87**	68	-	-	54	**89**	-	67
	MP2	?	**97**	-	**95**	57	**80**	86	62	-	-	-	**92**	54	62
	BI	**95**	93	-	**100**	71	**97**	**99**	**95**	66	-	76	**100**	-	**95**
Combined mtDNA	MP1	**85**	-	-	56	-	-	-	**80**	-	72	58	**100**	-	?
	MP2	?	**100**	-	**100**	-	-	55	**98**	-	71	59	**100**	-	?
	BI	62	-	-	55	-	87	-	77	-	**97**	64	**99**	-	?
Combined I	MP1	**96**	-	-	72	-	-	54	**78**	51	70	71	**97**	63	-
	MP2	?	**100**	-	**90**	-	-	68	**97**	-	64	69	**96**	61	-
	BI	**100**	**98**	**90**	**100**	90	**98**	**100**	**100**	93	**100**	94	**100**	**98**	**100**
Combined II	MP1	59	-	-	**86**	-	-	-	-	-	63	62	**90**	-	56
	MP2	?	**82**	-	**92**	-	-	-	-	52	71	65	**92**	-	62
	BI	**100**	**95**	-	**100**	**96**	**98**		94	93	94	81	**95**	-	**98**

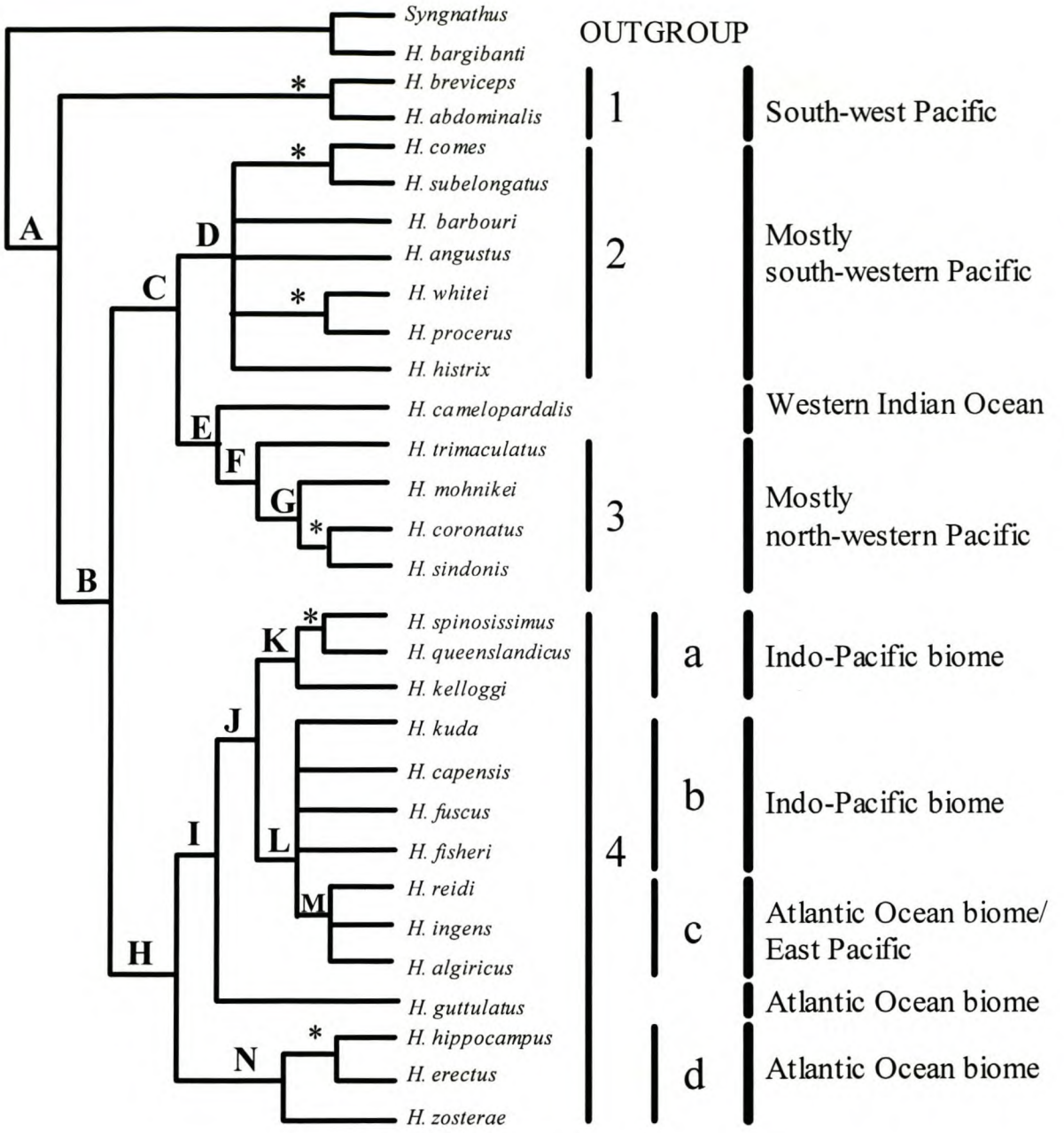

Fig. 1.4 Synthesis of phylogenetic information of MP and BI trees constructed by combining nodes from combined and individual partitions as shown in Table 1.10. Nodes of interest are labeled with letters A-N and correspond to those given in Table 1.10. Additional terminal nodes that were recovered in all analyses are marked with asterisks. The monophyly of clades 1-4 is supported by high bootstrap values and/or posterior probability values of the nodes defining them (Table 1.10). Grouping of seahorses in clade 4 into four subclades was based primarily on their associations with specific geographic regions.

As mentioned above, *H. bargibanti* was always placed basal in the phylogeny, and the south-west Pacific seahorses *H. breviceps* and *H. abdominalis* comprised the next basal clade (supported by node B). The data also suggest that the remaining species can be placed into three monophyletic lineages. The first of these (clade 2) mainly comprises species from the south-western Pacific, an exception being the more widespread *H. histrix* (Pacific basin to East Africa, Lourie *et al.* 1999). Note that all Australian seahorses have been considered Pacific Ocean species. Faunas even in Western Australia are often genetically more closely associated with those of the Pacific rather than the Indian Ocean (Berquist & Kelly-Borges 1995; McMillan & Palumbi 1995; Williams & Benzie 1998) because upwelling west of the Australian coast constitutes a biogeographic barrier (Fleminger 1986; Wells *et al.* 1994). The second assemblage (clade 3) comprises species mostly confined to the north-western Pacific, only *H. trimaculatus* is more widespread (west Pacific to eastern Indian Ocean; Lourie *et al.* 1999). The third and last assemblage (clade 4) includes species that together are distributed circumglobally. The species comprising subclades 4a and b are found exclusively in the Indo-Pacific, and those in subclades 4c and d are restricted to the Atlantic Ocean. The monophyletic lineage defined by node L (subclades 4b and c) is interesting in that several of its species are the sole representatives of the genus *Hippocampus* in their home region, including *H. ingens* on the American west coast, *H. algiricus* in West Africa, and *H. capensis* in estuaries located on the south coast of South Africa. The phylogenetic placement of two species remains uncertain. Firstly, the East African species *H. camelopardalis* was sometimes associated with the species of clade 3 (node E), but this association was well-supported in a single case only. Secondly, the Mediterranean species *H. guttulatus* was grouped with the species of subclades 4a-c (node I) in several analyses, but this association was never strongly supported. On one of the MP trees, this species was instead grouped with subclade 4d (combined mtDNA, 70% bootstrap support).

The monophyly of most species represented by more than one sample could not be challenged: although some specimens were closely associated with seahorses other than their conspecifics (Fig. 1.3b), such clades tended to be weakly supported (e.g. the three specimens of *H. comes* among clade 2 seahorses and the three specimens of *H. trimaculatus* among clade 3 seahorses). An exception was the West Atlantic species *H. reidi*: the specimen from the Gulf of Mexico was closely associated with *H. ingens* from the East Pacific, whereas the Brazilian specimen had a sister taxon relationship with the West African *H. algiricus*.

1.3.3 Molecular clock calibration

Three calibration points based on well-documented vicariance events were used. The emergence of the Isthmus of Panama, which isolated the tropical western Atlantic and eastern Pacific oceans (Jordan 1908; Gorman & Kim 1977; Keigwin 1978; Coates *et al.* 1992), was used as the first event to infer dates of divergence among taxa. A single seahorse species (*Hippocampus ingens*) is found along the western coastline of the Americas, and it is most closely related to the western Atlantic species *H. reidi*, suggesting that these two species diverged after the closure of the Central American Seaway. Ocean circulation was reorganised from 4.6 mya (Haug & Tiedemann 1998), closure was almost complete at 3.6 mya (Coates 1992), and complete closure was achieved at 2.7 mya (Marshall 1996). As seahorses live in shallow water and are likely to have been able to cross from one ocean basin to the next until complete closure occurred, a lower time constraint of 2.7 mya was used. Reorganisation of ocean circulation is unlikely to have affected the dispersal of coastal seahorse species, as opposed to pelagic teleosts. However, it must be kept in mind that this may overestimate the rate of divergence, as many marine organisms have been shown to have diverged prior to this date (Marko 2002). Hence, an upper time constraint of 3.6 mya was applied. An alternative colonisation scenario of the East Pacific by long-distance dispersal of West Atlantic seahorses via Cape Horne (which may have taken place after the closure of the Central American Seaway) seems unlikely: to date, there is no evidence in the literature that any tropical or sub-tropical genera of marine organisms have dispersed via this route. Moreover, *H. reidi* has not been recorded south of Rio de Janeiro, and *H. ingens* is not present south of Peru (Lourie *et al.* 1999). However, it must be noted that *H. erectus* was observed in the South Atlantic coast in Patagonia (San Matias Gulf, 42 degrees south) by Ortí and Gostonyi (pers. comm.).

The closure of the Tethyan Seaway between Africa and Eurasia was used as a second calibration point, as this major tectonic events seems to have resulted in the first phylogenetic split encountered in clade 4. On the basis of the distributions of the extant species associated with clade 4, we propose that the species of subclade 4d *(H. hippocampus*, *H. erectus*, *H. zosterae* and *H. guttulatus*) represent descendents of a western Tethyan/Atlantic/Caribbean lineage, whereas the eastern Tethyan/Indo-Pacific lineage comprises subclades 4a (*H. kelloggi*, *H. spinosissimus* and *H. queenslandicus*), 4b (*H. kuda* and others) and 4c (*H. ingens*, *H. reidi* and *H. algiricus*). Note that although the seahorses in subclade 4c are Atlantic Ocean species, their derived position on all phylogenetic trees suggests that they colonised this region more recently.

It seems that there were several closures and re-openings throughout the Oligocene and mid-Miocene (Rosen 1988), possibly because of climate variations resulting in sea level changes throughout this period (see Zachos *et al.* 2001). However, the link with the Indian Ocean was permanently severed approximately 14 mya (Hsü & Bernoulli 1978; Vrielynk *et al.* 1997), and this date was thus used as a lower bound for the divergence between subclades 4a-c and subclade 4d. Axelrod & Raven (1978) suggested that the seaway closed as early as 17-18 mya. Although there seems to be less support for this older date, we used 17 mya as the upper bound to account for the possibility that faunal divergence took place prior to the final closure of the seaway. Lastly, it cannot be ruled out that these bounds may result in an overestimate: Sonnenfeld (1985) suggested that intermittent contact between the Mediterranean and the Indian Ocean continued well in to the Messinian (6.7 mya). Migration of seahorses around South Africa represents an alternative scenario which cannot be excluded based on the available data, and which could have taken place both before and after the closure of the Tethyan Seaway. However, we feel that this solution is less parsimonious, because environmental conditions throughout the Tethys Sea are likely to have been more suitable for tropical and subtropical seahorses than they were in southern Africa, where sea surface temperatures are likely to have undergone extensive fluctuations from the mid-Miocene until the present (Lindesay 1998). Wilson & Rosen (1998) showed that zooxanthellae-containing corals were abundant in what is now the Mediterranean throughout the Miocene. As such reef-building corals grow in tropical and subtropical regions, and are sensitive even to slight fluctuations in temperature (Glynn 1993; Brown 1997; D'Croz & Robertson 1997; Aronson *et al.* 2000), environmental conditions must have remained fairly stable within the Tethyan Seaway throughout this epoch.

A third calibration point was based on Lourie (in press), who found a deep genetic split among cytochrome *b* lineages of Indo-West Pacific seahorses identified as *H. kuda*. The one lineage was associated with the Indian Ocean and eastern Indonesia, and the other with the Pacific Ocean, suggesting that this split is the result of a vicariance event that impacted on a previously continuously distributed lineage (see also chapter 2). The divergence event was estimated to have taken place approximately 2.28 mya, which was used as a lower bound to constrain the search for divergence ages.

We decided not to base molecular clock calibrations on fossil data, because of the highly fragmented fossil record of seahorses. To our knowledge, the oldest (and possibly only) fossilised seahorses have been found in Italy (Sorbini 1988) and were identified as *H. ramulosus* (a synonym for the extant species *H. guttulatus*). The fact that the deposits containing these specimens have been dated as being from the upper Miocene, supports the

presence of these seahorses in the western Tethys (today's Mediterranean) during the time when the Tethyan Seaway closed. All dates estimated in this chapter are at best speculative, as tectonic events are assumed to be reflected in phylogenetic patterns observed among extant seahorse lineages, and because of the uncertain dating of these events.

The tree topology used to estimate the age of divergence events in MULTIDIVTIME is shown in Fig. 1.5. It is based on Fig. 1.4 and includes species from all major lineages for which nearly complete sequence data were available. Note that the monophyly of clades 2 and 3 was accepted, even though it was strongly supported by posterior probabilities from Bayesian analyses only (Table 1.10). Also, the monophyly of *H. guttulatus* with the seahorses associated with subclade 4d has been accepted, as this hypothesis is confirmed by the mitochondrial phylogeny in chapter 2.

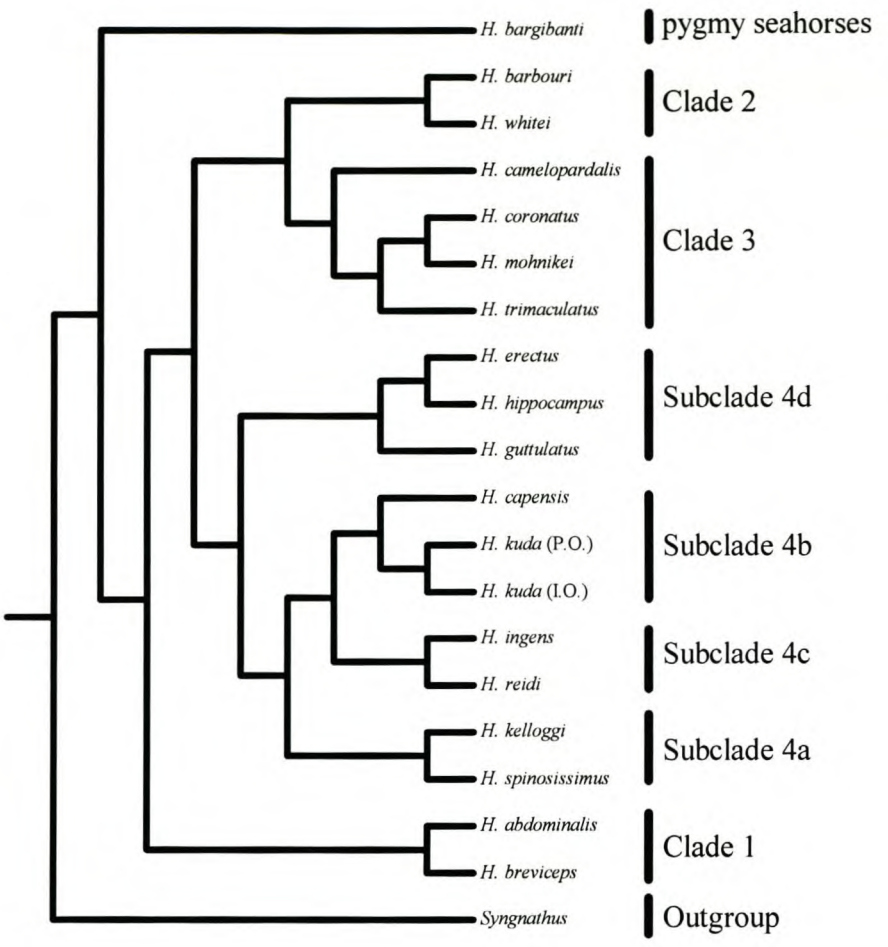

Fig. 1.5 The topology used to estimate divergence dates among lineages of data-base A in ESTBRANCHES and MULTIDIVTIME. Data-base B did not include *H. bargibanti*. Abbreviations in brackets behind *H. kuda* refer to the origin of the samples: P.O. = Pacific Ocean; I.O. = Indian Ocean.

As no Aldolase and cytochrome *b* data were available for the pygmy seahorse, *Hippocampus bargibanti*, the divergence date of this lineage with all other seahorses was estimated using RP1 and 16S rRNA sequences only (data-base A). Divergence events in the ingroup (i.e. clades 1-4) were estimated using both this data-base and a data-base comprising all four molecular markers (data-base B). Divergence estimates and their standard deviations, as well as 95% credibility intervals calculated using the two data-sets, are listed in Table 1.11. Initially, a prior ingroup node age (*rrtm*) of 30 million years (± 15 million years, S.D.) was specified. Prior rates (*rtrate*) at the ingroup root were 0.00737 ± 0.0037 (S.D.) in the case of data-base A and 0.0094 ± 0.0047 (S.D.) in the case of data-base B. Divergence estimates of

additional runs using the same specifications differed by no more than 0.99 million years (mean difference: 0.49 ± 0.36 [S.D.]), which suggests that the programme was run sufficiently long for the Monte carlo chains to converge. Using different prior specifications (*rrtm* = 20 million years ± 10 million years [S.D.] with a corresponding *rtrate* of 0.0111 ± 0.0055 [data-base A] and 0.0141 ± 0.0071 [data-base B] and *rrtm* = 40 million years ± 20 million years [S.D] with a corresponding *rtrate* of 0.0055 ± 0.0028 [data-base A] and 0.0071 ± 0.0035 [data-base B]) resulted in estimates that were consistently lower (*rrtm* = 20 million years: up 4.5 million years) or higher (*rrtm* = 40 million years: up to 1.5 million years), suggesting that the specifications of these priors influences estimates of divergence dates to a limited extent. The posterior estimates of the age of the root node using different *rttm* values converged onto a value of close to 26 mya (*rrtm* = 40 million years: 27.7 million years; *rrtm* = 20 million years: 23.7 million years; mean *rrtm* = 25.7), which was approximately obtained when setting *rttm* to 30 million years (Table 1.11). This suggests that specifying the age of the root node as 26 million years should result in the most accurate estimates. However, Table 1.11 shows that divergence age estimates calculated using *rttm* = 26 million years differ only slightly from those calculated with *rttm* = 30 million years.

The split between pygmy seahorses and true seahorses was estimed to have taken place approximately 26 million years ago, i.e. during the late Oligocene (Chattian: 29.3 – 23.3 mya; time scales used in this paragraph are based on Harland *et al.* 1989). Although it could be argued that this estimate is spurious because it was calculated using a data-base consisting of only two partitions, the fact that the other six divergence estimates were similar to those calculated using a data-base that comprised all four markers, indicates that the RP1 and 16S rRNA sequences may have been sufficient to obtain reasonable divergence estimates. All remaining divergence events were estimated to have taken place during the Miocene (23.3 – 6.7 mya), the oldest one being the divergence of the basal Australian lineage (clade 1) from clades 2-4 (approximately 22-23 mya). Clade 4 diverged from clades 2 and 3 approximately 19 mya and the first split within clade 4 (between subclades 4a-c and subclade 4d), which was assumed to be the result of a vicariance event caused by the closure of the Tethayn Seaway (estimated here as approximately 15.45 mya, i.e. preceding 14 mya, the date of final closure used as a lower bound), approximately coincided with the split between the primarily West Pacific clades 2 and 3. The split between subclades 4a and 4b-c was estimated to have occurred 12 mya, and the invasion of the Atlantic Ocean by Indo-Pacific seahorses of clade 4 (the split between subclades 4b and 4c) took place approximately 8.5 mya.

Table 1.11 Time estimates of major splitting events among *Hippocampus* lineages showing the mean estimate ± the mean of the corresponding standard deviations, as well as confidence intervals (in brackets) from three MULTIDIVTIME runs. Two different data-bases and corresponding tree topologies were used (the tree topology of data-base A is shown in fig. 1.5): A) RP1 and 16S rRNA sequences of representatives of the true seahorses (*H. abdominalis*, *H. breviceps*, *H. barbouri*, *H. whitei*, *H. camelopardalis*, *H. trimaculatus*, *H. mohnikei*, *H. coronatus*, *H. erectus*, *H. hippocampus*, *H. guttulatus*, *H. capensis*, *H. kuda* [Pacific Ocean lineage], *H. kuda* [Indian Ocean lineage], *H. ingens*, *H. reidi*, *H. kelloggi* and *H. spinosissimus*) and the pygmy seahorse (*H. bargibanti*); B) RP1, 16S rRNA, Aldolase and cytochrome *b* sequences of 'true' seahorses. In both cases, *Syngnathus* sp. was used as the outgroup taxon. As the prior specification of the age of the root node (*rttm*) was found to affect posteriors, the posterior estimate for the age of this split calculated using the prior *rttm* = 30 mya was used as the prior in a subsequent set of runs (*rttm* = 26 mya).

Split		Data-base			
		A		B	
		specified age of root node (in million years)			
First lineage	Second lineage	30	26	30	26
Pygmy seahorses	'True' seahorses	26.40 ± 7.24 (17.16 – 44.64)	25.60 ± 6.57 (16.95 – 41.76)	-	-
Clade 1	Clades 2-4	22.72 ± 5.71 (15.75 – 37.64)	22.38 ± 5.43 (15.68 – 36.40)	23.05 ± 5..75 (15.87 – 38.01)	22.28 ± 5.11 (15.67 – 32.30)
Clades 2 and 3	Clade 4	19.12 ± 4.05 (14.70 – 30.04)	18.71 ± 3.49 (14.18 – 27.91)	19.23 ± 3.82 (14.74 – 28.89)	18.85 ± 3.71 (14.66 – 28.87)
Clade 2	Clade 3	15.28 ± 4.59 (7.15 – 25.79)	14.97 ± 4.15 (7.005 – 24.23)	15..37 ± 4.46 (6.80 – 25.25)	15.14 ± 4.26 (7.25 – 24.42)
Subclades 4a-c	Subclade 4d	15.44 ± 0.87 (14.07 – 16.92)	15.44 ± 0.86 (14.07 – 16.92)	15.46 ± 0.86 (14.07 – 16.92)	15.42 ± 0.86 (14.06 – 16.91)
Subclade 4a	Subclade 4b and c	12.07 ± 2.66 (6.02 – 16.09)	12.15 ± 2.63 (5.98 – 16.13)	12.02 ± 2.71 (6.12 – 16.08)	12.01 ± 2.66 (5.96 – 16.05)
Subclade 4b	Subclade 4c	8.67 ± 2.95 (3.68 – 14.39)	8.68 ± 2.95 (3.67 – 14.32)	8.61 ± 2.92 (3.69 – 14.20)	8.66 ± 2.89 (3.74 – 14.27)

1.4 Discussion

1.4.1 Comparison of nuclear and mitochondrial markers

Even though nuclear genes were difficult to sequence in individual cases, they generally had a greater potential to recover the deeper nodes in the phylogeny. This is particularly true for the larger RP1 intron. The missing characters in some of the RP1 sequences seem to have had little effect on phylogenetic placement of the particular species. The well-supported sister taxon relationship of *Hippocampus sindonis* with *H. coronatus*, for example, was confirmed by the Aldolase and 16S rRNA phylogenies. In addition, the sister taxon relationship of *H. comes* and *H. subelongatus*, which both contained a high proportion of missing characters in their RP1 sequences (Appendix I, Table A2), was confirmed by both the cytochrome *b* and the 16S rRNA phylogenies. A particularly striking difference between nuclear and mitochondrial sequences was the genetic distance between ingroup and outgroup species. The comparatively small genetic distances among mitochondrial sequences of ingroup and outgroup species shows that homoplasy becomes considerably greater when comparing more distantly related species. Ortí & Meyer (1997) suggested a saturation plateau of about 20% pair-wise differentiation for ribosomal DNA. In case of the nuclear fragments used in this study, this level of differentiation is surpassed when comparing sequences of true seahorses, pygmy seahorses and pipefishes. However, nuclear data tend to be characterised by a low number of homoplasies and apparent lack of saturation even among fairly distantly related taxa (e.g. Groth & Barrowclough 1999; Matthee *et al.* 2001; Creer *et al.* 2003), a characteristic that was also observed in the present study. The notion that the pipefish performed satisfactorily as an outgroup in the case of nuclear genes was confirmed by the fact that MP analyses of nuclear data sets using clade 1 as alternative outgroup resulted in trees with identical topologies, and bootstrap support for individual nodes on these trees differed only slightly. In contrast, in the case of mitochondrial data, using clade 1 as outgroup resulted in different MP tree topologies (significant in one case), and bootstrap support for some nodes increased considerably.

1.4.2 Indo-Pacific origin of seahorses

The large genetic distance of the pygmy seahorse, *H. bargibanti*, to all other seahorses based on RP1 and 16S rRNA sequences suggests an ancient divergence of this group from the main clade of seahorses. *H. bargibanti* is widely distributed throughout the western Pacific, but the fact that this species is highly specialised to parasitising certain species of *Muricella* gorgonians (Kuiter 2000) suggests that it is unlikely to disperse readily beyond the region where the host species occur. Among the species associated with the main clade of seahorses, the most basal positions are occupied by *H. breviceps* and *H. abdominalis*. Both species are associated with the Australian continent, suggesting that this may be the region of origin for seahorses. An Australian or south-west Pacific origin of seahorses is also supported by the distributions of the three possible sister genera of the genus *Hippocampus*. Pygmy pipehorses of the genus *Amphelikturus* are restricted to the Atlantic biome, the genus *Acentronura* is widely distributed throughout the Indo-Pacific, and all known specimens of *Idiotropiscis* have been found in Australian waters. Among these three genera, the species of the genus *Idiotropiscis*, and particularly a recently discovered species from southern New South Wales, are most seahorse-like in appearance (Kuiter 2000).

1.4.3 Biogeography and evolutionary history

The divergence time between pygmy seahorses and 'true' seahorses was estimated to be approximately 26 mya (late Oligocene). The late Oligocene was a period characterized by global warming (Zachos *et al.* 2000). This may have resulted in an increased abundance and diversity of coral species in the Indo-West Pacific. A rich fauna of zooxanthellae-containing corals became established in parts of what is now southern Southeast Asia and New Guinea, during the late Oligocene and early Miocene (Hall 1996; Wilson & Rosen 1998). It is not certain whether the gorgonians parasitised by *H. bargibanti* have symbiotic zooxanthellae, but this is quite possible, as most gorgonians do (Sutton & Hoegh-Guldberg 1990). Other pygmy seahorses also seem to be exclusively associated with gorgonians (*H. minotaur*: Gomon 1997; *H. denise*: Lourie & Randall 2003), and it is thus appears that pygmy seahorses diverged from other seahores because of habitat specialization. The gorgonians may at first have served as holdfasts, but eventually, they also became a source of food.

The divergence time of 22 mya estimated for the split between clade 1 and all other 'true' seahorses coincides with the collision of the Australian continental margin with eastern

Indonesia (Ali & Hall 1995; Malaihollo & Hall 1996). Two possible reasons for the divergence can be proposed. Firstly, seahorses originated in southeast Asia and then colonized the Australian coastline. In addition to a certain extent of isolation by distance (which may have resulted in allopatric speciation), the split between the two early lineages may have been the result of the adaptation of the one lineage to cooler water temperatures. The central part of what is today the Great Barrier Reef (located off north-eastern Australia) did not experience a tropical climate until 15 to 10 mya, and the southern part only became tropical in the last few million years (Davies *et al.* 1987). This hypothesis is supported by the fact that both extant species in clade 1 are found on the southwest-coast of Australia, suggesting that the distribution of these cooler water seahorses may have shifted southwards as water temperatures in northern Australia increased. *H. breviceps* additionally occurs on the cooler Australian south-east coast, and *H. abdominalis* (or *H. bleekeri*) is found in New Zealand (Lourie et al. 1999; Kuiter 2000). Secondly, seahorses may have originated in Australia and subsequently colonized Southeast Asia. In this case, they would have adapted to tropical conditions. Although this scenario seems plausible, as the most seahorse-like pygmy pipehorses are associated with the Australian continent, the fact that pygmy seahorses (the sister taxon of 'true' seahorses) are likely to have diverged in a region where gorgonians containing symbiotic zooxanthellae occurred at the beginning of the Neogene, suggests that the first hypothesis is better supported. On the other hand, the pygmy seahorse *H. minotaur* occurs in cooler water off New South Wales, and the gorgonian it parasitises may not contain symbiotic zooxanthellae. Further research on pygmy seahorses and pygmy pipehorses is thus required to ascertain where true seahorses are more likely to have evolved.

Although several nodes of the molecular phylogeny of the genus *Hippocampus* presented were not supported by high bootstrap and posterior probabilities values, several novel insights can be proposed. It seems that subsequent to the origin of the genus *Hippocampus* in the Indo-Pacific biome, the main clade of seahorses split into three major lineages. Two of these remained in the Indo-Pacific (clades 2 and 3) and can be divided into a mostly south-western and a mostly north-western Pacific group. As bootstrap support for node C was low, it cannot be concluded with certainty whether these two clades are more closely associated with each other than either of them is with clade 4. However, as nodal support was fairly high in the case of a Bayesian analysis using data from up to four partitions, this is nevertheless plausible. Wilcox *et al.* (2003) found that posterior probabilites represented better phylogenetic accuracy than did bootstrap values under the conditions of their study. Cummings *et al.* (2003) suggested that posterior probabilities differ significantly from bootstap values in cases when long branches

are separated by a very short internode, a pattern that was also evident in the phylogeny presented in Wilcox *et al.* (2003). The divergence time of slightly above 15 mya estimated for the divergence of these clades is fairly close to the date estimated for the closure of the Indonesian Seaway (15.5 mya, Tsuchi 1997), and it is thus possible that this tectonic event is responsible for the split of these two lineages. Tsuchi (1997) further suggested that the closure of the Indonesian Seaway resulted in an intensification of the warm Kuroshio current, which transports water from Indonesia towards Japan. This in turn led to the establishment of the tropical 'Kadonosawa fauna' on most coasts of the Japanese Islands. It is thus possible that the seahorses associated with clade 3 (i.e. which includes four primarily north-western Pacific species, two of them endemic to Japan) became established in the region during that time. There was some support for an association of the East African species *H. camelopardalis* with clade 3, but the results remain inconclusive. The basal placement of this species (node E) suggests that it is part of a lineage that became geographically separated from seahorses associated with the West Pacific early during the evolutionary history of the genus.

In contrast to the first two assemblages, most of whose species remained in the West Pacific, the evolutionary history of the third major assemblage (clade 4) was characterised by dispersal events on a global scale. We are not aware of a major climatic or tectonic event that may have caused the divergence of clade 4 from clades 2 and 3, but it is likely that this group became genetically distinct because it established itself in the Atlantic biome, a scenario that is supported by the fact that the basal position of subclade 4d within this group was well supported on most trees and the fact that the next derived species, *H. guttulatus*, is also associated with the Atlantic biome. This notion is also supported by the fact that although there was no major barrier to gene flow during the Miocene (as the Tethyan and Central American Seaways were both open), differentiation of coral faunas into the two principal modern faunas (Caribbean and Indo-Pacific) is considered to have taken place during this period (Rosen 1988). The Caribbean has been a centre of origin for many of the species found throughout the tropical Atlantic (Briggs 1974), and it is possible that seahorses from this region then re-colonised the Indo-Pacific biome, giving rise to subclades 4a-c, a scenario depicted by Fig. 1.6a. Alternatively, the group defined by node J (subclades 4a-c) remained in the Indo-Pacific and was the source of a maximum of three colonisation events of the Atlantic biome (Fig. 1.6b). Additional scenarios set between these two extremes are equally plausible, and cannot be resolved in the absence of a more robust phylogeny.

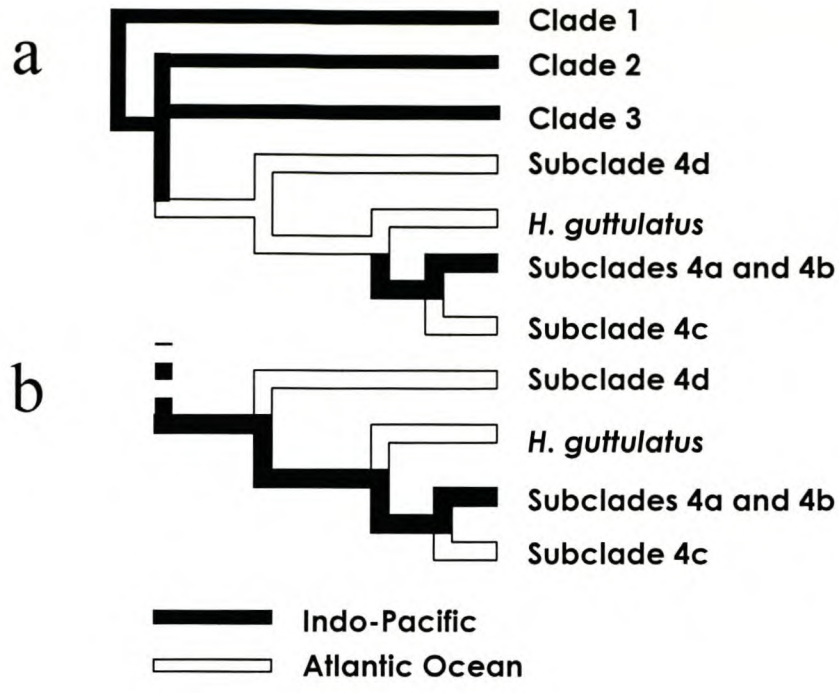

Fig. 1.6 Two alternative hypothesis regarding the history of colonisation of the Atlantic biome by members of the genus *Hippocampus*. In both cases, it is assumed that *H. guttulatus* is more closely associated with seahorses of subclades 4a and 4b than it is with those of subclade 4d.

Some additional deductions about the evolutionary history of seahorses can be made using estimated dates of the closures of the Tethyan and Central American Seaways (alternative scenarios, such as long distance colonisation of the Atlantic Ocean via South Africa, are also possibly, but were not explored further). The complications associated with the phylogenetic placement of *H. guttulatus* make a reconstruction of the early history of seahorses in the Atlantic Ocean difficult, but two hypotheses are likely, depending on whether this species is more closely associated with subclades 4a and 4b or with subclade 4d (note: this section option is well supported by the mitochondrial data-base used in chapter 2). In the first scenario, the split between subclade 4d and the remainder of the group preceded the closure of the Tethyan Seaway. East Atlantic seahorses may already have been distinct from West Atlantic seahorses, and their distribution may have extended into the Indian Ocean and beyond. This differentiation into two major groups is most likely to be the result of isolation by distance,

which resulted in speciation. A geographical division of the tropical Atlantic fauna into an east Atlantic and a tropical American province was evident as early as the Miocene (Le Lœuff & von Cosel 1998). The expanse of the Atlantic Ocean is a minor obstacle to gene flow between the western and eastern Atlantic faunas (Briggs 1974), and many tropical marine species are represented on both coasts (Lessios *et al.* 1999). However, the dispersal abilities of a species depend primarily on its early life-history stages: species with pelagic eggs and long-lived planktonic larvae have high dispersal capabilities (Palumbi 1992). Seahorses are likely to disperse poorly because of the lack of a pelagic larval stage: newborn seahorses are merely small versions of the adults and lack the egg yolk reserves required for long-distance dispersal. If, alternatively, *H. guttulatus* is more closely associated with subclade 4d than it is with subclade 4a and b, differentiation of Atlantic seahorses into a European and an American lineage is equally likely to have taken place after the closure of the Tethyan Seaway. The other European species, *H. hippocampus*, is presently distributed throughout the range occupied by *H. guttulatus* (Lourie *et al.* 1999), but its close association with the two American species in subclade 4d, and the small genetic distances among them suggests that this species may have recently diverged from an American ancestor. Such a colonisation event may have taken place due to an intensification of the Gulf Stream's current velocity as a result of the gradual rising of the Isthmus of Panama, which culminated 3.8 million years ago (Kaneps 1979; Keller & Barron 1983). The estimated divergence time of *H. erectus* and *H. hippocampus* of 5.2 mya calculated using the molecular clock precedes this date.

As mentioned previously, although subclade 4c comprises species endemic to the Atlantic Ocean, their derived position on all phylogenetic trees suggests that they colonised this region more recently. The split of this lineage from subclade 4b was estimated to have occurred approximately 9 mya. This second invasion of the Atlantic Ocean may have occurred either in a westward direction via Southern Africa, or in an eastward direction via the Pacific basin and the Central American Seaway (see also chapter 2).

1.4.5 Geminate species pairs in the Atlantic Ocean

The distribution of species (or species complexes) on both sides of the tropical Atlantic Ocean has been attributed to either vicariance (spreading of the Atlantic basin 65-20 mya and rise of the Panamanian isthmus; Rosen 1975) or recent dispersal events with the Caribbean as a centre of origin (Briggs 1974). Both of the Atlantic seahorse subclades have representatives on

both sides of the Atlantic Ocean: in subclade 4c, *H. reidi* and *H. ingens* occur in the west (West Atlantic and East Pacific), and *H. algiricus* in the east (West Africa); in subclade 4d, *H. erectus* and *H. zosterae* are West Atlantic species and *H. hippocampus* is an East Atlantic species. In both cases, genetic distances between members of these pairs are minimal (Fig. 1.2). This is particularly striking in case of the Aldolase sequences, which are identical for all three members of each pair (Fig. 1.2a). Phylogenies incorporating RP1 sequences did not recover the monophyly of *H. ingens* and *H. reidi*, but this may be due to incomplete lineage sorting. Resolution at the tips is higher in case of the mitochondrial genes than in the nuclear genes used in this study, but the two mitochondrial gene fragments are in conflict in case of the geminate species of subclade 4c: the 16S rRNA phylogeny indicates that *H. reidi* and *H. ingens* are more closely associated with each other than they are with *H. algiricus*, whereas the cytochrome *b* phylogeny indicates a closer association of *H. reidi* with *H. algiricus* than with *H. ingens*. Whichever associations are correct, the fact that East Atlantic and West Atlantic/East Pacific lineages are likely to have diverged after the closure of the Tethyan Seaway suggests that the dispersal hypothesis seems more appropriate to explain present-day distribution patterns of the geminate seahorse species. The same result was found in the genus *Ophiblennius* (Muss *et al.* 2001). However, the gene fragments used in this study do not provide sufficient resolution to identify whether the Caribbean was a centre of origin in both cases.

CHAPTER 2: THE EVOLUTIONARY HISTORY OF 'KUDAOID SEAHORSES'

2.1 Introduction

Phylogenetic analyses in chapter 1 revealed that the Knysna seahorse is part of a monophyletic lineage of non-spiny seahorses defined by node L (i.e. subclades 4b and 4c) on Fig. 1.4, which also includes the following taxa: *H. kuda*, *H. borboniensis*, *H. fuscus*, *H. algiricus*, *H. ingens*, *H. reidi* and *H. fisheri* (see also Casey 1999; Jones *et al.* 2003). Within this assemblage, it is suggested that the Knysna seahorse is most closely related to *H. kuda*, a species complex widely distributed throughout the Indo-Pacific (Casey 1999). As it is disputed which regional populations are associated with the *H. kuda* complex, and which are merely closely related species (see next paragraph), we refer to all of the evolutionary lineages defined by node L as 'kudaoid seahorses'.

The phylogenetic placement of the species associated with (or closely related to) the *H. kuda* complex is uncertain because of confusion regarding their taxonomy and distribution. Lourie *et al.* (1999) consider the range of *H. kuda* to encompass the Indian subcontinent, Thailand, Singapore, Vietnam, Hong Kong, Taiwan, the Philippines, Malaysia, Indonesia, Japan, as well as possibly northern Australia and some Pacific islands. Kuiter (2000), on the other hand, restricts the distribution of the species to the Maldives, Sri Lanka, Andaman Sea, Singapore and western Indonesia to Ryukyus, Japan. In addition, several seahorses regarded as *H. kuda* by Lourie *et al.* (1999) are given species status by Kuiter (2000), including *H. arnei* (southern China Seas and Philippines), *H. moluccensis* (Ambon and eastern Sulawesi), *H. polytaenia* (Flores Seas), and *H. taeniopterus* (Moluccen Seas to Sulawesi and Bali). Neither mentions the presence of *H. kuda* on the east coast of Africa, whereas Dawson (1986) states that the species occurs in Mozambique and Kenya. Lourie *et al.* (1999) report the presence of the closely related species *H. borboniensis* and possibly also *H. fuscus* in this region, whereas Kuiter (2000) considers only *H. borboniensis* a Western Indian Ocean species, and restricts the distribution of *H. fuscus* to the Red Sea and Arabian seas. The phylogenetic distinctness of *H. capensis* from some other populations of kudaoid seahorses has been questioned, on account of

the occurrence of morphologically similar marine seahorses. The only distinguishing morphological characteristic of *H. capensis* is a reduced or absent coronet, but this feature seems convergent because it has also been observed in marine seahorses from Tanzania (McPherson, pers. comm.[2]) and in seahorses from the Red Sea identified as *H. fuscus* (Gabr, pers. comm.[3]). Although *H. capensis* is considered the world's only seahorse confined to estuaries, other species, including *H. kuda*, *H. kampylotrachelos* and *H. dahli*, have also been reported to sometimes penetrate estuaries (Kuiter 2000; Sreepada, pers. comm.[4]). The uncertainties regarding the morphological and physiological distinctiveness of *H. capensis* render the validity of the high conservation status of this species questionable. At present, it cannot be ruled out that the Knysna seahorse may not be a distinct species, as conspecifics may actually occur farther north on the African east coast. If this holds, the South African populations of *H. capensis* may merely constitute a portion of the diversity present in the Western Indian Ocean.

In an attempt to clarify the taxonomic status of *H. capensis* and to investigate the phylogenetic uniqueness of this estuarine taxon among kudaoid seahorses, mtDNA markers were employed. In addition to the mitochondrial 16S rRNA data-base used previously (chapter 1), the mitochondrial control region (CR) was sequenced. The CR is involved in the control of mtDNA replication and RNA transcription and is the only portion of the mitochondrial gene that is not transcribed (Hoelzel *et al.* 1991). It is suggested that because of reduced functional constraints, some portions of the CR evolve much faster than the rest of the mitochondrial genome (Brown *et al.* 1986). As the CR is likely to provide good resolution among closely related organisms, it is the preferred mitochondrial marker for studying relationships among closely related species (e.g. Meyer *et al.* 1990; Arnason & Rand 1992; Lee *et al.* 1995) or at the population level (e.g. Nielsen *et al.* 1997; Houlden *et al.* 1999; Nyakaana & Arctander 1999; Flagstad *et al.* 2000). Although the CR has an overall high evolutionary rate, it also contains conserved regions that are presumably functionally constrained, notably a central conserved region believed to be critical for mitochondrial metabolism, although its exact function is not understood (Lee 1995). Despite rapid evolutionary rates of 'left' and 'right' domains flanking the central conserved region, Avise *et al.* (1987) considered the CR to 'bridge the gap' between phylogenetics and population genetics. In this study, a section of the rapidly evolving right domain of the mitochondrial CR was sequenced in as many specimens of kudaoid seahorses as could be obtained. Phylogenetic relationships among different monophyletic clades identified

[2] Jana McPherson, Department of Zoology, University of Oxford, Oxford, UK
[3] Howaida Gabr, Suez Canal University, Ismaila, Egypt

with this data-set were then investigated using a limited number of representative specimens from each clade and analysing three mitochondrial gene fragments: firstly, the complete mitochondrial CR was sequenced; secondly, the nearly complete 16S rRNA sequences of kudaoid seahorses generated in chapter 1 were used, and some additional specimens were added to this existing data-base; thirdly, the cytochrome *b* sequences available from GenBank (see chapter 1) were included in some analyses.

2.2 Materials and methods

2.2.1 Sample acquisition, DNA extraction, amplification and sequencing

In the case of 16S rRNA, two sequences of *H. kuda* from India and one sequence of *H. kuda* from Fiji were added to the existing data-base. In case of the right domain of the CR, 58 specimens from eight species of kudaoid seahorses and six outgroup taxa were sequenced (Table 2.1). DNA extraction procedures were performed using the methods described previously (Chapter 1). As no syngnathid CR primers were available, a set of primers was designed using primer walking and the cytochrome *b* sequences of *Hippocampus kuda* (Casey 1999, GenBank accession numbers AF192679-AF192686) and 12S rRNA sequences of four seahorse species from Japan (Mukai *et al.* 2001, GenBank accession numbers AB032027-AB032030). A section of the *Hippocampus capensis* mtDNA encompassing the cytochrome *b* gene (partial), $tRNA^{Thr}$, $tRNA^{Pro}$, complete CR, $tRNA^{Phe}$ and 12S rRNA (partial) was amplified. The forward primer (HCAL1: 5'-C T G T A T C T G G T T G A A T G G A G-3') anneals to the 5' end (heavy strand) of the cytochrome *b* gene (heavy strand) and the reverse primer (HCAH1: 5'-C C A C T C T T T A C G C C G A C T T Y T-3') to the 5' end of the 12S rRNA gene (light strand). This amplicon was approximately 1444 bp long. A consensus sequence was derived from five individuals by trimming the edges of the complete amplicons, which resulted in a sequence 1403 bp in length (deposited in GenBank, accession number AY149663,

[4] Anantha Sreepada, National Institute of Oceanography, Goa, India

see also Appendix I, Table A11). In *H. capensis*, the consensus sequence comprises 36 bp of cytochrome *b*, 85 bp of $tRNA^{Thr}$, 70 bp of $tRNA^{Pro}$, 887-888 bp of CR, 66 bp of $tRNA^{Phe}$ and 259 bp of 12S rRNA. As this large fragment did not amplify routinely, particularly when DNA was of poor quality, a second set of primers was designed to specifically amplify 533 nucleotides of CR right domain (forward primer: HCAL2: 5'-C A C A C T T T C A T C G A C G C T T-3'; reverse primer: HCAH2: 5'-T C T T C A G T G T T A T G C T T T A-3'). We are confident that the gene fragment sequenced is mitochondrial rather than a nuclear pseudogene because no heterozygous sequences were identified and no multiple bands were amplified.

Table 2.1 Specimens used in this study. Capital letters in species code names represent specimens' country of origin. Complete CR sequences and 16S rRNA sequences were obtained from selected specimens of some species, which are marked with an uppercase 1 or 2, respectively (see text for detail). As no complete specimens were available (thus no morphological species confirmation) of the seahorses from Tamil Nadu, India, provided by Sesh Serebiah (only fin clips were provided), these samples were coded as kudaTN (for Tamil Nadu).

Species name	Code	# specimens	Collection locality	Identified by	Collector/Source
Ingroup species					
H. algiricus	algiBN[1,2]	6	Benin	P. Teske	Z. Sohou
H. borboniensis	borbIN	1	Tamil Nadu, India	S. Lourie	(Project Seahorse)
H. capensis	capeZA[1,2]	10	Knysna Estuary, South Africa	P. Teske	P. Teske
H. fisheri	fishUS	2	Hawai'i	S. Lourie	(Project Seahorse)
H. fuscus	fuscEG[1,2]	7	Red Sea, Egypt	P. Teske	H. Gabr
	fuscTZ	1	Hamadi, Tanzania	S. Lourie	J. McPherson
	fusc??	1	unknown	S. Lourie	The Sealife Centre, UK
H. ingens	ingeMX	1	Mexico, Pacific coast	S. Lourie	J. Baum
	ingePE[2]	1	Peru	S. Lourie	Bronx Zoo
H. kuda	kudaFJ[1,2]	4	Fiji	H. Hamilton	H. Hamilton
	kudaID1	2	Lombok, Indonesia	S. Lourie	B. Haq and H. Haq
	kudaIN/TN	8	Tamil Nadu, India	A. Perry	A. Perry and S. Serebiah
	kudaMZ	2	Inhaca Island, Mozambique	P. Teske	P. Teske
	kudaPH[1,2]	6	Philippines	M. Santos	M. Santos
	kudaTW[1]	1	Taiwan	S. Lourie	L.-S. Fang
	kudaZA[1,2]	3	Durban Harbour, South Africa	P. Teske	Durban Aquarium
H. reidi	reidBR[1]	1	Brazil	P. Teske	Aquarium trade
	reidMX[2]	1	Mexico	S. Lourie	J. Baum
Outgroup species					
H. hippocampus	hippPG[2]	1	Portugal	J. Curtis	J. Curtis
H. guttulatus	guttPG[2]	1	Portugal	J. Curtis	J. Curtis
H. erectus	erecUS[2]	1	United States (captive)	G. Sirpenski	G. Sirpenski
H. kelloggi	kellVN[2]	1	Danang Sea, Vietnam	Long and Huang	S. Lourie
H. spinosissimus	spinPH[2]	1	Palawan, Philippines	S. Lourie	S. Lourie
H. queenslandicus	queeAU[1,2]	1	Queensland, Australia	M.L. Horne	M.L. Horne

The CR right domain of the kudaoid seahorses and the six outgroup specimens (Table 2.1) was amplified with the polymerase chain reaction (PCR). PCR procedures were performed using the methods described previously (chapter 1). The annealing temperature during cycle sequencing was 50°C. The 5' portion (light strand) of the fragment amplified contains a long repetitive series of up to 12 thymine nucleotides. Due to difficulties in sequencing through this array, the fragment was sequenced in the reverse direction only. PCR amplification and

subsequent sequencing reactions were repeated for every tenth sample and in all cases, the duplicate sequences were identical to the original sequences.

In an attempt to improve the phylogenetic resolution obtained from CR right domain data, up to five representative samples from as many of the major clades as possible were selected based on the quality of their DNA. The complete CR was amplified and sequenced in these individuals using the HCAL1 and HCAH2 primers (PCR conditions are identical to those described above for amplifying the right domain).

CR sequences were aligned using CLUSTALX (Thompson *et al.* 1997). 16S rRNA sequences were highly conserved among kudaoid seahorses and outgroup specimens and were aligned by eye. In the case of the CR right domain, a homologous region of 382 nucleotides was obtained for all individuals after trimming ambiguous ends. Fourty-one haplotypes were obtained. In the case of the complete CR sequences, the total length of aligned sequences was 885-888 nucleotides. A central portion maximally 47 bp in length located within the central conserved region was excluded because the identity of nucleotides was difficult to interpret on some of the chromatographs, as this section was far downstream from the annealing sites of the forward and reverse primers. As the central conserved region is unlikely to contain many variable sites among the different (closely related) kudaoid seahorses, it is improbable that the exclusion of this section would have resulted in significant loss of phylogenetic signal. Additionally, a section 12 bp in length located at the 3' end of the CR was excluded because the identity of nucleotides were difficult to interpret.

2.2.2 Phylogenetic analyses

A minimum spanning network of the right domain CR sequences was constructed using the MSN algorithm employed in ARLEQUIN version 2.000 (Schneider *et al.* 2000) and drawn in LOTUS FREELANCE GRAPHICS version 9 (1991-1998). The same sequences were used to construct a parsimony phylogram in PAUP*, which was used to identify major lineages of kudaoid seahorses. Parsimony searches and Bayesian inference analyses were performed using the procedure described previously (chapter 1), and as the results from parsimony analyses and Bayesian inference were congruent, nodal support obtained using the two methods was indicated on the parsimony phylogram. Trees were rooted using outgroup species that were previously identified as being closely related to the kudaoid seahorses (subclades 4a: *H. spinosissimus*, *H. queenslandicus* and *H. kelloggi*; and 4d: *H. guttulatus, H. erectus* and *H.*

hippocampus; chapter 1). Each of the indels within the ingroup was a single base-pair in length. Hence, these gaps were treated as 5th characters in the parsimony analysis. Due to a considerable amount of terminal branch swapping, parsimony searches were constrained to a maximum of 10 000 trees. A limited number of representatives from as many of the major geographic clades identified on the parsimony phylogram as possible were used for further analyses, irrespective of the level of nodal support from parsimony analysis and Bayesian inference: if possible, the complete CR and 16S rRNA were sequenced, and corresponding cytochrome *b* sequences of kudaoid seahorses were downloaded from GenBank. A minimum spanning network of 16S rRNA sequences was computed using the programme TCS version 1.06 (Clement *et al.* 2000), which is available at

http://bioag.byu.edu/zoology/crandall_lab/programs.htm.

TCS estimates genealogies from DNA sequences by implementing the statistical parsimony method described in Templeton *et al.* (1992). Unlike the conceptually simpler method in ARLEQUIN, which can be used to investigate relationships among fairly distantly related species, TCS networks are constructed with a cut-off limit of a certain number of steps beyond which the probability of a connection is less than 95% of being true. Phylogenetic relationships among clades were investigated individually for each partition by carrying out parsimony searches and Bayesian inference as described previously. Unfortunately, the taxonomic sampling of individual partitions was not compatible: firstly, several samples amplified only in case of one of the gene fragments used because of the poor quality of their DNA, and secondly, cytochrome *b* sequences were not available for all of the clades of kudaoid seahorses identified. Hence, a combined approach was also followed in order to investigate phylogenetic relationships among all clades of kudaoid seahorses. Data were analysed using two different data-bases. The first data-base included all 16S rRNA, CR and cytochrome *b* sequences available. The second data-base excluded cytochrome *b* sequences to limit the amount of missing data. Both data-bases were analysed using parsimony and Bayesian inference following the procedure described previously, and trees were drawn as cladograms. In the case of some of the outgroup taxa (representatives of subclades 4a and 4d, chapter 1), sequences of specimens originating from different regions were combined, e.g. *H. guttulatus* from Portugal (16S rRNA and CR sequences) and *H. guttulatus* from Italy (also referred to as *H. ramulosus*; cytochrome *b* sequences). As *H. spinosissimus* and *H. queenslandicus* were found to be closely related (chapter 1) and Lourie *et al.* (1999) did not regard them as sufficiently morphologically distinct to be considered different species, sequences of individual partitions from these two taxa were also combined

2.3 Results

The segment of the mitochondrial CR sequenced in 58 specimens from 8 species of kudaoid seahorses contained 67 variable sites (three indels, 55 transitions, 7 transversions and 2 sites that contained both transitions and transversions). These polymorphic sites defined 41 haplotypes/evolutionarily discreet lineages.

The network of kudaoid seahorse haplotypes (Fig. 2.1) is characterised by a centrally positioned group of specimens identified as *H. kuda*, *H. borboniensis* or *H. fuscus*, most of which were collected in the Indian Ocean (India, southern Indonesia or East Africa). The haplotypes of the specimens of *H. capensis* (South Africa), *H. fuscus* (Red Sea) and *H. algiricus* (West Africa) are closely associated with this group, whereas *H. ingens*, *H. fisheri* and *H. kuda* from the West Pacific (Taiwan, Fiji and the Philippines) are more distantly related. Three specimens collected in the Philippines had a haplotype that was also found in the Indian Ocean (Lombok, Indonesia). A single specimen of Indian Ocean *H. kuda*-type seahorses (collected in Durban harbour, South Africa) was found to be closely associated with the estuarine seahorse *H. capensis*, despite considerable morphological differences. The parsimony phylogram constructed using the same sequences (Fig. 2.2) recovered approximately six to seven geographic clades, the representatives of which were associated with the following regions: South African estuaries/western Indian Ocean; (*H. capensis* and the single *H. kuda* specimen from Durban; Red Sea (*H. fuscus*); Indian Ocean (*H. kuda/borboniensis/fuscus*); West Pacific (*H. kuda*; this lineage can possibly be divided into a north-western and a south-western Pacific clade); East Atlantic (*H. algiricus*); Americas (*H. reidi* and *H. ingens*); and Hawaii (*H. fisheri*). Nodal support from parsimony analysis and Bayesian inference was high for three major lineages: East Atlantic kudaoid seahorses (*H. algiricus*), West Pacific kudaoid seahorses (*H. kuda* from the Philippines, Fiji and Taiwan) and American/Hawaiian seahorses (*H. ingens*, *H. reidi* and *H. fisheri*).

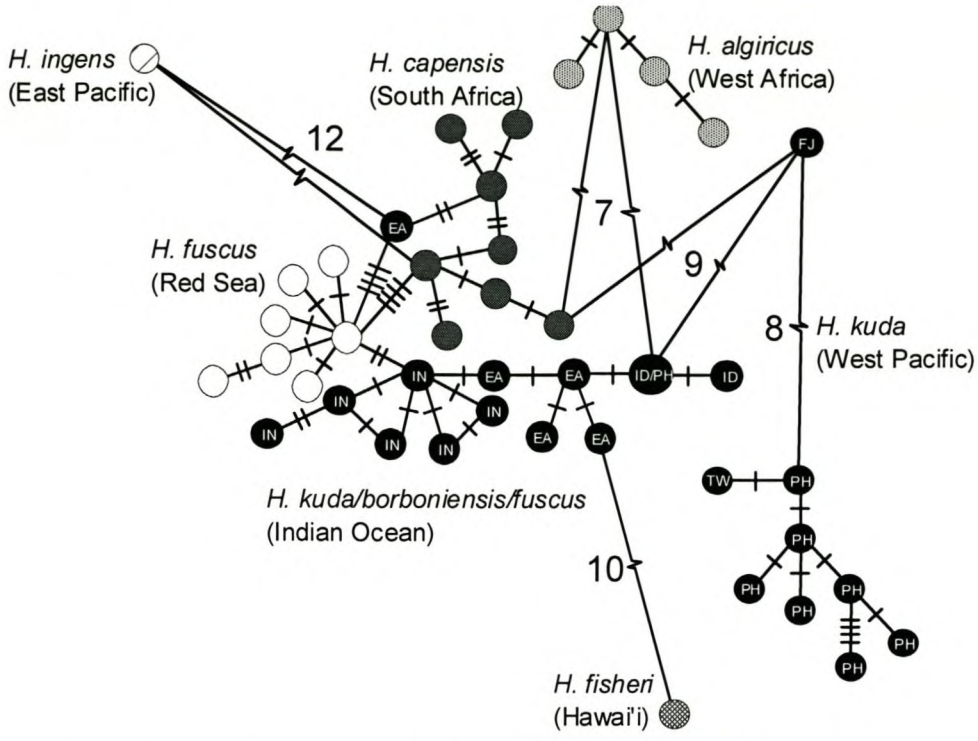

Fig. 2.1 Haplotype network constructed from sequences of the right domain of the mitochondrial CR of kudaoid seahorses. Each nucleotide substitution is indicated by a cross-bar, or, in cases in which a large number of substituions was found, by a number. Frequencies of individual haplotypes within each species are not indicated, as sample sizes differed considerably. Species associated with the *H. kuda* complex (*H. kuda/H. borboniensis/H. fuscus* from the Indian Ocean and *H. kuda* from the West Pacific) are represented by black circles. The origin of each of these specimens is indicated by the following acronyms: EA = East Africa and Southeast Africa; FJ = Fiji; IN = India; ID = Indonesia; PH = Philippines; TW = Taiwan. In the case of some clades of haplotypes distantly related to the central cluster, only a single haplotype is shown.

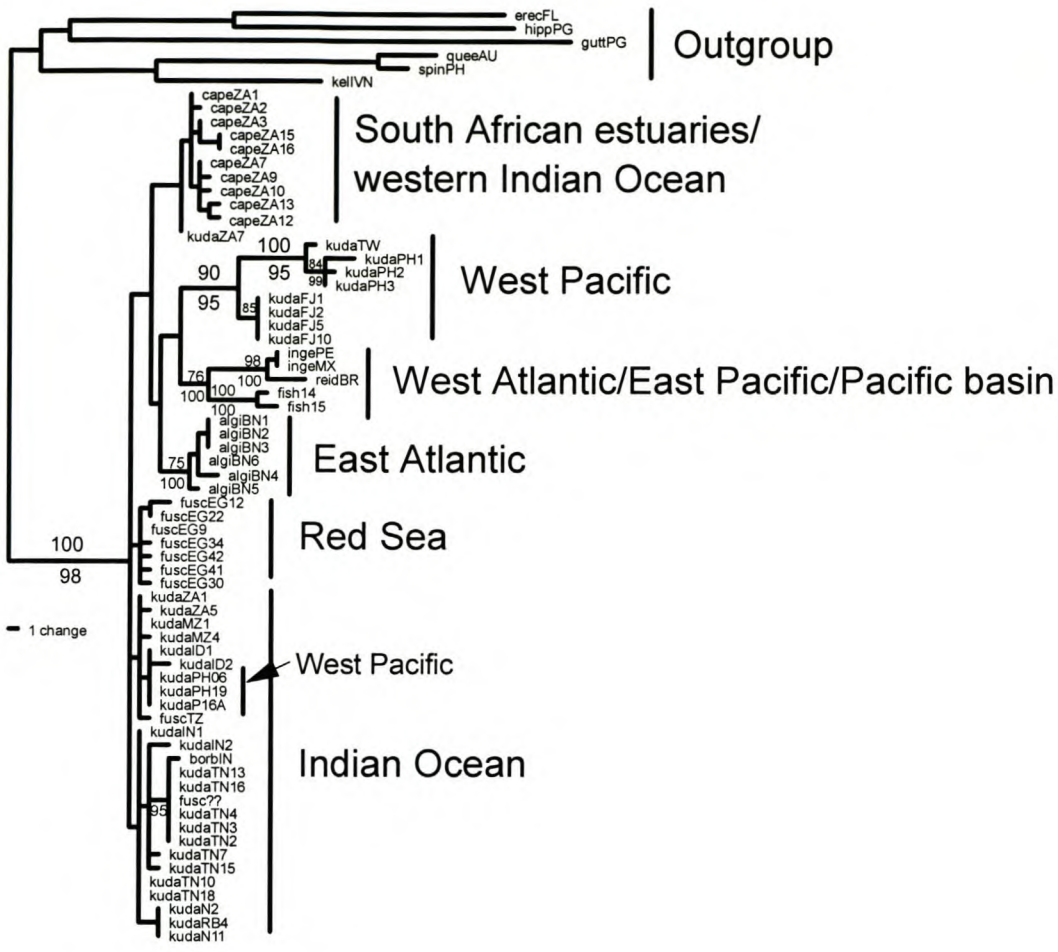

Fig. 2.2 Phylogram of one of 10 000 equally most parsimonious trees constructed from CR right domain sequences of all kudaoid seahorses analysed in this study. Bootstrap values from parsimony analysis and posterior probabilities from Bayesian inference have been added above and below branches, respectively (bootstrap values below 75% and posterior probabilities below 95% are not shown). Tree statistics from parsimony analysis: tree length: 263; RI: 0.831. Associations with geographical regions have been indicated for each of the clades identified. Species names were abbreviated as follows: *Hippocampus algiricus* = algi, *H. borboniensis* = borb, *H. capensis* = cape, *H. fisheri* = fish, *H fuscus* = fusc, *H. kuda* = kuda, *H. reidi* = reid. The following country codes were used: AU = Australia, BN = Benin, BR = Brazil, FJ = Fiji, IN and TN = India, MX = Mexico, PE = Peru, PG = Portugal, PH = Philippines, TZ = Tanzania, VN = Vietnam, and ZA = South Africa.

A comparison of the characteristics of the mitochondrial markers employed in this chapter based on MP analyses is shown in Table 2.2. 16S rRNA is the most conserved marker (lowest proportion of variable sites) whereas the CR, and particularly the right domain, is most variable. However, due to many homoplasies (low retention index), the CR sequences are not more parsimony informative than the more conserved cytochrome *b* sequences.

Table 2.2 Comparison of MP tree scores constructed using consensus sequences from five kudaoid seahorse taxa (*H. kuda* [India]; *H. kuda* [Philippines]; *H. capensis*; *H. algiricus*; *H. reidi*) and *H. queenslandicus/spinosissimus.*

Partition	Number of				Tree length	RI
	total characters	variable characters	informative characters	equiparsimonious trees		
16S rRNA	489	33 (6.75%)	5 (1.02%)	2	38	0.883
CR (right domain)	382	80 (20.95%)	12 (3.14%)	2	94	0.467
CR (complete)	936	149 (15.92%)	25 (2.67%)	2	173	0.500
Cytochrome *b*	1020	141 (13.82%)	32 (3.14%)	3	163	0.595
Combined	2448	323 (13.19%)	63 (2.57%)	3	378	0.532

A haplotype shared by one of the *H. kuda* specimens from India (kudaIN1) and *H. fuscus* from Egypt (fuscEG), was found to be basal in the minimum spanning network computed using 16S rRNA haplotypes of kudaoid seahorses, as it was in a pivotal position within the network (Fig. 2.3). The haplotypes of a second specimen of *H. kuda* from India (kudaIN2) and of *H. kuda* from South Africa (kudaZA) were closely associated with this basal haplotype (both differed by one mutational step). Haplotypes of West Pacific seahorses identified as *H. kuda*, as well as *H. capensis* and the West African species *H. algiricus*, all differed from the central haplotype by two mutational steps. The haplotypes of the two American species (*H. ingens* and *H. reidi*) were more distantly related (9 and 11 mutational steps, respectively).

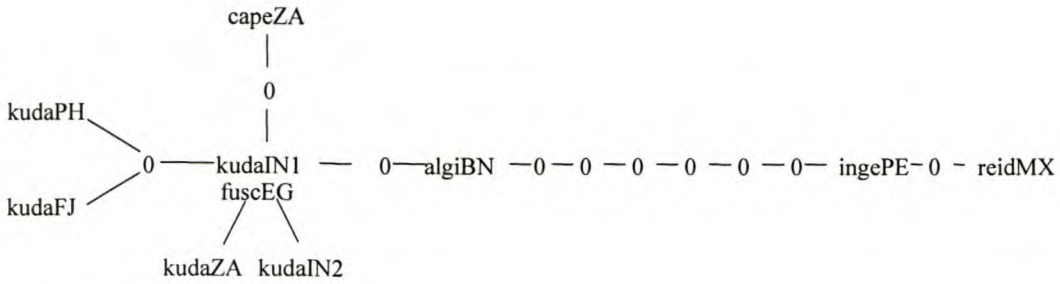

Fig. 2.3 TCS network constructed from 16S rRNA sequences of some of the kudaoid seahorse species. Zeros represent interior haplotypes that were not represented in the samples. Species names were abbreviated as follows: *Hippocampus algiricus* = algi, *H. capensis* = cape, *H. fuscus* = fusc, *H. ingens* = inge, *H. kuda* = kuda, and *H. reidi* = reid. Acronyms following appreviated species names represent the following countries: PH = Philippines, FJ = Fiji, ZA = South Africa, IN = India, EG = Egypt, BN = Benin, PE = Peru, MX = Mexico.

As trees reconstructed using parsimony and Bayesian inference mostly congruent, phylogenetic reconstructions were represented as parsimony phylograms, and bootstrap values from parsimony analyses and posterior probabilities from Bayesian inference were added above and below branches, respectively (Fig. 2.4). Due to the fact that sequence data available for the three individual partitions 16S rRNA, complete CR and cytochrome *b* were not compatible, it is difficult to compare the different phylogenies. Nodal support was low in most cases, with the following exceptions: In the case of 16S rRNA and cytochrome *b*, the monophyly of kudaoid seahorses was well-supported; in the case of complete CR, the clade comprising West Pacific seahorses associated with the *H. kuda* complex (kudaPH, kudaTW and kudaFJ) and the monophyly of kudaZA7 and *H. capensis* was well-supported. The cytochrome *b* and complete CR phylogenies supported a clade comprising *H. reidi* and *H. algiricus*. *H. ingens* clusters with *H. reidi* on the 16S rRNA tree, whereas this species is basal to all other kudaoid seahorse species on the cytochrome *b* tree. Note that because of reasons discussed in chapter 1, the latter arrangement is likely to be incorrect (*H. reidi* and *H. ingens* seem to be a geminate species pair which diverged after the closure of the Central American Seaway). No *H. ingens* samples were available for the complete CR, but the species clusters with *H. reidi* on the tree constructed using CR right domain sequences (Fig. 2.2), and bootstrap and posterior probability support for this association was very high.

Parsimony analyses and Bayesian inference of combined data-sets (a) including and (b) excluding cytochrome *b* resulted in phylogenetic trees that differed considerably with regard to resolution and nodal support (Fig. 2.5), and the topologies of all four trees (i.e. two MP phylogenies and two BI phylogenies) were significantly different from each other (Table 2.3).

However, the weakly supported basal position of *H. ingens* suggested by the cytochrome *b* phylogeny in Fig. 2.4c was lost in all phylogenies when combining cytochrome *b* sequences with the two other data sets, and the monophyly of *H. reidi* and *H. ingens* was always recovered. There was also good nodal support for the association of *H. fisheri* with the two American species, and the basal position of the West African *H. algiricus* within this clade was consistently recovered, although nodal support for this association was low in the case of the parsimony phylogeny including cytochrome *b* sequences. All four phylogenies recovered a monophyletic relationship among the two tip clades *H. kuda* (Philippines)/*H. kuda* (Fiji) with *H. capensis*/*H. fuscus*. The placement of *H. kuda* (India) and *H. kuda* (South Africa) was problematic, and any associations of these with other species tended to be weakly supported.

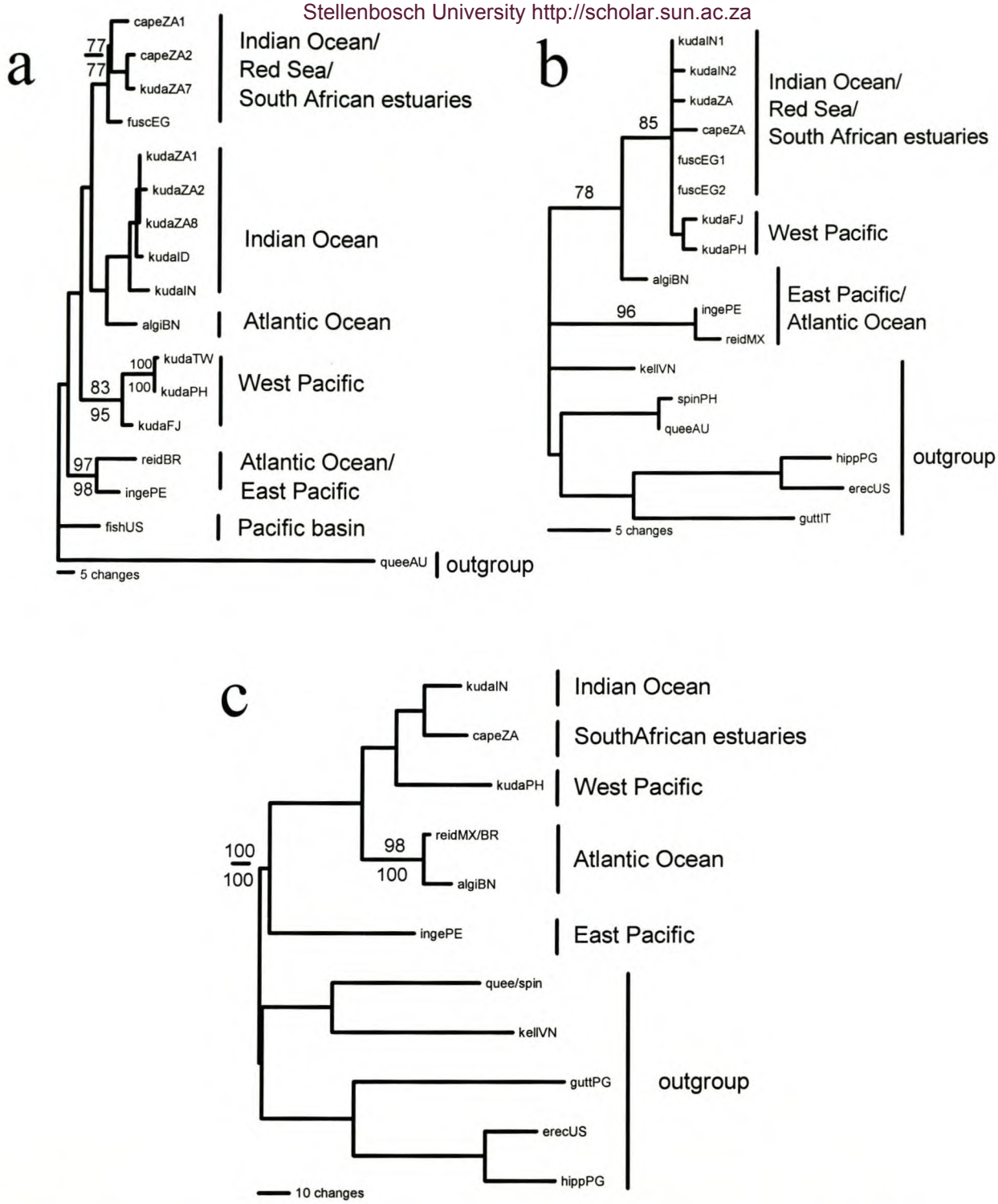

Fig. 2.4 Phylograms of one of a several equally most parsimonious trees constructed using three individual partitions; a) 16S rRNA; b) complete control region; c) cytochrome *b*. Bootstrap values from parsimony analyses and posterior probabilities from Bayesian inference are shown above and below branches, respectively (bootstrap values below 75% and posterior probabilities below 95% are not shown). Associations with geographical regions have been indicated for each clade. Tree statistics from parsimony analyses were: a) no. equiparsimonious trees: 3; tree length: 98; RI: 0.86; b) no. equiparsimonious trees: 2; tree length: 194; RI: 0.75; c) no. equiparsimonious trees: 2; tree length: 443; RI: 0.64.

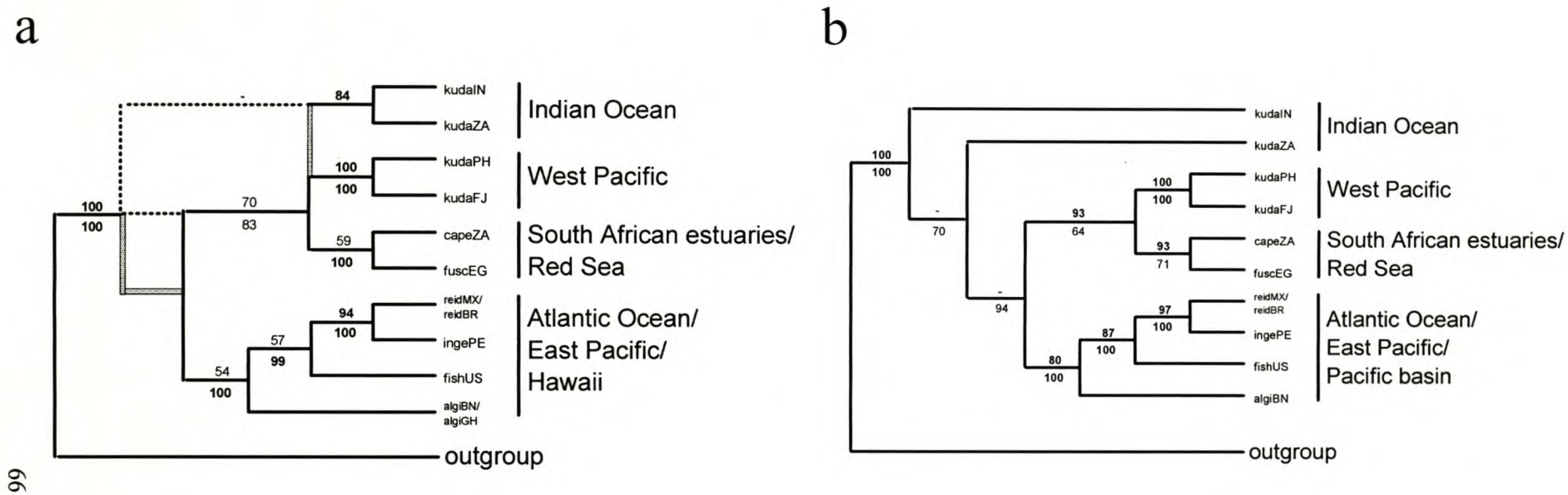

Fig. 2.5 Cladograms constructed using combined sequence data from kudaoid seahorses; a) 16S rRNA, CR and cytochrome *b*; b) 16S rRNA and CR. Nodal support from the two methods of phylogenetic reconstruction used is shown as follows: above branches: Bootstrap values from parsimony analyses; below branches: posterior probabilities from Bayesian inference. High nodal support (bootstrap values ≥75% and posterior probabilities ≥95%) is shown in boldface. The two trees recovered in a) were too different to be represented in the form of a single tree. Hence, different branching patterns are indicated for each tree individually: grey lines: branching patterns from MP analysis; broken lines: branching patterns from Bayesian inference. Associations with geographical regions have been indicated for each clade. Tree statistics from MP analyses were: a) no. equiparsimonious trees: 1; tree length: 402; RI: 0.737; b) no. equiparsimonious trees: 2; tree length: 859; RI: 0.664.

Table 2.3 Results of pair-wise SH tests comparing tree topologies of combined data sets of kudaoid seahorses using two different methods of phylogenetic reconstruction. MP = maximum parsimony coding gaps as 5th characters; BI = Bayesian interference using the optimal model for each data set as determined by MODELTEST (GTR + Γ in both cases). Data-base "a" includes sequences of 16S rRNA, CR and cytochrome *b*; data-base "b" excludes cytochrome *b* sequences. Δ -ln L denotes the difference in likelihood values between the two tree topologies.

Tree 1		Tree 2		Δ –ln L	p
Data-base	Method	Data-base	Method		
a	MP	a	BI	155.16	<0.01
a	MP	b	MP	18.68	0.02
a	BI	b	BI	66.06	<0.01
b	BI	b	MP	96.57	<0.01

2.4 Discussion

2.4.1 General phylogenetic patterns

Although phylogenetic trees constructed from mitochondrial DNA sequences using different methods and partitions (or combinations thereof) did not always have identical topologies, there was generally good support for a total of eight lineages of kudaoid seahorses (Figs. 2.2, 2.4 and 2.5). Their approximate distributions are shown in Fig. 2.6. Based on the analysis where comprehensive taxonomic sampling was used, these eight lineages could be grouped into four major clades, each of which was associated with specific geographic regions: firstly, an Indian Ocean lineage (including *H. kuda* from eastern Africa, India and Indonesia, as well as specimens identified as *H. fuscus* and *H. borboniensis* from these regions); secondly, a lineage comprising species from the periphery of the Indian Ocean (*H. fuscus* from the Red Sea and *H. capensis* from South Africa). These first two lineages seem to be phylogentically closely related as indicated by the identical 16S rRNA sequences of *H. fuscus* from the Red Sea and of *H. kuda* from India. Thirdly, a lineage of West Pacific seahorses (*H. kuda* from the Philippines, Taiwan and Fiji) was found. These taxa are morphologically similar to *H. kuda* lineages from the Indian Ocean (Lourie *et al.* 1999), but considerable genetic differences were found in the case of CR right domain sequences: the lowest number of mutational steps between these two assemblages was found between the haplotypes of *H. kuda* from Fiji and *H. kuda* from Indonesia/Philippines (9 steps), which is higher than the number of steps between *H. algiricus* and the genetically least distant Indian Ocean *H. kuda*-haplotype (7 steps). Lastly, a lineage exists with a distribution reaching from West Africa to Hawaii (*H. algiricus*, *H. reidi*, *H. ingens* and *H. fisheri*). Despite considerable genetic differences within this group, all methods of phylogenetic inference employed identified a monophyletic origin of these four evolutionary lineages, with the exception of the phylogenies constructed from the cytochrome *b* data-base.

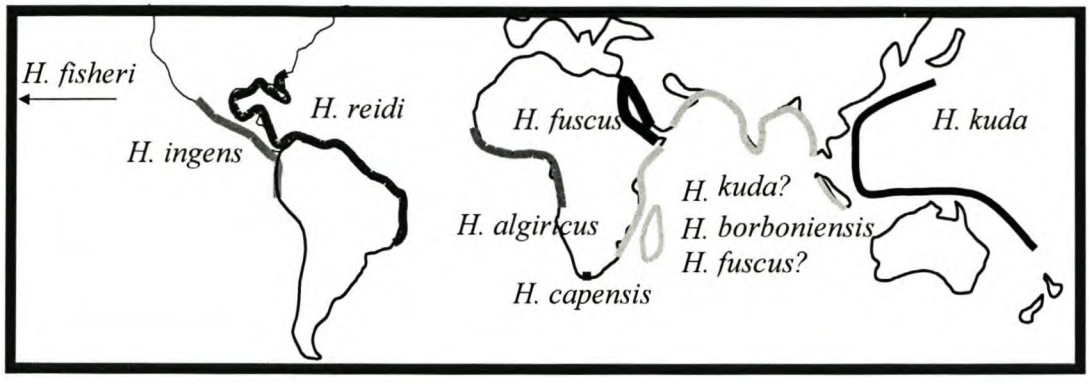

Fig. 2.6 Approximate distributions of the eight groups of kudaoid seahorses identified using mitochondrial DNA sequences.

2.4.2 Evidence for past vicariance in the Indo-West Pacific

The considerable differences between mitochondrial sequences of seahorses identified as *H. kuda* from the Indian Ocean (East Africa, India and southern Java) and those from the western Pacific (Taiwan, Philippines and Fiji) may be an indication that a biogeographic barrier existed between the two regions, which resulted in the genetic divergence of these morphologically very similar groups. The Indo-West Pacific region has a complicated history of faunal convergences and vicariance events resulting from tectonic movements, volcanism and changes in sea-level (Greenfield 1968; Shackleton & Opdyke 1977; Audley-Charles 1981; Woodland 1983; Winterbottom *et al.* 1984; Cane & Molnar 2001). Similar divergences between Indian Ocean and western Pacific sibling species have been identified in many other species of marine organisms and have been attributed to the isolation of formerly continuously distributed species (McManus 1985; McMillan & Palumbi 1995; Palumbi 1996; Barber *et al.* 2002). Barber *et al.* (2002) showed that distinct regional genetic patterns exist throughout the range of the mantis

shrimp *Haptosquilla pulchella*, despite this species' high dispersal potential in the high velocity Indonesian Throughflow. Genetic patterns that developed during historical vicariance events may thus persist long after the barrier to gene flow has disappeared, even in species that are considered to disperse readily. Although the majority of *H. kuda* specimens from Tayabas Bay in the Philippines had a haplotype typical of West Pacific specimens (i.e. clustering with haplotypes from Taiwan and Fiji), three specimens from this population had a haplotype found in a specimen from southern Java (i.e. clustering with haplotypes from India and eastern Africa, Fig. 2.2). The presence of this haplotype in a population of seahorses in the Philippines is difficult to explain if one assumes that the two groups diverged from each other following a vicariance event. Current flow between the Celebes Sea (south of the Philippines) and the Sulu Sea (southwards towards eastern Java) was constricted during periods of low sea-level (Voris 2000), and cold upwelling further isolated this region (Fleminger 1986). During high sea-levels, there is a strong current flow from the West Pacific into the Indian Ocean via the Makasar Strait, and seahorses from the Indian Ocean would thus have to swim against this current to reach the Philippines. Dispersal in the opposite direction seems more likely, and more intensive sampling throughout Indonesia should clarify whether dispersal of seahorses carrying the West Pacific haplotypes into the Indian Ocean has been significantly more common than eastwards dispersal of seahorses carrying the Indian Ocean haplotypes.

An alternative explanation may be that despite the geographical proximity of southern Indonesia and the Philippines, the seahorses carrying Indian Ocean haplotypes dispersed eastwards via northern Australia. This scenario is supported by the fact that the single CR haplotype found in Fiji is genetically less distant from Indian Ocean haplotypes than those found in the Philippines and Taiwan (Fig. 2.2). In a study on the starfish *Acanthaster planci*, Benzie (1998) found that there was a considerably higher proportion of an allozyme variant typical of the Indian Ocean in populations in the south-western Pacific, than in populations in the north-western Pacific and even Western Australia. More research is required to establish whether there was an ancient connection between the Indian Ocean and the south-western Pacific, and whether haplotypes from the south-western Pacific then gave rise to the north-western haplotypes.

The existence of two major lineages of seahorses identified as *H. kuda*, and the fairly large number of mutational steps between these at the control region level, suggests that it may be justified to consider these different species. This notion is supported by the fact that *Hippocampus capensis* and *H. fuscus* from Egypt are both more closely related to *H. kuda* from the Indian Ocean than is *H. kuda* from the West Pacific, and both are considered distinct

species. As some of the seahorses that clustered with the Indian Ocean lineage were identified as *H. borboniensis* Duméril, 1870 (a species originally described based on a specimen collected near the Indian Ocean island of Réunion, Lourie *et al.* 1999), it may be suitable to reserve this name for kudaoid seahorses from the Indian Ocean. The West Pacific lineage should be referred to as *H. kuda*, Bleeker, 1852, as this species was described based on a specimen from an 'East Indian archipelago' (Lourie *et al.* 1999), i.e. a location in the West Pacific.

2.4.3 Phylogeographic patterns in the Indian Ocean

Although there was some evidence for phylogeographic structure within the Indian Ocean (three major subclades were identified on the parsimony phylogram constructed from CR right domain sequences: Indian, Indonesian/West Pacific and East African), genetic differences between these were minimal, indicating that isolation by distance is more likely to be responsible for the patterns observed than any true barrier to gene flow. A larger sample size could reveal that some haplotypes are shared among these regions. The prevailing currents in the Indian Ocean (Fig. 2.7) indicate that most gene flow occurs from East Africa northwards towards the Indian subcontinent and Indonesia, as well as southwards towards South Africa, because in both cases, the currents flow along the coastline. Long-distance dispersal from south-east Asia towards Africa is theoretically possible, but gene flow of this type is likely to be rare. Although such a scenario seems to be supported by the fact that Indonesian/West Pacific haplotypes of *H. kuda* clustered with East African haplotypes rather than the haplotypes of the geographically more proximate Indian population (i.e. suggesting gene flow from south-east Asia to Africa via the Equatorial Currents), this pattern may be an artifact of a small sample size: additional samples from India are likely to reveal that haplotypes more closely associated with those from East Africa and Indonesia/West Pacific are also present in this region.

2.4.4 Phylogenetic distinctness of *H. capensis*

Haplotypes of the two peripheral populations from the Knysna Estuary (*H. capensis*) and the Red Sea (*H. fuscus*) were characterised by ancestral monophyly. The fact that large numbers of specimens were available from both populations (chapter 3), none of which were found to cluster with haplotypes from other regions, may indicate that these two populations are isolated from other populations of kudaoid seahorses in the Indian Ocean. However, closely associated

haplotypes may nevertheless exist in the region. A specimen from Durban harbour identified by Sara Lourie as *H. kuda* or *H. borboniensis* (kudaZA7), was genetically more closely associated with *H. capensis* than with any other seahorse, although *H. fuscus* (Egypt) is morphologically more similar to *H. capensis*. (Fig. 2.8). Hence, although *H. capensis* and *H. fuscus* tend to cluster together on most of the phylogenetic trees constructed, it is more likely that the populations of *H. capensis* were founded by seahorses present along the East African coast rather than by *H. fuscus* from the Red Sea. The latter scenario would additionally be improbable because of the estimated young age of the Red Sea population (discussed in chapter 3). Moreover, sampling in the lower reaches of the Red Sea and in the adjacent Persian Gulf and north-western Indian Ocean may reveal that the *H. fuscus* samples from the region near the Suez Canal merely appear phylogenetically distinct from seahorses in the Indian Ocean because of isolation by distance. Hence, whereas *H. capensis* is likely to be phylogenetically distinct because of the absence of suitable seahorse habitat southwards of KwaZulu-Natal province, South Africa, there is little reason to suggest similar isolation of the *H. fuscus* population from the Red Sea from seahorses present in adjacent regions. However, it is imperative that additional samples from southeastern Africa be obtained in order to determine from where the marine seahorses that gave rise to *H. capensis* originated.

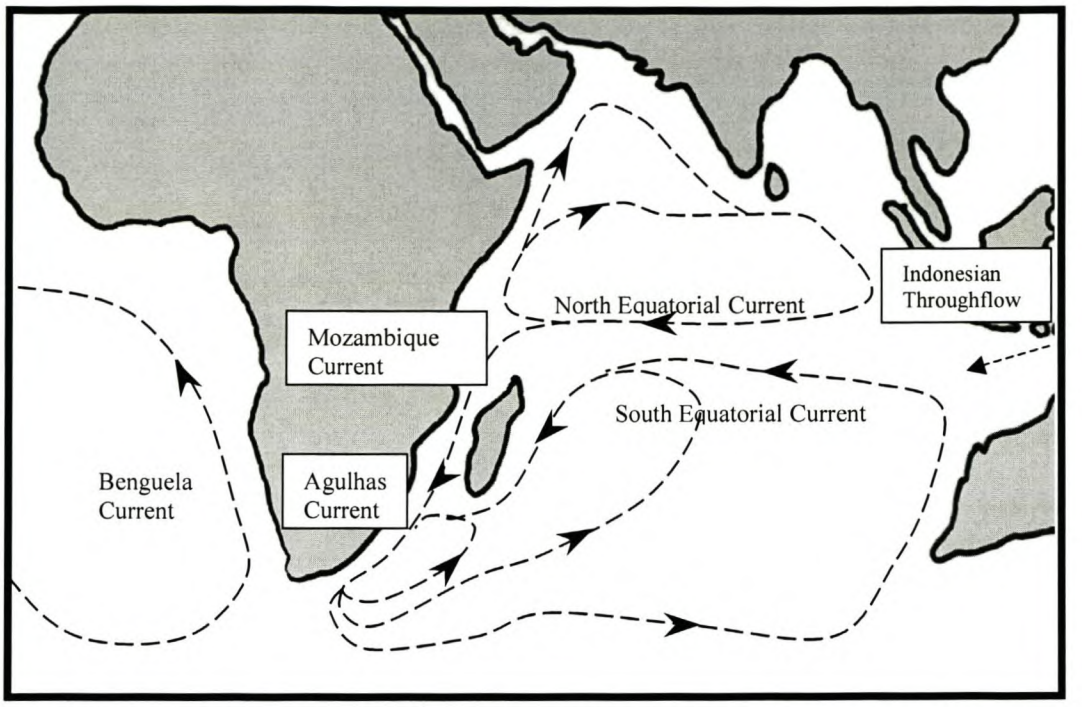

Fig. 2.7 Circulation map depicting the major currents in the Indian Ocean and adjacent regions. Redrawn from Branch & Branch (1992) and Cane & Molnar (2001).

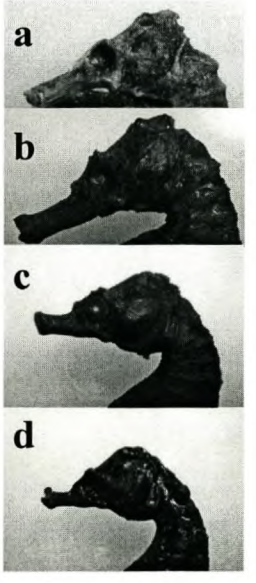

Fig. 2.8 Head morphologies of four specimens of kudaoid seahorses: a) *H. kuda* (Mozambique); b) *H. kuda* or *H. borboniensis* (South Africa); c) *H. fuscus* (Egypt); d) *H. capensis* (South Africa). Note the absence of a coronet in *H. fuscus* and *H. capensis*.

2.4.5 An Indo-West Pacific origin of kudaoid seahorses?

Several factors suggest that the kudaoid seahorse lineage originated in the Indo-Pacific: firstly, the Indo-Pacific species *H. spinosissimus*, *H. queenslandicus* and *H. kelloggi* are most closely associated with the ingroup. The Atlantic biome species *H. hippocampus*, *H. guttulatus* and *H. erectus*, on the other hand, are more distantly related. Secondly, one of the 16S rRNA haplotypes of *H. kuda* specimens from India had a pivotal position within the haplotype network, suggesting that seahorses from this region have given rise to closely related lineages such as *H. kuda* (West Pacific), *H. fuscus* (Red Sea) and *H. capensis*. The monophyletic relationship among the *H. capensis*/*H. fuscus* and *H. kuda* (West Pacific) clades was well supported on one of the phylogenies based on combined data (Fig. 2.5b), but it is possible that this is an artifact of the method of phylogenetic inference used: as most of the kudaoid seahorse species seem to have diverged from each other quite recently, a bifurcating tree may be less suitable to infer phylogenetic relationships among species than a haplotype network, which allows for multifurcations and does not assume that all ancestral haplotypes are extinct. On all haplotype networks, the *H. capensis* and *H. fuscus* haplotypes are most closely associated with *H. kuda* haplotypes from the Indian Ocean. As the 16S rRNA evolves very slowly, it is possible that the pivotal haplotype (kudaIN1 and fuscEG) is an ancestral haplotype that gave rise to the other haplotypes with which it is closely associated. Although the patterns evident on the haplotype networks seem to indicate that kudaoid seahorses originated somewhere in the Indian Ocean and then colonised all other regions, it cannot be ruled out that a basal lineage ancestral to both the Indian Ocean and West Pacific haplotypes was once present throughout both regions. Following a vicariance event in the Indo-West Pacific, the western (Indian Ocean linage) gave rise to the Atlantic lineage and, more recently, the populations on the periphery of the Indian Ocean, whereas the eastern (West Pacific) lineage colonised islands in the Pacific basin.

2.4.6 The colonisation of the Americas and the Pacific basin

Large genetic distances were found between the 16S rRNA sequences of the two American species (*H. reidi* and *H. ingens*) and the Indo-Pacific group. The fact that the haplotype of the West African species *H. algiricus* was located between the Indian Ocean group comprising *H. kuda*, *H. fuscus* and *H. capensis* and the American group on the haplotype network of 16S rRNA sequences (Fig. 2.3) suggests that Indian Ocean seahorses established themselves in the

eastern Atlantic and subsequently colonised the Americas. Phylogenetic trees seem to support this scenario. With the exception of the cytochrome *b* phylogeny (Fig. 2.4c), there was generally good support for a sister taxon relationship between the West African species *H. algiricus* and the two American species. The Benguela upwelling system along the west coast of southern Africa constitutes a biogeographic barrier between the warm water faunas of the Indian Ocean and the eastern Atlantic (Shannon 1985). Divergence between the Atlantic/East Pacific and the Indian Ocean species occurred prior to the closure of the Central American Seaway, because the close association between the east Pacific *H. ingens* and the west Atlantic *H. reidi* suggests that these two species diverged from a common ancestor due to the rising of the Panamanian Isthmus between the reorganisation of ocean circulation 4.6 mya (Haug & Tiedemann 1998) and final closure 2.7 mya (Marshall 1988). During this time, the Benguela upwelling system was in place, but chance dispersal events across this cold-water barrier cannot be ruled out. Lessios *et al.* (2001) showed that dispersal of tropical sea urchins of the genus *Diadema* across the Benguela barrier was possible until 1-2 mya. This phenomenon can be explained by the fact that the intensity of the Benguela Current has experienced cyclic fluctuations throughout the Pleistocene (Meyers *et al.* 1983; Shannon 1985). The intensity of the warm Agulhas Current, which flows southwards along the east coast of Africa, has also been characterised by substantial fluctuations (Bé & Duplessy 1976, Hutson 1980), and intrusions of warm water from the Indian Ocean into the Atlantic have occurred until the present (Shannon *et al.* 1990). Only a short stretch of CR right domain sequences was available for the Hawaiian species *H. fisheri*, but its sister taxon relationship with the two American species was well supported whenever this sequence was incorporated into a data-base of combined sequences. The fact that *H. fisheri* is closely associated with American species and distantly related to *H. kuda* from the West Pacific, suggests that seahorses that had established themselves in the Americas subsequently dispersed further in a westward direction to colonise islands in the Pacific basin. The possibility that *H. fisheri*, *H. ingens* and *H. reidi* are associated with a lineage that dispersed from the West Pacific eastwards needs to be explored. Although the monophyly of this clade with the East Atlantic species *H. algiricus* was well-supported on several trees, the fact that this latter species is more closely related to Indian Ocean kudaoid seahorses than to the American/Hawaiian lineage may be an indication that it could have diverged from Indian Ocean kudaoids more recently. Moreover, Leis (1984) stated that 55 shallow-water fish species from the Indo-West Pacific managed to cross the Pacific barrier to establish themselves in the East Pacific, but there is almost no evidence of successful migrations in the opposite direction (Briggs 1999). It would thus be surprising why seahorses, which are

likely to have poor dispersal abilities, would have managed to disperse from the East Pacific towards the Pacific basin. Scheltema (1988) found that the North Equatorial Current carries larval stages from the East Pacific to the Indo-West Pacific. Although young seahorses are unlikely to be truly planktonic (chapter 1), it thus seems theoretically possible that East Pacific species may establish themselves in the Pacific basin. Moreover, westward dispersal probably occurs much more readily, as both the northern and the southern equatorial currents flow westwards, and a weaker counter-current sandwiched between them flows in an eastward direction (Tsuchi & Ingle 1992). The absence of fishes of East Pacific origin in the Indo-West Pacific is not due to their inability to disperse in a westward direction, but rather the inability to establish themselves in a region that possesses a greater diversity than their region of origin (Briggs 1999). An invading species is thus unlikely to establish itself because of an inability to outcompete established and well-adapted species with which it competes for an ecological niche. In the case of *H. fisheri*, it is possible that ancestral seahorses originating from the East Pacific may have established themselves in Hawaii because of the absence of other seahorse species, but failed to establish themselves in the high-diversity East Indies further westwards.

It is interesting to note that *H. kuda* also occurs in Hawaii (Lourie *et al.* 1999), but unfortunately, no samples could be obtained from that region. It is possible that this species is a more recent arrival in this region. Genetically, this population is likely to be similar to the other two Pacific lineages of *H. kuda* (north-western Pacific: Philippines and Taiwan; south-western Pacific: Fiji). This may indicate that the ranges of two distantly related kudaoid seahorse lineages are overlapping in the Pacific basin. The expanse of the Pacific Ocean has been shown to be a barrier to the dispersal of many marine organisms (Ekman 1953; Scheltema 1988; Rohde & Hayward 2000). However, the presence of both *H. kuda* and *H. fisheri* in Hawaii indicates that the dispersal ability of seahorses seems to be considerably greater than one would assume on the basis of their poor swimming abilities.

2.4.7 High endemism of kudaoid seahorses

It is interesting to note that none of the seahorses associated with subclade 4d identified in chapter 1 (i.e. *H. erectus*, *H. zosterae* and *H. hippocampus*) are represented on the west coasts of America and Africa, despite the long evolutionary history of this lineage in the Atlantic Ocean. It is possible that they were replaced in these regions by kudaoid seahorses during the Pliocene or Pleistocene. Alternatively, the lower species richness in these regions could be

explained by the fact that the west coasts of continents are characterised by upwelling, resulting in environmental conditions less stable than those found on the east coasts of continents (Berrit 1973; Ingham 1970; Rébert 1983; Fleminger 1986; Wells *et al.* 1994). The East Pacific is additionally characterised by El Niño events and other Pacific-wide climatic phenomena (Henderson-Sellers & Robinson 1986). Such conditions may exclude seahorse species that tolerate only a narrow range of certain environmental conditions, and may also increase the chance of extinctions occurring. The comparatively large number of species associated with (and closely related to) the *H. kuda* complex, and the fact that they are widely distributed and present in regions where no other seahorses are found, may be an indication that the members of this group have better dispersal abilities than other seahorses. Additionally, they may be more capable of adapting to unfavourable environmental conditions, and are thus better suited to permanently establish themselves in new habitats.

CHAPTER 3: CONSERVATION GENETICS OF *HIPPOCAMPUS CAPENSIS*

3.1 Introduction

3.1.1 Ecology and distribution of *Hippocampus capensis*

Hippocampus capensis is usually found singly or in small groups in shallow water in association with aquatic vegetation such as the eelgrasses *Zostera capensis* and *Ruppia maritima* and the macroalgae *Codium* spp. and *Caulerpa filiformis* (Bell *et al.* 2003). Like other seahorses, the Knysna seahorse is a poor swimmer, and most individuals are found in regions of the estuaries where current velocities are low (Smith 1981). A prehensile tail is used to hold on to submerged objects to prevent the animals from being washed away by tidal currents (Whitfield 1995). Predators are avoided by means of cryptic colouring, which can be adapted to match the surrounding habitat. Most seahorses found are predominantly black, brown or yellow, but some green and white individuals have also been found (Teske & Lockyear, pers. obs.). Additionally, filament-like appendages sometimes grow out of the skin to improve camouflage (Teske, pers. obs.). The diet of *H. capensis* consists predominantly of small crustaceans, whereas juveniles feed exclusively on zooplanktonic organisms up to 0.75 mm in size (Genade & Hirst 1986).

Sexual maturity in the Knysna seahorse is attained within one year at a length of approximately 65 mm SL (SL = distance from mouth to top of operculum + distance from top of operculum to tail; Lourie *et al.* 1999). Breeding occurs in summer when water temperatures reach about 20° C (Smith 1981; Lockyear, pers. comm.[5]). Like in all other members of the genus *Hippocampus*, the male courts the female until she is ready to deposit her eggs in his brood pouch (Genade & Hirst 1986). Fertilisation occurs within the pouch (Grange & Cretchley 1995). After about 45 days the juveniles (9.10 mm SL), which may number between

[5] Jackie Lockyear, Knysna Seahorse Research Group, Knysna, South Africa

30 and 120, emerge singly from the pouch while the male experiences contractions to help expel them (Whitfield 1995). The newly released juveniles tend to swim in the horizontal position and some are carried by the ebb tidal currents into the sea, where their chances of survival are unknown (Smith 1981).

Extensive SCUBA surveys during 2000-2002 revealed that the current distribution of Knysna seahorses is restricted to the Knysna, Swartvlei and Keurbooms Estuaries on the south coast of the Western Cape Province, South Africa (Teske & Lockyear, pers. obs.; Fig 3.1). Reports of any additional populations in the Klein Brak, Breede, Duiwenhoeks and Goukou estuaries (Whitfield 1995; Grange pers. comm.[6]), could not be confirmed. Human settlement along all three estuaries (with the associated industrial, domestic and recreational activities) poses a severe threat to the survival of the species (Skelton 1987). The largest development in the area is presently the marina development on Thesen's Island (Fig. 3.1). Disturbances of submerged vegetation are expected only at the two mouths of the marina development, and would thus be quite localised (Adams, pers. comm.[7]). However, it is difficult to predict whether changing flow dynamics might result in significant losses or disturbances of the vegetation in the vicinity of the island. Apart from human activities, the seahorses are also exposed to natural hazards. For example, the Knysna Estuary experiences a freshwater flood every 10-12 years (Day *et al.* 1952), which affects particularly the marine fauna present in the system (Korringa 1956). A flood during November 1996 resulted in the mortality of substantial numbers of seahorses (Grange, pers. comm.). In the Swartvlei Estuary, similar flooding resulted in a mass mortality during 1991, when at least 3000 seahorses were removed from the population (Russell 1994). Floods during 2002 resulted in mortalities in the Swartvlei and Keurbooms Estuaries, but none were observed in the larger Knysna Estuary (Lockyear, pers. comm.). It is unclear how these mass mortalities influence the population dynamics of the species but given that freshwater floods are natural occurrences, the populations are probably able to recover quickly due to the high productivity characteristic of estuaries.

[6] Neil Grange, Eco-Africa environmental consultants, Observatory.
[7] Dr. Janine Adams, Botany Department, University of Port Elizabeth, Port Elizabeth, South Africa.

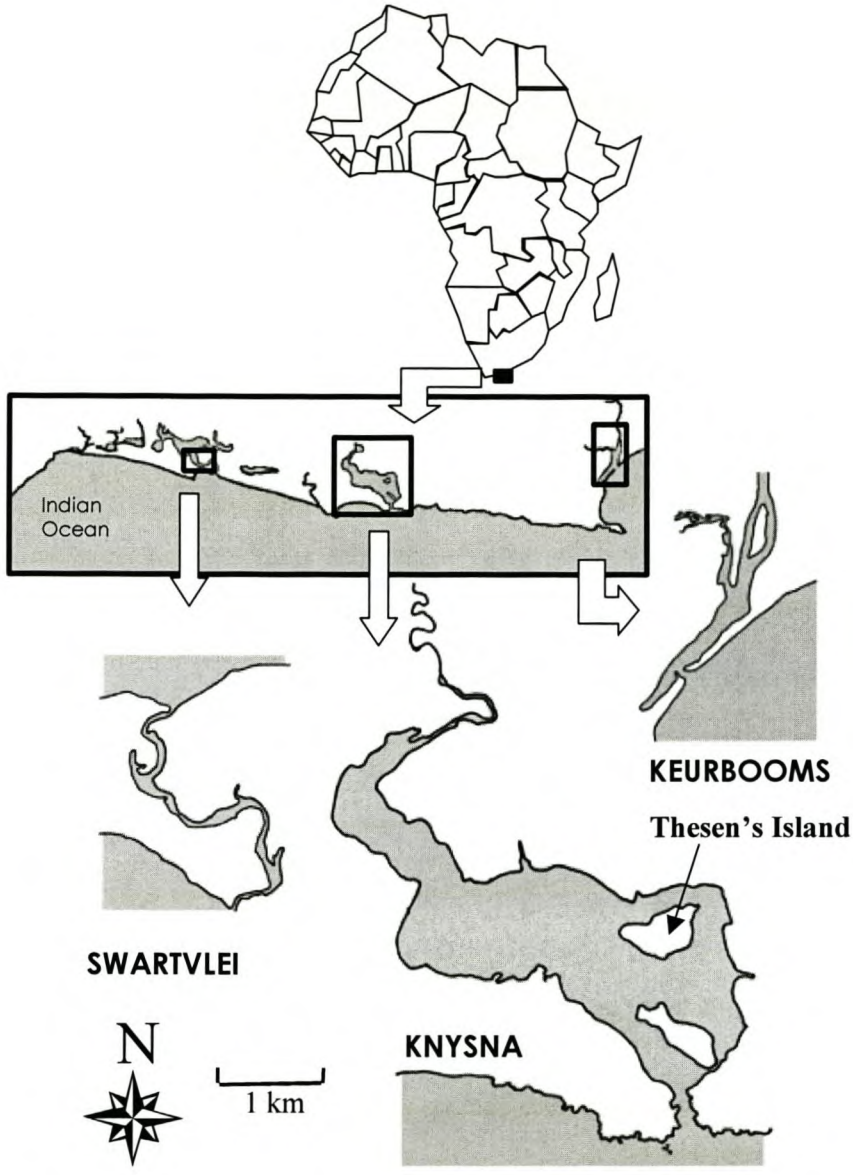

Fig. 3.1 Geographic localities of the three estuaries inhabited by *H. capensis*.

3.1.2 The estuarine environment

Estuaries lie at the interface between the ocean and the land, forming a meeting place of the saltwater regime of the sea and the freshwater flow of rivers. The oceanic input is driven primarily by the regular forcing of the tides, while the freshwater input is dependent on variable rainfall in the catchment areas of the rivers (Schumann *et al.* 1999). In comparison with many other regions of the world, the flow of freshwater into South African estuaries is limited, and

consequently the outflow into the sea does not appreciably affect conditions on the adjacent shelf regions, except during floods. The tidal range along the South African coastline is also small (Schumann *et al.* 1999), and hence, stratification is commonly observed within estuaries. The estuaries inhabited by *H. capensis* are located in a region that is characterised by humid and temperate conditions in which seasonal rainfall regimes predominate. Rainfall reaches an annual average of 400-700 mm (Heydorn & Tinley 1980). The estuaries in this region can be divided into two major types: tidal estuaries and non-tidal lagoons (or temporarily open/closed estuaries), the latter being temporarily separated from the sea by a sand bar that forms in their mouth areas due to low river flow and longshore winds (Whitfield *et al.* 1983). These latter estuaries may become tidal for a period of time following a breaching event. Generally, estuaries are periodically catastrophic systems, which are dominated by 'true estuarine species': animals with marine affinities which have adapted to tolerate the extensive fluctuations in environmental parameters typical of estuaries. Most true estuarine species are capable of living in the sea but are rare because of competition with marine animals (McLusky 1981). Likewise, marine species are able to penetrate estuaries when environmental conditions are suitable. During periods of high river flow, many of the marine species that have become established during drier periods are purged from the system, as they are not capable of tolerating the rapid changes in salinity and possibly also temperature. The occurrence of floods ensures that true estuarine species maintain their dominant position in the system. These are usually small-bodied species, with small crustaceans (isopods, amphipods, cumaceans and tanaids), polychaetes, gastropods and bivalves being the most common representatives of the macrobenthos (Teske & Wooldrige 2001). The availability of large quantities of small prey items makes estuaries ideal habitats for small fish, and many species of coastal teleosts consequently utilise estuaries as nursery areas (McLusky 1981). For the same reason, estuaries may also be attractive to syngnathids, which are highly adapted to stalking small prey items. Apart from the Knysna seahorse, two species of pipefish are found in South African estuaries, namely *Syngnathus temminckii* and *S. watermeyeri* (Kuiter 2000)

From a conservation perspective, it is important to consider the fact that differences exist among the three estuaries inhabited by *H capensis*. The Knysna Estuary (estuarine mouth: 34° 04' 35"S, 23° 03' 40" E) is by far the largest of the three estuarine systems. It covers an area of approximately 19 km^2 (Geldenhuys 1979) and has a tidal reach of about 10 km (Reddering & Esterhuysen 1984). The estuary consists of a lower marine dominated section, a central lagoonal section and an upper estuarine portion (Largier *et al.* 2000). The mouth of the estuary is characterised by a rock formation of Ordovician origin known as the Knysna heads (Toerin

1979), which maintains a large, permanently open estuary mouth. The large input of tidal marine water and the limited inflow of freshwater ensure that much of the Knysna Estuary's faunal diversity is of marine origin. The co-existence of marine, diadromous and endemic estuarine forms results in the Knysna Estuary having the highest biodiversity of any South African estuary (Grindley 1985), and the presence of at least two rare fish species (*H. capensis* and the goby *Pandaka silvana*) confers a high conservation value to this system. The benthic macrofauna includes 310 species, and over 50 species of fish have been identified (Day 1981a). The Knysna River becomes muddy only during flood conditions (Grindley 1985), but under normal conditions, siltation from external sources is virtually absent from the estuary. However, internal movement of sediment has created problems in areas where artificial stuctures, such as causeways, have been erected (Chunnett 1965).

The Keurbooms Estuary (34° 02' S; 23° 23' E) is located approximately 42 km east of the Knysna Estuary. This estuary is fed by the Keurbooms and Bitou Rivers, and covers an area of approximately 2.7 km^2 (Duvenage & Morant 1984). The highest mean annual run-off estimate for the Keurbooms and Bitou Rivers has been 1.6×10^8 m^3 (Noble & Hemens 1978), which is higher than the highest value measured for the Knysna Estuary (1.3×10^8 m^3, Pitman *et al.* 1981). This, in combination with a much smaller estuary area and a small inlet, suggests that the impact of freshwater floods on the Keurbooms Estuary's fauna is much more severe than in the Knysna Estuary.

The Swartvlei system (34° S, 22° 46' E) is located approximately 26 km west of the Knysna Estuary and consists of a lower estuarine zone and an upper lake. The estuary is the smallest of the three systems inhabited by *H. capensis*, with a length of approximately 7.2 km and water surface area of 2 km^2 (Liptrot 1978) and is periodically isolated from the sea by a sand bar. The mean annual runoff from this estuary's catchment area is comparatively low (6.6×10^7 m^3, Anonymous 1978), but mass mortalities occur when the sand bar is breached (Russell 1994; Teske, pers. obs.).

Most of South Africa's estuaries have originated during the Pliocene and Pleistocene epochs (Cooper *et al.* 1999). The presence of lagoonal deposits dated as being Tertiary in age north of the Knysna Estuary suggests that a lagoon or an estuary was present at approximately the same position as today's Knysna Estuary at some stage during the Tertiary (Du Toit 1966). The late Tertiary and particularly the Pleistocene were characterised by repeated and drastic changes in climate (Butzer & Helgren 1972), suggesting that some of the rivers feeding today's estuaries may have converged lower on the continental shelf during cold periods characterised by low sea

levels. The present morphology of the estuaries has remained relatively unchanged since the beginning of the present interglacial period approximately 13 000 years ago (Cooper 1991).

3.1.3 Population structure

From a phylogeographic perspective the distribution of the estuarine Knysna seahorse is interesting. As the distribution of *H. capensis* is closely linked to suitable habitat, with most individuals being restricted to subtidal vegetation in shallow water (Lockyear, pers. comm.), some population genetic structuring may exist among and even within estuaries. Gene flow among different demographic units of a species is an important factor to counteract genetic divergence among populations caused by drift (Ciofi & Bruford 1999), because it reduces inbreeding and allows the spread of favourable adaptations among populations (Wright 1951). Estimates of gene flow thus provide important information for any conservation initiative directed to the long-term persistence of a species in its natural habitat (Ciofi & Bruford 1999). Previous genetic studies on population structure in teleosts have shown that considerable differences exist depending on whether a species occurs in rivers, estuaries, or in the marine habitat. Primary freshwater fish, defined as being physiologically intolerant to marine conditions and considered to disperse via freshwater only (Myers 1938) are generally characterised by high levels of genetic structure (Waters & Burridge 1999; Waters *et al.* 2001), because they can disperse only via river captures or by the confluence of the lower reaches of rivers during marine regression or lowering of the sea level (Banarescu 1986). On the opposite end of the scale are marine species, which may be distributed over vast areas, although genetic analyses have shown that many seemingly continuously-distributed teleost species are characterised by isolation by distance (Pogson *et al.* 1995; Gold & Richardson 1998; Mamuris *et al.* 1999; Nesbø *et al.* 2000).

The extent of genetic divergence among populations of a species depends strongly on its ability to disperse. For example, Indo-Pacific trumpetfishes (*Aulostomus chinensis*) are strong swimmers and reveal only limited population structure across the Indian and Pacific Oceans (Bowen *et al.* 2001). In contrast, genetic studies on marine species associated with coral reefs suggest that dispersal is not usually as extensive as has been believed (Planes 2002). For example, individual populations of the tropical damselfish *Acanthochromis polyacanthus* inhabiting adjacent reefs often did not share any haplotypes (Planes *et al.* 2001). Studies on diadromous fish species (i.e. species whose life-cycles includes both freshwater and marine

phases) have revealed that these show significantly less geographic differentiation than primary freshwater species, and tend to display patterns more similar to those of readily dispersing marine forms (Gyllensten 1985), particularly in mobile species such as salmon (Thomas *et al.* 1986). Ayvazian *et al.* (1994) found significantly more structure among estuarine populations of the catfish *Cnidoglanis macrocephalus* than among marine populations. This was attributed to the fact that this species does not disperse readily and has specialised habitat requirements. Significant structure was found among populations of *Lates calcarifer*, a catadromous teleost (present in freshwater, estuaries and bays), which was attributed to its absence from the marine habitat and short larval phase (Chenoweth *et al.* 1998). The trends emerging from these and other studies suggest that the extent of genetic structure among teleost populations depends on a combination of factors, including the species' dispersal capabilities (Palumbi 1992), their degree of habitat specialisation (Smith & Fujio 1982) and the degree with which they are associated with either the freshwater or the marine habitat.

Distribution records of *H. capensis* in the Knysna Estuary suggest that this species cannot be placed into any of the above-mentioned categories. The species is found throughout the estuary, from the lower reaches to the estuarine head (Teske & Lockyear, pers. obs.). Generally, the head of an estuary is characterised by near freshwater salinities and is devoid of both euryhaline marine species and primary freshwater species (Branch & Branch 1992; Teske & Wooldridge 2003). The fact that *H. capensis* is highly adapted to tolerate drastic fluctuations in salinity and temperature characteristic of estuaries (Branch & Branch 1992) suggests that it is a true estuarine species, i.e. a species restricted to estuaries throughout its life (Day 1981b). This would suggest that the Knysna seahorse is part of little-studied group of species for which there is a paucity of information on population structure. The fact that Knysna seahorses are not capable of powerful independent locomotion, and are thus probably unable to actively disperse to other estuaries via the marine habitat, suggests that genetic differentiation among populations may be greater than in teleosts associated with estuaries that have been studied to date, and similar to that found in reef fishes. However, given the species' wide salinity tolerance range (see below), passive dispersal by means of currents can not be ruled out. Whitfield (1995) reported that on occasion newborn seahorses are carried away by ebb tidal currents into the ocean. This suggests that this species may have a planktonic phase, and, consequently, high dispersal capabilities (Palumbi 1992). However, newborn seahorses have no egg yolk to provide nourishment during a period of dispersal, start feeding immediately, attempt to hold on to vegetation or other objects and are mostly released in areas where the current is not strong (Lockyear, pers. comm.), which suggests that Whitfield's (1995) observations of marine

dispersal were atypical, and probably merely the result of young seahorses not always being able to withstand the currents. Whether or not young seahorses end up in the sea may thus be up to chance. It is unlikely that Knysna seahorses will establish themselves in the marine habitat because of a combination of low water temperatures associated with occasional upwelling events (Bower & Crawford 1981; Schumann 2000), scarcity of suitable habitat (Branch & Branch 1992) and scarceness of food, as the coastal region is far less productive than estuaries (Day *et al.* 1981; Branch & Branch 1992). A small portion of migrating seahorses may survive these adverse conditions for a sufficiently long period of time to allow them to colonise adjacent estuaries. *H. capensis* is able to survive in salinities ranging from 1-59 (Riley 1986), and captive individuals have reproduced successfully in seawater over several generations (Gunter, pers. comm.[8]), supporting the notion that high salinity does not present a barrier to gene flow. However, even if the level of gene flow is high, local adaptations (inferred by unique haplotypes or clades of haplotypes) may exist within the three populations, because the three estuaries inhabited by *H. capensis* differ considerably with regard to hydrological aspects.

It is possible that population structure exists not only among the different estuaries, but also within them, and particularly within the large Knysna Estuary. Not all areas in this system are equally suited to provide habitat for *H. capensis*, and hence, distribution is patchy (Teske, pers. obs., Lockyear, pers. comm.). There are large areas within the Knysna estuary that do not support plant growth, where there may be too much siltation, or where aquatic vegetation is excessively covered with stinging cnidarians (*Tubularia warreni*). Areas providing suitable habitat may thus be cut off from each other by large stretches of inhospitable environment. Small populations are at risk of extinction because of the accumulated effects of low population numbers. These include environmental stochasticity, natural catastrophes, demographic stochasticity and genetic stochasticity (Shaffer 1981) as well as alteration of the relative abundance of pre-existing species (Nunney & Campbell 1993). Determining whether sub-populations exist within the estuary impacts on the way in which population genetic parameters should be interpreted. A large number of haplotypes within the entire population, for example, may distract from the fact that levels of genetic diversity in each of the sub-populations may actually be very low.

[8] Marlene Gunter, Port Nolloth Sea Farms, Port Nolloth, South Africa

3.1.4 The conservation status of *H. capensis*

In order to devise adequate management strategies for an endangered species, such as the Knysna seahorse, it is important to investigate its population history, geographic partitioning throughout its range, and distribution of genetic diversity (Avise 1989; O'Brien 1994). Fish biologists have long recognised the need to identify differences among stocks and to manage them to conserve local adaptations (Taylor 1991; Carvalho 1993). Based on conventional meristics, morphometric work and limited sequencing of the mitochondrial cytochrome *b* gene, Toeffie (2000) found that a certain amount of variation may exist between the Knysna and Swartvlei populations, and that mixing of the two populations should be avoided to preserve local adaptations until more information becomes available. The Keurbooms population has not been included in Toeffie's study, and the status of this population is thus entirely unknown.

3.1.4.1 Genetic bottlenecks

As estuaries are periodically catastrophic systems in which large proportions of the fauna may be eliminated during a freshwater flood, it is particularly important to determine whether such events are likely to have resulted in genetic bottlenecks in any of the three populations. Populations that experience large reductions in effective size for prolonged periods are expected to lose genetic diversity. However, as floods take place over relatively short periods of time, and population numbers are likely to increase again rapidly after such events because of the high fecundity of seahorses (a single batch may contain up to 120 juveniles; Whitfield 1995), the effect of floods on genetic diversity may be negligible. A severe population bottleneck may also occur when a small group of migrants from an established population founds a new population; the accompanying random genetic drift is then known as a founder effect (Hartl & Clark 1988). Loss or initial lack of genetic variation can affect both immediate and long-term population viability (Wayne *et al.* 1986; Leberg 1991; Borlase *et al.* 1993).

3.1.4.2 Effective population size

Some of the main concerns of conservation plans for endangered species are to monitor population size and to avoid the loss of genetic variation (Lande 1995). The effective

population size (N_f for mitochondrial data) is defined as the number of females contributing gametes to the next generation. It is considered an essential concept when linking the theory of population genetics to the "real world" of natural populations (Wright 1969) and a key demographic parameter used to understand population viability (Anthony & Blumstein 2000). The effective size of an animal population reflects the influence of a number of different demographic and genetic processes, including mating structure, immigration, and recent population history (Rogers & Kidd 1996). In many animal populations, N_f is smaller than the census female population size (N_{cf}), as not all females in a population contribute gametes to the next generation with equal probability. Several factors that might impact on N_f , N_{cf} and the total census population size N in populations of *H. capensis* have been identified:

Firstly, in order to derive N from N_f, the sex ratio must be known. However, more demographic data may be required: mating systems affect N_f by determining which gametes are passed on to the next generation, by their effects on reproductive skew (Parker & Waite 1997). For this reason, it is also important to investigate the operational sex ratio by determining the proportion of individuals in both sexes that are sexually active during the breeding season. Seahorses are believed to be monogamous (Vincent & Sadler 1995; Vincent 1995) which suggests that reproductive skew does not need to be considered (but see Kvarnemo 2000). Nevertheless, due to the patchy distribution of suitable habitat particularly in the Knysna Estuary, it is possible that a portion of sexually mature individuals may not be able to find a mating partner during the breeding season. Knowledge of the proportion of adults not reproductively active during this period is thus important to determine how N_f relates to N.

Secondly, many species settle preferentially in a habitat already inhabited by other members of the same species (Anthony & Blumstein 2000). Conspecific attraction has been documented in a large variety of species (Alatalo *et al.* 1982; Stamps 1988; Podolsky 1990; Muller *et al.* 1997; Muller 1998). In seahorses, there would be two advantages resulting from this kind of behaviour: firstly, the presence of other members of the species serves as a cue to indicate suitability of the habitat during habitat selection, and secondly, the chances of finding a suitable mate are improved. However, it is not always correct to assume that occupied patches are of the highest quality (Reed & Dobson 1993). SCUBA surveys have shown that large areas of potentially suitable habitat in the Knysan lagoon are only sparsely inhabited by seahorses, even though they do not seem to differ from areas where seahorse densities are very high (Lockyear, pers. comm.). If further research including a large number of environmental variables fails to identify any distinct differences between such areas, it cannot be ruled out that conspecific

attraction may influence settlement decisions, and needs to be incorporated into considerations regarding N_f.

Thirdly, dispersal directly influences N_f by reducing population size and potentially by influencing the ability of a population to escape disturbance (Anthony & Blumstein 2000). Additionally, dispersal indirectly influences N_f through its influence on a population's genetic structure (Shields 1987). In the case of Knysna seahorses, it needs to be considered that a proportion of adult seahorses may be flushed out of the estuaries during floods. Other factors that may influence N_f are fluctuations in population size (e.g. as the result of floods), variations in progeny numbers, or the fact that regions within a single estuary may be subject to extinction and re-colonisation by sub-populations.

3.1.4.3 Comparative population study

Population densities of seahorses within the Knysna, Swartvlei and Keurbooms Estuaries are likely to be considerably higher than densities of seahorses in the marine habitat (e.g. Inhaca island, Mozambique: Teske, pers. obs.; Fiji: Hamilton, pers. comm.[9]). This may be explained by the fact that estuaries are highly productive habitats. However, it is likely that the levels of genetic diversity and total population size of *H. capensis* are small compared with those of marine seahorses because of the comparatively small area of occupancy available to this species. In order to explore how population genetic parameters of the Knysna population compare to those of marine seahorses, population history, diversity and other estimates can be compared among these two groups.

[9] Dr. Healy Hamilton, Center for Biodiversity Research, California Academy of Sciences, San Francisco, California, USA

3.2 Materials and methods

3.2.1 Sample acquisition

A total of 138 specimens were sampled from all three estuaries during 2000/2002. In most cases, tissue samples were obtained non-destructively by taking small fin clips (<1mm^2) from the lower edge of the dorsal fin of living seahorses, and these were stored in 70% ethanol. This method has previously been used by Kvarnemo *et al.* (2000) to obtain tissue samples from seahorses. Due to the high conservation status of *H. capensis*, we tested it thoroughly on seahorses bred in captivity and did not find any adverse effects (chapter 4). To test possible population sub-structure within the large Knysna Estuary, three sites were chosen in regions of the estuary which SCUBA surveys had identified as containing high densities of seahorses (Lockyear, pers. comm.; Fig. 3.2). Thirty seahorses were randomly sampled from each of these three sites. Site 1 was located at the head of the estuary, and was characterised by turbid water and low salinity. Site 2 was located in the middle section of the estuary, which is characterised by higher salinities and slower current velocities (Largier *et al.* 2000). Site 3 was located in close proximity to the mouth. This portion of the estuary is characterised by near seawater salinities, cooler water temperature and strong tidal currents (Largier *et al.* 2000). Seahorses found at the two upper sites were all adults, whereas those at site 3 were all juveniles. Tissue samples from seahorses in the Swartvlei Estuary consisted partly of fin clips obtained from living seahorses (6 individuals), and partly of entire pectoral fins obtained from seahorses that had died following a breaching event (24 individuals). Samples originated from a large area in the middle section of the estuary (Fig. 3.2). The Keurbooms population is considered the smallest of the three assemblages (Lockyear, pers. comm.), and was believed to be extinct until a number of individuals were found in 2002. Hence, a permit was granted to sample only 18 individuals. These originated from sites in the upper estuary, two of which were located in the lower Bitou River, and one was located in the upper part of the Keurbooms Estuary (Fig. 3.2).

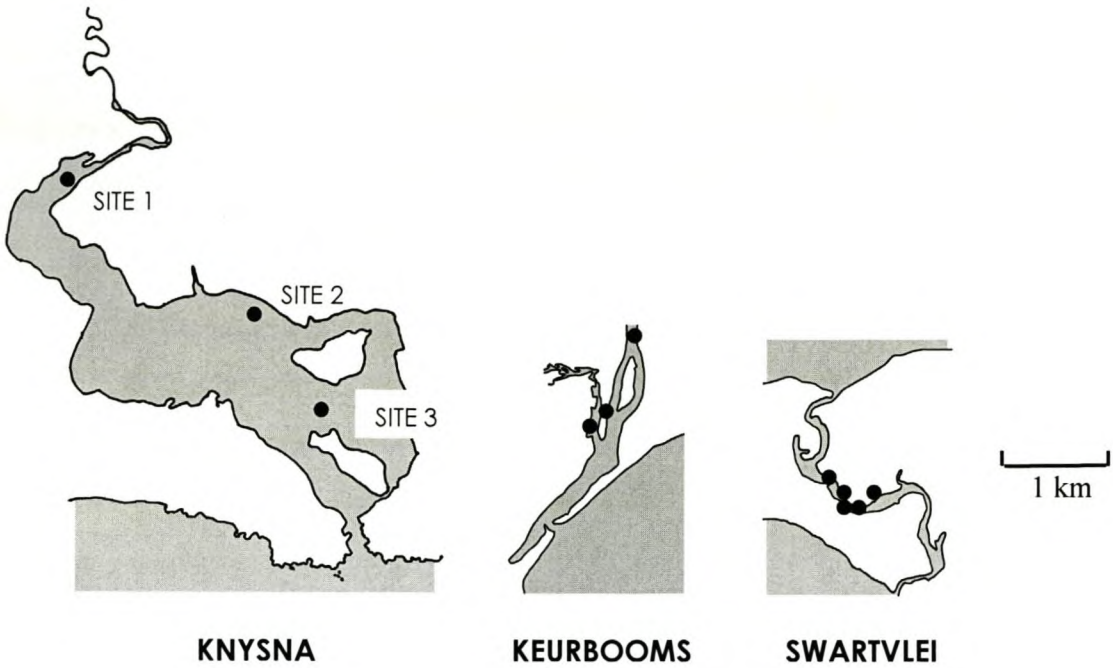

Fig. 3.2 Location of sampling sites in the three estuaries inhabited by *H. capensis*.

3.2.2 DNA extraction, amplification and sequencing

Genomic DNA was isolated by proteinase K digestion as described previously (chapter 1). Most fin clips were completely dissolved after approximately 5 h, after which a standard phenol-chloroform extraction followed (Sambrook *et al.* 1989). As no pellets were visible at the bottom of the tubes after centrifugation, most of the liquid portion was removed using a micropipette, and approximately 20 µl of fluid was left in the tube and evaporated. Samples were subsequently resuspended in 50 µl of TE buffer. Mitochondrial CR right domain sequences from 138 specimens of *H. capensis* were amplified and sequenced using the protocol described previously.

3.2.3 Data analysis

Sequences were aligned by eye in PAUP* version 4.0 beta 10 (Swofford 2002). A homologous region of 402 nucleotides was obtained for all individuals. Fifteen haplotypes were obtained, and these have been deposited in GenBank (accession numbers AY149664-AY149678).

3.2.3.1 Genetic diversity

To characterise genetic variation among sampling sites and populations, estimates of nucleotide diversity (π), haplotype diversity (h) and mean number of pair-wise differences (d) were obtained using the software package DNASP (Rozas & Rozas 1999). In order to minimise the effects of unequal sample sizes among populations, haplotypes of the Knysna and Swartvlei populations were sub-sampled. Each sub-sample consisted of 18 or 30 sequences randomly selected from the total set of sequences using a random number generator (http://www.randomizer.org). In order to calculate means and standard deviations of genetic indices, nine sub-samples were created in each case.

3.2.3.2 Intraspecific genealogy

An intraspecific genealogy for *Hippocampus capensis* was inferred using the programme TCS (Clement *et al.* 2000). The root of the genealogy is inferred by means of a common haplotype present within the population (referred to as 'most recent common ancestor', or MRCA). This haplotype not only tends to be the most abundant but also the most connected of all the haplotypes within a population.

3.2.3.3 Population structure

An analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) was conducted to investigate population structure within the Knysna Estuary and between the three estuaries. AMOVA uses the frequencies of haplotypes and the number of mutations between them to test the significance of the variance components associated with various hierarchical levels of genetic structure (e.g. within populations and among populations) by means of non-parametric

permutation methods (Excoffier *et al.* 1992). A distance matrix was constructed using the Tamura-Nei model (Tamura & Nei 1993), the corresponding gamma shape distribution parameter was calculated using maximum likelihood in PAUP*, and the significance of Φ-statistics was tested using 10 000 permutations. F_{ST} (Wright 1965), Φ_{ST} from AMOVAs and Φ_{ST} from pair-wise analyses were estimated as measures of population differentiation, using the programme ARLEQUIN.

3.2.3.4 Genetic bottlenecks

The populations were tested for evidence of recent genetic bottlenecks by implementing a graphical method introduced in Luikart *et al.* (1998). Haplotypes were assigned to frequency classes based on their particular frequency of occurrence within a population. Frequency classes were then plotted against the number of haplotypes assigned to each class. As rare alleles are more likely to become purged from a population during a bottleneck than intermediate or abundant alleles, fewer haplotypes should be found in the lowest frequency class than in intermediate frequency classes. A simulation programme (ALLELOCIDE, see Table A15 for pseudocode) was written to identify the size at which a randomly mating population which has undergone a recent catastrophic reduction in population size, e.g. due to a freshwater flood, is at risk of losing genetic diversity (i.e. undergoes a genetic bottleneck). In order to determine by how much the female effective population size of the Knysna population would have to decrease for this population to risk losing haplotypes, the following hypothetical model was employed: the starting population consisted of 10 000 individuals, haplotype proportions reflected those found in the Knysna population, and the number of individuals was reduced in increments of 20, each run being repeated 100 times.

3.2.3.5 Effective population size and population history

The concept of effective population sizes has played an important role in the conservation genetics of endangered species, because with such information, better decisions can be made about breeding designs and artificial migration events (Crandall *et al.* 1999). The population genetic estimator $\theta = 4\ N_e\mu$ (diploid data) or $\theta = 2\ N_f\mu$ (haploid data), where N_e/N_f is the effective population size (diploid data) or the effective female population size (mitochondrial data) and μ is the mutation rate leads to an expectation of the amount of genetic diversity in a

population. With a known mutation rate and an estimate of θ, one can estimate the effective population size of the population: $N_e = \theta/4\mu$ (diploid data) or $N_f = \theta/2\mu$ (haploid data). Until recently, effective population sizes were estimated by making pair-wise estimates of θ. Watterson (1975) estimated genetic diversity as a function of the number of segregating sites among sequences, denoted as K. Nei & Tajima (1981) introduced an alternative method based on the proportion of nucleotide differences per pair of sequences, π. Under the assumptions of the infinite-sites Wright-Fisher model (Fisher 1930), including no recombination, no selection, and constant effective population size, these estimators are unbiased. However, the variance associated with both estimators can be large. Felsenstein (1992a) and Fu & Li (1993a) demonstrated the inefficiencies of K and π, and the current trend is to incorporate genealogical information to obtain more efficient estimators of genetic diversity (Felsenstein 1992a,b; Fu 1994a,b; Kuhner *et al.* 1995).

The method developed by Kuhner *et al.* (1995) is a maximum likelihood approach that samples over a range of genealogies weighted by their likelihoods. It makes use of the Metropolis-Hastings sampling strategy: a repeated process of modifying a genealogy and accepting or rejecting it in proportion to the ratio of its probability to the probability of the previous genealogy (Metropolis *et al.* 1953; Hastings 1970). The method assumes no selection, constant population size, no recombination and no migration. A recent upgraded version (implemented in the software programme FLUCTUATE) accounts for growing (or declining) population sizes (Kuhner *et al.* 1998). If the population size has changed over time, the distribution of coalescence times will differ from its expectation in a population where θ is constant. For example, if a population has been growing, the most rootward branches should be relatively short, whereas if it has been shrinking, the most rootward branches will be relatively long (Kuhner *et al.* 1998). The programme FLUCTUATE makes a joint maximum likelihood estimate of θ and the growth rate g. Genealogical methods are at a particular advantage when g is low or negative, a case in which pair-wise methods tend to fail due to the confounding influence of the genealogical structure (Slatkin & Hudson 1991). The use of DNA sequence data with N_f estimators to detect contemporary population declines has limited practical utility because multiple, independent DNA sequences are necessary to obtain N_f estimates with reasonably narrow confidence intervals and recombination rates must be known (Schwartz *et al.* 1999). Mitochondrial DNA sequence data is potentially useful because it does not recombine, but results should nevertheless be accepted with caution, because they are based on a single marker only.

Historical demographic patterns of the Knysna population were investigated by making joint maximum likelihood estimates of the parameters θ ($=2N_f\mu$, where N_f is the effective female population size and μ is the hypothetical mutation rate for mtDNA CR right domain) and g (the exponential growth parameter) using FLUCTUATE version 1.3 (Kuhner *et al.* 1998). In order to comply with the data requirements of this programme, four rare haplotypes were removed, and two alternative methods were employed. For the first method, homoplasies were eliminated by changing nucleotides found at problematic positions to represent the same nucleotide found in the most closely related ancestral haplotype. To compensate for this, the changed nucleotide substitution was introduced elsewhere so that the net number of mutations remained unchanged. In the second method, the homoplasic site at position 344 was removed (see Results, Table 3.1), resulting in a total of only seven haplotypes. In both cases, randomly created input trees were used for each run. If the programme is run for sufficiently long, any bias created by the starting tree should be lost (Kuhner, pers. comm.[10]). Watterson's (1975) segregating sites estimate was used as the initial estimate of θ for each run, and the transition:transversion ratio was estimated using the maximum likelihood algorithm in PAUP*. Historical N_f values were calculated using the formula $N_t = \theta\, e^{-(g\mu)t}$, where N_t is the effective female population size at time t in the past (Kuhner *et al.* 1998). The present female effective population size was determined using the formula $N_f = \theta$ generation time^{-1} μ^{-1}. A divergence time of 1 million years per 2% sequence divergence has been widely used for bony fishes (Brown *et al.* 1979; Bermingham & Avise 1986; Grewe *et al.* 1990), but this mutation rate is based on the entire mtDNA molecule. The specific mutation rate for the non-coding CR is higher, and rates used in the recent literature range between 3.6% per million years (Donaldson & Wilson 1999) to 5-10% per million years (Brunner *et al.* 2001). As the calibration of a molecular clock is not possible in *H. capensis*, hypothetical estimates of θ and g were determined using three possible mutation rates: 2% million years^{-1}, 3.6% million years^{-1} and 5% million years^{-1}. The generation time of seahorses is probably no greater than 1-3 years; sexual maturity is attained within one year (Whitfield 1995), and seahorses kept in captivity live up to at least three years (Lockyear *et al.* 1997). Hence, results are reported using generation times of one, two and three years. For comparison, the parameter θ was also calculated using pair-wise estimations as implemented in the programme ARLEQUIN: the parameter θ_S (Watterson 1975; Tajima 1989) is based on the number of segregating sites and the parameter θ_π (Tajima 1983) is based on the number of nucleotide differences.

[10] Dr. Mary Kuhner, Department of Genetics, University of Washington, Washington, USA

3.2.3.6 Raggedness statistic and departure from selective neutrality

Harpending's raggedness statistic (Harpending 1994) was calculated in ARLEQUIN in order to determine whether there was significant genetic evidence for population growth or stasis. Departure from selective neutrality was tested using Fu's F_s, Fu and Li's D*, and Fu and Li's F* (Fu & Li 1993b; Fu 1997). Fu's F_s is particularly suited to detect departures from neutrality in non-recombining sequences characterised by a high frequency of rare haplotypes and recent mutations (Fu 1997). If a significant departure from selective neutrality is detected only when implementing Fu and Li's tests, this suggests that background selection is the more likely cause of this deviation from neutrality. If, on the other hand, only F_s is significant, departure from the assumption of neutrality is more likely due to population growth or hitchhiking (Fu 1997). Fu's F_s was calculated using the programme ARLEQUIN, whereas Fu and Li's tests were determined using DNASP.

3.2.3.7 Comparative population study

3.2.3.7.1 Sample acquisition

In order to compare population genetics and histories of *H. capensis* with those of closely related marine seahorses, 47 samples were obtained from a population of the Red Sea seahorse *H. fuscus*, and 35 samples were obtained from a population of *H. kuda* from Tayabas Bay in the Philippines. Samples of *H. kuda* consisted of either fin clips or tail clips, and specimens were used further for laboratory experiments, whereas specimens of *H. fuscus* were all sacrificed, and complete pectoral fins were used to obtain genomic DNA. DNA extraction and amplification of mitochondrial CR right domain was performed using the methods described previously. Intraspecific genealogies of the three populations from the Knysna Estuary, the Red Sea and Tayabas Bay were reconstructed using the programme TCS version 1.06 (Clement *et al.* 2000), which is available at http://bioag.byu.edu/zoology/crandall_lab/programs.htm. TCS estimates genealogies from DNA sequences by implementing the statistical parsimony method described in Templeton *et al.* (1992). Unlike the conceptually simpler method in ARLEQUIN, which can be used to investigate relationships among fairly distantly related species, TCS networks are constructed with a cut-off limit of a certain number of steps beyond which the probability of a connection is less than 95% of being true. As the number of sequences available from each

population differed (*H. capensis* = 90, *H. fuscus* = 47, *H. kuda* = 35), 35 sequences from the first two populations were randomly chosen by means of a random number generator (http://www.randomizer.org). To characterise genetic variation among populations, estimates of nucleotide diversity (π) and haplotype diversity (h) were obtained using ARLEQUIN.

3.2.3.7.2 Mismatch distribution

Episodes of population growth and decline leave characteristic signatures in the distribution of nucleotide site differences between pairs of individuals (Rogers & Harpending 1992). In histograms showing the relative frequencies of pairs of individuals who differ by *i* sites, where *i* = 0, 1, 2 etc., an episode of growth (either continuous or as the result of a sudden population explosion) generates a wave that over time travels to the right (Rogers & Harpending 1992). In contrast, the histogram of a stationary population is characterised by a decrease of relative frequency values from 0 towards the right, i.e. it is either multimodal or free of waves (Hudson & Slatkin 1991; Rogers & Harpending 1992). The simplest form of growth to consider is the model of sudden expansion (Rogers & Harpending 1992). An initial population, at equilibrium with $\theta = \theta_0$, is assumed to grow or shrink rapidly to a new size at which $\theta = \theta_1$, and this burst of growth is assumed to occur τ units of mutational time before the present. The value τ is defined as $\tau = 2ut$ (Li 1977), where t measures the time in generations and $u = 2\mu k$, where μ is the mutation rate of the gene fragment sequenced, and k is its length in nucleotides.

To infer historical demographic patterns of the three populations, observed distributions of pair-wise nucleotide differences (mismatch distributions) were compared with those expected for stationary (Watterson 1975; Slatkin & Hudson 1991; Rogers & Harpending 1992) and expanding (Rogers & Harpending 1992) populations by calculating pair-wise distances using the programme ARLEQUIN. The validity of the model was tested by obtaining the distribution of the test statistic SSD (sum of squared differences) between the observed and the estimated mismatch distribution (Excoffier & Schneider 1999) based on 1000 bootstrap replicates. The p-value of the SSD statistic is computed as the proportion of simulated cases that show an SSD value larger than the original (Schneider & Excoffier 1999). A significant SSD value is taken as evidence for departure from the estimated demographic model, which can be either a model of population expansion (if $\tau > 0$ and $\theta_1 > \theta_0$) or a model of population stationaity (if $\tau = 0$ or $\theta_1 = \theta_0$). The three parameters θ_0,(initial θ), θ_1 (final θ) and τ (moment estimator of the time to the expansion) estimated in ARLEQUIN were additionally used to obtain a graphic representation of

the expected pair-wise nucleotide differences for an expanding population using the programme DNASP. This programme was also used to plot frequencies of observed pair-wise differences, as well as the corresponding values expected under a model of population stasis. Due to the poor performance of this method to calculate population parameters and because of the comparatively small sample size as compared to the population study on the Knysna seahorse, we did not attempt to calculate the ages of the expansion events that characterised each of the three populations. Instead, values of τ were merely used to compare the relative ages of the populations with regard to the population age of the Knyna population calculated previously.

To contrast phylogenetic trends found in the different seahorse species with those of a closely related teleost, 28 CR sequences of the East Asian ninespine sticklebacks (*P. pungitius* and *P. sinensis*) were downloaded from GenBank (Takahashi & Goto 2001; accession numbers AB054320-AB054362), aligned with seahorse CR sequences in CLUSTALX, and then pruned to equal length. The two species comprise a single monophyletic mitochondrial lineage, and samples were collected from estuaries and bays throughout Japan. All of the tests and genetic indices carried out for the three seahorse populations were repeated for this data-base, with the exception of the reconstruction of an intraspecific phylogeny. Additionally, uncorrected p-distances were calculated for the stickleback data-base and compared with uncorrected p-distances of the kudaoid seahorse sequences used in chapter 2.

3.3 Results

3.3.1 Genetic diversity

The segment of the mitochondrial CR sequenced in 138 individuals of *H. capensis* contained 10 variable sites (one indel, eight transitions and one transversion; Table 3.1). These polymorphic sites defined 15 haplotypes, of which six were unique to the Knysna population, one was unique to the Keurbooms population, and one was found exclusively in the Swartvlei population. Six haplotypes were represented by single individuals. Although the sampling area in the Swartvlei Estuary was considerably larger than each individual sampling site within the Knysna Estuary, the number of haplotypes found was lower than at any one of the three sites within the Knysna Estuary. Haplotype 1 was the most abundant haplotype at each of the sites within the Knysna Estuary, and the second most abundant in the Keurbooms and Swartvlei populations.

Table 3.1 Frequency of occurrence of *Hippocampus capensis* CR right domain haplotypes in the Knysna Estuary (three sampling sites with 30 individuals each), Keurbooms Estuary (18 individuals) and Swartvlei Estuary (30 individuals). Segregating sites of derived haplotypes are compared with nucleotides at corresponding sites in haplotype 1 (in boldface). Nucleotides identical to the ones in haplotype 1 are marked with a dot. The position of each segregating site is indicated by a three digit number.

	n					Segregating sites									
	Knysna sites					0	0	1	1	1	1	2	2	3	3
						0	2	0	3	8	8	4	9	4	9
Haplotype	1	2	3	Keurbooms	Swartvlei	9	2	5	1	7	9	5	4	4	5
1	**15**	**13**	**8**	**3**	**9**	**T**	**C**	**A**	**T**	**C**	**-**	**A**	**C**	**T**	**G**
2	0	1	0	1	0	·	·	·	·	·	C	·	·	·	·
3	1	1	5	1	0	A	·	·	·	·	·	·	·	·	·
4	1	0	0	0	0	·	·	·	·	T	·	·	·	·	·
5	1	3	4	0	0	·	·	·	·	·	·	·	·	C	·
6	0	0	1	0	0	·	·	·	·	·	C	·	·	C	·
7	0	0	1	1	0	·	·	·	·	·	·	·	T	C	·
8	4	5	3	1	0	·	T	·	·	·	·	·	·	·	·
9	2	1	0	0	0	·	T	·	·	·	·	·	·	C	·
10	0	0	1	0	0	·	T	·	·	·	·	G	·	C	·
11	0	0	0	1	0	·	T	·	·	·	·	·	T	C	·
12	4	6	7	9	0	·	T	G	·	·	·	·	·	·	·
13	0	0	0	0	1	·	T	G	·	·	·	·	·	C	·
14	1	0	0	0	0	·	T	G	·	·	·	·	·	C	A
15	1	0	0	1	20	·	T	G	C	·	C	·	·	·	·

Haplotype diversity and expansion coefficients were similar in the Knysna and Keurbooms Estuaries, and lower in the Swartvlei Estuary, whereas nucleotide diversity was higher in the two smaller estuaries than in the Knysna Estuary because their haplotypes tended to be more divergent (Table 3.2). Magnitudes of haplotype and nucleotide diversities were not drastically affected by sub-sampling. This suggests that smaller sample sizes were sufficient to calculate fairly robust approximations of the genetic indices of each of the three populations.

Table 3.2 Genetic diversity indices and sample sizes (n) of the Knysna, Swartvlei and Keurbooms populations, as well as individual sampling sites within the Knysna Estuary. Indices include number of haplotypes (H); haplotype diversity (h); nucleotide diversity (π), number of polymorphic (segregating) sites (S); mean number of pair-wise nucleotide differences (d); expansion coefficient (S/d).

Site	n	H	h	π	S	d	S/d
Knysna combined	90	13	0.78	0.00353	10	1.41	7.1
Knysna combined*	30	7.22±1.20	0.77±0.04	0.0032±0.0004	5.9±1.2	1.19±0.37	5.9±3.9
Knysna 1	30	9	0.73	0.00348	8	1.40	5.7
2	30	7	0.76	0.00299	5	1.20	4.2
3	30	8	0.84	0.00425	7	1.71	4.1
Swartvlei	30	3	0.48	0.00461	5	1.86	2.7
Knysna combined*	18	5.78±1.30	0.75±0.08	0.0032±0.0006	4.9±1.3	1.27±0.23	3.8±0.6
Swartvlei*	18	2.55±0.53	0.46±0.05	0.0045±0.0005	4.5±0.5	1.80±0.20	2.9±0.5
Keurbooms	18	8	0.75	0.00458	7	1.84	3.8

*Values represent mean (±SD) calculated from nine subsamples. Each subsample consisted of a number of sequences randomly chosen from the original sample using a random number generator (http://www.randomizer.org).

3.3.2 Intraspecific genealogy

A star-like pattern was identified in the haplotype network constructed using the programme TCS (Fig. 3.3), which indicates recent ancestral monomorphism followed by a population expansion (Slatkin & Hudson 1991). Using the criteria outlined in Crandall & Templeton (1993), haplotype 1 has been designated as the root of the network. It is the most abundant haplotype in the Knysna Estuary, and it has the most pivotal position in the network. Due to several equally parsimonious solutions, the relationship between haplotype 1, and particularly the most derived haplotypes, is ambiguous. The relationships between such haplotypes were resolved using the criterion suggested by Crandall & Templeton (1993): whenever a derived haplotype was linked to two older haplotypes by an equal number of mutational steps, the derived haplotype was connected to the one older haplotype that was present in the population at greater frequency than the other.

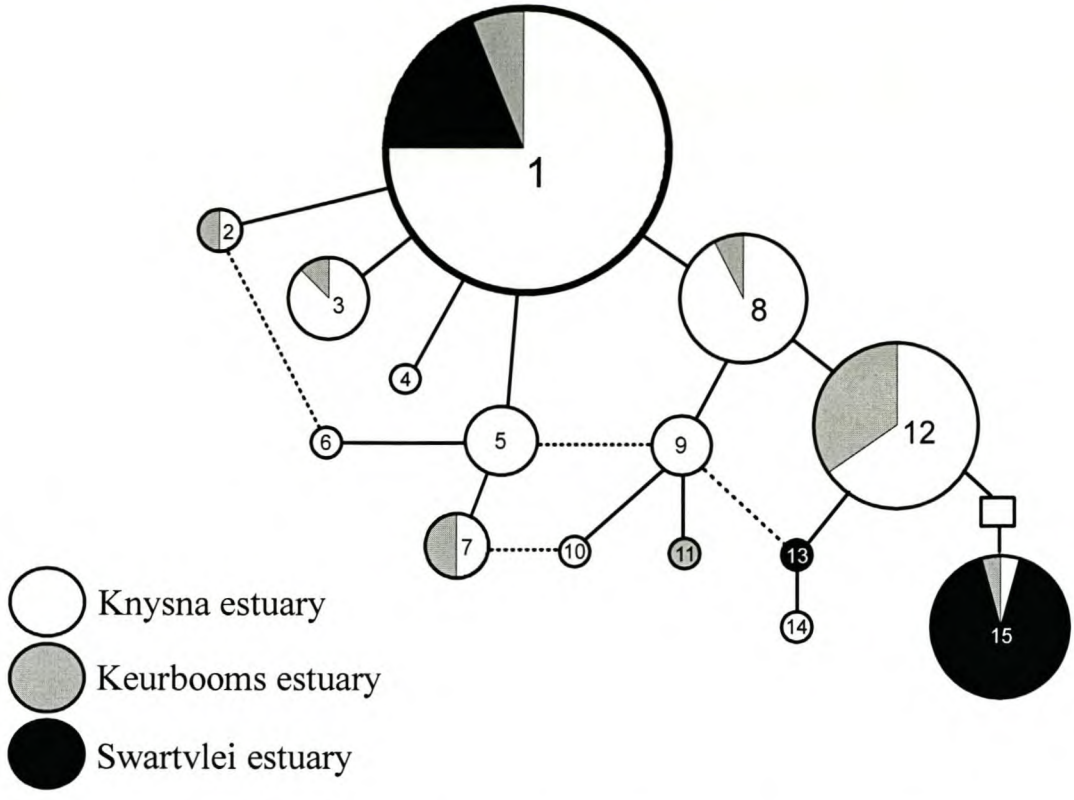

Fig. 3.3 Haplotype network of CR right domain haplotypes of *H. capensis*. Each haplotype is represented by a circle, the size of which indicates the frequency at which it was found. The square represents an internal node haplotype not present in the sample. Each line represents a single nucleotide substitution. Proportional representation of haplotypes in the different estuaries is indicated by subdivision of circles into up to three sections. All connections shown have a probability of ≥95% of being correct, but connections represented by solid lines are more highly supported than those represented by broken lines by virtue of criteria outlined in the text.

3.3.3 Population structure

No significant structure was found among the three sites within the Knysna Estuary using AMOVA (% variation among populations = 0 ; Φ_{ST} = 0; p = 0.8). Hence, the sequences from this population were combined in subsequent analyses. An AMOVA revealed significant structure between the three estuarine populations (% variation among populations = 29.49; Φ_{ST} = 0.331; P < 0.01), and pair-wise comparisons found significant structure among all three estuaries (Table 3.3).

Table 3.3 Pair-wise comparisons of genetic structure among the three estuaries; below diagonal: F_{ST} values; above diagonal: Φ_{ST} values. The level of significance is indicated by asterisks: *=0.05; **=0.01

	Knysna	Keurbooms	Swartvlei
Knysna		0.102**	0.432**
Keurbooms	0.075*		0.257**
Swartvlei	0.253**	0.343**	

3.3.4 Genetic bottlenecks

The large number of rare haplotypes in the Knysna population resulted in an L-shaped distribution of haplotype frequencies (Fig. 3.4), which suggests that this population has not experienced a recent genetic bottleneck (Luikart *et al.* 1998). In contrast, the presence of a single rare haplotype in the Swartvlei Estuary is indicative of a recent bottleneck or founder event in this population (Fig. 3.4b). A second plot of haplotypes found in the Knysna Estuary based on nine sets of 18 sequences each (Fig. 3.4c), shows that the L-shaped pattern is still recovered when using smaller sampling sizes. The Keurbooms population was also characterised by a high frequency of rare haplotypes (Fig. 3.4d).

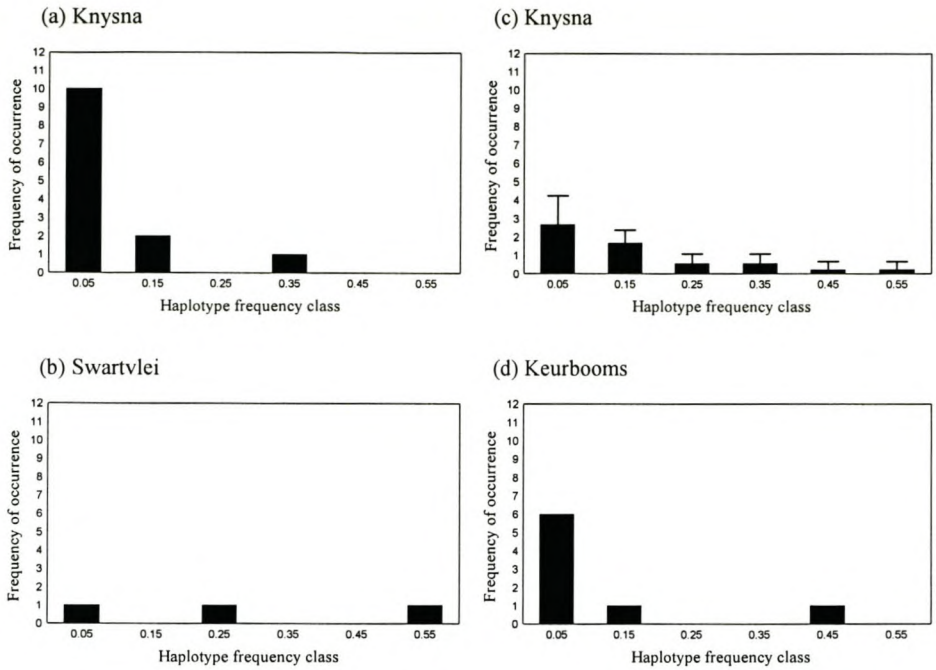

Fig. 3.4 CR right domain haplotype frequency distribution: a) Knysna population (90 individuals); b) Swartvlei population (30 individuals); c) Knysna population (18 individuals resampled nine times from original 90; bars indicate means, whiskers represent positive standard deviation); d) Keurbooms population (18 individuals).

Using the programme ALLELOCIDE, a reduction in population size to approximately 440 individuals resulted in the first rare haplotype/s being lost from the data set of Knysna haplotypes. The number of runs in which haplotypes were lost then rapidly increased as population size was reduced further (Fig. 3.5). Using a different initial population size of 30 000, 10 000, 5000 or 1000 individuals had no effect on the results.

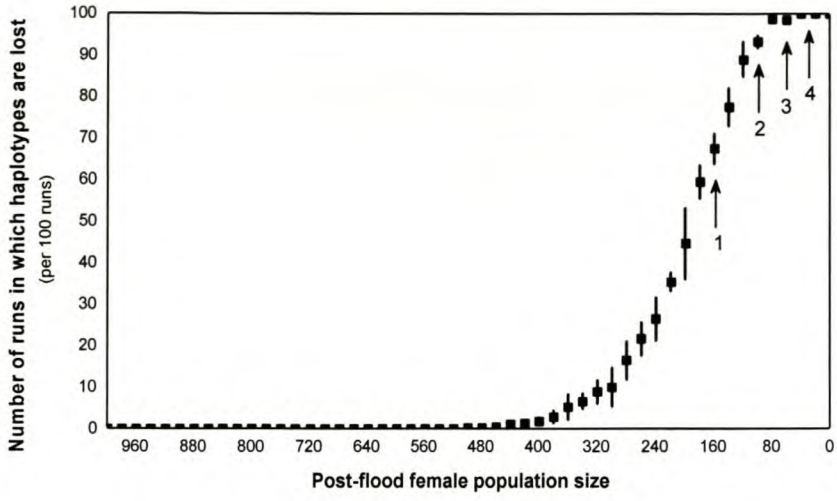

Fig. 3.5 Loss of haplotypes as a result of a catastrophic reduction in population size. Results from the Visual Basic simulation programme ALLELOCIDE were plotted; initial population size: 1000; increments: 20 individuals; number of replicates: 100. Values plotted are means (±SD) from ten repetitions. Numbers below some data points represent the mean number of haplotypes lost in 100 replicates of a particular run. See Appendix I, Tables A16 and A17 for data used to construct this diagramme.

3.3.5 Effective population size and population history

The population history of *H. capensis* was investigated using the Knysna population only. Apart from providing the largest data set and containing most of the haplotypes found, this population is most likely to be the most important from a conservation perspective (see Discussion). As the AMOVA results show a lack of subpopulation structure within the Knysna Estuary, the following tests were performed for a pooled data set of this population. A large number of short chains (1000 steps each) and long chains (10 000 steps each) was run for both methods in the programme FLUCTUATE, and sampling increments of 20 were used for both short and long chains. As results differed slightly depending on the number of chains run, the number of chains was increased from 15 to 40 short chains in increments of 5, and approximately half the number of long chains in each run. Θ and g are reported as the means from six runs. The θ value (± SD) calculated using the first method was 0.130 ± 0.021 with a simultaneously estimated exponential growth parameter (g) of 1399 ± 152. The results for the second method were as follows: θ (± SD) = 0.210 ± 0.124 and g (± SD) = 1999 ± 768. Θ values calculated using approaches based on pair-wise comparisons were θ_S = 0.0049 ± 0.0019 and θ_π = 0.0037 ± 0.0025.

The estimates of θ and g obtained using the first method were used to generate plots of historical population sizes over time (Fig. 3.6). Assuming that population growth was relatively constant, the approximate age of the Knysna population is defined as the point in time at which the historical N_f value is lower than 1% of the estimated present effective female population size. Calculated using the first method, this lies between 65 and 486 thousand years ago, whereas results using the second method placed the age of the population between 46 and 339 thousand years (not shown).

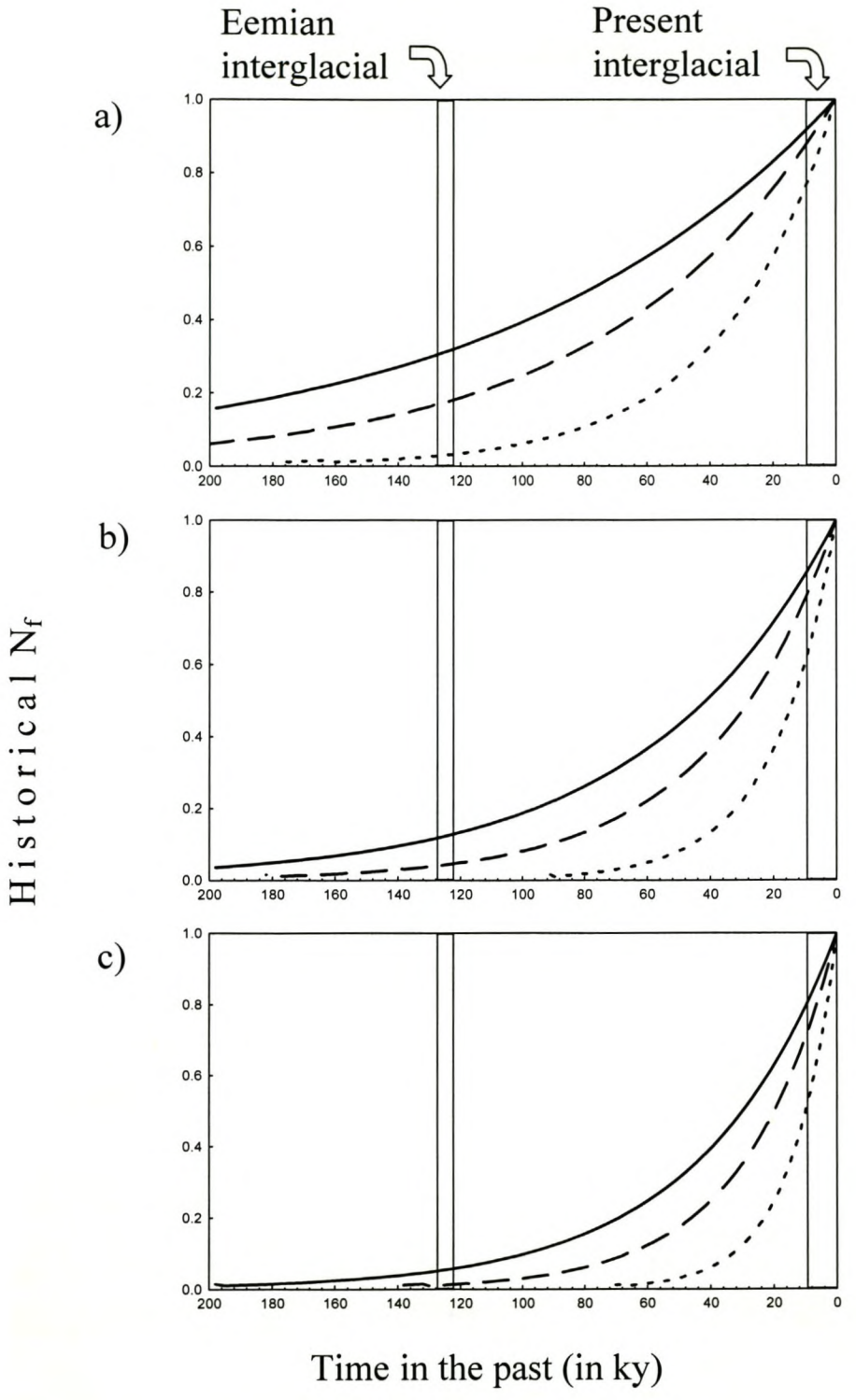

Fig. 3.6 Plots of historical effective female population sizes (N_f) at three putative CR right domain mutation rates: a) 2% my^{-1}; b) 3.6% my^{-1}; c) 5% my^{-1}. Three possible generation times are plotted: 1 year (dotted line), 2 years (broken line) and 3 years (solid line). Historical N_f values are expressed as proportions of present N_f.

3.3.6 Raggedness statistic and departure from selective neutrality

The recent population expansion suggested by the star-like phylogeny and the large growth rates was confirmed using the raggedness statistic, which did not reject the sudden expansion model ($r = 0.05$, $p = 0.41$). A negative and significant Fu's F_s test for the combined dataset from the Knysna population further supports the evidence for a demographic expansion of this population ($F_s = -7.34$, $p < 0.01$). Fu and Li's D* and F* for the Knysna population, on the other hand, were not significant ($D^* = -2.08$, $P > 0.05$, $F^* = -1.91$, $p > 0.05$). This excludes the possibility that background selection is responsible for the departure from neutrality, and justifies the application of the coalescent-based ML algorithm used in FLUCTUATE to investigate population history.

3.3.7 Comparative population study

Haplotypic diversity and nucleotide diversity were highest in the *H. kuda* population, lower in the *H. capensis* population residing in the Knysna Estuary, and lowest in the *H. fuscus* population. All tests for selective neutrality had a negative sign (i.e. indicating population expansion or background selection), but all three statistics calculated were significantly different from zero in the *H. fuscus* population only. In the case of *H. kuda*, only Fu's F_s was significant. The fact that no significant F_s statistic was found in the *H. capensis* population (contrary to the results obtained for the complete data set of *H. capensis* haplotypes) suggests that this is merely an artefact of small sample size. As Fu's F_s is more negative in the *H. fuscus* population than Fu and Li's tests, background selection can possibly be excluded as the reason for a departure from selective neutrality, and population growth is the more likely explanation (Fu 1997). The same is true of the *H. kuda* population and, to a lesser extent, the *H. capensis* population.

Table 3.4 Diversity indices, results of tests for selective neutrality (Fu & Li's D*, Fu & Lis's F* and Fu's Fs), and population parameters θ (= effective female population size × mutation rate) of three populations of kudaoid seahorses and Japanese sticklebacks (*Pungitius pungitius* and *P. sinensis*). Sample sizes were n = 35 for the three seahorse populations and n = 28 for the sticklebacks. H = number of haplotypes present in sample; h = haplotype diversity; π = nucleotide diversity; θ_S = θ based on the number of segregating sites; θ_π = θ based on nucleotide diversity. P-values below 0.05 are shown in italics.

Population	H	h (±S.D.)	π (±S.D.)	Fu and Li's D* (p-value)	Fu and Li's F* (p-value)	Fu's Fs (p-value)	θ_S (±S.D.)	θ_π (±S.D.)
H. capensis	6	0.571±0.084	0.00211±0.0017	-1.58 (>0.10)	-1.64 (>0.10)	-2.37 (0.06)	1.21±0.63	0.80±0.66
H. kuda	7	0.676±0.066	0.00341±0.0024	-1.63 (>0.10)	-1.79 (>0.10)	-3.02 (***0.02***)	1.46±0.71	1.28±0.92
H. fuscus	8	0.489±0.102	0.00175±0.0015	-2.63 (***<0.05***)	-2.83 (***<0.05***)	-5.57 (***<0.01***)	1.94±0.87	0.66±0.58
Pungitius sp.	28	1.000±0.010	0.04913±0.0200	-1.07 (>1.00)	-1.33 (>1.00)	-21.66 (***<0.01***)	14.65±4.90	19.06±9.69

Further support for population expansion in all three populations is evident in the star-like shapes of the haplotype networks constructed using TCS and the numerical dominance of the root haplotype in all three cases (Fig. 3.7), the non-significant SSD statistics in Table 3.5 (indicating no significant departure from the model of population expansion), and the unimodal shape of the distribution curves of observed pair-wise nucleotide differences in Fig. 3.8. Trends regarding θ values among the three populations differed considerably depending on whether they were calculated using nucleotide diversity θ_S or the number of segregating sites θ_π. This discrepancy may be an artefact of the high growth rate evident in all three populations of kudaoid seahorses, and the results are thus highly tentative. The τ values calculated in ARLEQUIN (Table 3.5) indicate that the *H. kuda* population is the oldest, the *H. capensis* population is of intermediate age, and the *H. fuscus* population is the youngest.

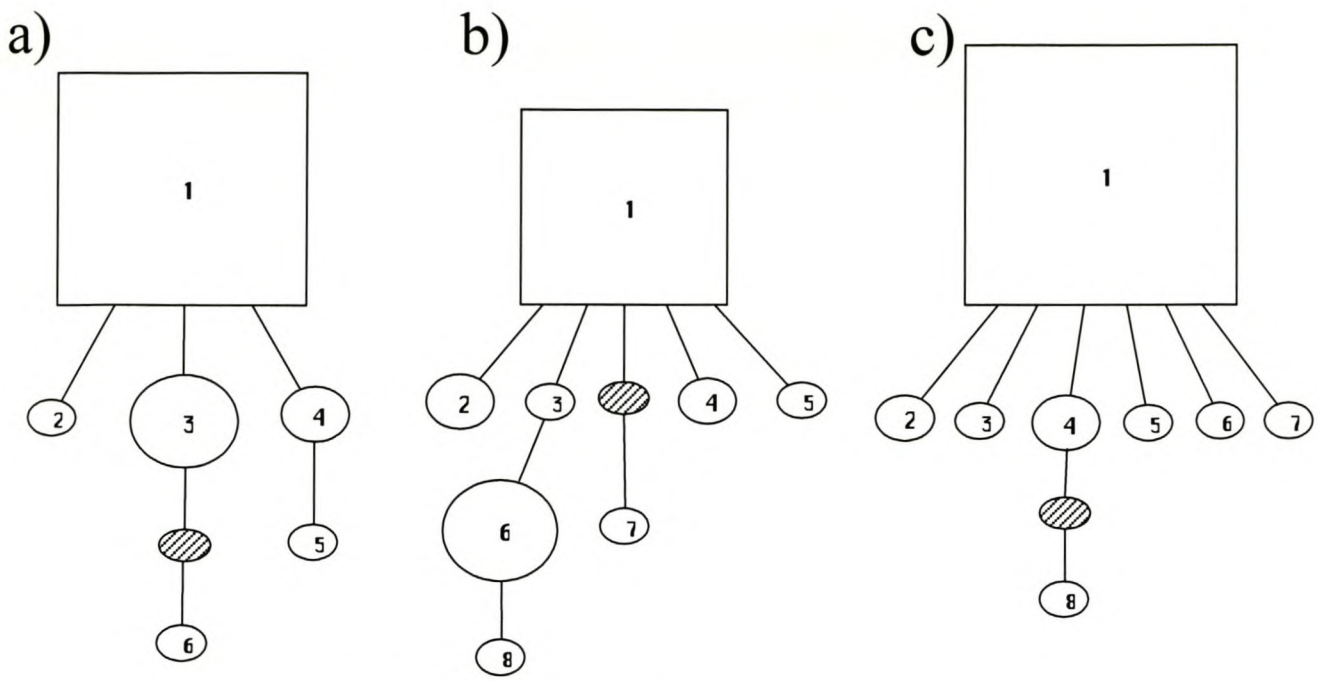

Fig. 3.7 TCS haplotype trees of three populations of kudaoid seahorses (in each case, n = 35): a) *Hippocampus capensis*, Knysna Estuary, South Africa; b) *H. kuda*, Tayabas Bay, Philippines; c) *H. fuscus*, Red Sea near Suez Canal. Squares represent haplotypes identified as basal, whereas ovals represent derived haplotypes; each line represents a single nucleotide substitution; shaded ovals represent interior node haplotypes not present in the samples.

Table 3.5 Estimated parameters for the sudden expansion model: θ_0 = population size before expansion; θ_1 = population size after expansion. τ = time passed since expansion. SSD: statistic testing for significant departure of observed data from sudden expansion model.

	θ_0	θ_1	τ	SSD	p
H. capensis	0.00	830.94	0.828	0.005	0.28
H. kuda	0.00	2.54	2.191	0.029	0.24
H. fuscus	0.00	426.17	0.666	0.005	0.27
Pungitius sp.	5.42	132.81	15.411	0.002	0.89

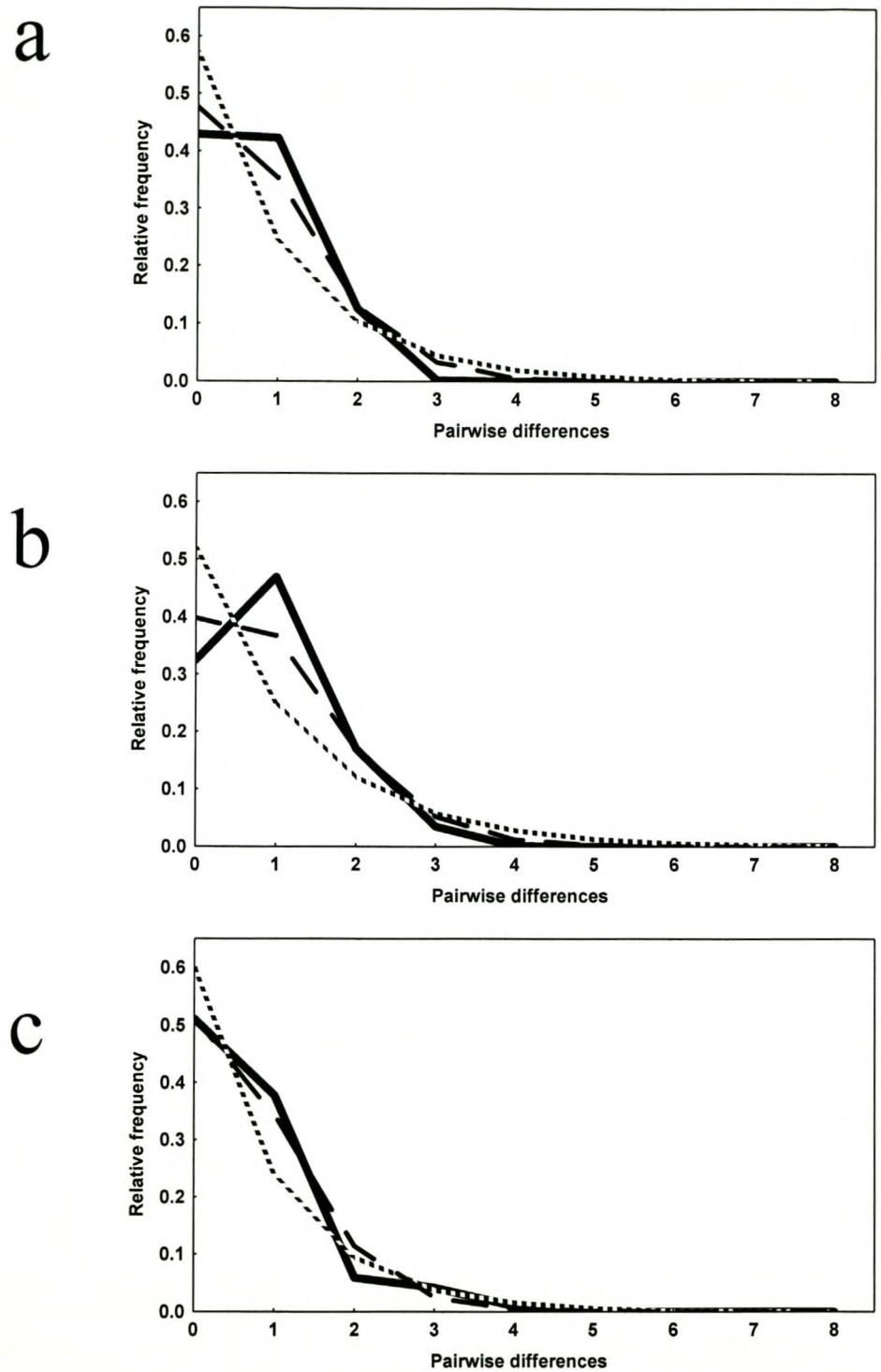

Fig. 3.8 Mismatch distribution plots of pair-wise nucleotide differences of the three seahorse populations compared with the frequency distributions expected under models of stationary and expanding populations. Solid line: observed distribution; broken line: distribution under a model of growth/decline; dotted line: distribution under a model of stasis. Populations: a) *Hippocampus capensis*; b) *H. kuda* (Philippines); c) *H. fuscus* (Red Sea).

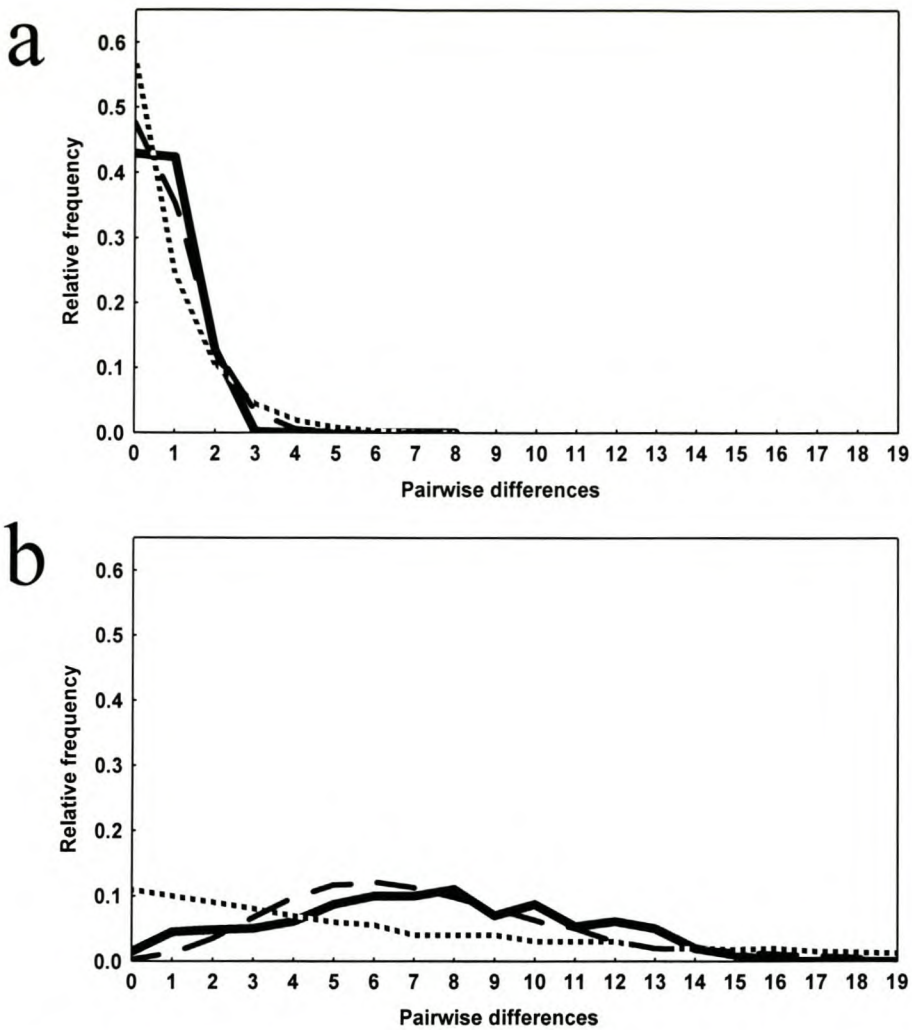

Fig. 3.9 Comparison of the mismatch distributions of (a) the Knysna Estuary population of *H. capensis* and (b) Japanese sticklebacks (*Pungitius pungitius* and *P. sinensis*) collected from estuaries and bays all over Japan (Takahashi & Goto 2001). Note that sample sizes differ: *H. capensis*: n=35; *P. pungitius/sinensis*: n = 28. Solid line: observed distribution; broken line: distribution under a model of growth/decline; dotted line: distribution under a model of stasis.

Mismatch distributions of the haplotypes of Japanese sticklebacks differ dramatically from those of the seahorse populations (Fig. 3.9). Although a signature of population expansion is still evident (the observed distribution is not significantly different from the distribution under a model of population expansion, Table 3.5) the wave has shifted to the right, is considerably flattened, and is not perfectly unimodal, which indicates that a considerable amount of time has passed since the expansion event. Note, however, that Japanese sticklebacks were sampled over a comparatively large area, and the samples are thus a combination of several populations.

3.4 Discussion

3.4.1 Genetic structure and gene flow

The samples from the Keurbooms and Swartvlei Estuaries each contained a single private haplotype (haplotypes 11 and 13), whereas the Knysna Estuary contained a total of six (Table 3.1). The remaining seven haplotypes were shared among at least two estuaries. Although it is possible that due to sampling efforts not all of the unique haplotypes present in the three populations were represented in the data set, it is interesting to note that clear haplotypic frequency differences exist among the estuaries. The significant structure (supported by both F_{ST} and Φ_{ST} values, Table 3.3) found among the three estuaries suggests that each of the populations constitutes a distinct management unit *sensu* Moritz (1994). However, the conclusion that the three populations may be evolving relatively independently under different stochastic processes, and that the population structure in the Swartvlei population differs considerably because this population experienced a recent population bottleneck, are at best tentative, because they are based on low sample sizes and a single neutral marker only. Although the right domain of the CR provides good resolution at the demographic level, Luikart *et al.* (1998) recommended that analyses for genetic bottlenecks should involve 5-20 independent loci. The statistics indicating a lack of population structure within the Knysna Estuary were calculated using a total sample size of 90, suggesting that this conclusion is fairly robust. Lack of genetic structure is also supported by the fact that site 3 was inhabited exclusively by juveniles and young, non-breeding adults, whereas the two upper sites were inhabited by adults. There may thus be considerable movement of seahorses within the estuary, and juveniles and adults may live spatially separated in preferred habitat types. The most apparent explanation for this pattern is the nature of the vegetation in the different areas. Site 3 is characterised by very dense patches of seagrass, which seem to be an ideal habitat for smaller seahorses, whereas larger seahorses would find it difficult to move within them. The more open vegetation in the estuary's upper reaches, on the other hand, may be more suitable for large seahorses.

3.4.2 Evolutionary history of *H. capensis*

Although it is difficult to date the exact time when *H capensis* diverged from its marine ancestors because the parameters estimated with FLUCTUATE are imprecise due to an uncertain genealogy and biased because they are based on only one locus (Kuhner *et al.* 1998), the results of this study can be used as a rough indication to date such an event. Using two approaches to determine haplotype relationships, three different mutation rates and three different generation times, the age of the Knysna population has been estimated to between 46 and 486 thousand years (late Pleistocene). Environmental conditions along the coast during the Pleistocene differed considerably from present day conditions, and several factors suggest that it was unlikely that tropical or sub-tropical marine seahorses that may have given rise to *H. capensis* were able to reach the Knysna Estuary during this time. Presently, the coastal waters of the east coast are dominated by the warm Agulhas Current, which aids in the southwards dispersal of tropical marine organisms (Heydorn 1978; Blaber 1981; Turpie *et al.* 2000). This dispersal occurs mainly during the austral summer, when sea surface temperatures are in the region of 24 ^{o}C and the Agulhas Current is more defined than during winter (Heydorn 1978), suggesting that they are facilitated both by water temperature and current strength. During the Pleistocene, sea-surface temperatures along South Africa's east coast were up to 4 ^{o}C cooler and the Agulhas Current was considerably weaker (Lindesay 1998). Moreover, although the Agulhas Current is deflected away from the coast in the Eastern Cape region because the continental shelf widens, transport of tropical species towards the south coast is possible as eddies may transport water from the Agulhas Current towards the shoreline (Branch & Branch 1992). In contrast, during much of the Pleistocene, the Agulhas Current was deflected eastwards just south of Madagascar (Lindesay 1998). However, a founder event is conceivable during the Eemian interglacial period (127-122 thousand years ago), a short warm phase within the Pleistocene during which sea surface temperatures along the south coast were higher than present-day temperatures (south west coast: +3.8^{o}C; south east coast: +0.9^{o}C) (Crowley & North 1991). The fossil record indicates that much of the south coast's Eemian coastal marine fauna included species presently restricted to warmer Indian Ocean currents off the Eastern Cape Province, KwaZulu Natal Province, or even Mozambique (Martin 1962; Davies 1971). This suggests that the tropical or sub-tropical marine seahorses that gave rise to the Knysna population may also have been considerably more abundant in the region during the Eemian interglacial than they are today. On the basis of the results of the first method to calculate θ and g, a generation time of 1-2 years

and a mutation rate of 3.6% my^{-1}or slightly higher, the possibility of a founder event during the Eemian interglacial is well supported (Fig. 3.6b).

As there is no suitable habitat along the south coast of South Africa, and *H. capensis* is highly specialised to survive in estuaries, physiological adaptation to the estuarine environment must have taken place in one of the three estuaries. If one assumes that a founder event took place during the Eemian interglacial, then environmental and biotic conditions characterising the three estuaries must have been similar to present conditions, as sea surface temperatures were similar, and the sea level was only 4 m higher than it is today (Gribnitz & Kent 1989). The Knysna Estuary, with its highest haplotypic diversity, seems the most likely location for this event to have taken place. Firstly, the marine-dominated lower reaches of this estuary provide an optimal transition zone for the adaptation of marine species to the estuarine environment. Secondly, the large size of the estuary reduces the detrimental effects of freshwater floods, and can also support a larger population of seahorses. This improves the prospect for an estuarine population to establish itself and ensures its long-term survival. Thirdly, because of its permanently open mouth, the estuary is readily accessible to seahorses migrating along the coastline. The notion that the Knysna population is the oldest is also supported by the fact that a rare estuarine goby, *Pandaka silvana*, is endemic to this estuary (Penrith & Penrith 1972). In contrast, no endemics have been found in the other two estuaries. The evolutionary history of *P. silvana* may be similar to that of *H. capensis*, because like the majority of seahorses, all other species in the genus *Pandaka* are exclusively tropical and marine.

Convergence of the Knysna, Keurbooms and Swartvlei rivers lower on the continental shelf during the colder period between the Eemian and present interglacial, and possibly a vicariance event brought about by a subsequent rise in sea levels, may eventually have resulted in three extant populations. The lower genetic diversity of the population residing in the Swartvlei Estuary can be explained by subsequent loss of haplotypes due to genetic bottlenecks. However, this scenario does not explain why the equally small population in the Keurbooms Estuary has a significantly higher genetic diversity. An alternative explanation for the observed haplotype pattern is small scale migration of seahorses between the estuaries. Assuming that the source population resides in the Knysna Estuary, it is more likely that seahorses that have been flushed out of this system end up in the Keurbooms Estuary rather than the Swartvlei Estuary, because the prevailing coastal current flows eastwards (Branch & Branch 1992). As mentioned previously, the fact that site 3 in the Knysna Estuary was inhabited exclusively by juveniles and subadults, and that no population sub-structure was found within this estuary,

suggests that juveniles and adults may be spatially separated. Dispersal within the estuary is probably accomplished passively through tidal currents, and it is likely that some juvenile seahorses may get flushed out to the sea before finding suitable habitat. This may provide the two smaller estuaries with an infrequent but continuous input of new colonists.

Although the Keurbooms population is characterised by high genetic diversity, the different ratio of haplotypes to that in the Knysna population (and particularly the low abundance of the common haplotype 1), suggests that this population may nevertheless undergo severe fluctuations in population size. The estuary experiences floods of substantial magnitude (Duvenage & Morant 1984), which are likely to be more detrimental to the fauna than floods of similar magnitude in the Knysna Estuary, on account of the Keurbooms Estuary's smaller size. Lockyear (pers. comm.) reported a mass mortality in the Keurbooms and Swartvlei estuaries after a freshwater flood in 2002. No seahorses were subsequently been found in the Keurbooms Estuary, and population densities in the Swartvlei Estuary had decreased considerably. In contrast, no mortalities were recorded in the Knysna Estuary during that period, and population densities have remained unchanged. For that reason, the Keurbooms population may be dependent on gene flow from the Knysna population in order to maintain its high genetic diversity. Lastly, it cannot be ruled out that the population structure detected among the three populations has been facilitated by humans. This is particularly plausible in the case of the Swartvlei population: although anecdotal evidence suggests that the species was present in the Swartvlei Estuary for several decades, this estuary was not included in the list of estuaries inhabited by *H. capensis* in 1986 (Dawson 1986).

3.4.3 Population size and conservation implications

The high number of rare haplotypes in the Knysna assemblage suggests that this population is sufficiently large to tolerate floods without being at risk of undergoing a genetic bottleneck. This notion is supported by survey work. The census population size of adult seahorses has been estimated to be approximately 60 000 (Lockyear & Teske, unpubl. data). The population has an even sex ratio and >90% of the males were found to be pregnant during the breeding season (Lockyear & Teske, pers. obs.). This suggests that the effective female population size (N_f) may be close to 30 000. Behavioural work on the Australian seahorse *H. whitei* has shown that these form long-term faithful pair bonds (Vincent & Sadler 1995). Although it is not known whether this is also the case in *H. capensis*, the fact that male seahorses do not collect

eggs from more than one female at a time (Vincent 1995), suggests that there is no justification for incorporating reproductive skew into calculations of N_f.

Methods to calculate effective population sizes based on genetic data yield higher values than the survey results (in all cases $\mu = 0.036\ my^{-1}$ and generation time = 1.5 years were used to calculate N_f from θ, and standard deviations are given). Approaches based on pair-wise comparisons of the number of segregating sites or the number of nucleotide differences, arrived at values approximately twice to three times as high as N_f calculated from census estimates (N_f [from θ_S] = 90 740 ± 35 185; Nf [from θ_π]= 68 519 ± 46 296). Even higher values of N_f were calculated using FLUCTUATE. When jointly estimating θ and g, effective female population sizes of $2.4 \times 10^6 \pm 0.4 \times 10^6$ (first method) and $3.8 \times 10^6 \pm 1.2 \times 10^6$ (second method) were determined. The considerable differences between the estimates of N_f calculated using different methods suggests that these should be interpreted cautiously because of uncertainties about input variables and the use of a single neutral marker. Given the caveat that the survey prediction is accurate, the fact that N_f values calculated using genetic methods were higher may be an indication that the population size of *H. capensis* in the Knysna Estuary has decreased considerably during the species' short evolutionary history. However, even a drastic decline in population numbers is unlikely to be detected using the genetic methods employed in this paper, because it is unlikely to result in a loss of rare haplotypes. Unless severe and for a prolonged period of time (which is not applicable in the case of a freshwater flood), it is also unlikely to have an effect on the proportional abundance of each haplotype within the population. The results of the simulation programme ALLELOCIDE illustrate this point. A one per cent chance of losing the first rare haplotype in the Knysna population (i.e. haplotype/s lost in one out of 100 runs, Fig. 3.5) exists when population numbers have been reduced to 400 individuals, and the number of runs in which haplotypes are lost then increases rapidly as population size is reduced further. If we assume a present effective female population size of 30 000 indivduals, then the removal of up to 99% of adult females would theoretically not result in the loss of any haplotypes. Although it could be argued that an Allee effect (Allee 1938) resulting from the low population density could result in the population size at which haplotypes are lost being substantially greater than estimated, the fact that adult seahorses were often found in aggregations (see Introduction) suggests that an Allee effect is unlikely to affect the entire remnant population: individuals that have survived a freshwater flood may be confined to some densely populated regions of the estuary, and potential mating partners are thus much more likely to encounter one another than they would be if their distribution was random.

The difference between theoretical and observed effective female population sizes may be an indication that anthropogenic pressures during the past decades could already have had a significant negative impact on the Knysna population. The present rate at which construction developments and other human activities are increasing along the estuary is all the more alarming in the light of these findings. The resulting habitat degradation may make recovery of the population after a naturally-occurring disaster such as a freshwater flood increasingly difficult. Whichever scenario resulted in the observed population structure in the two smaller estuaries (i.e. vicariance followed by genetic bottlenecks in the Swartvlei Estuary versus migration resulting in recent founder events, or a combination of the two), the Knysna Estuary has the greatest potential to ensure long-term survival of this species on account of the large size of its population and the less detrimental effect of freshwater floods. Hence, it is imperative that conservation efforts for this population be prioritised. Although the three populations of *H. capensis* constitute individual management units, this conclusion was reached mainly on the basis of differences in haplotype frequencies among the populations rather than a large proportion of private haplotypes. Due to the absence of distinct monophyletic clades of haplotypes unique to individual populations and the generally good support for the migration hypothesis, there is little reason to discourage the translocation of seahorses among the different estuaries, should this become necessary. The lack of population structure within the Knysna Estuary suggests that areas temporarily affected by habitat degradation due to construction developments in the vicinity, are likely to become repopulated on their own once they have reverted to their original state, provided this is still possible. Translocations in the Knysna Estuary are feasible in cases where the degradation of formerly populated areas is expected or documented, in which case seahorses could be collected and released in more pristine parts of the estuary.

3.4.4 Assessment of the conservation status of *H. capensis*

Vane-Wright *et al.* (1991) suggested that the more ancient and evolutionarily unique a lineage is, the greater its contribution to the world's extant evolutionary diversity. Hence, distinct lineages should be of greater conservation worthiness than lineages that have diverged recently and that are thus closely associated with other such lineages. Avise (1994) questioned this reasoning and suggested that because each biological species is genetically distinct from other species (and "every species alive today traces back through an unbroken chain of ancestry

over thousands of millions of years, irrespective of how many speciation events have intervened along its phylogenetic journey"), there is essentially no justification to prefer some species to others. Moreover, although a general trend exists to measure the conservation status of a taxonomic unit by the amount of sequence divergence from other such units, a population that is indistinguishable from similar populations on the basis of genetic data may nevertheless be characterised by recent adaptive changes that warrant a high conservation status (Greenberg *et al.* 1998).

Based on these considerations, it should not be of primary concern to determine how phylogenetically distinct *H. capensis* is from other kudaoid seahorse species in order to justify its high conservation status, but merely to determine whether it is truly genetically isolated from these taxa. The fact that the Knysna seahorse seems to be a true estuarine species (see chapter 3) and may thus be physiologically isolated from closely related marine seahorses (to which the estuarine habitat is likely to be inhospitable), may be sufficient to justify the species' high conservation status. However, in the absence of survival and cross-breeding experiments, both the reciprocal monophyly identified in the case of mitochondrial markers, and the absence of haplotypes that are shared with marine seahorses, gives support to the phylogenetic distinctness and thus the present conservation status of the Knysna seahorse.

3.4.5 Comparative population study

It is presently unknown whether the three kudaoid seahorse populations that were compared in this study may be of similar size. The small size of the Knysna Estuary suggests that this population is considerably smaller than the two marine populations, but based on sampling success per time unit, high densities of seahorses have been found in this system as compared to Inhaca Island, Mozambique (Teske, pers. obs.) and Fiji (Hamilton, pers. comm.). Due to the high productivity of estuaries and the large quantities of suitable prey items (see Introduction), it is likely that the Knysna Estuary is able to support a substantially larger number of seahorses than its comparatively small size may suggest. Although the values of θ calculated in ARLEQUIN are tentative because of substantial differences depending on the method used and large standard deviations, they may be suitable to theoretically compare the relative effective sizes of the populations. If one assumes that the four populations are characterised by equal generation times and CR right domain mutation rates, then the following N_f-values can be calculated (as previously, a generation time of 1.5 years and a mutation rate of 3.6% per million

years was used; N_f values are reported as ranges, which were calculated as: smaller value of θ – standard deviation and greater value of θ + standard deviation): *H. capensis*: 6 833 - 89 907; *H. kuda*: 17 592 - 106 037; *H. fuscus*: 3 907 - 137 296; Japanese sticklebacks: 465 351 - 1 372 185. These results indicate that the three seahorse populations are approximately equal in size, whereas the stickleback lineage may be up to ten times as large.

However, it must be borne in mind that the genetic parameters obtained for the stickleback lineage and those estimated for the marine seahorse populations cannot be directly compared, as the latter may represent only a fraction of regional diversity (e.g. the population from Tayabas Bay may not be representative of an entire regional lineage inhabiting the Philippines). The fact that the CR right domain haplotype of a single specimen of *H. kuda* collected in Taiwan was genetically very similar to the haplotypes collected in the Philippines suggests that these samples may be representative of a north-western Pacific regional population, as gene flow among different populations is likely to be high due to the absence of geographical barriers and an abundance of suitable habitat. However, additional data from other populations will be required to address this question, which is beyond the scope of this study. Irrespective of these considerations, the fact that population genetic parameters in the Knysna seahorse are approximately the same as those measured for other, non-endangered seahorse populations, is encouraging regarding the conservation of this species: despite a small area of occupancy and complete lack of genetic input from seahorse populations other than the ones residing in the small Keurbooms and Swartvlei Estuaries, there is no indication of genetic impoverishment.

If one assumes that each of the haplotypes identified in the three populations as MRCA is the founding haplotype that gave rise to the populations, then the values of τ estimated should be proportional to the ages of the populations. Genealogical methods of calculating population ages and effective sizes have been shown to be more efficient than pair-wise methods (Felsenstein 1992a,b; Fu 1994a,b; Kuhner *et al.* 1995), but they require strictly bifurcating genealogies. Approximations of these parameters can be obtained by transforming multifurcating genealogies into bifurcating genealogies by removing some rare haplotypes, a procedure which was followed in case of the complete *H. capensis* data-base. However, this is only possible in cases where large numbers of derived haplotypes exist, and/or when populations that were initially characterised by rapid expansion are approaching equilibrium and the number of haplotypes closely related to the MRCA is decreasing as a result of lineage extinctions caused by genetic drift. In all three seahorse populations, the majority of derived haplotypes differ from the MRCA by a single nucleotide substitution, and transforming starlike phylogenies into bifurcating phylogenies is not feasible. Several other problems may have

influenced the quality of the results. Firstly, rather than undergoing a single population expansion following a founder event that gave rise to the three seahorse populations (as the sudden expansion model assumes), it is likely that initial population growth was followed by fluctuations in population size. This seems particularly likely in case of the Knysna population, where freshwater floods are known causes of mortalities of seahorses and other estuarine species. However, Rogers & Harpending (1992) showed that the theoretical mismatch distribution is remarkably insensitive to violations of the model of sudden expansion. When an initial expansion is followed by later expansions or minor bottlenecks of population size, the theoretical mismatch distribution is affected only slightly. The empirical mismatch distribution is also robust when the initial population is small, and is not subdivided. Another problem that may be encountered with the data used is the fact that Rogers & Harpending's method incorporates Kimura's (1971) infinite sites model, which assumes that sites do not mutate more than once. The haplotype network constructed from 402 bp sequences of 138 *H. capensis* haplotypes contained several alternative connections among haplotypes, indicating the presence of homoplasies.

Effective female population sizes and population ages are only rough estimates, as calculations were based on pair-wise comparisons, results are based on a single molecular marker only, and violations of the sudden expansion model cannot be excluded. However, as Rogers & Harpending's method has been shown to be relatively insensitive to such violations, comparisons between the three populations may thus nevertheless be quite informative. Results for each population should, however, be considered relative to the other two populations rather than in absolute terms.

The different values of τ calculated suggest that the *H. kuda* population from Tayabas Bay is the oldest, the Knysna population of intermediate age, and the Red Sea population of *H. fuscus* the youngest. Using a large data-base for the Knysna seahorse, it was established that the most likely age of the Knysna population is approximately 120 000 years. A population expansion during the Eemian interglacial is also likely in the case of the *H. kuda* population, although the slightly higher value of τ calculated for this population may be an indication that its age precedes the Eemian interglacial. As temperatures in tropics were usually no more than 3 ^{o}C lower during the last ice age than they are today (Bard *et al.* 1998), environmental conditions in the Philippines may have been favourable for seahorses throughout the last glacial and interglacial phases. The younger age of the *H. fuscus* population can be explained by the fact that the Red Sea was isolated from the Indian Ocean during the last ice age and during that time was characterised by cool water and high salinity (Por 1978). Such conditions are unfavourable

for seahorses, most of which are restricted to tropical and sub-tropical regions (Lourie *et al.* 1999). It thus seems appropriate to place the age of the Red Sea population at the beginning of the present interglacial, approximately 10 000 years ago.

3.4.6 Is 'founder-effect speciation' common in seahorses?

Mayr (1954; 1963) formulated a speciation model according to which the probability of speciation is enhanced when a few migrant individuals colonising a new habitat start a new population. According to this model, a genetic structuring of the population is likely to ensue as the population adapts to its new habitat under the conditions of genetic depauperation caused by the founder event. The concept of 'founder-effect speciation' has subsequently been refined by several other authors (Carson 1971, 1975; Templeton 1980; Carson & Templeton 1984). Although the model has been criticised (Lande 1980; Barton & Charlesworth 1984; Rice & Hostert 1993; Coyne 1994) and it is highly disputed whether laboratory experiments have succeeded in corroborating it (Ringo *et al.* 1985; Moya *et al.* 1995; Templeton 1999), it nevertheless remains possible that new species arise quickly from populations established by a small number of founders in remote and isolated habitats (Moya *et al.* 1995).

Templeton (1980; 1981) expressed founder-effect speciation in genetic terms, which he termed 'genetic transilience'. A population that develops after a founder event usually differs considerably in genetic composition from its ancestral population, because the genetic bottleneck can lead to an accumulation of inbreeding and the induction of gametic disequilibrium. This then causes alleles to be selected more for their homozygous fitness effects (selective bottleneck). This shift in "genetic environment" thus directly leads to the speciation event. In this way, genetic transilience can lead directly to changes in morphology, physiology, life history and development. Speciation by genetic transilience is particularly attractive to explain speciation of the estuarine Knysna seahorse, a species living in a habitat inhospitable to its marine sister taxa because of severe fluctuations in salinity and temperature. Carson (1975) suggested that the number of loci affected by founder-effect speciation may be relatively small, as it does not involve alleles that are not affected by selection pressure. If the speciation event is relatively recent, no appreciable differences may thus be detected between two sister species at the CR level. In the case of *H. capensis*, it may have been sufficient if only genes involved in regulating cell volume in response to changes in external salinity had been affected by genetic transilience.

The feasibility of an alternative speciation scenario, sympatric speciation, needs to be explored. As the lower reaches of the Knysna Estuary are bay-like or lagoonal and provide habitat for many marine species, it is possible that a colony of marine seahorses established itself over time. Some of these seahorses then gradually adapted to tolerate fluctuations in salinity, until they had developed into a fully estuarine species that was physiologically distinct from the ancestral species, and was able to utilise both the marine and estuarine portions of the system. *H. capensis* and its marine ancestor may have co-existed in the estuary, but as the latter was not able to tolerate the rapid decrease in salinity associated with occasional freshwater floods, it eventually became extinct. Although this scenario seems attractive because the Knysna Estuary is unusual in southern Africa because it is marine- rather than river-dominated, flooding does occur approximately every 10 years, and physiological adaptations to rapid changes in salinity would have to take place quickly for a newly established population to survive in this type of habitat. It was established that the most likely period for a founder event to take place in the Knysna Estuary was the Eemian interglacial, approximately 120 000 years ago. Although water surface temperatures were higher than they are today (which would favour subtropical marine seahorses), the period was possibly also moister (Lindesay 1998), suggesting that freshwater floods took place more frequently than they do today, in which case speciation must have been rapid rather than gradual.

According to Templeton (1980) several attributes of a founder population are likely to increase the possibility of genetic transilience occurring. These include, among others, a large number of offspring and the potential for rapid population growth, low initial density, an initially subdivided population structure, overlapping generations, assortative mating, sexual selection on the mate recognition system, and imprinted or partially learned sexual behaviour. Also, a founder population originating from a large, panmictic population, is more likely to change than one from closely related individuals, because in a population that lacks genetic variability, there are few alternative alleles on which selection can act. Many of these attributes are applicable to the Knysna seahorse, and the founder event giving rise to this species may have occurred as follows: it is unlikely that several adult seahorses arrived at the estuary simultaneously and even if they did, it is unlikely that they would have been able to locate each other in the large habitat. However, if mating occurred in the source population, a single pregnant male has the potential to establish a localised population. Batches of up to 120 offspring have been recorded (Whitfield 1995). Assuming that the vegetation in the mouth area has not changed appreciably since speciation, the patchy nature of seagrass beds initially ensured that the offspring did not disperse over large distances. This, and the fact that seahorse

populations overlap, resulted in high levels of homozygosity. Establishment of the seahorse population may be explained following either of two hypotheses. In the first scenario, several individuals established sub-populations in different areas within the estuary. The initially subdivided nature of the Knysna population ensured that inbreeding levels were high within each sub-population, leading to an increase in homozygosity within each, and resulting in the fixation of alleles that in the ancestral population could not be selected for because of the high heterozygosity levels. The presence of several sub-populations, on the other hand, ensured that overall levels of genetic variablity within the Knysna population were not necessarily much below those in the ancestral population, allowing for the selection of alleles that resulted in increased fitness in the new habitat. An alternative scenario is the establishment of the Knysna population by the offspring of a single male, whose haplotype may have been identical or very similar to the one identified as MRCA. Such colonisation attempts may have taken place several times in the past, but failed because the alleles exposed by the selective bottleneck did not improve the colonisers' fitness. However, high genetic variability in the ancestral population ultimately ensured successful colonisation of the estuary. Colonisation due to a single founding individual is supported by the presence of a single haplotype displaying a star-like phylogeny, as well as the low θ_0 (0.00) determined using the method by Rogers (1995). Rapid population growth following colonisation was assured not only because of the large number of offspring produced, but also because of the abundance of food in the highly productive estuarine habitat.

The evolutionary histories of the other two populations studied, *H. kuda* from Tayabas Bay in the Philippines, and *H. fuscus* from the Red Sea, seem to have been similar to that of the Knysna seahorse in terms of the number of founders being very low, although it is as yet not certain whether the samples are merely suitable to represent local populations (e.g. Tayabas Bay) or whether they can be considered representative of entire lineages (e.g. *H. kuda* from the Philippines or the north-western Pacific). Genetic transilience does not seem necessary to explain how they became established in their new habitats, as environmental conditions in source and sink habitats possibly did not differ greatly. Conventional Darwinian adaptive radiation seems to be sufficient to explain their genetic distinctness, although it is as yet not established whether the different kudaoid seahorses are sufficiently distinct to be referred to as true 'biological species' according to Mayr's (1963) definition. The species status of the Knysna seahorse, however, seems secure: its populations are not only geographically isolated from populations of its sister species, but because *H. capensis* seems to be a 'true estuarine species'

living in a habitat that is inhospitable to marine seahorses, it is also likely to be physiologically isolated.

CHAPTER 4: THE EFFECT OF FIN CLIPPING ON GROWTH AND SURVIVAL OF *HIPPOCAMPUS CAPENSIS*

4.1 Introduction

4.1.1 Methods of teleost tissue acquisition

The use of genetic methods to study ecological and demographic characteristics of fish populations is widespread, and is fast becoming an alternative to traditional sampling methods. Comparatively small sample sizes are required to study fundamental population parameters such as effective population sizes, migration rates and population subdivision. Particularly in cases where non-genetic methods could previously not be employed at all, for example in rare fish (which have a low recapture rate) or fry (which are too small to be tagged), the advantages of genetic methods are evident.

In fisheries management studies, it has been common practice to sacrifice fish in order to obtain sufficient tissue for analysis (Whitmore *et al.* 1992). In the case of rare, vulnerable, or endangered fish species, however, only non-destructive methods should be considered. The development of the polymerase chain reaction (PCR) has facilitated the application of various minimally invasive sampling techniques to obtain genetic material. Several sources of tissue have been targeted in different fish species, including blood (Carmichael *et al.* 1986), scales (Carmichael *et al.* 1986; Whitmore *et al.* 1992), mucous samples (Robbins *et al.* 1989), and fin clips (Carmichael *et al.* 1986; Robbins *et al.* 1989; Wilson & Donaldson 1998). Seahorses and other syngnathids neither have scale epithelium, nor does a mucous layer protect their skin. Fin clipping is consequently the only possible method for sampling these and related teleosts, and it has already been applied by Kvarnemo *et al.* (2000) to obtain genetic material from male seahorses of the species *Hippocampus subelongatus*.

4.1.2 Notes on fin clipping

Fin clipping has traditionally been used as a simple and efficient means of marking fish, and fins were either removed partially or completely (e.g. Haines & Modde 1996). A completely removed fin will not regenerate, but partial fin removal produces a temporary mark until the fin regenerates (Nielsen 1992). Unlike the fins of elasmobranchs and the flippers of marine mammals, the bony rayed fins of teleosts contain few blood vessels, and as long as a fin clip is taken sufficiently far from the fin's base, injury is minimal. Wounds caused by partial fin clipping heal quickly (Nielsen 1992). A teleost fin is composed of multiple fin rays. Growth is a continuous process and occurs by addition of ray segments to the end of the fin, rather than by increase in length of the established ray segments. During regeneration, fin growth occurs in a similar fashion. Following amputation and wound healing, each fin ray establishes an independent blastema (Goss & Stagg 1957). Regeneration of the fin begins almost immediately after injury, and complete regeneration of partially clipped fins may occur within a few months (Churchill 1963). However, regenerated fins often can be recognised by deformities (Nielsen 1992).

Studies on the effect of fin clipping have shown that even though the procedure does not affect growth, survival of fin clipped fish has been shown to be lower than that of non-clipped controls, and fish less than 90 mm in length are more susceptible than larger fish (Coble 1967). However, fin clipping as a means of obtaining genetic material may require the removal of considerably less tissue than is necessary for marking, where it is imperative that a previously clipped fin still be recognisable as such after a certain period of time. For that reason, the above considerations are not equally applicable to the present investigation. Duke (1986) found that fungal growth occurs on fins that are clipped too deeply. This suggests that limiting the fin clip to the edge of the fin, can reduce the risk of diseases and mortality.

4.1.3 Rationale of this study

Even though the use of fin clipping as a non-destructive method to obtain genetic material from fishes is well established, in the case of rare and endangered teleosts such as the Knysna seahorse, field sampling should, however, be preceded by a detailed study on the effects of this method. Unforeseen complications may have a significant negative impact on small populations. In most species of teleosts, fins that are more important for movement or balance (major median fins and pectoral fins) are less suitable for clipping than clips of the adipose or

pelvic fins. However, which fin is targeted should be decided upon by considering anatomical peculiarities of each species. Seahorses have only four fins, namely a dorsal fin, two pectoral fins, and a reduced anal fin. The anal fin is too small and too close to sensitive abdominal tissue, and is therefore unsuitable for clipping. Tissue removal from one of the pectoral fins results in asymmetry that may affect manoeuvring, whereas clipping more than one fin is likely to increase the risk of mortality. The dorsal fin is consequently the only fin that can be considered for clipping. In seahorses, this fin is used to provide forward thrust, and is therefore of greater importance to the individual than in most other fish, where it is merely used to provide stability (Aleev 1969) or aids in locomotion (Drucker & Lauder 2001). It is thus imperative that the dorsal fin is clipped in a manner that reduces its functionality as little as possible.

Experiments carried out in this study focus on the effects of fin clipping on growth and survival. The effect on vulnerability to predation was not investigated, as seahorses do not normally employ escape as a means of predator avoidance, but instead rely on camouflage. Previous research has shown that fin clipping has no effect on growth in fish (Brynildson & Brynildson 1967; Coble 1971). However, as the mode of feeding in seahorses differs from that of most other fish, it was considered necessary to investigate this issue. Seahorses stalk their prey, and once a prey item is in close proximity to the tube-like pair of fused jaws, it is rapidly sucked into the buccal cavity. A deformed dorsal fin may potentially decrease the success of ambush and the execution of this intricate procedure of prey capture.

4.2 Materials and methods

4.2.1 Experimental design

The individuals used for these experiments were bred in captivity at Port Nolloth Sea Farms, located on the north west coast of South Africa, and were offspring of seahorses originally collected in the Knysna Estuary.

Groups of seahorses were kept in aquaria measuring 30 cm × 32 cm × 38 cm. Water was pumped from the sea into a large reservoir (100 m × 50 m × 1.50 m), and from there through the seahorses' holding tanks. The water was at ambient temperature and during the study period ranged between 14-16 °C during the first three months of the study (July-October), and between 18-22 °C during the last three months (November-January). This is within the temperature range the species experiences in its natural habitat (Lockyear, pers. comm.). Before reaching the seahorse tanks, the water was filtered using a sand filter, but not treated with UV light. The possibility that pathogens and parasites potentially present in the water were able to enter the holding tanks cannot be excluded, because fouling tubeworms of the genus *Spirorbis* became established in all tanks. Seahorses were fed exclusively on an excess diet of mysids, *Gastrosaccus* sp.

Two experiments were carried out to investigate the effects of fin clipping. In both cases, the following procedure was carried out. A pair of fingernail scissors (ZWILLING SOLINGEN) was disinfected with ethanol prior to treatment, and was left to dry for 5 min. Seahorses were caught individually with a scoop net, and taken out of the water for fin clipping. The animals were held between thumb and index finger, facing downwards, with the seahorses' head pointing towards the laboratory worker's arm. The dorsal fin was allowed to fold onto the lower blade of the scissors, and a small piece was then cut off from the edge of the fin's lower end. The cut was not disinfected. Seahorses were not tagged, as this method had previously been shown to negatively affect the animals (Teske & Lockyear, pers. obs.).

Experiment 1: small fin clips

Seahorses of approximately the same age were chosen from four different batches from different parents. Thirty young adult seahorses of both sexes from two batches from different

parents (batch 1 and 2) were used to investigate the effects of fin clipping on growth and survival. Fifteen arbitrarily chosen individuals from both groups were clipped (these were named "batch 1 clipped" and "batch 2 clipped"; fin clips were ≤ 2 mm^2 in size), while the remaining 15 individuals from each group were kept as controls ("batch 1 non-clipped" and "batch 2 non-clipped"). Seahorses from the control groups were also taken out of the water for approximately as much time as it took to clip a seahorse. The resulting four groups were each kept in a separate holding tank.

Experiment 2: large fin clips

An additional 15 seahorses from a third batch were used to investigate the effect of taking a considerably larger fin clip than in the previous cases (> 5 mm^2). This group was named "batch 3 clipped". It was considered necessary to include this investigation, as it could not be ruled out that during fieldwork, unnecessarily large fin clip samples would accidentally be obtained in individual cases. As these seahorses were subsequently compared with the clipped seahorses from the first two batches only, no controls were used in this case.

Body length and body mass were measured in order to assess whether non-clipped seahorses grew significantly faster than clipped seahorses (experiment 1), and to assess whether the size of the fin clip taken had an effect on growth (experiment 2). Body length is defined here as the distance from coronet to tail and was measured using MITUTOYA calipers. Body mass was measured by blot drying the seahorse with a towel, and then placing it on a METTLER P 1200 N scale. To allow for significant changes in length and mass, measurements were taken only at the start of the experiment, after 3 months, and after 6 months. Clipped fins of steelheads in Idaho have been shown to heal after 3-4 weeks, but wound healing is positively correlated with water temperature (Duke 1986). This suggests that periods of 3 months used in this study are appropriate to investigate the effects of fin clipping.

4.2.2 Data analysis

Statistical tests were conducted using SIGMASTAT 2.0 for Windows (Jandel Corporation 1995). Significance level for all univariate statistics was $p < 0.05$. At the beginning of the experiments, differences in body length and body mass of individuals from the two non-clipped

groups, the two groups from which small fin clips were obtained, and the group from which large fin clip were obtained, were assessed using a one-way analysis of variance (ANOVA).

Data measured after three months and after six months were treated as follows. Data from the two non-clipped groups and their clipped counterparts (experiment 1) were compared using a two-way ANOVA. This statistical analysis tests simultaneously whether individuals from the two different batches grew at a different rate and whether non-clipped seahorses grew at a different rate as compared to clipped seahorses. If necessary, data were log-transformed prior to analysis in order to fulfil the requirements for parametric tests (i.e. normality and equal variance).

The two groups from which small fin clips had been obtained were then pooled, and the resulting group was compared with the group from which large fin clips had been obtained (experiment 2) using a Student t-test. Whenever the requirements for parametric statistics were violated, a Mann-Whitney rank sum test was used.

4.3 Results

4.3.1 Fin clip experiments

A one-way ANOVA showed that at the start of the experiment, the individuals chosen from the first three batches did not differ significantly in length ($F = 2.08$, $p = 0.122$) and mass ($F = 2.27$, $p = 0.09$). Even though the power of these tests was low (0.32 and 0.37), the groups were considered sufficiently homogenous to be used for the experiments.

Fins regenerated quickly, with even the dorsal fins from which tissue material > 5 mm^2 in area had been removed being indistinguishable from non-clipped fins after approximately 6-8 weeks. No fatalities were recorded, and none of the seahorses used in the experiment was ever diagnosed as being ill or infected, even though fungal infections, tuberculosis, and mortalities were occasionally observed in seahorses not included in the experiments. Pregnancies occurred in all six tanks, but the size, mass, and number of offspring were not recorded.

Experiment 1: small fin clips

After three months and after six months, no significant differences in body length and body mass were found between the different batches used or between clipped and non-clipped seahorses (Fig. 4.1; Table 4.1). There was also no relationship between batches and the status of seahorses' fins.

Experiment 2: large fin clips

No significant differences in length and mass were found between the two groups of clipped seahorses after 3 months (length comparison: $T = 406$, $p = 0.15$ [Mann-Whitney test]; mass comparison: $t = -1.32$, $df = 43$, $p = 0.19$, power = 0.12) and after 6 months (length comparison: $t = 0.95$, $df = 43$, $p = 0.35$, power = 0.05; mass comparisons: $t = -1.3$, $df = 43$, $p = 0.20$, power = 0.12) (Fig. 4.2).

Table 4.1 Two-way analyses of variance investigating differences in body length (mm) and body mass (g) between seahorses from different batches, differences between clipped and non-clipped seahorses (fin status), and the effect of fin status on the two batches ($\alpha = 0.05$).

Time after clipping	data used	comparison batches			comparison fin status			relationship batches × fin status		
		F	p	power	F	p	power	F	p	power
3 months	length	1.84	0.18	0.14	1.75	0.19	0.13	1.75	0.19	0.13
	mass	0.93	0.34	0.05	0.58	0.45	0.05	0.83	0.37	0.05
6 months	length	1.45	0.23	0.09	3.02	0.09	0.23	2.03	0.16	0.16
	mass	0.10	0.75	0.05	2.64	0.11	0.23	0.46	0.50	0.05

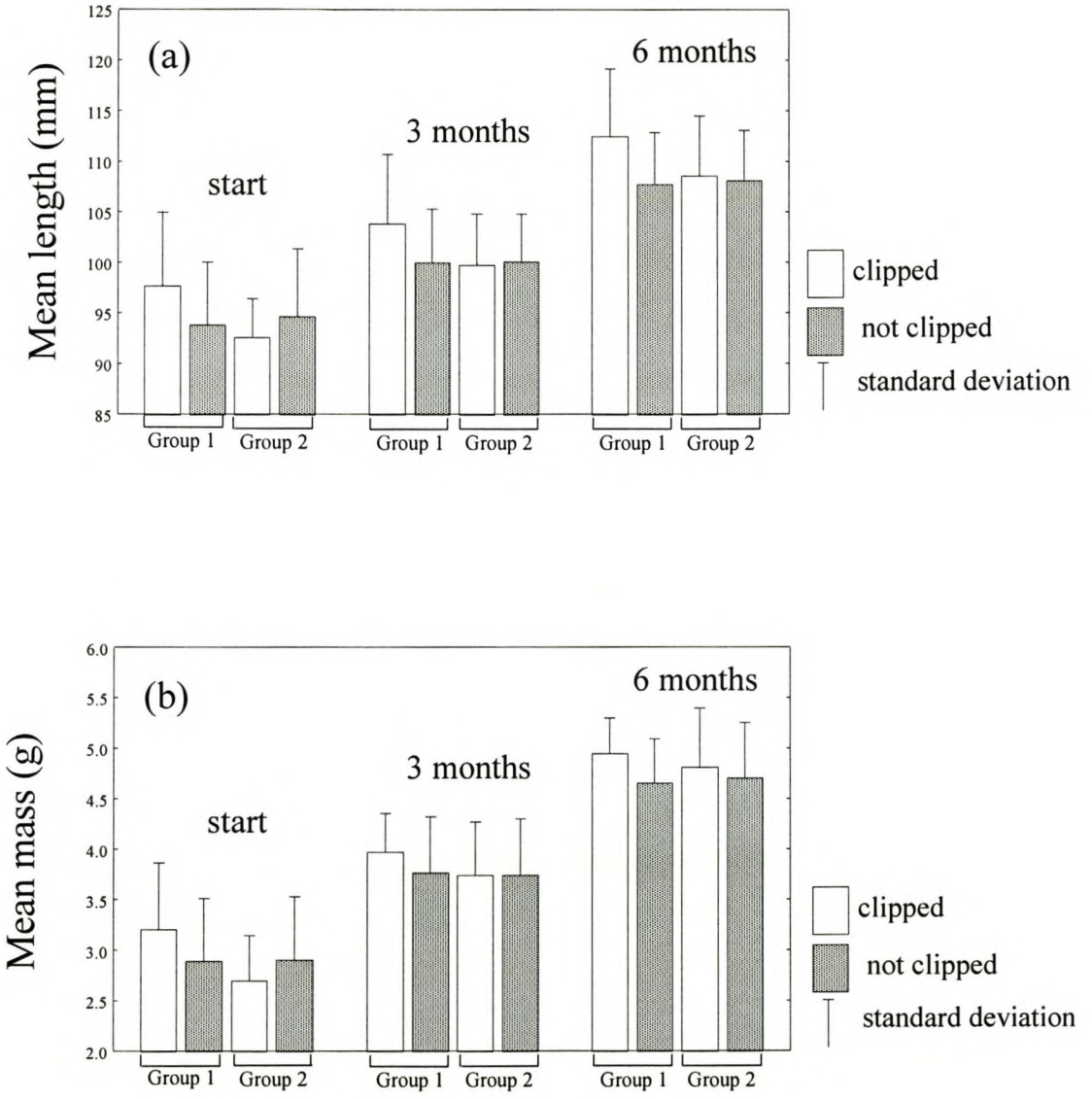

Fig. 4.1 Increase in (a) length (+S.D.) and (b) mass (+S.D.) of clipped and non-clipped seahorses during the study period

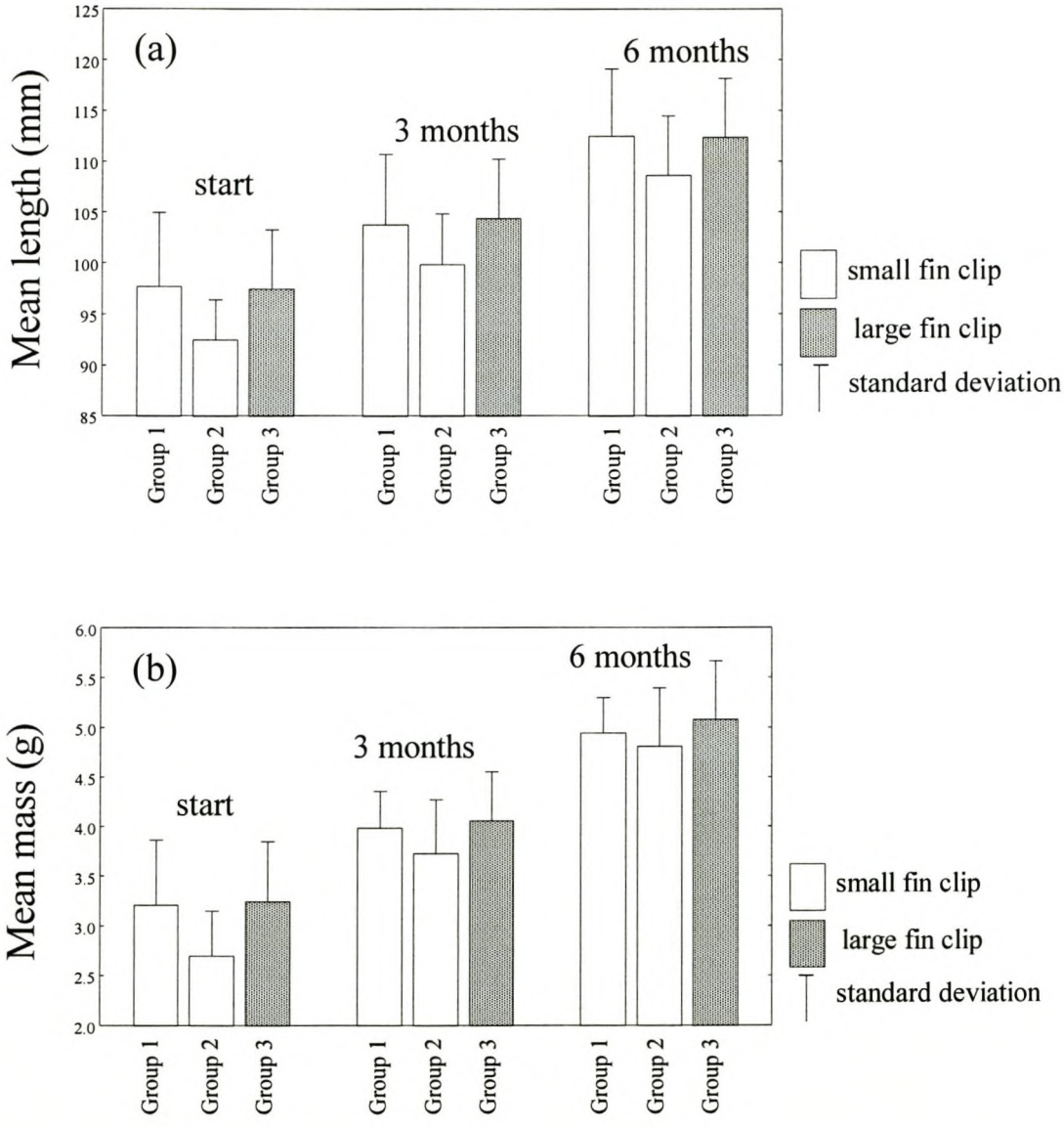

Fig. 4.2 Increase in (a) length (+S.D.) and (b) mass (+S.D) of individuals from which small fin clips (<2 mm^2, Groups 1 and 2) and large fin clips (>5 mm^2, Group3) were obtained.

4.4 Discussion

4.4.1 Assessment of the results

Even under the unfavourable conditions to which the seahorses were subjected (i.e. potential access of pathogens and no disinfection of the cut), fin clipping had no detectable effect on growth and survival of the seahorses. Despite low sample sizes, there is little reason to suggest that the effects of this method would be different if used in the field. It would have been desirable to work with a larger population size in order to increase the power of the results. Additionally, this could have been achieved by carrying out paired tests. However, due to the fact that tagged individuals in the wild were negatively affected by the strings used to attach the tags to their necks (Teske & Lockyear, pers. obs.), we decided to refrain from applying this method.

4.4.2 Suitability of fin clipping for sampling of wild populations

Although the procedures used to analyse the effects of fin clipping on growth of seahorses could be improved upon, the fact that none of the study individuals died was considered sufficient to sample *H. capensis* in the wild without risking to negatively impact on the populations. As mentioned previously, most of the samples from the three populations of the Knysna seahorse were obtained in this manner. In order to reduce the impact of sampling as much as possible, fin clips were taken while the study animals remained submerged in a bowl of water, and they were subsequently returned to the location of capture. The small amount of DNA present in the fin clips was sufficient to amplify all of the genes used in chapters 1-3, and in fact, even genomic DNA obtained from fin clips that were taken from young seahorses (site 3, chapter 3), and was barely visible on agarose gels, amplified more readily than DNA obtained from dried specimens and some of the older ethanol-stored specimens. Fin clipping thus seems to be an ideal method to obtain genetic material even from endangered teleosts, and for the study of seahorses, it is recommended unless the taxonomic identity of a particular specimen is uncertain. In general, any genetic study using small, endangered teleosts should be preceded by a study similar in design to the one presented here. Apart from being able to establish whether

the species under investigation is susceptible to fin clipping, it also provides an opportunity for researchers to optimise their methodology.

GENERAL DISCUSSION

The present study on the population genetics of the endangered Knysna seahorse and its phylogenetic placement within the genus *Hippocampus*, has made significant contributions towards the scientific understanding of this well-known but little-studied group of teleosts. It has given insight into evolutionary processes and detected previously unknown phylogenetic and biogeographic patterns.

Although research concerning the deep phylogeny of teleosts is presently trailing behind similar studies on mammals because of the availability of relatively few primers for the amplification of nuclear markers, the phylogeny of the genus *Hippocampus* obtained in this study is fairly robust, and is unlikely to change much as information from additional markers and/or samples is added. In future studies, the use of nuclear sequences more conserved than RP1 is recommended to calibrate a molecular clock for the deep phylogeny of seahorses. Also, some additional groups of syngnathids should be included. Due to the fact that most teleost specimens are routinely fixed in formaldehyde, many of the seahorse specimens available from museums are unsuitable for genetic analyses. Additionally, many seahorse species are only rarely found, and are thus not represented in the reference collections of most museums. Due to these problems, several important species could not be sequenced for this study, including *Hippocampus minotaur*, *H. denise* and other small-bodied seahorses that seem to be closely associated with the pygmy seahorse, *H. bargibanti*, as well as pygmy pipehorses (genera *Acentronura*, *Amphelikturus* and *Idiotropiscis*). Additionally, as the fossil record of pipefishes is much less fragmented than that of seahorses, an expanded analysis should also include a number of pipefish specimens of the closely related genus *Syngnathus* in order to provide additional calibration points for a molecular clock.

No attempt was made to date the comparatively recent divergence events among kudaoid seahorses, because data-bases of mitochondrial markers are comparatively incomplete at this stage, and the phylogeny of this lineage is as yet not fully resolved. Moreover, there was uncertainty regarding the calibration of evolutionary rates for each partition: it is not certain whether the haplotypes of the West Atlantic and East Pacific species that diverged after the closure of the Central American Seaway are likely to be basal or derived within each species. Haplotype networks of individual populations (chapter 3) were characterised by basal haplotypes (MRCAs) that were present in the majority of individuals within each population. A similar pattern may exist in *H. reidi* and *H. ingens*, and the chance that any two haplotypes

collected from either species may have arisen before or shortly after the closure of the Central American Seaway rather than three million years later is thus very high. For future research, larger sample sizes should be obtained from both species, and only the most derived haplotypes of each should be used in the calibration of a molecular clock.

Genetic diversity of the Knysna seahorse was not substantially lower than in tropical marine seahorses, although it is likely that the latter assemblages are characterised by considerable gene-flow between local populations. Moreover, all three populations sampled were characterized by a star-like phylogeny, in which a single ancestral haplotype had given rise to a large number of derived haplotypes. A comparison of the breeding strategies employed by seahorses and other syngnathids with those of sticklebacks suggests that long-distance colonisation and founder-effect speciation are considerably less likely to occur in sticklebacks: male syngnathids attach fertilised eggs to their bodies (pipefishes) or store them in a brood pouch (seahorses), whereas male sticklebacks build nests. Future studies could focus on whether all syngnathid species and/or populations are characterised by single MRCAs, whereas recently founded stickleback species/populations (e.g. populations founded after the last Pleistocene glaciation) are more likely to have several equally old MRCAs, which may not be closely related to one another. Pregnant male syngnathids are theoretically able to give rise to populations, whereas both a male and a female stickleback are required to colonise a new habitat. In the case of a new habitat distantly located from the source habitat (where allopatric speciation may take place because of little or no additional gene flow from the source population), the chance of two displaced sticklebacks of either sex to locate one another and mate is slim. This suggests that gradual population expansion is more common in sticklebacks, whereas long-distance colonisation (and rapid morphological differentiation) may be more common in syngnathids. The potential for long-distance colonisation in syngnathids may explain why kudaoid seahorses have a circumglobal distribution, although the maximum p-distance among their CR sequences is only 0.066 (among *H. fisheri* and one of the *H. kuda* haplotypes from the Philippines), as compared to 0.065 in a sample size of 28 individuals of Japanese sticklebacks (indicating a comparatively young age of the kudaoid seahorse lineage). In most other marine species studied to date, long-distance dispersal takes place during larval phases, which explains why species with pelagic eggs and long-lived, planktonic larvae tend to be much more widely distributed than species that lack larval phases (Palumbi 1992). As mentioned previously, seahorses do not have a planktonic larval phase (but see Kuiter 2000), and as young seahorses need to feed immediately, it is unlikely that they may disperse over great distances as part of the plankton. However, the fact that some adult seahorses have been

found associated with floating seaweed (Kuiter 2000) suggests that seahorses may disperse by rafting. Small marine fish or invertebrates using this raft for shelter may serve as food for a displaced seahorse, which may enable it to survive for a period of time until a new habitat is reached. Using this mode of transport, long-distance dispersal by both juvenile and adult seahorses seems feasible. However, the successful establishment of a new population is most likely when the dispersing individual is an adult, pregnant male seahorse.

REFERENCES

ABOUHEIF, E., ZORDOYA, R. & MEYER, A. 1998. Limitations of metazoan 18S rRNA sequence data: Implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. *Journal of Molecular Evolution* **47:** 394-405.

ADAMS, C.G. 1983. Speciation, phylogenesis, tectonism, climate and eustasy: factors in the evolution of Cenozoic larger foraminiferal bioprovinces. In: *Evolution, time and space: the emergence of the biosphere*, (eds) R.W. Sims, J.H. Price & P.E.S. Whalley, pp. 255-289. Whalley Academic Press, London.

AKAIKE, H. 1973. Information theory as an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*, (eds) B.N. Petrov & F. Csaki, pp. 267-281. Akademiai Kiado, Budapest.

ALATALO, R.V., LUNDBERG, A. & BJÖRKLUND, M. 1982. Can the song of male birds attract other males? An experiment with the pied flycatcher *Ficedula hypoleuca*. *Bird Behaviour* **4:** 42-45.

ALI, J.R. & HALL, R. 1995. Evolution of the boundary between the Philippine Sea Plate and Australia: paleomagnetic evidence from eastern Indonesia. *Tectonophysics* **251:** 251-275.

ALLEE, W.C. 1938. *The social life of animals*. Norton, New York.

ALVES-GOMES, J.A., ORTI, G., HAYWOOD, M., HEILIGENBERG, W. & MEYER, A. 1995. Phylogenetic analysis of the South American electric fishes (order Gymnotiformes) and the evolution of their electrogenic system: a synthesis based on morphology, electrophysiology, and mitochondrial sequence data. *Molecular Biology and Evolution* **12:** 298-318.

ANONYMOUS 1978. *Hidrouliese studie van die Swartvlei estuarium: effek van bestaande en voorgestelde brue op die waterbeweging*. Vol. I. Ingeneurs verslag. In opdrag van die dept. van beplanning en die omgewing. Stellenbosch. WNNR Verslag C/SEA 7805/1.

ANTHONY, L.L. & BLUMSTEIN, D.T. 2000. Integrating behaviour into wildlife conservation: the multiple ways that behaviour can reduce N_e. *Biological Conservation* **95:** 303-315.

ARBOGAST, B. S., EDWARDS, S. V., WAKELEY, J., BEERLI, P. & SLOWINSKI, J. B. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics* **33**:707-740.

ARNAUD, S., BONHOMME, F. & BORSA, P. 1999. Mitochondrial DNA analysis of the genetic relationships among populations of scad mackerel (*Decapterus macarellus*, *D. macrosoma*, and *D. russelli*) in South-East Asia. *Marine Biology* **135:** 699-707.

ARONSON, R.B., PRECHT, W.F., MACINTYRE, I.G. & MURDOCH, T.J.T. 2000. Ecosystems: coral bleach-out in Belize. *Nature* **405:** 36.

AVISE, J.C. 1989. A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution* **4:** 279-281.

AVISE, J.C. 1994. *Molecular markers, natural history and evolution.* Chapman & Hall, New York.

AVISE, J.C., ARNOLD, J., BALL, R.M., BERMINGHAM, E., LAMB, T. & NIGEL, J.E. 1988. Intraspecific phylogeography: Mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18:** 489-522.

AXELROD, D.I. & RAVEN, P.H. 1978. Late Cretaceous and Tertiary vegetation history of Africa. In: *Biogeography and Ecology of Southern Africa*, (ed) M.J.A. Werger. Junk, The Hague, pp. 77-130.

AYALA, F.J. 1986. On the virtues and pitfalls of the molecular evolutionary clock. *Heredity* **77:** 226-235.

AYVAZIAN, S.G., JOHNSON, M.S. & MCGLASHAN, D.J. 1994. High levels of genetic subdivision of marine and estuarine populations of the estuarine catfish *Cnidoglanis macrocephalus* (Plotosidae) in southwestern Australia. *Marine Biology* **118:** 25-31.

BAKER, R.H. & DESALLE, R. 1997. Multiple sources of character information and the phylogeny of Hawaiian *Drosophila*. *Systematic Biology* **46:** 654-673.

BANARESCU, P. 1986. "Vicariance" and dispersal in freshwater animals. *Hydrobiology* **19:** 257-268.

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

BARD, E., ARNOLD, M., HAMELIN, B., TISUERAT-LABORDE, N. & CABIOCH, G. 1998. Radiocarbon calibration by means of mass spectrometric Th-230/U-234 and C-14 ages of corals: an updated database including samples from Barbados, Mururoa and Tahiti. *Radiocarbon* **40:** 1085-1092.

BARTON, N.H. & CHARLESWORTH, B. 1984. Genetic revolutions, founder effects and speciation. *Annual Review of Ecology and Systematics* **15:** 133-164.

BAVERSTOCK, P. R. & MORITZ, C. 1996. Project design. In: *Molecular systematics*, (eds) D.M. Hillis, C. Moritz & B.K. Mable, pp. 17-27. Sinauer, Massachusetts.

BÉ, A.W.H. & DUPLESSY, J.-C. 1976. Subtropical convergence fluctuations and Quaternary climates in the middle latitudes of the Indian Ocean. *Science* **194:** 419-422.

BELL, E.M., LOCKYEAR, J.F., MCPHERSON, J.M., MARSDEN, A.D. & VINCENT, A.C.J. 2003. First field studies on an endangered South African seahorse, *Hippocampus capensis*. *Environmental Biology of Fishes* **67:** 35-46.

BENZIE, J.A.H. 1998. Genetic structure of marine organisms and SE Asian biogeography. In: *Biogeography and geological evolution of SE Asia*, (eds) R. Hall & J.D. Holloway, pp. 197-209. Backhuys Publishers, Leiden, The Netherlands.

BERMINGHAM, E. & AVISE, J.C. 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* **113:** 939-965.

BERMINGHAM, E., MCCAFFERTY, S.S. & MARTIN, A.P. 1997. Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: *Molecular systematics of fishes*, (eds) T.D. Kocher & C.A. Stepien, pp. 113-128. Academic Press, San Diego, California.

BERNARDI, G. 1997. Molecular phylogeny of the Fundulidae (Teleostei, Cyprinodontiformes) based on the cytochrome *b* gene. In: *Molecular systematics of fishes*, (eds) T.D. Kocher & C.A. Stepien, pp. 189-197. Academic Press, San Diego, California.

BERNARDI, G., SORDINO, P. & POWERS, D.A. 1993. Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. *Proceedings of the National Acadamey of Sciences of the USA* **90:** 9271-9274.

BERRIT, G.R. 1973. *Recherches hydroclimatiques dans les régions côtières de l'Atlantique oriental tropical*. État des connaissances et perspectives, Bulletin du Musée d'Histoire Naturelle de Paris. 3^{e} séries, Écologie générale 4, pp. 279-285.

BLABER, S.J.M. 1981. The zoogeographic affinities of estuarine fishes in south-east Africa. *South African Journal of Science* **77:** 305-307.

BLOCK, B.B., FINNERTY, J.R., STEWARD, A.F.R. & KIDD, J. 1993. Evolution of endothermy in fish: Mapping physiological traits on a molecular phylogeny. *Science* **260:** 210-214.

BORLASE, S.C., LOEBEL, D.A., FRANKHAM, R., NURTHEN, R.K., BRISCOE, D.A. & DAGGARD, G.E. 1993. Modeling problems in conservation genetics using captive *Drosophila* populations: consequences of equalisation of family sizes. *Conservation Biology* **7:** 122-131

BOWEN, B.W., BASS, A.L., ROCHA, L.A., GRANT, W.S. & ROBERTSON, D.R. 2001. Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. *Evolution* **55:** 1029-1039.

BOWER, D. & CRAWFORD, R. 1981. A black south-easter. *Custos* **10**: 20-21.

BRANCH, G. & BRANCH, M. 1992. *The living shores of Southern Africa*. Struik Publishers, Cape Town, South Africa.

BRIGGS, J.C. 1961. The East Pacific barrier and the distribution of marine shore fishes. *Evolution* **15:** 545-554.

BRIGGS, J.C. 1974. *Marine zoogeography*. McGraw-Hill, New York.

BRIGGS, J.C. 1995. *Global biogeography*. Elsevier, Amsterdam.

BRIGGS, J.C. 1999. Coincident biogeographic patterns: Indo-West Pacific Ocean. *Evolution* **53:** 326-335.

BRITTEN, R.J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231:** 1393-1398.

BROWN, B.E. 1997. Coral bleaching: causes and consequences. *Coral Reefs* (Supplement) **16:** S129-S138.

BROWN, W.M., GEORGE, M. JR. & WILSON, A.C. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the USA* **76:** 1967-1971.

BROWN, W.G., GADALETA, G., PETE, G., SACCONE, C. & SBISA, E. 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *Journal of Molecular Biology* **192:** 503-511.

BRUNNER, P.C., DOUGLAS, M.R., OSINOV, A., WILSON, C.C. & BERNATCHEZ L 2001. Holartic phylogeography of Artic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution* **55:** 573-586.

BUCKLEY, T.R. 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. *Systematic Biology* **51:** 509-523.

BUCKLEY, T.R., ARENSBURGER, P., SIMON, C. & CHAMBERS, G.K. 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Systematic Biology* **51:** 4-18.

BURRIDGE, C.P. & WHITE, R.W.G. 2000. Molecular phylogeny of the antitropical subgenus *Goniistius* (Perdiformes: Cheilodactylidae: Cheilodactylus): evidence for multiple transequatorial divergences and non-monophyly. *Biological Journal of the Linnean Society* **70:** 435-458.

BUTZER, K.W. & HELGREN, D.M. 1972. Late cenozoic evolution of the Cape Coast between Knysna and Cape St. Francis, South Africa. *Quaternary Research* **2:** 143-169.

CAMPBELL, B.C., STEFFEN-CAMPBELL, J.D. & GILL, R.J. 1994. Evolutionary origin of whiteflies (Hemiptera: Sternorrhyncha: Aleyrodidae) inferred from 18S rDNA sequences. *Insect Molecular Biology* **3:** 73-88.

CANE, M. A. & MOLNAR, P. 2001. Closing of the Indonesian seaway as a precursor to east African aridification around 3-4 million years ago. *Nature* **411:** 157-162.

CARMICHAEL, G.J., WILLIAMSON, J.H., SCHMIDT, M.E. & MORIZOT, D.C. 1986. Genetic marker identification in largemouth bass with electrophoresis of low-risk tissues. *Transactions of the American Fisheries Society* **115:** 455-456.

CARR, S.M. & MARSHALL, H.D. 1991. Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome b gene of Atlantic cod (*Gadus morhua*) by the polymerase chain reaction. *Canadian Journal of Fisheries and Aquatic Sciences* **48:** 48-52.

CARR, S.M., SNELLEN, A.J., HOWSE, K.A. & WROBLEWSKI, J.S. 1995. Mitochondrial DNA sequence variation and genetic stock structure of Atlantic cod (*Gadus morhua*) from bay and offshore locations on the Newfoundland continental shelf. *Molecular Ecology* **4:** 79-88.

CARSON, H.L. 1971. Speciation and the founder principle. *Stadler Genetic Symposia Series* **3:** 51-70.

CARSON, H.L. 1975. The genetics of speciation at the diploid level. *The American Naturalist* **109:** 83-92.

CARSON, H.L. & TEMPLETON, A.R. 1984. Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annual Review of Ecology and Systematics* **15:** 97-131.

CARVALHO, G.R. 1993. Evolutionary aspects of fish distribution: genetic variability and adaptation. *Journal of Fish Biology* **43:** 53-73.

CASEY, S.P. 1999. *A phylogenetic study of seahorses using the cytochrome b gene of mitochondrial DNA*. Ph.D. thesis, University of London, United Kingdom.

CHENOWETH, S.F., HUGHES, J.M., KEENAN, C.P. & LAVERY, S. 1998. Concordance between dispersal and mitochondrial gene flow: isolation by distance in a tropical teleost, *Lates calcarifer* (Australian barramundi). *Heredity* **80:** 187-197.

CHOU, Q., RUSSELL, M., BIRCH, D.E., RAYMOND, J. & BLOCH, W. 1992. Prevention of pre-PCR mispriming and primer dimerisation improves low-copy-number of amplifications. *Nucleic Acids Research* **20:** 1717-1723.

CHOW, S. & HAZAMA, K. 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology* **7:** 1255-1256.

CHUNNETT, E.P. 1965. *Siltation problems in the Knysna Lagoon.* Pretoria, CSIR Report MEG 353.

CHURCHILL, W.S. 1963. The effect of fin clip removal on survival, growth, and vulnerability to capture of stocked walleye fingerlings. *Transactions of the American Fisheries Society* **92:** 298-300.

CIOFI, C. & BRUFORD, M.W. 1999. Genetic structure and gene flow among Komodo dragon populations inferred by microsatellite loci analysis. *Molecular Ecology* **8:** S17-S30.

CLEGG, M.T., GAUT, B.S., LEARN, G.H. Jr. & MORTON, B.R. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences of the USA* **91:** 6795-6801.

CLEMENT, M., POSADA, D. & CRANDALL, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657-1660.

COATES, A.G. JACKSON, J.B.C., COLLINS, L.S., CRONIN, T.M., DOWSETT, H.J., BYBELL, L.M., JUNG, P. & OBANDO, J. 1992. Closure of the Isthmus of Panama: The near-shore marine record of Costa Rica and western Panama. *Geological Society of America Bulletin* **104:** 814-828.

COBLE, D.W. 1967. Effects of fin clipping on mortality and growth of yellow perch with a review of similar investigations. *Journal of Wildlife Management* **31:** 173-180.

COBLE, D.W. 1971. Effects of fin clipping and other factors on survival and growth of smallmouth bass. *Transactions of the American Fisheries Society* **3:** 460-565.

COGNATO, A.I. & VOGLER, A.P. 2001. Exploring data interaction and nucleotide alignment in a multiple gene analysis of Ips (Coleoptera: Scolytinae). *Systematic Biology* **50:** 758-780.

COLBORN, J., CRABTREE, R.E., SHAKLEE, J.B., PFEILER, E. & BOWEN, B.W. 2001. The evolutionary enigma of bonefishes (*Albula* spp.): cryptic species and ancient separations in a globally distributed shorefish. *Evolution* **55:** 807-820.

COOPER, J.A.G. 1991. *Sedimentary models and geomophological classification of river mouths on a subtropical, wave-dominated coast, Natal, South Africa*. Ph.D. thesis, University of Natal, Durban.

COOPER, J.A.G., WRIGHT, C.I. & MASON, T.R. 1999. Geomorphology and sedimentology. In: *Estuaries of South Africa*, (eds) B.R. Allanson & D. Baird, pp. 5-25. Cambridge University Press, United Kingdom.

COYNE, J.A. 1994. Ernst Mayr and the origin of species. *Evolution* **48:** 19-30.

CRANDALL, K.A. & TEMPLETON, A.R. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* **134:** 959-969.

CRANDALL, K.A., POSADA, D. & VASCO, D. 1999. Effective population sizes: missing measures and missing concepts. *Animal Conservation* **2:** 317-319.

CREER, S., MALHOTRA, A. & THORPE, R.S. 2003. Assessing the phylogenetic utility of four mitochondrial genes and a nuclear intron in the Asian pit viper genus, Trimeresurus: separate, simultaneous, and conditional data combination analyses. *Molecular Biology and Evolution* **20:** 1240-1251.

CROWLEY, T.J. & NORTH, G.R. 1991. *Paleoclimatology - Oxford Monographs on Geology and Geophysics. No. 18*. Oxford University Press, Inc. New York, NY.

CUMMINGS, M.P., HANDLEY, S.A., MYERS, D.S., REED, D.L., ROKAS, A. & WINKA, K. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Systematic Biology* **52:** 477-487.

DARLINGTON, P.J. Jr. 1957. *Zoogeography: the geographical distribution of animals*. John Wiley, New York.

DARLINGTON, P.J. Jr. 1959. Area, climate and evolution. *Evolution* **13:** 488-510.

DAVIES, O. 1971. Pleistocene shorelines in the southern and southeastern Cape Province (Part 1). *Annals of the Natal Museum* **20:** 403-442.

DAVIES, P.J., SYMONDS, P.A., FEARY, D.A. & PIGRAM, C.J. 1987. Horizontal plate motion: a key allocyclic factor in the evolution of the Great Barrier Reef. *Science* **238:** 1697-1700.

DAWSON, C.E. 1986. Family No. 145: Syngnathidae. In: *Smith's sea fishes*, (eds) M.M. Smith & P.C. Heemstra, pp. 445-458. Macmillan South Africa, Johannesburg.

DAY, J.H. 1981a. Summaries of current knowledge of 43 estuaries in southern Africa. In: *Estuarine ecology with particular reference to southern Africa*, (ed) J.H. Day, pp. 251-329. A.A. Balkema, Cape Town.

DAY, J.H. 1981b. Fauna. In: *Estuarine ecology with particular reference to southern Africa*, (ed) J.H. Day, pp. 147-178. A.A. Balkema, Cape Town.

DAY, J.H., MILLARD, N.A.H. & HARRISON, A.D. 1952. The ecology of South African estuaries. Part III Knysna: A clear open estuary. *Transactions of the Royal Society of South Africa* **33:** 367-413.

DAY, J.H., Blaber, S.J.M. & Wallace, J.H. 1981. Estuarine fishes. In: *Estuarine ecology with particular reference to southern Africa*, (ed) J.H. Day, pp. 197-221. A.A. Balkema, Cape Town.

D'CROZ, L & ROBERTSON, B.R. 1997. Coastal oceanographic conditions affecting coral reefs on both sides of the isthmus of Panama. *Proceedings of the 8th international coral reef symposium* **2:** 2053-2058.

DON, R.H., COX, P.T., BAKER, K. & MATTICK, J.S. 1992. "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Reserarch* **19:** 4008.

DONALDSON, K.A. & WILSON Jr., R.R. 1999. Amphi-panamic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA control region of fishes. *Molecular Phylogenetics and Evolution* **13:** 208-213.

DOWNTOWN, M. & AUSTIN, A.D. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy - the behavior of the incongruence length difference test in mixed-model analyses. *Systematic Biology* **51:** 19-30.

DRUCKER, E.G. & LAUDER, G.V. 2001. Locomotor function of the dorsal fin in teleost fishes: experimental analysis of wake forces in sunfish. *The Journal of Experimental Biology* **204:** 2943-2958.

DU TOIT, A.L. 1966. *The geology of South Africa.* 3rd edition. Northampton, John Dickens.

DUKE, R. 1986. *Protection of wild adult steelhead in Idaho by adipose fin removal.* Annual Report, BPA Fish and Wildlife Publications.

DUVENAGE, I.R. & MORANT, P.D. 1984. *Estuaries of the Cape. Part II: Synopses of available information on individual systems. Report No. 31: Keurbooms/Bitou System (CMS 19), Piesang (CMS 18).* CSIR Research Report 430 (eds Heydorn AEF, Grindley JR). Stellenbosch, South Africa.

EKMAN, S. 1953. *Zoogeography of the sea.* Sidgwick and Jackson, London.

ESCHMEYER, W.N. 1998. *Catalog of fishes. Volumes 1-3.* San Francisco: California Academy of Sciences.

ESPOSTI, M.D., DE VRIES, S., CRIMI, M., GHELLI, A., PATARNELLO, T. & MEYER, A. 1993. Mitochondrial cytochrome *b*: Evolution and structure of the protein. *Biochymica et Biophysica Acta* **1143:** 243-271.

EXCOFFIER, L., SMOUSE, P. & QUATTRO, J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131:** 479-491.

FARRIS, J.S., KALLERSJÖ, M., KLUGE, A.G. & BULT, C. 1994. Testing significance of congruence. *Cladistics* **10:** 315-320.

FARRIS, J.S., KALLERSJÖ, M., KLUGE, A.G. & BULT, C. 1995. Constructing a significant test for incongruence. *Systematic Biology* **44:** 570-572.

FELSENSTEIN, J. & CHURCHILL, G.A. 1996. A hidden Markov model approach to variation among sites in rate of evolution. *Molecular Biology and Evolution* **13:** 93-104.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39:** 783-791.

FELSENSTEIN, J. 1992a. Estimating effective population size from samples of sequences: inefficiency of pair-wise and segregating sites as compared to phylogenetic estimates. *Genetical Research* **59:** 139-147.

FELSENSTEIN, J. 1992b. Estimating effective population size from samples of sequences: a bootstrap Monte Carlo integration method. *Genetical Research* **60:** 209-220.

FISHER, R.A. 1930. *The genetical theory of natural selection.* Clarendon Press, Oxford.

FLAGSTAD, Ø., SYVERTSEN, P.O., STENSETH, N.C., STACY, J.E., OLSAKER, I., RØED, K.H. & JAKOBSEN, K.S. 2000. Genetic variability in Swayne's Hartebeest, an endangered antelope of Ethiopia. *Conservation Biology* **14:** 254-264.

FLEMINGER, A. 1986. The Pleistocene equatorial barrier between the Indian and Pacific Oceans and a likely cause for Wallace's Line. *UNESCO Technical Paper in Marine Science* **49:** 84-97.

FOURIE, D. 1997. *Sex-roles and the influence of size on mating competition in the Knysna seahorse, Hippocampus capensis.* Honours thesis, University of Stellenbosch.

FRICKE, R., 1988. *Systematik und historische Zoogeographie der Callionymidae (Teleostei) des Indischen Ozeans.* Ph.D. thesis, Albert-Ludwigs-Universität, Freiburg im Breisgau, Germany.

FRITZSCHE, R.A. 1980. Revision of the eastern Pacific Syngnathidae (Pisces: Syngnathiformes), including both recent and fossil forms. *Proceedings of the California Academy of Sciences* **42:** 181-227.

FU, Y.-X. 1994a. A phylogenetic estimator of effective population size or mutation rate. *Genetics* **136:** 685-692.

FU, Y.-X. 1994b. Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. *Genetics* **138:** 1375-1386.

FU, Y.-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147:** 915-925.

FU, Y.-X. & LI, W.-H. 1993a. Maximum likelihood estimation of population parameters. *Genetics* **134**: 1261-1270.

FU, Y.-X. & LI, W.-H. 1993b. Statistical tests of neutral mutations. *Genetics* **133:** 693-709.

GATESY, J., HAYASHI, C., CRONIN, M.A. & ARCTANDER, P. 1996. Evidence from milk casein genes that cetaceans are close relatives of hppopotamid artiodactyls. *Molecular Biology and Evolution* **13:** 954-963.

GATESY, J., MILINKOVITCH, M., WADDELL, V. & STANHOPE, M. 1999. Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. *Systematic Biology* **48:** 6-20.

GATESY, J., MATTHEE, C.A., DESALLE, R. & HAYASHI, C. 2002. Resolution of a supertree / supermatrix paradox. *Systematic Biology* **51:** 652-664.

GAUT, B.S., MUSE, S.V., CLARKE, W.D. & CLEGG, M.T. 1992. Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. *Journal of Molecular Evolution* **35:** 292-303.

GELDENHUYS, N.D. 1979. *Guidelines for the future utilisation of Knysna estuary*. Pretoria, Dept. Environmental Planning and Energy. Unpublished report.

GENADE, A.B. & HIRST, A.L. 1986. 'n Perdjie van 'n ander kleur. *Custos* **15:** 34-36.

GEYER, C.J. & THOMPSON, E.A. 1995. Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association* **90:** 909-920.

GLADSTEIN, D.S. & WHEELER, W.C. 1997. *POY - The optimisation of alignment characters*. Program and Documentation. American Museum of Natural History, New York, NY. Available at ftp.amnh.org/pub/molecular/poy.

GLYNN, P.W. 1993. Coral reef bleaching: ecological perspectives. *Coral Reefs* **12:** 1-17.

GOLD, J.R. & RICHARDSON, .LR. 1998. Mitochondrial DNA diversification and population structure in fishes from the Gulf of Mexico and Western Atlantic. *Journal of Heredity* **89:** 404-414.

GOMON, M.F. 1997. A remarkable new pygmy seahorse (Syngnathidae: *Hippocampus*) from southeastern Australia, with a rediscription of *H. bargibanti* Whitley from New Caledonia. *Memoirs of the Museum of Victoria* **56:** 245-253.

GOREMYKIN, V., HANSMANN, S. & MARTIN, W. 1997. Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: Revised molecular estimates of two seed plant divergence times. *Plant Systematics and Evolution* **206:** 337-351.

GORMAN. G.C. & KIM, Y.J. 1977. Genotype evolution in the face of phenotypic assertiveness: *Abudefduf* (Pomacentridae) from the Atlantic and Pacific sides of Panama. *Copeia* **1977:** 694-697.

GOSLING, E.M. 1994. Speciation and species concepts in the marine environment. In: *Genetics and Evolution of Aquatic Organisms*, (ed) A.R. Beaumont, pp. 1-14. Chapman & Hall, London.

GOSS, R.J. & STAGG, M.W. 1957. The regeneration of fins, and fin rays *in Fundulus heteroclitus*. *Journal of Experimental Zoology* **136:** 487-508.

GRANGE, N. & CRETCHLEY, R. 1995. A preliminary investigation of the reproductive behaviour of the Knysna seahorse, *Hippocampus capensis* Boulenger, 1900. *South African Journal of aquatic Science* **21:** 103-104.

GRANT, W.S. & LESLIE, R.W. 2001. Inter-ocean dispersal is an important mechanism in the zoogeography of hake (Pisces: *Merluccius* spp.). *Journal of Biogeography* **28:** 699-721.

GRAYBEAL, A. 1993. The phylogenetic utility of cytochrome *b*: Lessons from bufonid frogs. *Molecular Phylogenetics and Evolution* **2:** 256-269.

GREENBERG, R., CORDERO, P.J., DROEGE, S. & FLEISCHER, R.C. 1998. Morphological adaptation with no mitochondrial DNA differentitation in the Coastal Plain Swamp Swallow. *Auk* **115:** 706-712.

GREENFIELD, D.W. 1968. The zoogeography of *Myripristis* (Pisces: Holocentridae). *Systematic Zoology* **17:** 76-87.

GREWE, P.M., BILLINGTON, N. & HEBERT, P.D.N. 1990. Phylogenetic relationships among Salvelinus as inferred from mitochondrial DNA divergence. *Canadian Journal of Fisheries and Aquatic Sciences* **47:** 984-991.

GRIBNITZ K.-H., KENT, L.E. 1989. Marine and fluviatile terraces and colluvial fans reflecting Quaternary climate changes, Gough Island, South Atlantic. *South African Journal of Antarctic Research* **13:** 2-9.

GRINDLEY, J.R. 1985. *Estuaries of the Cape; Part II: Synopses of avaliable information on individual systems; Report No. 30: Knysna (CMS 13)*. CSIR research report 429, Stellenbosch, South Africa.

GROTH, J.G. & BARROWCLOUGH, G.F. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. *Molecular Phylogenetics and Evolution* **12:** 115-123.

GYLLENSTEN, U. 1985. The genetic structure of fish: Differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous and freshwater species. *Journal of Fish Biology* **26:** 691-699.

HAINES, G.B. & MODDE, T. 1996. Evaluation of marking techniques to estimate population size and first-year survival of Colorado Squawfish. *North American Journal of Fisheries Management* **16:** 905-912.

HALL, R. 1996. Reconstructing Cenozoic SE Asia. In: *Tectonic evolution of Southeast Asia*, (eds) R. Hall & D.J. Blundell, pp. 153-184. Geological Society of London Special Publication 106.

HARLAND, W.B., COX, A.V., LLEWELLYN, P.G., PICKTON, C.A.G, SMITH, A.G. & WALTERS, R. 1989. *A geologic time scale*. Cambridge University Press, United Kingdom.

HARPENDING, H.C. 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66:** 591-600.

HARTL, D.L. & CLARK, A.G. 1988. *Principles of population genetics*. Second edition. Sinauer Associates, Inc., Sunderland, Massachusetts.

HASTINGS, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57:** 97-109.

HAUG, H.H. & TIEDEMANN, R. 1998. Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. *Nature* **393:** 673-676.

HENDERSON-SELLERS, A. & ROBINSON, P.J. 1986. *Contemporary Climatology*. Wiley & Sons Inc., New York.

HERALD, E.S. 1959. From pipefish to seahorse – a study of phylogenetic relationships. *Proceedings of the California Academy of Sciences* **29:** 465-473.

HEYDORN, A.E.F. 1978. Ecology of the Agulhas Current region: an assessment of biological responses to environmental parameters in the south-west Indian Ocean. *Transactions of the Royal Society of South Africa* **43:** 151-190.

HEYDORN, A.E.F. & TINLEY, K.L. 1980. *Estuaries of the Cape, Part 1: Synopsis of the Cape coast, natural features, dynamics and utilisation*. CSIR Research Report No. 380.

HILLIS, D. M. & DIXON, M. T. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Quarterly Review of Biology* **66:** 411-453.

HILLIS, D.M. & HUELSENBECK, J.P. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *Journal of Heredity* **83:** 189-195.

HILTON-TAYLOR, C. 2000. *2000 IUCN Red List of Threatened Species*. IUCN, Gland, Switzerland and Cambridge, UK.

HOELZEL, A.R., HANCOCK, J.M. & DOVER, G.A. 1991. Evolution of the cetacean mitochondrial D-loop region. *Molecular Biology and Evolution* **8:** 475-493.

HORNE, M.L. 2001. A new seahorse species (Syngnathidae: *Hippocampus*) from the Great Barrier Reef. *Records of the Australian Museum* **53:** 243-246.

HOULDEN, B.A., COSTELLO, B.H., SHARKEY, D., FOWLER, E.V., MELZER, A., ELLIS, W., CARRICK, F., BAVERSTOCK, P.R. & ELPHINSTONE, M.S. 1999. Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinerus* (Goldfuss 1817). *Molecular Ecology* **8:** 999-1011.

HSÜ, K.J & BERNOULLI, D. 1978. *Genesis of the Tethys and the Mediterranean.* Initial Report DSDP.

HUELSENBECK, J.P., LARGET, B. & SWOFFORD, D. 2000. A compound Poisson process for relaxing the molecular clock. *Genetics* **154:** 1879-1892.

HUELSENBECK, J.P. & BOLLBACK, J.P. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. *Systematic Biology* **50:** 351-366.

HUELSENBECK, J.P. & RONQUIST, F. 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17:** 754-755.

HUELSENBECK, J.P. & IMENNOV, N.S. 2002. Geographic origin of human mitochondrial DNA: accommodating phylogenetic uncertainty and model comparison. *Systematic Biology* **51:**155-165.

HUTSON, W.H. 1980. The Agulhas current during the late Pleistocene: analysis of modern faunal analogs. *Science* **207:** 64-66.

INGHAM, M.C. 1970. Coastal upwelling in the northwestern Gulf of Guinea. *Bulletin of Marine Science* **20:** 2-34.

JANDEL CORPORATION 1995. SIGMASTAT for Windows version 2.0 statistical software, San Rafael, CA.

JONES, A.G., MOORE, G.I., KVARNEMO, C., WALKER D. & AVISE, J.C. 2003. Sympatric speciation as a consequence of male pregnancy in seahorses. *Proceedings of the National Academy of Sciences of the USA* **100:** 6598-6603.

JORDAN, D.S. 1908. A law of geminate species. *American Naturalist* **42:** 73-80.

JORGENSEN, R.A. & CLUSTER, P.D. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. *Annals of the Missouri Botanical Garden* **75:** 1238-1247.

KANEPS, A.G. 1979. Gulf Stream: velocity fluctuations during the late Cenozoic. *Science* **204:** 297-301.

KEIGWIN, L.D. 1978. Pliocene closing of the isthmus of Panama based on biostratigraphic evidence from nearby Pacific and Caribbean sea cores. *Geology* **6:** 630-634.

KELLER, G. & BARRON, J.A. 1983. Paleoceanographic implications of Miocene deep-sea hiatures. *Geological Society of America Bulletin* **97:** 590-613.

KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16:** 111-120.

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

KOCHER, T.D., THOMAS, W.K., MEYER, A., EDWARDS, S.V. PÄÄBO, S., VILLABLANCA, F.X. & WILSON, A.C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the USA* **86:** 6196-6200.

KORRINGA, P. 1956. *Oyster culture in South Africa. Hydrological, biological and ostreological observations in the Knysna lagoon, with notes on conditions in other South African waters.* Investigational Report of the Division of Sea Fisheries of South Africa 20.

KUHNER, M., YAMAMOTO, J. & FELSENSTEIN, J. 1995. Estimating effective population size and mutation rates from sequence data using Metropolis-Hastings sampling. *Genetics* **140:** 1421-1430.

KUHNER, M., YAMATO, J. & FELSENSTEIN, J. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149:** 429-434.

KUITER, R.H. 2000. *Seahorses, pipefishes and their relatives – a comprehensive guide to Syngnathiformes*. TMC Publishing, Chorleywood, UK.

KUITER, R.H. 2001. Revision of the Australian seahorses of the genus *Hippocampus* (Syngnathiformes: Syngnathidae) with a description of nine new species. *Records of the Australian Museum* **53:** 293-340.

KUMAR, S., TAMURA, K., JAKOBSEN, I.B. & NEI, M. 2001. MEGA version 2.1. Molecular Evolutionary Genetics Analysis software, Arizona State University, Tempe, Arizona, USA.

KVARNEMO, C., MOORE, G.I., JONES, A.G., NELSON, W.S. & AVISE, J.C. 2000. Monogamous pair bonds and mate switching in the Western Australian seahorse *Hippocampus subelongatus*. *Journal of Evolutionary Biology* **13:** 882-888.

KWON, O.Y., OGINO, K. & ISHIKAWA 1991. The largest 18S ribosomal RNA ever known. Nucleotide sequence and presumed secondary structure of the 18S rRNA of the pea aphid, Acyrthosiphon pisum. *European Journal of Biochemistry* **202:** 827-833.

LANDE, R. 1980. Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist* **116:** 463-479.

LANDE, R. 1995. Breeding plans for small populations based on the dynamics of quantitative genetic variance. In: *Population management for survival and recovery*, (eds) J.D. Ballou, M. Gilpin & T. Foose, pp. 318-340. Columbia University Press, New York.

LARGIER, J.L., ATTWOOD, C., HARCOURT-BALDWIN, J.L. 2000. The hydrographic character of the Knysna Estuary. *Transactions of the Royal Society of South Africa* **55:** 107-122.

LAVOUÉ, S., SULLIVAN, J.P. & HOPKINS, C.D. (submitted). Phylogenetic utility of the first two introns of the S7 ribosomal protein gene in African electric fishes (Teleostei: Mormyroidea) and congruence with other molecular markers. *Molecular Phylogenetics and Evolution.*

LE CHEMINANT, J. 2000. *An investigation to establish the suitability of Opalithplattchen (VIFE) tagging systems for marking the Knysna seahorse, Hippocampus capensis.* M.Sc. thesis, Bournemouth University, United Kingdom.

LE LŒUFF, P. & VON COSEL, R. 1998. Biodiversity patterns of the marine benthic fauna on the Atlantic coast of tropical Africa in relation to hydroclimatic conditions and paleogeographic events. *Acta Oecologica* **19**: 309-231.

LEE, W., CONROY, J., HOWELL, W.H. & KOCHER, T.D. 1995. Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution* **41:** 54-66.

LEBERG, P.L. 1991. Effects of genetic variation on the growth of fish populations: conservation implications. *Journal of Fish Biology* **37:** 193-195.

LESSIOS, H.A., KESSING, B.D., ROBERTSON, D.R. & PAULAY, G. 1999. Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. *Evolution* **53:** 806-817.

LESSIOS, H.A., KESSING, B.D. & PEARSE, J.S. 2001. Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution* **55:** 955-975.

LI, W.-H. 1977. Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics* **85:** 331-337.

LINDESAY, J. 1998. Past climates of southern Africa. In: *Climates of the Southern Continents: present, past and future*, (eds) J.E. Hobbs, J.A. Lindesay & H.A. Bridgman, pp. 161-206. Wiley & Sons Inc., New York.

LIPTROT, M.R.M. 1978. *Community metabolism and phosphorous dynamics in a seasonally closed South African estuary*. M.Sc. thesis, Rhodes University, Grahamstown.

LOCKYEAR, J.F., KAISER, H. & HECHT, T. 1997. Studies on captive breeding of the Knysna seahorse (*Hippocampus capensis*). *Aquarium Sciences and Conservation* **1:** 129-136.

LOURIE, S.A., VINCENT, A.C.J. & HALL, H. J. 1999. *Seahorses: an identification guide to the world's species and their conservation.* Project Seahorse, London, UK.

LOURIE, S.A. & RANDALL, J.E. 2003. A new pygmy seahorse, *Hippocampus densise* (Teleostei: Syngnathidae) from the Indo-Pacific. *Zoological studies* **42:** 284-291.

LOURIE, S.A. (in press). Signatures of Sundaland: Phylogeography of the seahorse *Hippocampus kuda* (Teleostei: Syngnathidae) in Southeast Asia. *Evolution.*

LUIKART, G., ALLENDORF, F.W., CORNUET, J.-M., & SHERWIN, W.B. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. *The Journal of Heredity* **89:** 238-247.

LYDEARD, C. & ROE, K.J. 1997. The phylogenetic utility of the mitochondrial cytochrome b gene for inferring relationships among Actinopterygian fishes. In: *Molecular systematics*

of fishes, (eds) T.D. Kocher & C.A. Stepien, pp. 285-303. Academic Press, San Diego, California.

MADDISON, W.P. & MADDISON, D.R. 2003. MacClade: analysis of phylogeny and character evolution, version 4.06. Sinauer, Sunderland, MA.

MALAIHOLLO, J.F.A. & HALL, R. 1996. The geology and tectonic evolution of the Bacan region, east Indonesia. In : *Tectonic evolution of southeast Asia*, (eds) R. Hall & D. Blundell. Geological Society Special Publication No. 106, pp. 483-497.

MAMURIS, Z., STAMATIS, C. & TRIANTAPHYLLIDIS, C. 1999. Intraspecific genetic variation of striped red mullet (*Mullus surmuletus* L.) in the Mediterranean Sea assessed by allozyme and random amplified polymorphic DNA (RAPD) analyses. *Heredity* **83:** 30-38.

MARINARI, E. & PARISI, G. 1992. Simulated tempering: a new Monte Carlo scheme. *Europhysics Letters* **19:** 451-458.

MARSHALL, L.G. 1988. Land mammals and the Great American Interchange. *American Scientist* **76:** 380-388.

MARTIN, A.R.H. 1962. Evidence relating to the Quaternary history of the Wilderness Lakes. *Transactions of the Proceedings of the Geological Society of South Africa* **65:** 19-42.

MARTIN, A.P., KESSING, B.D. & PALUMBI, S.R. 1990. Accuracy of estimating genetic distance between species from the short sequences of mitochondrial DNA. *Molecular Biology and Evolution* **7:** 485-488.

MATTHEE, C.A. & DAVIS, S.K. 2001. Molecular insights into the evolution of the family Bovidae: a nuclear DNA perspective. *Molecular Biology and Evolution* **18:** 1220-1230.

MATTHEE, C.A., BURZLAFF, J.D., TAYLOR, J.F. & DAVIS, S.K. 2001. Mining the mammalian genome for Artiodactyl systematics. *Systematic Biology* **50:** 367-390.

MATTHEW, W.D. 1915. Climate and evolution. *Annals of the New York Academy of Sciences* **24:** 171-318.

MAYR, E. 1954. Change of genetic environment and evolution. In: *Evolution as a process*, (eds) J. Huxley, A.C. Hardy & E.B. Ford, pp. 157-180. Allen & Unwin, London.

MAYR, E. 1963. *Animal species and evolution.* Harvard University Press, Cambridge, MA.

MCMANUS, J.W. 1985. Marine speciation, tectonics and sea-level changes in southeast Asia. *Proceedings of the 5th international Coral Reef Congress* **4:** 133-138.

MCMILLAN, W.O. & PALUMBI, S.R. 1995. Concordant evolutionary patterns among Indo-West Pacific butterflyfishes. *Proceedings of the Royal Society of London, B Series* **260:** 229-236.

MENON, A.G.K. 1977. A systematic monograph of the tongue soles of the genus *Cynoglossus* Hamilton-Buchanan (Pisces: Cynoglossidae). *Smithsonian Contributions to Zoology* **283:** 1-129.

METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H. & TELLER, E. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21:** 1087-1092.

MEYER, A. 1994. Shortcomings of the cytochrome *b* gene as a molecular marker. *Trends in Ecology and Evolution* **9:** 278-280.

MEYER, A., KOCHER, T.D., BASASIBWAKI, P. & WILSON, A.C. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347:** 550-553.

MEYERS, P.A., BRASSELL, S.C., HUC, A.Y., BARRON, E.J., BOYECE, R.E., DEAN, W.E., HAY, W.W., KEATING, B.H., MCNULTY, C.L., NOHARA, M., SCHALLREUTER, R.E., SIBUET, J.-C., STEINMETZ, J.C., STOW, D. & STADNER, H. 1983. Organic geochemistry of sediments recovered by DSDP/IPOD leg 75 from under the Benguela Current. In: *Coastal upwelling: its sedimentary record Part B, Sedimentary records of ancient and coastal upwelling*, (eds) J. Thiede & E. Suess, pp. 453-466. Plenum Press, New York.

MILINKOVITCH, M. C., LEDUC, R. G., ADACHI, J., FARNIR, F., GEORGES, M. & HASEGAWA, M. 1996. Effects of character weighting and species sampling on phylogeny reconstruction: A case study based on cetaceans. *Genetics* **144:** 1817-1833.

MILINKOVITCH, M. C. & LYONS-WEILER, J. 1998. Finding optimal ingroup topologies and convexities when the choice of outgroup is not obvious. *Molecular Phylogenetics and Evolution* **9:** 348-357.

MINDELL, D.P. & HONEYCUTT, R.L. 1990. Ribosomal RNA in vertebrates: Evolution and phylogenetic implications. *Annual Review of Ecology and Systematics* **21:** 541-566.

MIYA, C.M. & NISHIDA, M. 1997. Speciation in the open ocean. Nature 389: 803-804.Miyamoto, M.M. and Fitch, W.M. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology* **44:** 236-249.

MIYAMOTO, M.M. & FITCH, W.M. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology* **44:** 236-249.

MORAN, N.A., MUNSON, M.A., BAUMANN, P. & ISHIKAWA, H. 1993. A molecular clock in endosymbiontic bacteria calibrated using the insect host. *Proceedings of the Royal Society of London, B Series* **253:** 167-171.

MORITZ, C. 1994. Defining 'Evolutionary Significant Units' for conservation. *Trends in Ecology and Evolution* **9:** 373-375.

MORITZ, C., DOWLING, T.E. & BROWN, W.M. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18:** 269-292.

MOYA, A., GALIANA, A. & AYALA, F.J. 1995. Founder-effect speciation theory: failure of experimental corroboration. *Proceedings of the National Academy of Sciences of the USA* **92:** 3983-3986.

MUKAI, T., TSUIHIJI, T., SATO, T. & MORISAWA, M. 2001. Mitochondrial DNA divergence in the seahorse, *Hippocampus coronatus* (Syngnathiformes: Syngnathidae), collected from Sagami Bay. *Japanese Journal of Ichthyology* **47:** 139-143.

MULLER, K.L. 1998. The role of conspecifics in habitat settlement in a territorial grasshopper. *Animal Behaviour* **56:** 479-485.

MULLER, K.L., STAMPS, J.A., KRISHNAN, V.V. & WILLITS, N.H. 1997. The effects of conspecific attraction and habitat quality on habitat selection in territorial birds (*Troglodytes aedon*). *The American Naturalist* **150:** 650-661.

MURPHY, W.J., EIZIRIK, E., JOHNSON, W.E., ZHANG, Y.P., RYDER, O.A., O'BRIEN, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* **409:** 614-618.

MURRELL, A., CAMPBELL, N.J. & BARKER, S.C. 2001. A total-evidence phylogeny of ticks provides insights into the evolution of life cycles and biogeography. *Molecular Phylogenetics and Evolution* **21:** 244-258.

MUSS, A., ROBERTSON, D.R., STEPIEN, C.A., WIRTZ, P. & BOWEN, B.W. 2001. Phylogeography of *Ophioblennius*: the role of ocean currents and geography in reef fish evolution. *Evolution* **55:** 561-572.

MYERS, G.S. 1938. Fresh-water fishes and West Indian zoogeography. *Smithsonian Report* **1937:** 339-364.

NEI, M. & TAJIMA, F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* **97:** 145-163.

NESBØ, C.L., RUENESS, E.K., IVERSON, S.A., SKAGEN, D.W. & JAKOBSEN, K.S. 2000. Phylogeography and population history of Atlantic mackeral (*Scomber scombrus* L.): a genealogical approach reveals genetic structuring among the eastern Atlantic stocks. *Proceedings of the Royal Society of London, Series B* **267:** 281-292.

NEWMAN, W.A., JUMARS, P.A. & ROSS, A. 1976. Diversity trends in coral-inhabiting barnacles (Cirripedia, Pyrgomatinae). *Micronesica* **12:** 69-82.

NIELSEN, L.A. 1992. *Methods of Marking Fish and Shellfish.* American Fisheries Society Special Publication 23. Bethesda, Maryland, USA.

NIELSEN, J.L., FOUNTAIN, M.C. & WRIGHT, J.M. 1997. Biogeographic analysis of Pacific trout (*Oncorhynchus mykiss*) in California and Mexico based on mitochondrial DNA and nuclear microsatellites. In: *Molecular systematics of fishes*, (eds) T.D. Kocher & C.A. Stepien, pp. 53-69. Academic Press, San Diego, California.

NIXON, K.C. 1998. DADA version 1.4.1. Cornell University, Ithaca, New York.

NIXON, K.C. 1999-2002. WINCLADA version 1.0000. Published by the author. Ithaca, New York.

NOBLE, R.G. & HEMENS, J. 1978. *Inland water ecosystems in South Africa - a review of research needs*. South African Natural Science Programmes Report 34, Pretoria.

NUNNEY, L. & CAMPBELL, K.A. 1993. Assessing minimum viable population size: demography meets population genetics. *Trends in Ecology and Evolution* **8:** 234-239.

NYAKAANA, S. & ARCTANDER, P. 1999. Population genetic structure of the African elephant in Uganda based on variation at mitochondrial and nuclear loci: evidence for male-biased gene flow. *Molecular Ecology* **8:** 1105-1115.

O'BRIEN, SJ 1994. Genetic and phylogenetic analyses of endangered species. *Annual Review of Genetics* **28:** 467-489.

OCHMAN, H. & WILSON, A.C. 1987. Evolution of bacteria: evidence for a universal substitution rate in cellular genomes. *Journal of Molecular Evolution* **26:** 74-86.

ORTÍ, G., PETRY, P., PORTO, J.I.R., JE'GU, M. & MEYER, A. 1996. Patterns of nucleotide change in mitochondrial ribosomal RNA genes and the phylogeny of Piranhas. *Journal of Molecular Evolution* **42:** 169-182.

ORTÍ, G. & MEYER, A. 1997. The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. *Systematic Biology* **46:** 75-100.

PALUMBI, S.R. 1992. Marine speciation on a small planet. *Trends in Ecology and Evolution* **7:** 114-118.

PALUMBI, S.R. 1996. Nucleid acids II: The polymerase chain reaction. In: *Molecular Systematics*, (eds) D.M. Hillis, C. Moritz & B.K. Mable, pp. 205-247. Sinauer Associates Inc., Sunderland, Maryland.

PATTERSON, C. 1993. Chapter 36. Osteichthyes: Teleostei. In: *The Fossil Record 2*, (ed) M.J. Benton, pp. 621-656. Chapman & Hall, London, England.

PENRITH, M.J. & PENRITH, M.L. 1972. Redescription of *Pandaka silvana* (Barnard) (Pisces: Gobiidae). *Annals of the South African Museum* **60:** 105-108.

PIGRAM, C.J. & DAVIS, H.L. 1987. Terranes and the accretion history of the New Guinea region. *BMR Journal of Australian Geology and Geophysics* **10:** 193-211.

PITMAN, W.V. 1981. *Surface water resources in South Africa.* Vol. IV. Drainage regions EGHJKL The Western Cape. Part 1 (Text) Part 2 (Appendices). Johannesburg. Hydrological Research Unit Report No. 13/81.

PLANES, S. 2002. Biogeography and larval dispersal inferred from population genetic analysis. In: *Coral reef fishes. Dynamics and diversity in a complex ecosystem*, (ed) P.F. Sale, pp. 201-220. Academic Press, San Diego, California.

PLANES, S., DOHERTY, P.J., BERNARDI, G. 2001. Strong genetic divergence among populations of marine fish with limited dispersal, *Acanthochromis polyacanthus*, within the Great Barrier Reef and Cora Sea. *Evolution* **55:** 2263-2273.

PODOLSKY, R.H. 1990. Effectiveness of social stimuli in attracting laysan albatross to new potential nesting sites. *Auk* **107:** 119-125.

POGSON, G.H., MESA, K.A. & BOUTILIER, R.G. 1995. Genetic population structure and gene flow in the Atlantic Cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics* **139:** 375-385.

POSADA, D. & CRANDALL, K.A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14:** 817-818.

POWERS, D.A., LAUERMAN, T., CRAWFORD, D. & DIMICHELE, L. 1991. Genetic mechanisms for adapting to a changing environment. *Annual Review of Genetics* **25:** 629-659.

RAMBAUT, A. & BROMHAM, L.B. 1998. Estimating divergence dates from molecular sequences. *Molecular Biology and Evolution* **15:** 442-448.

REBERT, J.P. 1983. *Hydrologie et dynamique des eaux du plateau continental sénégalais.* Document Scientifique CRODT 89.

REDDERING, J.S.V & ESTERHUYSEN, K. 1984. *Sedimentation in the Knysna estuary.* ROSIE Report No. 9. University of Port Elizabeth, Dept. of Geology.

REED, J.M. & DOBSON, A.P. 1993. Behavioural constraints and conservation biology: conspecific attraction and recruitment. *Trends in Ecology and Evolution* **8:** 253-255.

RICE, W.R. & HOSTERT, E.E. 1993. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* **47:** 1637-1653.

RILEY, A.K. 1986. *Aspekte van die soutgehalte toleransie van die Knysna seeperdjie, Hippocampus capensis (Boulenger, 1900) in die Knysna estuarium.* Unpublished Report.

RINGO, J., WOOD, D., ROCKWELL, R. & DOWSE, H. 1985. An experiment testing two hypotheses of speciation. *The American Naturalist* **126:** 642-661.

ROBBINS, L.W., TOLLIVER, D.K., & SMITH, M.H. 1989. Nondestructive methods for obtaining genotypic data from fish. *Conservation Biology* **3:** 88-91.

RODRÍGUEZ, F., OLIVER, J.F., MARÍN, A. & MEDINA, J.R. 1990. The general stochastic model of nucleotide substituion. *Journal of Theoretical Biology* **142:** 485-501.

ROGERS, A.R. & HARPENDING, H. 1992. Population growth makes waves in the distribution of pair-wise genetic differences. *Molecular Biology and Evolution* **9:** 552-569.

ROGERS, J. & KIDD, K.K. 1996. Nucleotide Polymorphism, effective population size, and dispersal distances in the yellow baboon (*Papio hamadryas cynocephalus*) of Mikumi National Park, Tanzania. *American Journal of Primatology* **38:** 157-168.

RÖGL, F. & STEININGER, F.F. 1983. Vom Zerfall der Tethys zu Mediterran und Paratethys. Die neogene Palaogeographie and Palinspastik des zirkum-mediterranen Raumes. *Annalen des naturhistorischen Museums Wien* **85:** 135-163.

RÖGL, F. & STEININGER, F.F. 1984. Neogene Paratethys, Mediterranean and Indo-Pacific seaways. Implications for the paleobiogeography of marine and terrestrial biotas. In: *Fossils and climate*, (ed) P.J. Brenchley, pp. 171-200. Wiley, Chichester.

ROHDE, K. & HAYWARD, C.J. 2000. Oceanic barriers as indicated by scombrid fishes and their parasites. *International Journal for Parasitology* **30:** 579-583.

ROSEN, B.R. 1975. A vicariance model of Caribbean biogeography. *Systematic Zoology* **24:** 431-464.

ROSEN, B.R. 1984. Reef coral biogeography and climate through the late Cainozoic: just islands in the sun or a critical pattern of islands? In: *Fossils and climate*, (ed) P.J. Brenchley, pp. 201-262. Wiley, Chichester.

ROSEN, B.R. 1988. Progress, problems and patterns in the biogeography of reef corals and other tropical marine organisms. *Helgoländer Meeresuntersuchungen* **42:** 269-301.

ROUX, M. 1987. Evolutionary ecology and biogeography of recent stalked crinoids as a model for the fossil record. In: *Echinoderm studies*, volume 2, (eds) M. Jangoux & J.M. Lawrence, pp. 103-137. A.A. Balkema, Rotterdam, The Netherlands.

ROZAS, J. & ROZAS, R. 1999. DNASP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15:** 174-168.

RUSSELL, I.A. 1994. Mass mortality of marine and estuarine fish in the Swartvlei and Wilderness Lake systems, Southern Cape. *South African Journal of aquatic Sciences* **20:** 93-96.

SAMBROOK, J., FRITSCH, E. & MANIATIS, T. 1989. *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SANDERSON, M.J 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14:** 1218-1232.

SANDERSON, M. J. & DONOGHUE, M. J. 1989. Patterns of variation in levels of homoplasy. *Evolution* **43:** 1781-1795.

SCHELTEMA, R.S. 1988. Initial evidence for the transport of teleplanic larvae of benthic invertebrates across the east Pacific barrier. *Biological Bulletin* **174:** 145-152.

SCHNEIDER, S. & EXCOFFIER, L. 1999. Estimation of demographic parameters from the distribution of pair-wise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152:** 1079-1089.

SCHNEIDER, S., ROESSLI, D. & EXCOFFIER, L. 2000. *ARLEQUIN version 2000: A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

SCHUMANN, E.H. 2000. Oceanic exchanges and temperature variability in the Knysna Estuary. *Transactions of the Royal Society of South Africa* **55:** 123-128.

SCHUMANN, LARGIER, J & SLINGER, J. 1999. Estuarine hydrodynamics. In: *Estuaries of South Africa*, (eds) B.R. Allanson & D. Baird, pp. 27-52. Cambridge University Press, Cambridge.

SENSABAUGH, G.F. 1994. DNA typing of biological evidence material. In: *Ancient DNA*, (eds) B. Herrman & S. Hummel, pp. 141-148. Springer Verlag, New York.

SHACKLETON, N.J. & OPDYKE, N.D. 1977. Oxygen isotope and paleomagnetic evidence for early northern hemisphere glaciation. *Nature* **270:** 216-219.

SHAFFER, M.L. 1981. Minimum population sizes for species conservation. *BioScience* **31:** 131-134.

SHANNON, L.V. 1985. The Benguela ecosystem. I. Evolution of the Benguela physical features and processes. *Oceanography and Marine Biology: an annual review* **23:** 105-182.

SHIELDS, W.M. 1987. Dispersal and mating systems : investigating their causal connections. In: *Patterns of dispersal among mammals and their effect on the genetic structure of populations*, (eds) D. Chepko-Sade & Z.T. Halpin, pp. 3-24. University of Chicago Press, Chicago.

SHIMODAIRA, H. & HASEGAWA, M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16:** 1114-1116.

SIMON, C., FRATTI, F., BECKENBACK, A., CRESPI, B., LIU, H. & FLOOK, P. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomology Society of America* **87:** 651-701.

SKELTON, P.H. 1987. South African Red Data Book - Fishes. *South African National Scientific Programmes Report 137.*

SLATKIN, M. & HUDSON, R.R. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129:** 555-562.

SMITH, M.M. 1981. The seahorse of the Knysna Lagoon. *African Wildlife* **35:** 5-9.

SMITH, P.J. & FUJIO, Y. 1982. Genetic variation in marine teleosts: high variability in habitat specialists and low variability in habitat generalists. *Marine Biology* **69:** 7-20.

SMITH, M.W., CHAPMAN, R.W. & POWERS, D.A. 1998. Mitochondrial DNA analysis of Atlantic Coast, Chesapeake Bay, and Delaware Bay populations of the teleost *Fundulus heteroclitus* indicates temporally unstable distributions over geologic time. *Molecular Marine Biology and Biotechnology* **7:** 79-87.

SONNENFELD, P. 1985. Models of upper Miocene evaporite genesis in the Mediterranean region. In: *Geological evolution of the Mediterranean Basin*, (eds) D.J. Stanley & F.C. Wezel, pp. 323-346. Springer Verlag, New York.

SORBINI, L. 1988. Biogeography and climatology of Pliocene and Messinian fossil fish of Eastern Central Italy. *Bollettino del Museo civico di storia naturale di Verona* **14:** 1-85.

SPECHT, R.L. 1981. Biogeography of halophytic angiosperms (saltmarsh, mangrove and sea-grass). In: *Ecological biogeography of Australia*, (ed) A. Keast, pp. 557-589. W. Junk, The Hague, The Netherlands.

STAMPS, J.A. 1988. Conspecific attraction and aggregation in territorial species. *The American Naturalist* **131:** 329-347.

STEHLI, F.G. & WELLS, J.W. 1971. Diversity and age patterns in hermatypic corals. *Systematic Zoology* **20:** 115-126.

STEININGER, F.F. & RÖGL, F. 1984. Paleogeography and palinspastic reconstruction of the Neogene of the Mediterranean and Paratethys. In: *The geological evolution of the eastern Mediterranean*, (eds) J.E. Dixon & A.H.F Robertson, pp. 659-668. Blackwell, Oxford.

STEPIEN, C.A. & KOCHER, T.D. 1997. Molecules and morphology in studies of fish evolution. In: *Molecular systematics of fishes*, (eds) T.D. Kocher & C.A. Stepien, pp. 1-11. Academic Press, San Diego, California.

SUTTON, D.C. & HOEGH-GULDBERG, O. 1990. Host-zooxanthellae interactions in four temperate marine invertebrate symbioses: assessment of effect of host extracts on symbionts. *The Biological Bulletin* **178:** 175-186.

SWOFFORD, D.L. 2002. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 beta 10*. Sinauer Associates, Sunderland, Massachusetts.

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

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

TAKAHASHI, H. & GOTO, A. 2001. Evolution of East Asian sticklebacks as shown by mitochondrial DNA control region sequences. *Molecular Phylogenetics and Evolution* **21:** 135-155.

TAMURA, K. & NEI, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10:** 512-526.

TAYLOR, B.E. 1991. A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture* **98:** 185-297.

TEMPLETON, A.R. 1980. The theory of speciation via the founder principle. *Genetics* **94:** 1011-1038.

TEMPLETON, A.R. 1981. Mechanisms of speciation - a population genetics approach. *Annual Review of Ecology and Systematics* **12:** 23-48.

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

TEMPLETON, A.R. 1999. Experimental tests of genetic transilience. *Evolution* **53:** 1628-1632.

TEMPLETON, A.R., CRANDALL, K.A. & SING, C.F. 1992. A cladistic analysis of phentoypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132:** 619-633.

TESKE, P.R. & WOOLDRIDGE, T.H. 2001. A comparison of the macrobenthic faunas of permanently open and temporarily open/closed South African estuaries. *Hydrobiologia* **464:** 227-243.

TESKE, P.R. & WOOLDRIDGE, T.H. 2003. What limits the distribution of subtidal macrobenthos in permanently open and temporarily open/closed South African estuaries? Salinity vs. sediment particle size. *Estuarine, Coastal and Shelf Science* **57:** 225-238.

THOMAS, W.K., WITHLER, R.E. & BECKENBACH, A.T. 1986. Mitochondrial DNA analysis of Pacific salmonid evolution. *Canadian Journal of Zoology* **64:** 1059-1064.

THOMPSON, J.D., GIBSON, T.J, PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24:** 4876-4882.

THORNE, J.L., KISHINO, H. & PAINTER, I.S. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* **15:** 1647-1657.

THORNE, J.L. & KISHINO, H. 2002. Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* **51:** 689-702.

TOEFFIE, Z. 2000. *A preliminary examination of the morphology and genetic structure within populations of the Knysna seahorse, Hippocampus capensis (Pisces: Syngnathidae), from two South African estuaries*. MSc thesis, University of the Western Cape.

TOERIN, D.K. 1979. *The geology of the Oudtshoorn area.* Explanation to sheet 3322. Department of mines, Pretoria. Government Printer.

TOPS, S. 1999. *An investigation into the effects of habitat complexity and food types on the behaviour of the Knysna seahorse – Hippocampus capensis*. Honours thesis, University of St. Andrews.

TSUCHI, R. & INGLE, J.C. Jr. 1992. *Pacific Neogene: Environment, evolution and events.* University of Tokyo Press, Tokyo.

TURPIE, J.K., BECKLEY, L.E. & KATUA, S.M. 2000. Biogeography and the selection of priority areas for conservation of South African coastal fishes. *Biological Conservation* **92:** 59-72.

VAN DE PEER, Y., BALDAUF, S.L., DOOLITTLE, W.F. & MEYER, A. 2000. A comprehensive rate-calibrated phylogeny of (crown) eukaryotes using rRNA sequences and comparison with protein data. *Journal of Molecular Evolution* **51:** 565-576.

VANE-WRIGHT, R.I., HUMPHRIES, C.J. & WILLIAMS, P.H. 1991. What to protect - systematics and the agony of choice. *Biological Conservation* **55:** 235-254.

VINCENT, A.C.J. 1995. A role for daily greetings in maintaining seahorse pair bonds. *Animal Behaviour* **49:** 258-260.

VINCENT, A.C.J. & SADLER, L.M. 1995. Faithful pair bonds in wild seahorses, *Hippocampus whitei*. *Animal Behaviour* **50:** 1557-1569.

VORIS, H.K. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* **27:** 1153-1167.

VRIELYNK, B., ODIN, G.S. & DERCOURT, J. 1997. Reconstruction of Mediterranean Late Cenozoic hydrography by means of carbon isotope analyses. In: *Reconstruction of Marine Paleoenvironments*, (ed) J.E. Meulenkamp, pp. 25-47. Utrecht Micropaleontological Bulletin.

WATERS, J.M. & BURRIDGE, C.P. 1999. Extreme intraspecific mitochondrial DNA sequence divergence in *Galaxias maculatus* (Osteichthyes: Galaxiidae), one of the world's most widespread freshwater fish. *Molecular Phylogenetics and Evolution* **11:** 1-12.

WATERS, J.M., CRAW, D., YOUNGSON, J.H. & WALLIS, G.P. 2001. Genes meet geology: fish phylogeographic pattern reflects ancient, rather than modern, drainage connections. *Evolution* **55:** 1844-1851.

WATTERSON, G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7:** 256-276.

WAYNE, R.K., FOREMAN, L., NEWMAN, A.K., SIMONSON, J.K. & O'BRIEN, S.J. 1986. Genetic monitors of zoo populations: morphological and electrophoretic assays. *Zoo Biology* **5:** 215-232.

WELLS, P.E., WELLS, G.M., CALI, J. & CHIVAS, A. 1994. Response of deep-sea benthic foraminifera to late Quaternary climate changes, southeast Indian Ocean, offshore Western Australia. *Marine Micropalaeontology* **24:** 185-229.

WHEELER, W.C. 1996. Optimisation alignment: the end of multiple sequences alignment in phylogenetics? *Cladistics* **12:** 1-9.

WHEELER, W.C. & HONEYCUTT, R.L. 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. *Molecular Biology and Evolution* **5:** 90-96.

WHITFIELD, A.K. 1995. Threatened fishes of the world: *Hippocampus capensis* Boulenger, 1990 (Syngnathidae). *Environmental Biology of Fishes* **44:** 362.

WHITFIELD, A.K., ALLANSON, B.R. & HEINECKEN, T.J.E. 1983. *Estuaries of the Cape. Part II: Synopses of available information on individual systems.* Report No. 22: Swartvlei (CMS 11), CSIR Research Report 421. Stellenbosch, South Africa.

WHITMORE, D.H., THAI, H. & CRAFT, C.M. 1992. Gene amplification permits minimally invasive analysis of fish mitochondrial DNA. *Transactions of the American Fisheries Society* **121:** 170-177.

WILCOX, T.P., ZWICKL, D.J., HEATH, T.A. & HILLIS, D.M. 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution* **25:** 361-371.

WILLIAMS, S.T. & BENZIE, J.A.H. 1998. Evidence of a phylogeographic break between populations of a high-dispersal startfish: congruent regions within the Indo-West Pacific defined by colour morphs, mtDNA and allozyme data. *Evolution* **52:** 87-99.

WILSON, R.R. Jr. & DONALDSON, K.A. 1998. Restriction digest of PCR-amplified mtDNA from fin clips is an assay for sequence genetic "tags" among hundreds of fish in wild populations. *Molecular Marine Biology and Biotechnology* **7:** 39-47.

WILSON, M.E.J & ROSEN, B.R. 1998. Implications of paucity of corals in the Paleogene of SE Asia: plate tectonics or center of origin? In: *Biogeography and geological evolution of SE Asia*, (eds) R. Hall & J.D. Holloway, pp. 165-195. Backhuys Publishers, Leiden, The Netherlands.

WILSON, A.B., VINCENT, A., AHNESJÖ, I. & MEYER, A. 2001. Male pregnancy in seahorses and pipefishes (Family Syngnathidae): Rapid diversification of paternal brood pouch morphology inferred from a molecular phylogeny. *The Journal of Heredity* **92:** 159-166.

WILSON, A.B., AHNESJÖ, I., VINCENT, A. & MEYER, A. 2003. The dynamics of male brooding, mating patterns, and sex roles in pipefishes and seahorses (family Syngnathidae). *Evolution* **57:** 1374-1386.

WINTERBOTTOM R., REIST, J.D. & GOODCHILD, C.D. 1984. Geographic variation in *Congrogadus subducens* (Teleostei, Perciformes, Congrogadidae). *Canadian Journal of Zoology* **62:** 1605-1617.

WOODLAND, D.J. 1983. Zoogeography of the Siganidae (Pisces): an interpretation of distribution and richness patterns. *Bulletin of Marine Science* **33:** 713-717.

WRIGHT, S. 1951. The genetical structure of populations. *Annals of Eugenics* **15:** 323-354.

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

WRIGHT, S. 1969. *Evolution and the genetics of populations, Vol. 2: The theory of gene frequencies*. University of Chicago Press, Chicago.

YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **13:** 555-556.

YANG, Z. & YODER, A.D. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species. *Systematic Biology* **52:** 705-716.

YODER, A.D. & YANG, Z. 2000. Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution* **17:** 1081-1090.

YODER, A.D., IRWIN, J.A. & PAYSEUR, B.A. 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. *Systematic Biology* **50:** 408-424.

ZACHOS, J., PAGANI, M., SLOAN, L., THOMAS, E. & BILLUPS, K. 2001. Trends, rhythms, and abberations in global climate 65 Ma to present. *Science* **292:** 686-693.

ZARDOYA, R. & MEYER, A. 2001. Vertebrate phylogeny: limits of inference of mitochondrial genome and nuclear rRNA sequence data due to an adverse phylogenetic signal / homoplasy ratio. In: *Major events in early vertebrate evolution: paleontology, phylogeny and development*, (ed) P. Ahlberg, pp. 135-155. Taylor & Francis, London.

ZHANG, D.X. & HEWITT, G.M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12:** 563-584.

ZHU, D., JAMIESON, B.G.M., HUGALL, A. & MORITZ, C. 1994. Sequence evolution and phylogenetic signal in control region and cytochrome *b* sequences of rainbowfishes (Melanotaeniidae). *Molecular Biology and Evolution* **11:** 672-683.

ZUCKERKANDL, E. & PAULING, L. 1962. Molecular disease, evolution, and genetic heterogeneity. In: *Horizons in Biochemistry*, (eds) M. Kasha & B. Pullman, pp. 189-225. Academic Press, New York.

ZUCKERKANDL, E. & PAULING, L. 1965. Evolutionary divergence and convergence in proteins. In: *Evolving genes and proteins*, (ed.) V. Bryson & H.J. Vogel, pp. 97-166. Academic Press, San Diego, California.

APPENDIX I

Table A1. Coding of indels from aligned RP1 sequences.

Specimen code	Presence/absence of indel
temmZA	00010
bargID	00010
comePH	01?11
subeAU	0????
whitAU1	01111
whitAU2	01111
procAU	01111
trimID	10000
cameTZ	10011
barbID1	00011
barbID2	00011
histMZ	00011
mohnJP1	10001
mohnJP2	10001
mohnVN	10001
coroJP	10011
sindJP	10001
brevAU1	00010
brevAU2	00010
abdoNZ	00010
abdoAU	00010
spinPH	00001
queeAU	00001
kudaIN	00001
kudaZA	00001
kellVN	00001
kellIN	00001
reidMX	00000
reiBR2	00000
algiBN	00000
kudaPH	00001
capeZA1	00000
capeZA2	00000
ingePE	00000
ingeMX	00000
fuscEG	00000
fuscSL	00001
fusc??	00000
fishUS	00000
guttIT	00001
hippPG	00001
erecUS	00001

Table A2. RP1 sequence alignment using POY software.

temmZA	TGGACTATTTTGCCCGATAATGTGGCAAAATAAGCTTTAATGGACACGTTTTTGTCTCGTCCAC---ATTCGTACATGCATGATGGCTTAGATTGAGCGC
bargID	TGGTCT-TTATTCTTCATAATATGGCCAAATAAGCTTGAATTGACATATTTTTGTAGCGTGAAC--AATTTTTACACACGTGATGGCTTAGATTGAGTAC
brevAU	TGGTCT-TTATTCTTCATAATGTGGCCAAATAAGCTTGATTCGACATGTTTTTGTAGCGTAAAC--TTTTTTTACATAAGTGATGGCTTAGATTGAGTTT
breAU2	TGGTCT-TTATTCTTCATAATGTGGCCAAATAAGCTTGATTCGACATGTTTTTGTAGCGTAAAC--TTTTTTTACATAAGTGATGGCTTAGATTGAGTTT
abdoNZ	TGGTCT-TTATTCTTCATTATGTGGCCAAATAAGCTTGATTCGACATGTTTTTGTAGCGTAAAC--CTTTTTTACATATGTGATGGCTTAGATTGAGTTC
abdoAU	TGGTCT-TTATTCTTCATTATGTGGCCAAATAAGCTTGATTCGACATGTTTTTGTAGCGTAAAC--CTTTTTTACATATGTGATGGCTTAGATTGAGTTC
comePH	TGGTCT-TTATTCTTCATAAAGTGGCCAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGACTAC
subeAU	TGGTCT-TTATTCTTCATAAAGTGGCCAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAAC--TTTTTTTACATATGTGATGGCTTAGATTGACTAC
whitAU1	TGGTCT-TTATTCTTCGTAATGTGGCCAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGACTAC
whitAU2	TGGTCT-TTATTCTTCGTAATGTGGCCAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGACTAC
procAU	TGGTCT-TTATTCTTCGTAATGTGGCCAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGACTAC
barbID1	TGGTCT-TTATTCTTCATAA?GTGGCCAAATAAGCTTGATCCGACATGTCTWTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGACTAC
barbID2	TGGTCT-TTATTCTTCATAA?GTGGCCAAATAAGCTTGATCCGACATGTCTWTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGACTAC
histMZ	TGGTCT-TTATTCTTCATAATGTGGCCAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAACT-TTTTTTTACACATGTGATGGCTTAGATTGACTAC
cameTZ	TGGTCT-TTATTCTTCATAATGTGGCCGAATAAGCTTGATCCGACAAGTCTTTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGAGTAC
trimID	TGGTTT-TTATTCTTCGTAATGTGGCCAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAAC--TTTTTTTACATATGTGATGGCTTAGATTGAGTAC
mohnJP1	TGGTCT-TTATTCTTCATAATATGGCCAAATAAGCTTGATCCGACATGTCTTTTTTGCGTAAAC--ATTTTTTACATATGTGATGGCTTAGATTGAGTAC
mohnJP2	TGGTCT-TTATTCTTCATAATGTGG?CAAATAAGCTTGATCCGACATGTCTTTTTTGCGTAAAC--ATTTTTTACATATGTGATGGCTTAGATTGAGTAC
mohnVN	TGGTCT-TTATTCTTCATAAT?TGGCCAAATAAGCTTGATCCGACATGTCTTTTTTGCGTAAAC--ATTTTTTACATATGTGATGGCTTAGATTGAGTAC
coroJP	TGGTCT-TTATTCTTCATAATGTGGCCAAATAAGCTTGATCCGACATGTCTTTTTTGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGAGTAC
sindJP	TGGTCT-TTATTCTTCATAATGTGGCCAAATAAGCTTGATCCGACATGTCTTTTTTGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGAGTAC
spinPH	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAACC-TTTTTTTACATATGTGATGGCTTAGATTGAGTAC
queeAU	TGGTCT-TTATTCTTCaTAATGTGGCTaaaTAAGCTTGATccGACATGTCTTTGTAGCGtAAACC-TTaTTTTACATATGTGATGGCTTAGATTGAGTAC
kellVN	TGGTAT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGAC--GTCTTTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGAGTAC
kellIN	TGGTAT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGAC--GTCGTTGTAGCGTAAACTTTTTTTTTACATATGTGATGGCTTAGATTGAGTAC
kudaIN	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGTAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
kudaPH	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGTAGCGTAAACT-TATTTTTACA?ATGTGATGGCTTAGATTGAGTAC
kudaZA	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGTAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
reidMX	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
reidBR	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
algiBN	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
capeZA1	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
capeZA2	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
ingePE	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAAC?-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
ingeMX	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAAC?-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
fuscEG	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGTAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
fuscSL	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
fusc??	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTATGGAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
fishUS	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGGTTGATCCGACATGTCTATGGAGCGTAAACT-TATTTTTACATATGTGATGGCTTAGATTGAGTAC
guttIT	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTTTGTAGCGTTAACT-TTTTTTTACATATGTGATGGCTTAGATTGAGTAC
HippPG	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGAGTAC
erecUS	TGGTCT-TTATTCTTCATAATGTGGCTAAATAAGCTTGATCCGACATGTCTTTGTAGCGTAAACT-TTTTTTTACATATGTGATGGCTTAGATTGAGTAC

Table A2. RP1 sequence alignment using POY software (continued).

temmZA	TTGGTGACGACGGATTGCAGTCTTAGAAGACACAACAAAGCGTGGTTGACTCGCTCAGTTGAATGC----T?GAT--------CCGC-CGAGCAG----G
bargID	TTTGTAGTGTCGGATGGCAGTTTTAAACAACGCAACAAAGCGTGGT---GTAGCTCGAGTGAATGC----TAGTTGCAATGAGTCGC-GGCGTTG----C
brevAU	TATGTGATGGCGGATGGCAGTTTTAAACAACACAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAAAGGGGCTC-GGAGCAG----G
breAU2	TATGTGATGGCGGATGGCAGTTTTAAACAACACAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAAAGGGGCTC-GGAGCAG----G
abdoNZ	TATGTGATGGCGGATGGCAGTTTTAAACAACACAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
abdoAU	TATGTGATGGCGGATGGCAGTTTTAAACAACACAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
comePH	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAATAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
subeAU	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTCGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
whitAU1	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCCAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
whitAU2	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCCAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
procAU	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCCAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
barbID1	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGCTAGTTAGTTGCAATGGGGCTC-GGAGCAG----G
barbID2	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGCTAGTTAGTTGCAATGGRGCTC-GGAGCAG----G
histMZ	TATGTGACGGCGGATGGCGGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTAGCAATGGGGCTC-GGAGCAG----G
cameTZ	TATGTGATGGCGGATGGCAGTTTTAAACGACGCATCAAAGCGTGGTGGAGTAGCTTGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
trimID	TATGTGCTGGCGGATGGCAGTTTTAAACGACGCTACAAAGCGTGGTGGTGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
mohnJP1	TATGTGAAGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GAAGCAG----G
mohnJP2	TATGTGAAGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GAAGCAG----G
mohnVN	TATGTGAAGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GAAGCAG----G
coroJP	TATGTGAAGGTGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
sindJP	TATGTGAAGGTGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
spinPH	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----CAGTTGCAATGGAGCTC-GGAGCAG----G
queeAU	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----CAGTTGCAATGGAGCTC-GGAGCAG----G
kellVN	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAAGAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAG-TGCAATGGGGCTC-GGAGCAG----G
kellIN	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAAGAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAG-TGCAATGGGGCTC-GGAGCAG----G
kudaIN	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
kudaPH	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTTGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
kudaZA	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
reidMX	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
reidBR	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTMGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
algiBN	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
capeZA1	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
capeZA2	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
ingePE	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
ingeMX	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGG?GCTC-GGAGCAG----G
fuscEG	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
fuscSL	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
fusc??	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
fishUS	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGTGCTC-GGAGCAG----G
guttIT	TATGTGATGGCGGATGGCTGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
hippPG	TATGTGATGGCGGATGGCAGTTTTAAACGACGAAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTC-GGAGCAG----G
erecUS	TATGTGATGGCGGATGGCAGTTTTAAACGACGCAACAAAGCGTGGTGGAGTAGCTCGAGTGAATGC----TAGTTGCAATGGGGCTCGGRACCAGCCCTG

Table A2. RP1 sequence alignment using POY software (continued).

temmZA	CCGTGCATGTGTTCGCTTATAACAGCGACTTTGGAGACCGTAGCGTGGATACCCGATAGCGCTTTTCCCCCTCCTC--AAGTCGCTGATTTAATCTTTCG
bargID	CCCTGCATGTGTTCGATTTAGACAGTTATTTTTAAGACCTTACCGCGGATAAAGAACAG-ATTGTTTTAAATGAGT--CCGTCG-TGTTTTGTTTAAATG
brevAU	CCGTGCGTGTGTTCCATTTAAACAGCTATTTTTAAGTCCGTACCGCGGGTAAAGGACAG--TAGTTTTAAATCAGC--GATTCG-CGTTATGGTGTGAAG
breAU2	CCGTGCGTGTGTTCCATTTAAACAGCTATTTTTAAGTCCGTACCGCGGGTAAAGGACAG--TAGTTTTAAATCAGC--GATTCG-CGTTATGGTGTGAAG
abdoNZ	CCGTGCGTGTGCTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TAGTTTTAATTTAGC--GATTC?-GGTTTTGGTGTGAAG
abdoAU	CCGTGCGTGTGCTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TAGTTTTAATTTAGC--GATTC?-GGTTTTGGTGTGAAG
comePH	CCGTGCGTGTGTTCAATTTAAGCAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTT-ATT-AGC--GAGTCG-CGTTTTGGTGTGAAG
subeAU	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTT-ATTGAGC--GAGTCG-CGTTTTGGTGTGAAG
whitAU1	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
whitAU2	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGAC----TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
procAU	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
barbID1	CCGTGCGTGTGTTCAATTCAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
barbID2	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
histMZ	CCGTCCGTGTGTTCAATTTAAACAGCTATTTTTCAGTCCGGACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
cameTZ	CCGTGCGTGCGTTCAATTTAAACAGCTATTTTTAAGTCCGTATCGCGGATAAAGGACAG--TAGTTTTAATTAAGC--GAGTCG-CGGTTTGGTGTGAAG
trimID	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTGCCGCGGATAAAGGACAG--TACTTTTGATTCAAC--GAGTCG-CGTTTTGGTGTGAAG
mohnJP1	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGGCAG--TACTTTTAGTTCAACGAGAGTCG-CGATTTGGTGTGAAG
mohnJP2	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGGCAG--TACTTTTAGTTCAACGAGAGTCG-CGATTTGGTGTGAAG
mohnVN	CCGT?CGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGGCAG--TACTTTTAGTTCAACGAGAGTCG-CGATTTGGTGTGAAG
coroJP	CTGTGCGTGTGTTCAATTTAAACAGCTATTTTTTAGTCCATATCGCGGATAAAGGACAG--TACTTTTAATTCGAC--GAGTCG-CGTTTTGGTGTGAAG
sindJP	CTGTGCGTGTGTTCAATTTAAACAGCTAT??
spinPH	CCGTGCGTGTGTTCAATTGAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACGG--TGCTTTTAATTCAGC--GAGTCG-CGTTACGGTGTGAAG
queeAU	CCGTGCGTGTGTTCAATTGAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACGG--TGCTTTTAATTCAGC--GAGTCG-CGTTATGGTGTGAAG
kellVN	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTGCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
kellIN	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTGCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
kudaIN	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGT--GAGTCG-CGTTTTGGTGTGAAG
kudaPH	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGT--GAGTCG-CGTTTTGGTGTGAAG
kudaZA	CCGTGTGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
reidMX	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGTGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
reidBR	CCGTGCGTGTGTTCAATTTAARCAGCTATTTTTAAGTCCGTACCGTGGATAAAGGACAG--TTCTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
algiBN	CCGTGCGTGTGTTCAATTTAAGCAGCTATTTTTAAGTCCGTACCGTGGATAAAGGACAG--TTCTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
capeZA1	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGTGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
capeZA2	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGTGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
ingePE	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTA?GTCCGTACCGTGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
ingeMX	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGTGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
fuscEG	CCGTGCGTGTGTTCAATTTAAACCGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
fuscSL	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGTGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
fusc??	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGTGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
fishUS	CCGTGCGTGTGTTCAATTTAAACCGCTATTTTTGAGTCGGTGCCGTGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
guttIT	CCGTGCTTGTGTTCAATTTAAACAGCTATTTTTAAGTCCGTACCGCGGATAAAGGACAG--TACTTTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
hippPG	CCGTGCGTGTGTTCAATTTAAACAGCTATTTTTAAGTCC-----GCGGATAAAGGCCAG--TACTGTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG
erecUS	CCT??TAAAGGCCAG--TACTGTTAATTCAGC--GAGTCG-CGTTTTGGTGTGAAG

Table A2. RP1 sequence alignment using POY software (continued).

temmZA	--AAAGTGCTTCTGGTAAAAGTATTTTC----GCATCAGGATTTGTTTAGCGTTGCCAAGCT-CCTTGTCGCAGCGTTAGCTTAGCAATATGGCTAACTG
bargID	TAATCCTGCTAGT--TAGACGTCCATGCAATTGCCACAGTGTATGCTGCGGTTTGGCTAACT-GCTTACCAAGGCAGCAGTCTCACTAATATGCTAACTA
brevAU	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAATGCGTGCTTCGGTTTGGCTCTCT-GCTTGCCATGGCAGTAGTTTAGCTAATCGGCTAACTG
breAU2	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAATGCGTGCTTCGGTTTGGCTCTCT-GCTTGCCATGGCAGTAGTTTAGCTAATCGGCTAACTG
abdoNZ	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAAC--TGTGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGTTTAGCTAATTGGCTAACTG
abdoAU	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAAC--TGTGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGTTTAGCTAATTGGCTAACTG
comePH	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTATCTTAGCTAATTGGCTAACTG
subeAU	--AAAATGCTCGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAACTTAGCTAATTGGCTAACTG
whitAU1	--AAAATGCTAGT--TAATAATCCTGGTCATGGCAACAGTGTGTGCTTCGGGTTGGCTGTCT-GCTTGCCAAGGCAGTATCATAGCTAATTGGCTAACTG
whitAU2	--AAAATGCTAGT--TAATAATCCTGGTCATGGCAAC--TGTGTGCTTCGGGTTGGCGGTCT-GCTTGCCAAGGCAGTATCATAGCTAATTGGCTAACTG
procAU	--AAAATGCTAGT--TAATAATCCTGGTCATGGCAACAGTGTGTGCTTCGGGTTGGCTGTCT-GCTTGCCAAGGCAGTATCATAGCTAATTGGCTAACTG
barbID1	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGCTT-GCTTGCCAAGGCAGTAACTTAGCTAATTGGCTAACTG
barbID2	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGCTT-GCTTGCCAAGGCAGTAACTTAGCTAATTGGCTAACTG
histMZ	--AAAATGCTAGT--TAATAATCCTGGTAATTGCAATAGTGCGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGACAGTAGCTTAGCTAATTGGCTAACTG
cameTZ	--AAAATGCTAGT--TAATAGTCCTAGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
trimID	--AAAATGCTAGT--TAATACTCCTGGCAATGGCGAAAGTGTGCGCTCCGGTTAGTCTGTCTGGCTTGCCAAGGCATTAGCTTGGCTAATTGGCTAACTG
mohnJP1	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGCGTGCGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGTAGTAGCTTAGCTAATTGGCTAACTG
mohnJP2	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGCGTGCGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGTAGTAGCTTAGCTAATTGGCTAACTG
mohnVN	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGCGTGCGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGTAGTAGCTTAGCTAATTGGCTAACTG
coroJP	--AAAAGGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGCGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
sindJP	??
spinPH	--AAAATGCTAGT--TAATCGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTGT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
queeAU	--AAAATGCTAGT--TAATCGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTGT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
kellVN	--AAAATGCTAGT--TCATAGTCCTGGCAATGG--------TGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
kellIN	--AAAATGCTAGT--TCATAGTCCTGGCAATGG--------TGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
kudaIN	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGACT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
kudaPH	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGACT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
kudaZA	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTTGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
reidMX	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTCCTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
reidBR	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
algiBN	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
capeZA1	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
capeZA2	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
ingePE	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTCCTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
ingeMX	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTCCTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
fuscEG	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
fuscSL	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTC--CTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
fusc??	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACAGTGTGTCCTTCGGTTTGGCTGTCT-ACTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
fishUS	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAACTGTGTGTGCTTCGGTTTGGCTGTCT-?CTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG
guttIT	--AAAATGCTAGT--TAATAGTCCTGGCAATGGCAAC--TGTGTGCTTCGGTTTGGCTGTCT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCT-ACTG
hippPG	--AAAATGCTCGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTATGGCTGTCT-GCTTGCCAATGCAGTAGCTTAGCTAATTGGCTAACTG
erecUS	--AAAATGCTCGT--TAATAGTCCTGGCAATGGCAACAGTGTGTGCTTCGGTATGGCTGTCT-GCTTGCCAAGGCAGTAGCTTAGCTAATTGGCTAACTG

Table A2. RP1 sequence alignment using POY software (continued).

temmZA	CATCAA-CTCCTTG--AA-----GGTATAGCGTGTAC-------TATAT--TGAC-GAT-CAAATAATTTGGTAGT?CAC----AGGT-TAGAATTCGGA
bargID	GCTC-A-TTTTTT-T-AA-----TGTATACT-TTTAA-------TTTTTAAAGACATACCCATAAACTATCTTAGCCAAT----TACA-GACCACA-ATA
brevAU	GCTC-T-TTTTTTGT-TA-----TGTATACTGTGCAA-------TTTTCAAAGACTTGTCCATAAAATGGCTTCGCTAAG----TACA-TGCCTTATGCT
breAU2	GCTC-T-TTTTTTGT-TA-----TGTATACTGTGCAA-------TTTTCAAAGACTTGTCCATAAAATGGCTTCGCTAAG----TACA-TGCCTTATGCT
abdoNZ	GCTC-T-TTTTTT-T-AA-----TGTATACTGTGCAA-------TTTTCAAAGCCTTATCCATAAAATGGCTTAGCTAAG----TACA-TACCTTATGCG
abdoAU	GCTC-T-TTTTTT-T-AA-----TGTATACTGTGCAA-------TTTTCAAAGCCTTATCCATAAAATGGCTTAGCTAAG----TACA-TACCTTATGCG
comePH	GCTT-A-TTTTCT----A-----TGTATACTGTGCAATGTGCAA???
subeAU	GCTC-A-TTTTTT-T-AA-----TGTATACTGTGCAAT???
whitAU1	TCTT-A-TTTTTT---AA-----TGTAGTGTGTATATTGTGCAATTTTCAAAGACTTATCCATAAAATAGCTTAGCCAAGTACAT?CAGTACCTTAGGCT
whitAU2	TCTT-A-TTTTTT---AA-----TGTAGTGTGTATATTGTGCAATTTTCAAAGACTTATCCATAAAATAGCTTAGCCAAGTACATACAGTACCTTAGGCT
procAU	TCTT-A-TTTTTT---AA-----TGTAGTGTGTATATTGTGCAATTTTCAAAGACTTATCCATAAAATAGCTTAGCCAAGTACATACAGTACCTTAGGCT
barbID1	GCTC-?-TTTTTT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCCA?AAAA----------TA?CTTAGGCT
barbID2	GCTC-A-TTTTTT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCCATAAAA----------TAACTTAGGCT
histMZ	GCTC-A-TGTTTT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCCT?AAAA----------TAACTTGGGCT
cameTZ	GCTT-A-ATTTTT---AATGTACTGTATATTGTGCAA-------TTTTCAAAGACTTATACATAAAATAGCTTAGCCAGG----TACA-TACCTTGGGCT
trimID	TCTT-A-TTTTTT--CAATGTACTGTATATTGTGCAA-------TTTTCAAAGACTTATCCATAAAATAGCTTAGCCAAG----TACA-TACCTGAGGCT
mohnJP1	GCTT-A-TTTTTT---AATGCACTGTATATTGTGCAA-------TTTTCAAAGACTTATCCATAAAATAGCTTAGCCAAG----TACA-TACCTTA-GCT
mohnJP2	GCTT-A-TTTTTT---AATGTACTGTATATTGTGCAA-------TTTTCAAAGACTTATCCATAAAATAGCTTAGCCAAG----TACA-TACCTT?GGCT
mohnVN	GCTT-A-TTTTTT---AATGCACTGTATATTGTGCAA-------TTTTCA?AGACTTATCCATAAAATAGCTTAGCCAAG----TACA-TACCCTAGGCT
coroJP	GCTC-A-TTTTTT---AATGTACTGTATATTCTGCCA-------TTTTCAAAGACTTATCCATAAAATAGTTTAGCCAAG----TACA-TACCTTAGGCC
sindJP	???????TTTTTT---AATGTACTGTATATTCTGCAA-------TTTTCAAAGACTTATCCATAAAATAGTTTAGCCAAG----TACA-TACCTTAGGCT
spinPH	GCTC-A-TTTTCT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTGTCTATAAAATAGCTTAGCCAAG----TACA-TGCCTTAGGCT
queeAU	GCTC-A-TTTTCT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTGTCTATAAAATAGCTTAGCCAAG----TACA-TACCTTAGGCT
kellVN	GCTC-A-TTTTCT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTAGCCAAG----TACA-TACCTTAGGCT
kellIN	GCTC-A-TTTTCT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTAGCCAAG----TACA-TACCTTAGGCT
kudaIN	GCTC-A-TTTTCT---TA-----TGTGTACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTCAGCCAAG----TACA-TACCTTGGGCT
kudaPH	GCTC-A-TTTTCT---AA-----TGTGTACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTCAGCCAAG----TACA-TACCTTGGGCT
kudaZA	GCTC-A-TTTTCT---AA-----TGTGTACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTCAGCCAAG----TACA-TACCTTGGGCT
reidMX	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
reidBR	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
algiBN	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
capeZA1	GCTC-A-TGTATT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
capeZA2	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
ingePE	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
ingeMX	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
fuscEG	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
fuscSL	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTCGCCAAG----TACA-TACCTTGGGCT
fusc??	GCTC-A-TTTTCT---CA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTAGCCAAG----TACA-TACCTTGGGCT
fishUS	GCTC-A-TTTTCT---CA-----TGTATA??????????????????TTCAAAGACTTATCTATAAAATAGCTTCGCCA?G----TACA-TACCTTGGGCT
guttIT	GCTC-A-TTTACT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAACTTAGCCAAG----TACA-TACTTTAGGCT
hippPG	GCTC-A-TTTTCT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTAGCCAAG----TACA-TACCTTAGGCT
erecUS	GCTC-A-TTTTCT---AA-----TGTATACTGTGCAA-------TTTTCAAAGACTTATCTATAAAATAGCTTAGCCAAG----TACA-TACCTTAGGCT

Table A2. RP1 sequence alignment using POY software (continued).

temmZA	ATCCGTGTGA-T-CTG--GTGTGTG--CGTT?-CGTTGCTC--ACTGCT-TAAAA?CCAAATCTTTAAACG-TGGTTG----TTTTGTTTGCAGA-GACA
bargID	TTGCTTGCCAGT-CT-TTGCTTGTA-TCGTAA-C-TAACTC?AGCTGCTGTGTTGTGTTGAAGTCCAAATGAAACGTG----ATTTGTTTGCAGA-GACA
brevAU	TTGCTTACCAGT-CTGTTGTTTGTAGACGTAA-C-TAGCTCGA-----TGTGTTGTGTTGAAGTCTAAATTAAACGCG----ATTTGTTTGCAGA-GACA
breAU2	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGA-----TGTGTTGTGTTGAAGTCTAAATTAAACGCG----ATTTGTTTGCAGA-GACA
abdoNZ	TTGCTTACCAGT-CTGTTGTTT?TAGTCGTAA-C-TAGCTCGA-----TGTGTTGTGTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
abdoAU	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGA-----TGTGTTGTGTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
comePH	????????????????---TGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTCTTGCGTTGAAGTCTAAATTAAACGTGATTTATTTGTTTGCAGA-GACA
subeAU	??
whitAU1	TTGCTTACCAGT-C---TGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTCTTGCGTTGAAGTCTAAATTAAACGTGATT?ATTTGTT?GCAGA-GACA
whitAU2	TTGCTTACCAGT-C---TGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTCTTGCGTTGAAGTCTAAATTAAACGTGATT?ATTTGTT?GCAGA-GACA
procAU	TTGCTTACCAGT-C---TGTYTGTAGTCGTAA-C-TAGCTCGATGTGCTGTCTTGCGTTGAAGTCTAAATTAAACGTGATTTATTTGTTTGCAGA-GACA
barbID1	TTGCTTACTACT-CTGTTGTTTGTCGTCGTAA-C-TAGCTCGATGTGCTGTCTTGCGTTGAAGTCTAAATTAAACGTGATTTATTTGTTTGCAGA-GACA
barbID2	TTGCTTACTACT-CTGTTGTTTGTCGTCGTAA-C-TAAcTC?ATGTGCTGTCTTGCGTTGAAGTCTAAATTAAACGTGATTTATTTGTTTGCAGA-gAc?
histMZ	TTGCTTACTAGT-CTGTTGTTTGTAGTCGTAA-C-TAACTCGATGTGCTGTCTTGCGTTGAAGTCTAAATTAAACGTGATTTATTTGTTTGCAGA-GACA
cameTZ	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTGTTAAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-G???
trimID	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAACC-TAGCTCGA-----TGTGTTGTGTTGAAGTCAAAATCAAACGTGATTTATTTGTTTGCAGA-GACA
mohnJP1	TTGTTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGA-----TGTGTTGCGTTGAAGTCTAAATTAAACGTGATTTATATGTTTGCAGA-GACA
mohnJP2	TTGTTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGA-----TGTGTTGCGTTGAAGTCTAAATTAAACGTGATTTATATGTTTGCAGA-GACA
mohnVN	TTGTTTACCAGTACTGTTGT?TGTAGTCGTAA-C-TAGCTCGA-----TGT?TTGCGTTGAAGTCTAAATTAAACGTGATTTATATGTTTGCAGA-GACA
coroJP	TTGCTTACCGGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GACA
sindJP	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGA-----TGTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GACA
spinPH	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GACA
queeAU	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGT-GCAGA-GACA
kellVN	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTCTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GACA
kellIN	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTCTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GACA
kudaIN	TTGCTTACCAGT-CTGTTGTTTGTGGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GACA
kudaPH	TTGCTTACCAGT-CTGTTGTTTGTGGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GACA
kudaZA	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GGCA
reidMX	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
reidBR	TTGCTTGCCART-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
algiBN	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
capeZA1	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
capeZA2	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TA?CTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
ingePE	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
ingeMX	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
fuscEG	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
fuscSL	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
fusc??	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
fishUS	TTGCTTGCCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTGTGTTGTTTTGAAGTCTAAATTAAACGTG----ATTTGTTTGCAGA-GACA
guttIT	TTGCTTACCAGT-C---TGTTTGTAGTGGTAA-C-TAGCTCGATGTGCTGTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGGTTGCAGA-GACA
hippPG	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTCTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGTTTGCAGA-GACA
erecUS	TTGCTTACCAGT-CTGTTGTTTGTAGTCGTAA-C-TAGCTCGATGTGCTCTGTTGTGTTGAAGTCTAAATTAAACGTGATTTATTTGTTTGCAGA-GACA

Table A2. RP1 sequence alignment using POY software (continued).

temmZA	ACGGCCATG-TT-CAGCACGAGTGCTAAGATA
bargID	ACAGCCATG-TT-CAGCACGAGTGCAAAAATT
brevAU	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
breAU2	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
abdoNZ	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
abdoAU	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
comePH	ACGGCCATG-TT-CAGCAC?AGTGCAAAAATT
subeAU	????????????????????????????????
whitAU1	ACGGCCATG-TT-CCGCACGAGTGCGAAAAAG
whitAU2	ACGGCCATG-TTC-CGCACGAGTGCGAAAAAG
procAU	ACGGCCATG-TTC-AGCACGAGTGCAAAAATT
barbID1	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
barbID2	AC?GCCATG-TT-C?GCACKAGTGCAAAAaTT
histMZ	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
cameTZ	????????????????????????????????
trimID	ACGGCCATG-?T-CAGCACGAGTGC???????
mohnJP1	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
mohnJP2	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
mohnVN	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
coroJP	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
sindJP	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
spinPH	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
queeAU	ACGGCCAtG-TT-CAGCACGAGTGCAAAAATT
kellVN	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
kellIN	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
kudaIN	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
kudaPH	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
kudaZA	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
reidMX	ACGGCCATG-TT-?AGCACGAGTGCAAAAATT
reidBR	ACGGCCATG-TT-cAGCACGAGTGCAAAAATT
algiBN	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
capeZA1	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
capeZA2	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
ingePE	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
ingeMX	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
fuscEG	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
fuscSL	ACGGCCATG-Tt-CAGCACGAGtGCAAAAATt
fusc??	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
fishUS	?CGGCCATG-TT-CAGCACGAGTGCAAAAATT
guttIT	ACGGCCATG-TTCCAGCACGAGTGCAAAAATT
hippPG	ACGGCCATG-TT-CAGCACGAGTGCAAAAATT
erecUS	ACGGCCATG-TT-????????????????????

Table A3. 16S rRNA sequence alignment using POY software.

Sacus	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTTAACTGTCT
bargID	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAAGGCTAAACTGTCT
brevAU	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCCGTATGAATGGCATAACGAGGGCTAAACTGTCT
abdo??	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCCGTATGAATGGCATAACGAGGGCTAAACTGTCT
comePH	GTTTAACGGCCGCGG--ATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
subeAU	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
comeVN	GTTTAACGGCCGCGGCTATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
whitAU	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
barb??	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
trimHK	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
mohnJP	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
coroJP	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
sindJP	GTTTAACGGCCGCGG--ATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAAGGCTAAACTGTCT
cameTZ	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
spinPH	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
queeAU	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
kellVN	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
kudaZA	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
kudaPH	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
reidMX	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
algiBN	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
ingePE	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
capeZA	GTTTAACGGCCGCGG--ATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
fuscEG	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
erec??	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAGAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT
hippPG	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAGAGACCTGTATGAATGGCATAACGAGGGCTAAGCTGTCT
guttIT	GTTTAACGGCCGCGG-TATTTTGACCGTGCAAAGGTAGCGCAATCACTTGTCTTTTAAATAAAGACCTGTATGAATGGCATAACGAGGGCTAAACTGTCT

Table A3. 16S rRNA sequence alignment using POY software (continued).

Sacus	CCTTTACCTAGTCAATGAAATTGATCTCCCCGTGCAGAAGCGGGGATATTAATATAAGACGAGAAGACCCTGTGGAGCTTGAGACAAT-AGACGGATTT--T
bargID	CCTTACCTTTATCAATGAAATTGATCCCCCCGTGCAGAAGCGGGGATTATCACATAAGACGAGAAGACCCTGTGGAGCTTTAAGCAAA---ATTGATTTATT
brevAU	CCTTACCCAGGTTAATGAAATTGATCTTTCCGTGCAGAAGCGGGAATAAACCCATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAATAAATGAACTTATT
abdo??	CCTTACCCAGGTTAATGAAATTGATCTTTCCGTGCAGAAGCGGGAATTACACCATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAATAAATGAATTTATT
comePH	CCTTACTCCAGTTAATGAAATTGATCTTTCCGTGCAGAAGCGGGAATTGACTTATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAATAGACGAGTTTATT
subeAU	CCTTACTCCAGTTAATGAAATTGATCTTTCCGTGCAGAAGCGGGAATTGACTTATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAATAGACGAGTTTATT
comeVN	CCTTACTCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACTCATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAATAGACGAGTTTATT
whitAU	CCTTACCCCGGTTAATGAAATTGATCCTCCCGTGCAGAAGCGGGAATTAACCCATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAATAGACGAGTTTATT
barb??	CCTCACCCTAGTTAATGAAATTGATCCTCCCGTGCAGAAGCGGGAATTAACTCATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAACAGATGAATTTGTT
trimHK	CCTTACCTAAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACCCATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAATAGATGAATTAATA
mohnJP	CCTTACCCAAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAATCCATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAGCGGATGAATTAATT
coroJP	CCTTACCTAGGTTAATGAAATTGATCTTCCCGTGCAGAAGCGAGAATTAACTCATAAGACGAGAAGACCCTGTGGAGCTTTAGATAAATAAATGAACTAATC
sindJP	CCTTACCTAGGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACTCATAAGACGAGAAGACCCTGTGGAGCTTTAGATGAATAAATGAATTGATT
cameTZ	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGAGAATTAACCCATAAGACGAGAAGACCCTGTGGAGCTTTAGATAAATAGATGATTTTATT
spinPH	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTTAGATAACTGGATGAATTTATT
queeAU	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTTAGATAACTAGATGAATTTATT
kellVN	CCTTACCCCAGTTAATGAAATTGATCCTCCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTTAGATAAATAGATGAATTTATT
kudaZA	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTCAGAC-AATAGATGAATTTATT
kudaPH	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTCAGAC-AATAGATGAATTTATT
reidMX	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTGGCCCATAAGACGAGAAGACCCTGTGGAGCTTCAGACAAATAGATGAATTTATT
algiBN	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACCCATAAGACGAGAAGACCCTGTGGAGCTTCAGAC-AATAGATGAATTTATT
ingePE	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTGACCCATAAGACGAGAAGACCCTGTGGAGCTTCAGACAAATAGATGAA-TTATT
capeZA	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTCAGAC-AATAGATGAATTTATT
fuscEG	CCTCACCCCAGTTAATGAAATTGATCTTCCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTCAGAC-AATAGATGAATTTATT
erec??	CCTCACCCCAGTTAATGAAATTGATCTTTCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTTAGATAAATAAATGAATTTATT
hippPG	CCTCACCCCAGTTAATGAAATTGATCTTTCCGTGCAGAAGCGGGAATTAACACATAAGACGAGAAGACCCTGTGGAGCTTTAGATGTATAAATGAATTTATT
guttIT	CCTTACTCCAGTTAATGAAATTGATCTCCTCGTGCAGAAGCGGGGATTAACCCATAAGACGAGAAGACCCTGTGGAGCTTTAGAT-GATAGATGAATTTATT

Table A3. 16S rRNA sequence alignment using POY software (continued).

Sacus	ATAAGTTTATACTCCCTCCA-ATGTCTTCGGTTGGGGCGACCGCGGAGTAAAGAGCAACCTCCATGAGGACTAAGGT-G-TAACCTTAAATCCCAGAAC
bargID	ATAAAATAA-AAATCATTAT-ATGCTTTTAGTTGGGGCGACCGCGGAGTACTAAAAAACCTCCGTGAGGATTGAGGT-A-AACCCTTAAAACTAAGAAA
brevAU	AAACATATA-ACCTCATTTTAATATCTTTAGTTGGGGCGACCGCGGAGTAAAACAAAACCTCCGTGAGGACTGAGGTAA-AAACCTTATACCTGAGAAA
abdo??	AAGCCAGTA-ACCTCATTTTAATATCTTTAGTTGGGGCGACCGCGGAGTAAAACAAAACCTCCGTGAGGATTGAGGT-A-AAACCTTATACCTAAGAAA
comePH	AAATCTATA-ACCTCATCTTAATATCTTTAGTTGGGGCGACCGCGGAGTAAAATAAAACCTCCGTGAGGATTGAGGTGA-AAACCTTACACCCAAGAAA
subeAU	AAATCTATA-ACCTCATCTTAATATCTTTAGTTGGGGCGACCGCGGAGTAAAATAAAACCTCCGTGAGGATTGAGGTGA-AAACCTTACACCCTAGAAA
comeVN	AAATCTATA-ACCTCATCTTAATATCTTTAGTTGGGGCGACCGCGGAGTAAAACAAAACCTCCGTGAGGATTGAGATGA-AAACCTTACACCCAAGAAA
whitAU	AAACCTATA-ACCTCATCTTTATATCTTTAGTTGGGGCGACCGCGGAGTAAAACAAAACCTCCGTGAGGATTGAGGTGA-AAACCTTACACCTAAGAAA
barb??	AAACCTACA-ACCTCATCTTAATATCTTCAGTTGGGGCGACCGCGGAGTAAAACAAAACCTCCGTGAGGATTGAGGTGA-AAACCTTATACCCAAGAAA
trimHK	AAATCTATA-ACCTCATCTTGATATCTTTAGTTGGGGCGACCGCGGAGTAAAACAAAACCTCCGTGAGGATAGAGGT-A-AAACCTTATACCTAAGAAT
mohnJP	AAACCAATA-ACCTCATCCTAATATCTTTAGTTGGGGCGACCGCGGAGTAGAACAAAACCTCCGTGAGGATTGAGGT-A-AAACCTTACACCTAAGAAT
coroJP	AAACCAACA-CCCTCATTTTAATATCTTTAGTTGGGGCGACTGCGGAGTAAAACAAAACCTCCATGAGGATTGAGGT-A-AAACCTTTCACCTAAGAAT
sindJP	AAACCAATA-ACCTCATTTTAATATCTTTAGTTGGGGCGACCACGGAGTAAGACAAAGCCTCCGTGAGGATTAAGGT-A-AAACCTTTCACCTAAGAAT
cameTZ	AAACCAATA-ACCTCACCTT-GCATCTTCAGTTGGGGCGACCGCGGAGTAAAAAAAAACCTCCGTGAGGAATGAGGT-A-AAACCTTACACCTAAGAAT
spinPH	AAAACAATA-ACCTCATCTTAA-ATCTTTAGTTGGGGCGACCGCGGAGCAAAACAAAACCTCCGTGAGGATTGAGGTAA-AAACCTTATACCCAAGAAC
queeAU	AAAACAATA-ACCTCATCTTAA-ATCTTTAGTTGGGGCGACCGCGGAGCAAAACAAAACCTCCGTGAGGATTGAGGT-A-AAACCTTATACCCAAGAAC
kellVN	TAAACAATA-ACCTCATCTTAA-ATCTTCAGTTGGGGCGACCGCGGAGCAAAACAAAACCTCCGTGAGGATTGAGGCAA-AAGCCTTACACCCAAGAAT
kudaZA	AAAACAATA-ACCTCATCCTAA-GTCTTTAGTTGGGGCGACCGCGGAGCAAAACAAAACCTCCGTGAGGAATGAGGTAA-AAACCTTATACCTAAGAAT
kudaPH	AAAACAATA-ACCTCATCCTAA-GTCTTTAGTTGGGGCGACCGCGGAGCAAAACAAAACCTCCGTGAGGAATGAGGTAA-AAACCTTACACCCAAGAAT
reidMX	AAAACAATA-ACCTCATCCTAA-GTCTTTAGTTGGGGCGACCGCAGAGCAAAATAAAACCTCCGTGAGGAATGAGGTAA-AAACCTTATACCCAAGAAT
algiBN	AAAACAMTA-ACCTCATCCTAA-GTCTTTAGTTGGGGCGACCGCGGAGCAAAAACAAACCTCCGTGAGGAATGAGGTAA-AAACCTTATACCCAAGA?T
ingePE	AAAACAATA-ACCTCATCCTAA-GTCTTTAGTTGGGGCGACCGCGGAGCAAAATAAAACCTCCGTGAGGAATGAGGTAA-AAACCTTATACCCAAGAAT
capeZA	AAAACAATA-ACCTCATCCTAA-GTCTTTAGTTGGGGCGACCGCGGAGCAAAACAAAACCTCCGTGAGGAATGAGGTAA-AAACCTTATACCCAAGAAT
fuscEG	AAAACAATA-ACCTCATCCTAA-GTCTTTAGTTGGGGCGACCGCGGAGCAAAACAAAACCTCCGTGAGGAATGAGGTAA-AAACCTTATACCCAAGAAT
erec??	AAAACAATA-ACCTCATTTTAA-ATCTTTAGTTGGGGCGACCGCGGAGTAAAACAAAACCTCCGTGAGGACTGAGGTGATAAACCTTATACCTAAGAAT
hippPG	AAAACAATA-ACCTCATTTTAA-ATCTTTAGTTGGGGCGACCGCGGAGTAAAACAGAACCTCCGTGAGGACTGAGGTGA-AAACCTTATACCTAAGAAT
guttIT	AAAACGACG-ACCTCATCTTAA-ATCTTTAGTTGGGGCGACCGCGGAGTAAAACAAAACCTCCGTGAGGATTGAGATGA-AAACCTTATACCCAAGAAT

Table A3. 16S rRNA sequence alignment using POY software (continued).

Sacus	GACAGTTCAAAGAAGCAAAATTTTTGACCTA-AGGATCCGGCA-CA-GCCGATCAACGAACCGAGTTACCCCAGGGATAACAGCGCAATCCCCTTCAAGA
bargID	AACACTTCAAAGTATTAAAATATTTAACCTA-AAGATCCGGTA-AT-ACCGATTAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTTTTTAAGA
brevAU	GTCATTTCTAAGTACCAAAATATTTGACCCA-AAGATCCGGCA-ATAGCCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCCTTTTGAGA
abdo??	GTCATTTCTAAGCACCAAAATATTTGACCTA-AAGATCCGGC--ATAGCCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA
comePH	GTCATTTCTAAGTACCAAAATATTTGACCAA-TAGATCCGGCA-CTGACCGATCAACGAACCTAGTTACCCCAGGGAT-ACAGCGCAATCCTCTTTAAGA
subeAU	GTCATTTCTAAGTACCAAAATATTTGACCAT-AAGATCCGGCA-CT??CCGATCAACGAACCT-GGTACCCCAGGGATAACA?CGCAATCCTCTTGAAGA
comeVN	GTCATTTSTAAGTACCAAAATATTTGACCAA-CAGATCCGGCA-CTGACCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTAAGA
whitAU	GTCATTTCTAAGTACCAAAACATTTGACCCA-TAGATCCGGCA-ATGACCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTAAGA
barb??	GTCATTTCTAAGTACCAAAAAATTTGACCTA-TAGATCCGGCA-ATG-CCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTAAGA
trimHK	GYCATTTCTAAGAACCAAAATATTTGACCTA-AAGATCCGGTA-ATAACCGATCAACGAACATAGTTACCCCAGGGATAACAGCGCAATCCTCTTTAAGA
mohnJP	GTCATTTCTAAGTACCAAAATATTTGACCTT-TAGATCCGGCA-ATGACCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTCAAGA
coroJP	GTCATTTCTAAGCACCAAAATTTTTGACCTA-AAGATCCGGTA-ATAGCCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTCAAGA
sindJP	GTCACTTCTAAGCACCAAAATCTTTGACCTA-AAGATCCGGTA-ATAACCGATCGACGAACCTAGTTACACCAGGGATAACAGCGCAATCCTCTTCAAGA
cameTZ	GTCATTTCTAAGCACCAAAATATTTGACCTA-AAGATCCGGTA-ATAACCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTCAAGA
spinPH	GTCATTTCTAAGCACCAAAATATTTGACCTATTAGATCCGGCA-ACAGCCGATCAACGGACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA
queeAU	GTCATTTCTAAGCACCAAAATATTTGACCTATTAGATCCGGCA-ACAGCCGATCAACGGACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA
kellVN	GTCATTTCTAAGTACCAAAACATTTGACCTA-TAGATCCGGCATACAGCCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA
kudaZA	GTCATTTCTAAGTACCAAAATATTTGACCCA-TAGATCCGGCA-ACAGCCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATTCTCTTTGAGA
kudaPH	GTCATTTCTAAGTACCAAAATATTTGACCCA-T-GATCCGGCA-ACAGCCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATTCTCTTTGAGA
reidMX	GTCATTTCTAAGTACCAAAATATTTGACCTA-TAGATCCGGCA-ATAACCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA
algiBN	GTCATTTCTAAGTACCAAAATATTTGACCTA-TAGATCCGGCA-ACAACCGATCAACGAMCCTGGTTACCCCAGGGAA????????????????????
ingePE	GTCATTTCTAAGTACCAAAATATTTGACCTA-TAGATCCGGCA-ATAACCGATCAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA
capeZA	GTCATTTCTAAGTACCAAAATATTTGACCCA-CAGATCCGGCA-ACAGCCGATCAACGAACCTAGTTACCCCAGGGAT-ACAGCGCAATTCTCTTTGAGA
fuscEG	GTCATTTCTAAGTACCAAAATATTTGACCCA-TAGATCCGGCA-ACAGCCGATCAACGAACCT?GTTACCCCAGGGATAACA????????????????
erec??	GTCATTTCTAAGTGCCAGAATATCTGACCCG-TAGATCCGGC--ATAGCCGATTAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA
hippPG	GTCATTTCTAAGTGCCAAAATATTTGACCCA-TAGATCCGGCA-ATAGCCGATTAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA
guttIT	GTCATTTCTAAGTACCAAAATATTTGACCAA-TTGATCCGGCA-ACAGCCGATTAACGAACCTAGTTACCCCAGGGATAACAGCGCAATCCTCTTTGAGA

Table A3. 16S rRNA sequence alignment using POY software (continued).

Sacus	GTCCCTATCGACAAGGGGGTTTACGACCTCGATGTTGGATCAGGGTAT-CCTAATGGTGTAG
bargID	GTCCATATCGACAAAAGGGTTTACGACCTCGATGTTGGATCAGGATAT-CCTAATGGTGTAG
brevAU	GTCCCTATCGACAAGGGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGCAG
abdo??	GTCCCTATCGACAAGGGGGTTTACGACCTCGATGTTGGATCAGGACAT--CTAATGGTGTAG
comePH	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
subeAU	GTCCCTATCGACAAGAGGGTTTACGACCTC?ATGTTGGATCAGGACAT-CC-AATGGTGTAG
comeVN	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
whitAU	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
barb??	GTCCCTATCGACAAGGGGGTTTACGACCTCGATGTTGGATCAGGATAT-CCTAATGGTGTAG
trimHK	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
mohnJP	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
coroJP	GCCCTTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAMGACAT-CCCAATGGTGTAG
sindJP	GCCC-TATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
cameTZ	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
spinPH	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
queeAU	GTCCCTATCGACAAGAGGGTTTACGACCTC?ATGTTGGATCAGGACAT-CCTAATGGTGTAG
kellVN	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
kudaZA	GTCCCTATCGACAAGAGAGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
kudaPH	GTCCCTATCGACAAGAGAGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
reidMX	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
algiBN	??-?????????????
ingePE	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
capeZA	GTCCCTATCGACAAGAGAGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
fuscEG	??-?????????????
erec??	GTCCCTATCGACAAGGGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG
hippPG	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACATCCCTAATGGTGTAG
guttIT	GTCCCTATCGACAAGAGGGTTTACGACCTCGATGTTGGATCAGGACAT-CCTAATGGTGTAG

Table A4. Aldolase sequence alignment using ClustalX software.

```
temmZA    TCCCTCAAGACTGGCCATCATTGAGAACGCCAACGTTCTGGCCCGCTATGCCAGCATCTGCCAGATGGTAGGA-----CGAGGAGGGTCCGTATCCG---------G
brevAU    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGAATCGCCTACTCTCGTCGCTGTGCAATTTCGTATT
abdoAU    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
abdoNZ    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
barbID    CCCCACACCTCTTGCCATTGCCGAAAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCCACTCTCGTCGCTGTGCA-TGTGGTGTT
anguAU    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCCACTCTCGTCGTTGTGCA-TGTGGTGTT
subeAU    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCCACTCTCGTCGTTGTGCA-TGTGGTGTT
comePH    CCCCACACCTCTTGCCATTGCCGAAAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCCACTCTCGTCGCTGTGCA-TGTGGTGTT
comeVN    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCCACTCTCGTCGCTGTGCA-TGTGGTGTT
whitAU    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCCACTCTCGTCGTTGTGCA-TGTGGTGTT
cameTZ    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
trimVN    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGAATCGCCTACTCTCGTCTCTGTGCA-TGTGGTGTT
trimID    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGAATCGCCTACTCTCGTCTCTGTGCA-TGTGGTGTT
mohnJP    CCCCACACCTCTTGCCATTGCTGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
coroJP    CCCCACACCTCTTGCAATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCAACGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
sindJP    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCAACGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
spinPH    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
queeAU    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTATTCTCGTCGCTGTGCA-TGTGGTGTT
kellVN    CCCCACACCTCTTGCCATTGCCGAGAATGCAAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
kellIN    CCCCACACCTCTTGCCATTGCCGAGAATGCAAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
kudaZA    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
kudaPH    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
capeZA    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
ingePE    CCCCACACCTCTTGCCATTGCCGAGAATGCCAATGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
algiBN    CCCCACACCTCTTGCCATTGCCGAGAATGCCAATGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
reidMX    CCCCACACCTCTTGCCATTGCCGAGAATGCCAATGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
fusc??    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
borbTZ    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
borbMG    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TATGGTGTT
fishUS    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTAGTTGTGCA-TGTGSTGTT
guttIT    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
guttPG    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
hippPG    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
erecUS    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
zostUS    CCCCACACCTCTTGCCATTGCCGAGAATGCCAACGTGTTGGCTAGATATGCCAGCATCTGTCAACAGGTTGGCATCGCCTACTCTCGTCGCTGTGCA-TGTGGTGTT
```

Table A4. Aldolase sequence alignment using ClustalX software (continued).

```
temmZA    TGACATCTCCCGAGGCGCTTCCTAACGCTCTCC-TGCATCCCTCAGCACGGCATCGTGCCCATTGTGGAGCCTGAAATTCT
brevAU    TGTCTTCCTTGGAGGATGTAATTAAATCACATC-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
abdoAU    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
abdoNZ    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
barbID    TGTCTTCCTTGGAAGATGTAATTCAAACACATT-ATCTCTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
anguAU    TGTCTTCCTTGGAAGATGTAATTCAAACACATT-ATCTCTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
subeAU    TGTCTTCCTTGGAAGATGTAATTCAAACACATT-ATCTCTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
comePH    TGTCTTCCTTGGAAGATGTAATTCAAACACATT-ATCTCTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
comeVN    TGTCTTCCTTGGAAGATGTAATTCAAACACATT-ATCTCTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
whitAU    TGTCTTCCTTGGAAGATGTAATTCAAACACATT-ATCTCTGATTAGAATGGACTAGTGCCTATTGTGGAGCCTGAAATTCT
cameTZ    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
trimVN    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
trimID    TGTCTTCCTTGGAGGGTGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
mohnJP    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGACTGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
coroJP    TGTCTTCGTTGGAGGATATAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
sindJP    TGTCTTCGTTGGAGGATATAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
spinPH    TGTCTTCCTTGGAGGATGTAATTCAAACGCATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
queeAU    TGTCTTCCTTGGAGGATGTAATTCAAACGCATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
kellVN    TGTCTTCCTTGGAGGATGTAATTCAAACGCATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
kellIN    TGTCTTCCTTGGAGGATGTAATTCAAACGCATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTTC
kudaZA    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
kudaPH    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
capeZA    TGTCTTCCTTGGAGGATGTAATTCAAACACATTTATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
ingePE    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
algiBN    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
reidMX    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
fusc??    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
borbTZ    TGTCTTCCTTGGAGGATGTAATTCAAACACATTTATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
borbMG    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
fishUS    TGTCTTCCTTGGAGGATGTAATTCAAACACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
guttIT    TGTCTTCCTTGGAGCATGTAATTCAAACACATT-GTCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
guttPG    TGTCTTCCTTGGAGCATGTAATTCAAACACATT-GTCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
hippPG    TGTCTTCCTTGGAGGATGTAATTCAACCACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
erecUS    TGTCTTCCTTGGAGGATGTAATTCAACCACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
zostUS    TGTCTTCCTTGGAGGATGTAATTCAACCACATT-ATCTTTGATTAGAATGGACTGGTGCCTATTGTGGAGCCTGAAATTCT
```

Table A5. Mitochondrial sequence amplified in *H. capensis* using the HCAL1 and HCAH1 primers. The underlined portion of the control region represents the CR right domain sequence used in most phylogenetic analyses in this study.

cytochrome *b* (partial):
CCTGTATCTGGGTGAATGGAGAATAAAAATCTTNAATGAAACT

$tRNA^{Thr}$:
GCATTTGNARTTTGTAGCTMAGTATWWGAGCGCCGGTTTTGTAAACCGGAGGCCGGAAGTTAAACCCCTC
CCTAGTGC

$tRNA^{Pro}$:
TCAGAAAAAGGAGAATCGAACTCCCACCGCCGGCTCCCAAAGCTGGTATTCTAGGTTAAAATATTTTCTG

control region:
GTACATATATGAAATATCCTTATATATATATATAGTGCATATTATTAATTCCTTAGAACATTTTATGTATAA
ACACCATTAATTTATATTAAACATAAAATAGTAACATAAAACTAAGGTAGACATAAACCACTAATTTAAG
AATACATAAATAGTTATAAATGCTGAGAGATTTAAGAATCAGCACAAAACTATTGGACAAAGATATACCA
AGACTCCAAATATTATTAATTTAAGTATTCGATGCAGTAAGAGCCTACCAAACAATCTATTTCTTAAGGAT
AACGGTTCTTGATGGTCAGGAGCCCTAATTGAAGGGtTGTTACCTAAAAGGTGAATTATTCCTGGCATCTC
CTCCTTTTTCAGGTCCATTAATTTATTAATCCGCACACTTTCATCGACGCTTACATTGACTAATGGTGTTAA
ACCTTCGACTCGTTACCCCCCAAGCCGGGCGTTCTCTCCACAGGGGCAGCTGGTTTCTTTTTTTTTTC**TCCC**
TTCATGAACATTTCAGAGTGCACACGGCCCTATAATTGAAGGTTGAACATTCGTTCCTTATTGAGTA
ATATAAATGTAATGTTGAAATTACATTACTTAAGAATTGCATAAATCTCTTTCTAGAGCATAATAGAT
AACATTTTACCTATAAGTCGCCCCCCCCTTCTGTTTTTAACTGAAAAACCCCCCAACCCCCCTAAGG
CCCTAAAGTAACTAACAATTCAACAAAATCACTATAAACAATAAAACCCCAAGATTATCAAGTATCC
AACAGATTGTACAGTATTGGTGATGTGTAATTTCGTATTTATGTATTATCACTTTTTAAAAAATACAT
CGTATCACACCCATTTATTACAATTAACCAAGATAAAATACTAGAAGGCCAT

$tRNA^{Phe}$:
CGTAGCTTAAATCTTAAAGCATAACACTGAAGATGTTATTATGAACCCTAGAAAGTTCCGAAAGCA

12S rRNA (partial):
CAAAGGCTTGGTCCTAGCTTTACTATTATTTATAACCAAACTTACACATGCAAGCATCCGCACTCCCGTGA
GAATGCCCTTAACCCTCTTATGAGATCAAGGAGCTGGTATCAGGTACAAATAATTGCCCATAACACCTTGC
TTAGCCACACCCCCAAGGGAATTCAGCAGTGATAAACATTAAGCCATAAGTGTAAACTTGACTTAGTTAA
GGTTTTTAGAGCCGGTAAAACTCGTGCCAGCCACCGCGGTTATCCTG

Table A6. ALLELOCIDE code

```
'Set ConnWrite = New ADODB.Connection
'Set cmdWrite = New ADODB.Command
'Set rsWrite = New ADODB.Recordset
'
'With ConnWrite
'  .Provider = "MicroSoft.Jet.OleDb.4.0"
'  .ConnectionString = "c:\haplodata\hapolodata.mdb"
'  .Open
'End With
'
'With cmdWrite
'  .ActiveConnection = ConnWrite
'  .CommandType = adCmdText
'  .CommandText = "delete from lifetable"
'End With
'cmdWrite.Execute
'
'"Set ConnCheck = New ADODB.Connection
'Set cmdCheck = New ADODB.Command
'Set rsCheck = New ADODB.Recordset
'
'With ConnCheck
'  .Provider = "MicroSoft.Jet.OleDb.4.0"
'  .ConnectionString = "c:\haplodata\hapolodata.mdb"
'  .Open
'End With
'
" rsCheck.Open "Select * from HaploTypes", Conn, adOpenKeyset, adLockBatchOptimistic
' Label3.Caption = rsCheck.RecordCount
' rsCheck.MoveFirst
' For i = 1 To rsCheck.RecordCount
'    Selected1 = Selected1 + rsCheck!Percentage
'    rsCheck.MoveNext
' Next
' Label4.Caption = FormatNumber(100 - Selected1, 2) & "%"
'
'ConnCheck.Close
'Set ConnCheck = Nothing

'Set ConnSurvive = New ADODB.Connection
'Set CmdSurvive = New ADODB.Command
'Set RsSurvive = New ADODB.Recordset
'
'With ConnSurvive
'  .Provider = "MicroSoft.Jet.OleDb.4.0"
'  .ConnectionString = "c:\haplodata\hapolodata.mdb"
'  .Open
'End With
'
" RsSurvive.Open "Select * from HaploTypes", Conn, adOpenKeyset, adLockBatchOptimistic
' Label3.Caption = RsSurvive.RecordCount
' RsSurvive.MoveFirst
' For i = 1 To RsSurvive.RecordCount
'    Selected1 = Selected1 + RsSurvive!Percentage
'    RsSurvive.MoveNext
' Next
' Label4.Caption = FormatNumber(100 - Selected1, 2) & "%"
'
'ConnSurvive.Close
'Set ConnSurvive = Nothing
```

Table A16. Results from five sets of ALLELOCIDE runs. The starting population size was 1000 individuals.

% individuals removed from population	No. runs in which haplotypes are lost from the population (per 100 runs)					Mean no. runs in which haplotypes are lost (±SD)	
	1	**2**	**3**	**4**	**5**		
50	1	0	0	0	0	0.2	(±0.45)
52	1	0	0	0	0	0.2	(±0.45)
54	1	0	1	0	0	0.4	(±0.55)
56	1	1	1	1	1	1	(±0.00)
58	1	1	1	1	2	1.2	(±0.45)
60	3	1	1	2	2	1.8	(±0.84)
62	3	1	2	5	3	2.8	(±1.48)
64	3	1	7	8	7	5.2	(±3.03)
66	5	8	5	6	9	6.6	(±1.82)
68	5	7	10	12	10	8.8	(±2.77)
70	6	8	15	15	6	10	(±4.64)
72	24	12	16	16	14	16.4	(±4.56)
74	19	23	19	19	28	21.6	(±3.97)
76	29	30	31	19	23	26.4	(±5.18)
78	35	32	36	38	36	35.4	(±2.19)
80	42	42	52	33	54	44.6	(±8.53)
82	64	62	60	54	57	59.4	(±3.97)
84	66	71	62	68	70	67.4	(±3.58)
86	81	80	70	76	80	77.4	(±4.56)
88	86	87	85	94	93	89	(±4.18)
90	93	93	94	91	95	93.2	(±1.48)
92	98	98	99	98	100	98.6	(±0.89)
94	99	100	100	98	95	98.4	(±2.07)
96	100	100	100	100	100	100	(±0.00)
98	100	100	100	100	100	100	(±0.00)
100	100	100	100	100	100	100	(±0.00)

Table A17. Mean number of haplotypes lost in five sets of ALLELOCIDE runs. The starting population size was 1000 individuals.

% individuals removed from population	Mean no. haplotypes lost in 100 runs					Mean no. haplotypes lost in 5 repetitions (±SD)	
	1	2	3	4	5		
50	0.01	0.00	0.00	0.00	0.00	0.00	(±0.00)
52	0.01	0.00	0.00	0.00	0.00	0.00	(±0.00)
54	0.01	0.00	0.01	0.00	0.00	0.00	(±0.00)
56	0.01	0.01	0.01	0.01	0.01	0.01	(±0.00)
58	0.01	0.01	0.01	0.01	0.02	0.01	(±0.00)
60	0.03	0.01	0.01	0.02	0.02	0.02	(±0.01)
62	0.03	0.01	0.02	0.05	0.03	0.03	(±0.01)
64	0.03	0.01	0.07	0.08	0.07	0.05	(±0.03)
66	0.05	0.08	0.05	0.06	0.09	0.07	(±0.02)
68	0.05	0.07	0.10	0.12	0.10	0.09	(±0.02)
70	0.06	0.08	0.17	0.17	0.07	0.11	(±0.05)
72	0.28	0.14	0.16	0.18	0.16	0.18	(±0.05)
74	0.20	0.25	0.19	0.20	0.31	0.23	(±0.04)
76	0.32	0.33	0.35	0.23	0.27	0.30	(±0.04)
78	0.46	0.38	0.45	0.48	0.39	0.43	(±0.04)
80	0.55	0.52	0.80	0.36	0.65	0.57	(±0.14)
82	0.90	0.88	0.71	0.69	0.68	0.77	(±0.10)
84	0.92	1.03	0.99	1.04	0.96	0.99	(±0.04)
86	1.30	1.39	1.17	1.27	1.35	1.30	(±0.08)
88	1.72	1.70	1.64	1.89	1.68	1.73	(±0.09)
90	2.23	2.01	2.14	2.30	2.05	2.15	(±0.11)
92	2.55	2.75	3.01	2.99	3.05	2.87	(±0.19)
94	3.66	3.69	3.58	3.75	3.19	3.58	(±0.20)
96	4.80	4.45	4.79	4.85	4.79	4.74	(±0.14)
98	6.50	6.58	6.38	6.68	6.38	6.50	(±0.12)
100	13.00	13.00	13.00	13.00	13.00	13.00	(±0.00)

APPENDIX II

Molecular Ecology (2003) 12, 1703–1715 doi: 10.1046/j.1365-294X.2003.01852.x

Population genetics of the endangered Knysna seahorse, *Hippocampus capensis*

P. R. TESKE, M. I. CHERRY and C. A. MATTHEE
Department of Zoology, University of Stellenbosch, Private Bag XI, Matieland 7602, South Africa

Abstract

The evolutionary history of the endangered Knysna seahorse, *Hippocampus capensis*, and the extent of gene flow among its three known populations, were investigated using 138 mitochondrial DNA control region sequences. Similarly high levels of genetic diversity were found in two of the populations (Knysna and Keurbooms Estuaries), whereas diversity in the third population (Swartvlei Estuary) was lower. Although most haplotypes are shared between at least two populations, based on the haplotype frequency distributions the three assemblages constitute distinct management units. The extant population structure of *H. capensis* suggests that the Knysna seahorse originated in the large Knysna Estuary. The presence of seahorses in the two smaller estuaries is either the result of a vicariance event at the beginning of the present interglacial period, colonization of the estuaries via the sea, or a combination of the two.

Keywords: control region, gene flow, estuaries, genetic structure, *Hippocampus capensis*, mtDNA

Received 1 October 2002; revision received 29 January 2003; accepted 19 February 2003

Introduction

The Knysna seahorse, *Hippocampus capensis* Boulenger, 1900, was the first seahorse species to be listed as endangered on the IUCN Red Data List (Hilton-Tyler 2000). This status is derived from its limited distribution and, consequently, its low global abundance and vulnerability. The species is endemic to South Africa and is the only seahorse species known to exclusively inhabit estuaries.

Extensive SCUBA surveys in the region during 2000–2002 revealed that the current distribution of Knysna seahorses is restricted to the Knysna, Swartvlei and Keurbooms Estuaries on the south coast of the Western Cape Province, South Africa (Lockyear, Seahorse Research Group Knysna, personal communication; Fig. 1). Reports of any additional populations, e.g. in the Klein Brak Estuary (Whitfield 1995), could not be confirmed. Human settlement along all three estuaries (with the associated industrial, domestic and recreational activities) poses a severe threat to the survival of the species (Skelton 1987). Apart from human activities, the seahorses are also exposed to natural hazards. Freshwater floods regularly result in seahorse mortality (Grange, personal communication, Russell 1994) and tend to be more severe in the Swartvlei and Keurbooms Estuaries than in the larger Knysna Estuary (Lockyear, personal communication).

From a phylogeographical perspective the distribution of the estuarine Knysna seahorse is interesting. As the distribution of *H. capensis* is closely linked to suitable habitat, with most individuals being restricted to subtidal vegetation in shallow water (Lockyear, personal communication), some population genetic structuring may exist among and even within estuaries. Previous genetic studies on population structure in teleosts have dealt extensively with freshwater species (Waters & Burridge 1999; Waters *et al.* 2001), marine species (Pogson *et al.* 1995; Gold & Richardson 1998; Mamuris *et al.* 1999; Nesbø *et al.* 2000; Bowen *et al.* 2001; Planes *et al.* 2001; Planes 2002) and diadromous species (i.e. species whose life cycles include both freshwater and marine phases; Gyllensten 1985; Thomas *et al.* 1986; Ayvazian *et al.* 1994; Chenoweth *et al.* 1998), but a paucity of information is available on estuarine endemics. Therefore, our study is filling an important gap.

The fact that Knysna seahorses are not capable of powerful independent locomotion and are thus unable to actively disperse to other estuaries via the marine habitat suggests that genetic differentiation among populations may be high. However, passive dispersal by means of currents cannot be ruled out. It is unlikely that Knysna seahorses may establish themselves in the marine habitat because of a combination of low water temperatures

Correspondence: P. R. Teske. Fax: +27 (0)21 808 2405; E-mail: ptl@sun.ac.za.

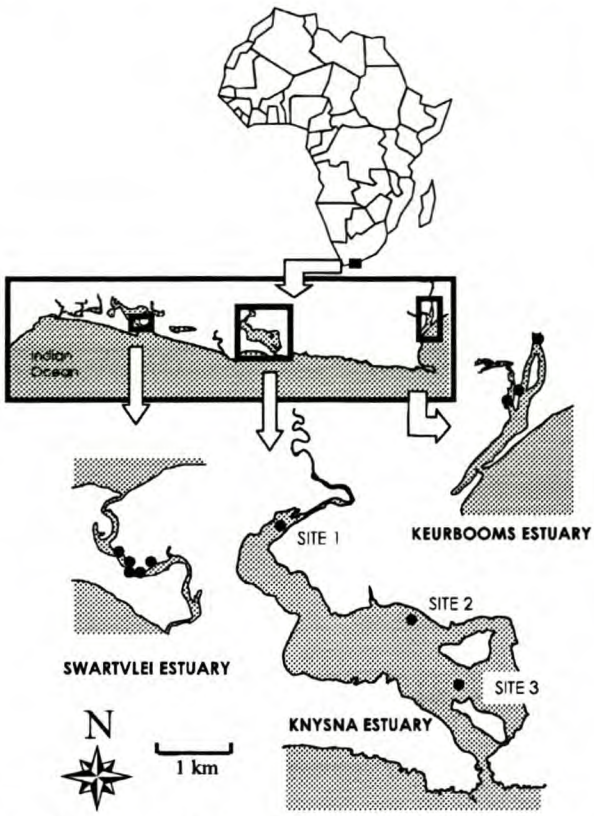

Fig. 1 Geographic localities and relative sizes of estuaries sampled in this study.

associated with occasional upwelling events (Bower & Crawford 1981; Schumann 2000), scarcity of suitable habitat (Branch & Branch 1992) and scarceness of food, as the coastal region is far less productive than estuaries (Day *et al.* 1981; Branch & Branch 1992), but a small portion of migrating seahorses may survive these adverse conditions for a sufficiently long period of time to allow them to colonize new habitats. *H. capensis* is able to survive salinities ranging from one to 59 (Riley, unpublished data, cited in Whitfield 1995), and captive individuals have reproduced successfully in seawater over several generations (Gunter, Port Nolloth Sea Farms, personal communication), supporting the notion that high salinity does not present a barrier to gene flow.

However, even if the potential for gene flow is high, unique haplotypes or differences in haplotype frequencies may exist among the three populations as a result of genetic divergence due to differences in environmental conditions among the three estuarine systems. The three estuaries inhabited by *H. capensis* differ considerably with regard to hydrological aspects. The Knysna Estuary (estuarine mouth: 34°04′ S, 23°03′ E) is by far the largest of the three estuarine systems, covering a water surface area of ≈ 19 km² (Geldenhuys 1979). The mouth of the estuary is characterized by a rock formation of Ordovician origin known as the Knysna heads (Toerin 1979), which maintains a large, permanently open estuary mouth. The large input of tidal marine water and the limited inflow of freshwater ensure that much of the Knysna Estuary's faunal diversity is of marine origin. The coexistence of marine, diadromous and endemic estuarine forms results in the Knysna Estuary having the highest biodiversity of any South African estuary (Grindley 1985), and the presence of at least two rare fish species (*H. capensis* and the goby *Pandaka silvana*) confers a high conservation value to this system. The mouth of the Keurbooms Estuary (34°02′ S, 23°23′ E) is located ≈ 42 km east of the Knysna Estuary. This estuary is fed by the Keurbooms and Bitou rivers, and covers an area of ≈ 2.7 km² (Duvenage & Morant 1984). The highest mean annual runoff estimate for the Keurbooms and Bitou Rivers has been 1.6×10^8 m³ (Noble & Hemens 1978), which is higher than the highest value measured for the Knysna Estuary (1.3×10^8 m³, Pitman 1981). This, in combination with a much smaller estuary area and a small inlet, suggests that the impact of freshwater floods on the Keurbooms Estuary's fauna is much more severe than in the Knysna Estuary. The Swartvlei system (34° S, 22°46′ E), which is located ≈ 26 km west of the Knysna Estuary, consists of a lower estuarine zone and an upper lake. The estuary is the smallest of the three systems with a water surface area of 2 km² (Liptrot 1978) and is periodically isolated from the sea by a sand bar, which forms in its mouth area due to a combination of low freshwater input and longshore winds (Whitfield *et al.* 1983). The mean annual runoff from this estuary's catchment area is comparatively low (6.6×10^7 m³, Anonymous 1978), but mass mortalities occur when the sand bar is breached (Russell 1994; Teske personal observation).

In order to devise adequate management strategies for an endangered species, it is important to investigate its population history, geographical partitioning throughout its range, and distribution of genetic diversity (Avise 1989; O'Brien 1994). Based on conventional meristics, morphometric work and limited sequencing of the mitochondrial cytochrome *b* gene, Toeffie (2000) found that a certain amount of variation may exist between the Knysna and Swartvlei populations, and that mixing of the two populations should be avoided to preserve local adaptations until more information becomes available. In this study, mitochondrial DNA (mtDNA) control region (CR) rapidly evolving right domain sequences were used to investigate the above issues in greater detail. We were particularly interested in identifying areas of high genetic diversity, determining the level of gene flow among populations inhabiting the three estuaries (in order to assess whether they constitute distinct management units according to the definition of Moritz 1994) and to investigate whether there is any evidence of genetic substructuring within the large

Knysna Estuary. In the absence of any additional data, this mtDNA information will be important as a first step in identifying areas of high conservation value and to provide preliminary guidelines for management strategies regarding the translocation of seahorses among or within estuaries.

Materials and methods

Sample acquisition

A total of 138 specimens were sampled from all three estuaries (Fig. 1) during 2001–2002. In most cases, tissue samples were obtained nondestructively by taking small fin clips (< 1 mm²) from the lower edge of the dorsal fin of living seahorses, and these were stored in 70% ethanol. This method has been used previously to obtain tissue samples from seahorses by Kvarnemo *et al.* (2000). Because of the high conservation status of *Hippocampus capensis*, we tested it thoroughly on seahorses bred in captivity and did not find any adverse effects (additional information is available from the authors on request). To test possible population substructure within the large Knysna Estuary, three sites were chosen in regions of the estuary which SCUBA surveys had identified as containing high densities of seahorses (Lockyear, personal communication; Fig. 1). Thirty seahorses were arbitrarily sampled from each of these three sites. Site 1 was located at the head of the estuary, and was characterized by turbid water and low salinity. Site 2 was located in the middle section of the estuary, which is characterized by higher salinities and slower current velocities (Largier *et al.* 2000). Site 3 was located in close proximity to the mouth. This portion of the estuary is characterized by near seawater salinities, cooler water temperature and strong tidal currents (Largier *et al.* 2000). Seahorses found at the two upper sites were all adults, whereas those at site 3 were all juveniles and subadults. Tissue samples from seahorses in the Swartvlei Estuary consisted partly of fin clips obtained from living seahorses (6 individuals), and partly of entire pectoral fins obtained from seahorses that had died following a breaching event (24 individuals). Samples originated from a large area in the middle section of the estuary (Fig. 1). The Keurbooms population is considered the smallest of the three assemblages (Lockyear, personal communication), and was believed to be extinct until a number of individuals were found in 2002. Hence, a permit was granted to sample only 18 individuals. These originated from sites in the upper estuary, two of which were located in the lower Bitou River, and one was located in the upper part of the Keurbooms Estuary (Fig. 1).

DNA extraction, amplification and sequencing

Genomic DNA was isolated from fin clips by proteinase K digestion followed by a standard phenol–chloroform extraction procedure (Sambrook *et al.* 1989). Samples were subsequently resuspended in 50 μL of TE buffer.

A set of primers was designed to amplify 533 nucleotides of CR right domain (forward primer: HCAL2: 5′-CACACTTTCATCGACGCTT-3′; reverse primer: HCAH2: 5′-TCTTCAGTGTTATGCTTTA-3′). This portion of the control region was chosen because it amplified most readily. Primers designed to amplify the left domain or the whole control region did not routinely amplify polymerase chain reaction (PCR) product, which was a particular problem in the case of small fin clips from juveniles and degraded DNA from dead specimens. We are confident that the gene fragment sequenced is mitochondrial rather than a nuclear pseudogene because no heterozygous sequences were identified and no multiple bands were amplified.

The DNA of 138 specimens of *H. capensis* was amplified with PCR. Each 50 μL PCR contained ≈ 1 ng/mL of total genomic DNA, 0.2 μM of each dNTP, reaction buffer including 100 mM NaCl, 0.1 mM EDTA and 20 mM Tris–HCl (pH 8.0), 0.4 μM of each primer, 2.5 mM of $MgCl_2$ and 1 unit of thermostable polymerase. The PCR profile consisted of an initial denaturation step (5 min at 94 °C), followed by 35 cycles of denaturation (30 s at 94 °C), annealing (1 min at 50 °C) and extension (1 min at 72 °C), and a final extension step (10 min at 72 °C). PCR products were directly cycle-sequenced using a BigDye sequencing kit (Applied Biosystems) and the data were analysed on a 3100 AB automated sequencer. The 5′ portion (light strand) of the fragment amplified contains a long repetitive series of up to 12 thymine nucleotides. Owing to difficulties in sequencing through this array, the fragment was sequenced in the reverse direction only. As a control, PCR amplification and subsequent sequencing reactions were repeated for every tenth sample. In all cases, the duplicate sequences were identical to the original sequences.

Data analysis

Sequences were aligned by eye in PAUP Version 4.0 beta 10 (Swofford 2002). A homologous region of 402 nucleotides was obtained for all individuals. Fifteen haplotypes were obtained, and these have been deposited in GenBank (Accession nos AY149664–AY149678). To characterize genetic variation among sampling sites and populations, estimates of nucleotide diversity (π), haplotype diversity (h) and mean number of pairwise differences (d) were obtained using the software package DNASP (Rozas & Rozas 1999). In order to minimize the effects of unequal sample sizes among populations, haplotypes of the Knysna and Swartvlei populations were subsampled. Each subsample consisted of 18 or 30 sequences randomly selected from the total set of sequences using a random number generator (http://www.randomizer.org). In order to calculate means and standard deviations of genetic indices,

nine subsamples were created in each case. An intraspecific phylogeny for *H. capensis* was inferred using the program TCS (Clement *et al.* 2000), and an analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) was conducted to investigate population structure within the Knysna Estuary and among the three estuaries. A distance matrix was constructed using the Tamura Nei model (Tamura & Nei 1993), the corresponding gamma shape distribution parameter was calculated using maximum likelihood in PAUP, and the significance of Φ-statistics was tested using 10 000 permutations of a nonparametric permutation approach described in Excoffier *et al.* (1992). F_{ST} (Wright 1965), Φ_{ST} from AMOVAS and Φ_{ST} from pairwise analyses were estimated as measures of population differentiation, using the program ARLEQUIN Version 2.000 (Schneider *et al.* 2000).

The populations were tested for evidence of recent genetic bottlenecks by implementing a graphical method introduced in Luikart *et al.* (1998). Haplotypes were assigned to frequency classes based on their particular frequency of occurrence within a population. Frequency classes were then plotted against the number of haplotypes assigned to each class. As rare alleles are more likely to become purged from a population during a bottleneck than intermediate or abundant alleles, fewer haplotypes should be found in the lowest frequency class than in one or more intermediate frequency classes. A simulation programme (ALLELOCIDE, available from the authors on request) was written to identify the size at which a randomly mating population which has undergone a recent catastrophic reduction in population size, e.g. due to a freshwater flood, is at risk of losing genetic diversity (i.e. undergoes a genetic bottleneck). In order to determine by how much the female effective population size of the Knysna population would have to decrease for this population to risk losing haplotypes, the following hypothetical model was employed. The starting population consisted of 30 000 individuals, haplotype proportions reflected those found in the Knysna population, and the number of individuals was reduced in increments of 100, each run being repeated 100 times. The starting population size was then gradually decreased to approach the value at which the first haplotypes were lost (10 000 individuals, 5000 individuals and 1000 individuals). Ten runs were then carried out with a starting population of 1000 individuals, the number of individuals being decreased in increments of 20 and each run being repeated 100 times. The percentage of replicates per run in which haplotypes were lost and the mean number of haplotypes lost in 100 replicates were recorded.

Historical demographic patterns of the Knysna population were investigated by making joint maximum likelihood estimates of the parameters θ (= $2N_f\mu$, where N_f is the effective female population size and μ is the hypothetical mutation rate for mtDNA control region right domain) and *g* (the exponential growth parameter) using the program FLUCTUATE Version 1.3 (Kuhner *et al.* 1998). In order to comply with the data requirements of this program, four rare haplotypes were removed, and two alternative methods were employed. For the first method, homoplasies were eliminated by changing nucleotides found at problematic positions to represent the same nucleotide as those found in the most closely related ancestral haplotype. To compensate for this, the changed nucleotide substitution was introduced elsewhere so that the net number of mutations remained unchanged. For the second method, the homoplasic site at position 344 was removed, resulting in a total of only seven haplotypes. In both cases, randomly created input trees were used for each run. If the program is run for sufficiently long, any bias created by the starting tree should be lost (Kuhner personal communication). Watterson's (1975) segregating sites estimate was used as the initial estimate of θ for each run, and a transition/transversion ratio of 12.9 was estimated using the maximum likelihood algorithm in PAUP. Historical N_f values were calculated using the formula $N_t = \theta\, e^{-(g\mu)t}$, where N_t is the effective female population size at time t in the past (Kuhner *et al.* 1998). The present female effective population size was determined using the formula $N_f = \theta$ generation time^{-1} μ^{-1}. A divergence time of 1 Myr per 2% sequence divergence has been widely used for bony fishes (Brown *et al.* 1979; Bermingham & Avise 1986; Grewe *et al.* 1990), but this mutation rate is based on the entire mtDNA molecule. The specific mutation rate for the noncoding control region is higher, and rates used in the recent literature range between 3.6%/Myr (Donaldson & Wilson 1999) and 5–10%/Myr (Brunner *et al.* 2001). As calibration of a molecular clock is not possible in *H. capensis*, hypothetical estimates of θ and *g* were determined using three possible mutation rates: 2, 3.6 and 5%/Myr. The generation time of seahorses is probably no greater than 1–3 years; sexual maturity is attained within 1 year (Whitfield 1995), and seahorses kept in captivity live up to at least 3 years (Lockyear 1997). Hence, results are reported using generation times of 1, 2 and 3 years. For comparison, the parameter θ was also calculated using pairwise estimations as implemented in ARLEQUIN. The parameter θ_S (Watterson 1975; Tajima 1989) is based on the number of segregating sites and the parameter θ_π (Tajima 1983) is based on the number of nucleotide differences.

Harpending's raggedness statistic (Harpending 1994) was calculated in order to determine whether there was significant genetic evidence for population growth or stasis using ARLEQUIN. Departure from selective neutrality was tested using Fu's F_s, Fu and Li's D^*, and Fu and Li's F^* (Fu & Li 1993; Fu 1997). Fu's F_s is particularly suited to detect departures from neutrality in nonrecombining sequences characterized by a high frequency of rare haplotypes and recent mutations (Fu 1997). If a significant departure from selective neutrality is detected only when

Table 1 Frequency of occurrence of *Hippocampus capensis* CR right domain haplotypes in the Knysna Estuary (three sampling sites with 30 individuals each), Keurbooms Estuary (18 individuals) and Swartvlei Estuary (30 individuals). Segregating sites of derived haplotypes are compared with nucleotides at corresponding sites in haplotype 1 (in bold). Nucleotides identical to the ones in haplotype 1 are marked with a dot. The position of each segregating site is indicated by a three-digit number

	n					Segregating sites									
	Knysna sites					0	0	1	1	1	1	2	2	3	3
						0	2	0	3	8	8	4	9	4	9
Haplotype	1	2	3	Keurbooms	Swartvlei	9	2	5	1	7	9	5	4	4	5
1	**15**	**13**	**8**	**3**	**9**	**T**	**C**	**A**	**T**	**C**	**—**	**A**	**C**	**T**	**G**
2	0	1	0	1	0	.	.	.	.	.	C	.	.	.	.
3	1	1	5	1	0	A	.	.	.	.	.	.	.	.	.
4	1	0	0	0	0	.	.	.	.	T	.	.	.	.	.
5	1	3	4	0	0	.	.	.	.	.	.	.	.	C	.
6	0	0	1	0	0	.	.	.	.	.	C	.	.	C	.
7	0	0	1	1	0	.	.	.	.	.	.	.	T	C	.
8	4	5	3	1	0	.	T	.	.	.	.	.	.	.	.
9	2	1	0	0	0	.	T	.	.	.	.	.	.	C	.
10	0	0	1	0	0	.	T	.	.	.	.	G	.	C	.
11	0	0	0	1	0	.	T	.	.	.	.	.	T	C	.
12	4	6	7	9	0	.	T	G	.	.	.	.	.	.	.
13	0	0	0	0	1	.	T	G	.	.	.	.	.	C	.
14	1	0	0	0	0	.	T	G	.	.	.	.	.	C	A
15	1	0	0	1	20	.	T	G	C	.	C	.	.	.	.

implementing Fu and Li's tests, this suggests that background selection is the more likely cause of this deviation from neutrality. If, however, only F_s is significant, departure from the assumption of neutrality is more likely due to population growth or hitchhiking (Fu 1997). Fu's F_s was calculated using ARLEQUIN with 10 000 simulated samples, whereas Fu and Li's tests were determined using DNASP.

Results

The segment of the mitochondrial control region sequenced in 138 individuals of *Hippocampus capensis* contained 10 variable sites (1 indel, 8 transitions and 1 transversion; Table 1). These polymorphic sites defined 15 haplotypes, of which 6 were unique to the Knysna population, 1 was unique to the Keurbooms population and 1 was found exclusively in the Swartvlei population. Six haplotypes were represented by single individuals. Although the sampling area in the Swartvlei Estuary was considerably larger than each individual sampling site within the Knysna Estuary, the number of haplotypes found was lower than at any one of the three sites within the Knysna Estuary. Haplotype 1 was the most abundant haplotype at each of the sites within the Knysna Estuary, and the second most abundant in the Keurbooms and Swartvlei populations. Haplotype diversity and expansion coefficients were similar in the Knysna and Keurbooms estuaries, and lower in the Swartvlei Estuary, whereas nucleotide diversity was higher in the two smaller estuaries than in the Knysna Estuary because their haplotypes tended to be more divergent (Table 2). Magnitudes of haplotype and nucleotide diversities were not drastically affected by subsampling. This suggests that smaller sample sizes were sufficient to calculate good approximations of the genetic indices of each of the three populations.

No significant structure was found among the three sites within the Knysna Estuary using AMOVA (% variation among populations = 0; $\Phi_{ST} = 0$; $P = 0.8$). Hence, the sequences from this population were combined in subsequent analyses. An AMOVA revealed significant structure between the three estuarine populations (% variation among populations = 29.49; $\Phi_{ST} = 0.331$; $P < 0.01$), and pairwise comparisons found significant structure among all three estuaries (Table 3).

A star-like pattern was identified in the haplotype network constructed using the program TCS (Fig. 2), which indicates recent ancestral monomorphism followed by a population expansion (Slatkin & Hudson 1991). Using some of the criteria outlined in Crandall & Templeton (1993), haplotype 1 has been designated as ancestral. It is the most abundant haplotype in the Knysna Estuary, and it has the most pivotal position in the network. Owing to several equally parsimonious solutions, the relationship between haplotype 1, and particularly the most derived haplotypes, is ambiguous. The relationships between such haplotypes were resolved using the criterion suggested by

Table 2 Genetic diversity indices and sample sizes (n) of the Knysna, Swartvlei and Keurbooms populations, as well as individual sampling sites within the Knysna Estuary. Indices include number of haplotypes (H); haplotype diversity (h); nucleotide diversity (π), number of polymorphic (segregating) sites (S); mean number of pairwise nucleotide differences (d); expansion coefficient (S/d)

Site	n	H	h	π	S	d	S/d
Knysna combined	90	13	0.78	0.00353	10	1.41	7.1
Knysna combined*	30	7.22 ± 1.20	0.77 ± 0.04	0.0032 ± 0.0004	5.9 ± 1.2	1.19 ± 0.37	5.9 ± 3.9
Knysna							
1	30	9	0.73	0.00348	8	1.40	5.7
2	30	7	0.76	0.00299	5	1.20	4.2
3	30	8	0.84	0.00425	7	1.71	4.1
Swartvlei	30	3	0.48	0.00461	5	1.86	2.7
Knysna combined*	18	5.78 ± 1.30	0.75 ± 0.08	0.0032 ± 0.0006	4.9 ± 1.3	1.27 ± 0.23	3.8 ± 0.6
Swartvlei*	18	2.55 ± 0.53	0.46 ± 0.05	0.0045 ± 0.0005	4.5 ± 0.5	1.80 ± 0.20	2.9 ± 0.5
Keurbooms	18	8	0.75	0.00458	7	1.84	3.8

*Values represent mean (± SD) calculated from nine subsamples. Each subsample consisted of a number of sequences randomly chosen from the original sample using a random number generator (http://www.randomizer.org).

Table 3 Pairwise comparisons of genetic structure among the three estuaries; below diagonal: F_{ST} values; above diagonal: Φ_{ST} values

	Knysna	Keurbooms	Swartvlei
Knysna		0.102**	0.432**
Keurbooms	0.075*		0.257**
Swartvlei	0.253**	0.343**	

*P = 0.05; **P = 0.01.

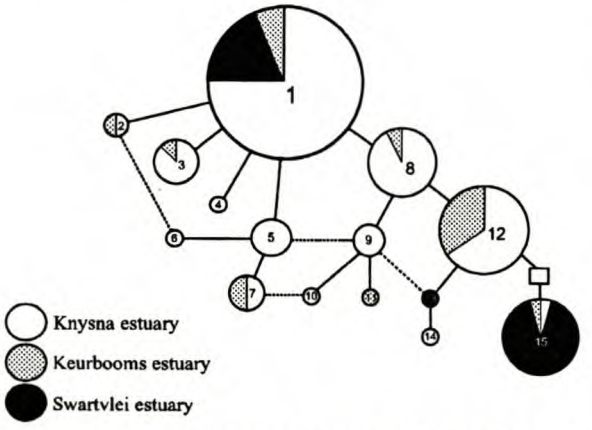

Fig. 2 Haplotype network of CR rapidly evolving right domain haplotypes of *Hippocampus capensis*. Each haplotype is represented by a circle, the size of which indicates the frequency at which it was found. The square represents an internal node haplotype not present in the sample. Each line represents a single nucleotide substitution. Proportional representation of haplotypes in the different estuaries is indicated by subdivision of circles into up to three sections. All connections shown have a probability of ≥ 95% of being true, but connections represented by solid lines are more highly supported than those represented by broken lines by virtue of criteria outlined in the text.

Crandall & Templeton (1993): whenever a derived haplotype was linked to two older haplotypes by an equal number of mutational steps, the derived haplotype was connected to the older haplotype that was present in the population at greater frequency than the other.

The large number of rare haplotypes in the Knysna population resulted in an L-shaped distribution of haplotype frequencies (Fig. 3a), which suggests that this population has not experienced a recent genetic bottleneck (Luikart *et al.* 1998). In contrast, the presence of a single rare haplotype in the Swartvlei Estuary is indicative of a recent bottleneck or founder event in this population (Fig. 3b). A second plot of haplotypes found in the Knysna Estuary based on 9 sets of 18 sequences each (Fig. 3c), shows that the L-shaped pattern is still recovered when using smaller sampling sizes. The Keurbooms population was also characterized by a high frequency of rare haplotypes (Fig. 3d), but haplotype 1 was the second most frequently encountered haplotype in this population. Using the program ALLELOCIDE, a reduction in population size to ≈ 500 individuals resulted in the first rare haplotype/s being lost from the data set of Knysna haplotypes. The number of runs in which haplotypes were lost then increased rapidly as population size was reduced further (Fig. 4). A mean number of one haplotype was lost at a population size of ≈ 150 individuals. Using a different initial population size of 30 000, 10 000 or 5000 individuals had no effect on the results.

The population history of *H. capensis* was investigated using the Knysna population only. Apart from providing the largest data set and containing most of the haplotypes found, this population is most likely to be the most important from a conservation perspective (see Discussion). As the AMOVA results show a lack of subpopulation structure within the Knysna Estuary, the following tests were performed for a pooled data set of this population. A large number of short (1000 steps each) and long (10 000 steps

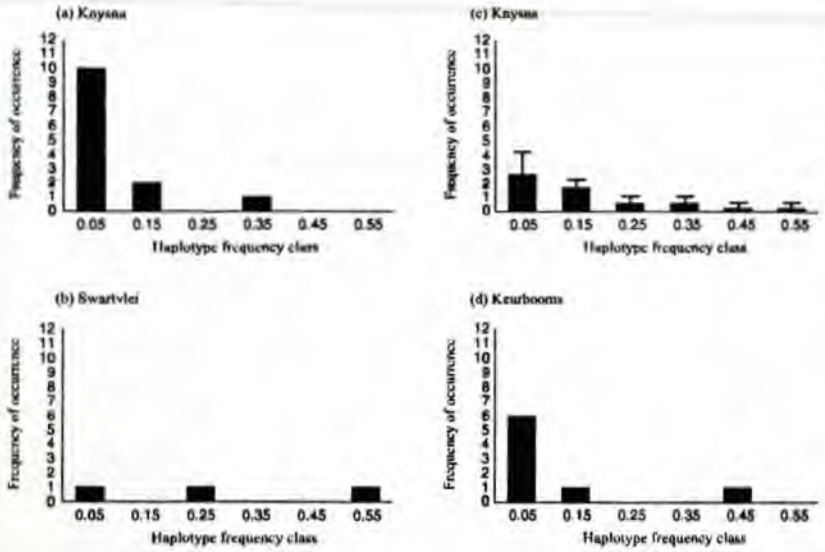

Fig. 3 Control region right domain haplotype frequency distribution; (a) Knysna population (90 individuals); (b) Swartvlei population (30 individuals); (c) Knysna population (18 individuals resampled nine times from original 90; bars indicate means, whiskers represent positive standard deviation); (d) Keurbooms population (18 individuals).

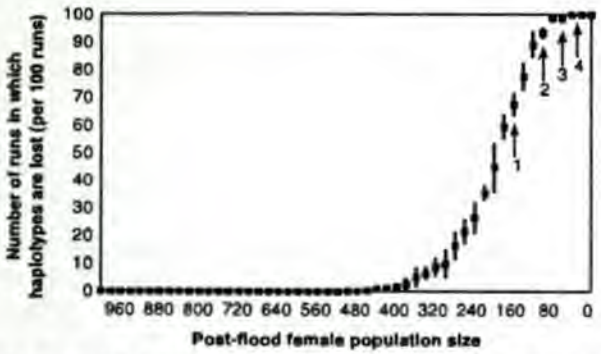

Fig. 4 Loss of haplotypes as a result of a catastrophic reduction in population size. Results plotted from Visual Basic simulation program ALLELOCIDE; starting population size: 1000; increments: 20 individuals; number of replicates: 100. Values plotted are means (± SD) from 10 repetitions. Numbers below some data points represent the mean number of haplotypes lost in 100 replicates of a particular run.

each) chains was run for both methods in the program FLUCTUATE, and sampling increments of 20 were used for both short and long chains. As results differed slightly depending on the number of chains run, the number of chains was increased from 15 to 40 short chains in increments of 5, and approximately half the number of long chains in each run. We report θ and g as the means of six runs. The θ-value (± SD) calculated using the first method was 0.130 ± 0.021 with a simultaneously estimated exponential growth parameter (g) of 1399 ± 152. The results for the second method were as follows: θ (± SD) = 0.210 ± 0.124 and g (± SD) = 1999 ± 768. Values of θ, calculated using approaches based on pairwise comparisions, were $\theta_S = 0.0049 \pm 0.0019$ and $\theta_\pi = 0.0037 \pm 0.0025$.

The recent population expansion suggested by the star-like phylogeny and the large growth rates was confirmed using the raggedness statistic, which did not reject the sudden expansion model ($r = 0.05$, $P = 0.41$). A negative and significant Fu's F_s test for the combined data set from the Knysna population further supports the evidence for a demographic expansion of this population ($F_s = -7.34$, $P < 0.01$). Fu and Li's D^* and F^* for the Knysna population, however, were not significant ($D^* = -2.08$, $P > 0.05$, $F^* = -1.91$, $P > 0.05$). This excludes the possibility that background selection is responsible for the departure from neutrality, and justifies the application of the coalescent-based ML algorithm used in FLUCTUATE to investigate population history. Estimates of θ and g obtained using the first method were used to generate plots of historical population sizes over time (Fig. 5). Assuming that population growth was relatively constant, the approximate age of the Knysna population is defined as the point in time at which the historical N_f value is < 1% of the estimated present effective female population size. Calculated using the first method, this lies between 65 000 and 486 000 years ago, whereas results using the second method placed the age of the population between 46 000 and 339 000 years (not shown).

Discussion

Genetic structure and gene flow

The samples from the Keurbooms and Swartvlei estuaries each contained a single private haplotype (haplotypes 11 and 13), whereas the Knysna Estuary contained a total of six (Table 1). The remaining seven haplotypes were shared among at least two estuaries. Although it is possible that

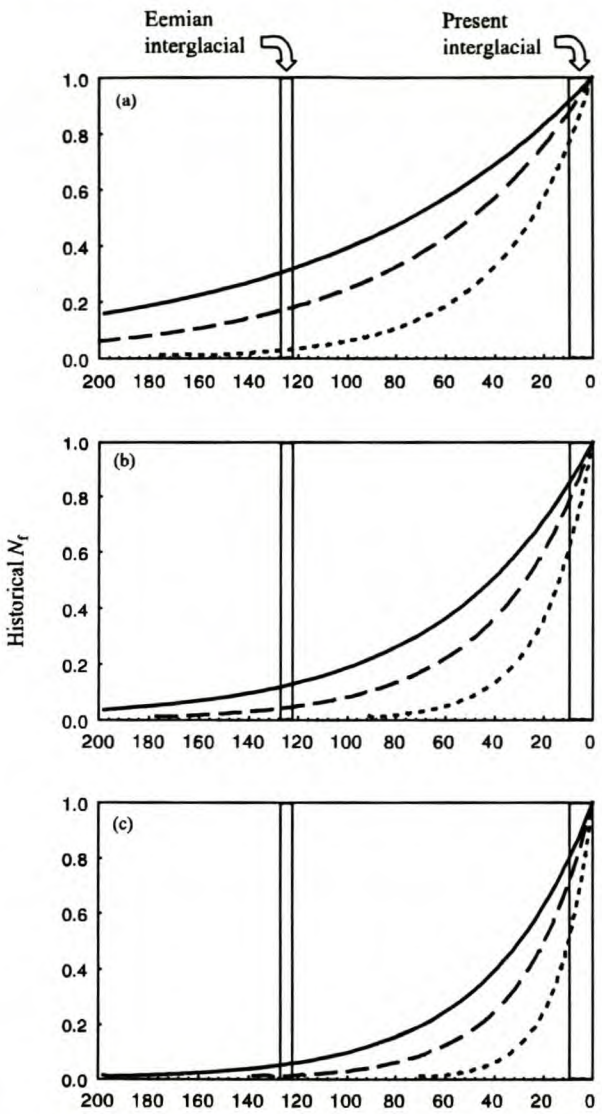

Fig. 5 Plots of historical effective female population sizes (N_f) at three putative CR right domain mutation rates: (a) 2%/Myr; (b) 3.6%/Myr; (c) 5%/Myr. Three possible generation times are plotted: 1 year (dotted line), 2 years (broken line) and 3 years (solid line). Historical N_f values are expressed as proportions of present N_f.

due to sampling efforts not all of the unique haplotypes present in the three populations were represented in the data set, it is interesting to note that clear haplotypic frequency differences exist among the estuaries. The significant structure (supported by both F_{ST} and Φ_{ST} values, Table 3) found among the three estuaries suggests that each of the populations constitutes a distinct management unit *sensu* Moritz (1994). However, the conclusion that the populations in the three estuaries may be evolving relatively independently under different stochastic processes, is at best tentative, because it is based on low sample sizes and a single neutral marker only. The same can be said about the conclusion that the population structure in the Swartvlei population differs considerably from that detected in the other two populations because this population experienced a recent population bottleneck, founder event or loss of haplotypes due to random genetic drift due to small population size: although the right domain of the control region provides good resolution at the demographic level, Luikart *et al.* (1998) recommended that analyses for genetic bottlenecks should involve 5–20 independent loci. We find no evidence for population substructure within the Knysna Estuary and these analyses were based on a total sample size of 90 individuals, suggesting that this conclusion is fairly robust. Lack of genetic structure in the Knysna Estuary is also supported by the fact that site 3 was inhabited exclusively by juveniles and young, nonbreeding adults, whereas the two upper sites were inhabited by adults. Juveniles and adults may live spatially separated in preferred habitat types. The most apparent explanation for this pattern is the nature of the vegetation in the different areas. Site 3 is characterized by very dense patches of seagrass, which seem to be an ideal habitat for smaller seahorses, whereas larger seahorses would find it difficult to move within them. The more open vegetation in the estuary's upper reaches, however, may be more suitable for large seahorses.

Evolutionary history of Hippocampus capensis

Although it is difficult to date the exact time when *Hippocampus capensis* diverged from its marine ancestors because the parameters estimated with FLUCTUATE are imprecise due to an uncertain genealogy and biased because they are based on only one locus (Kuhner *et al.* 1998), the results of this study can be used as a rough indication to date such an event. Using two approaches to determine haplotype relationships, three different mutation rates and three different generation times, the age of the Knysna population has been estimated to between 46 000 and 486 000 years (late Pleistocene). Environmental conditions along the coast during the Pleistocene differed considerably from present-day conditions, and several factors suggest that it was unlikely that tropical or subtropical marine seahorses that may have given rise to *H. capensis* were able to reach the Knysna Estuary during this time. At present, the coastal waters of the east coast are dominated by the warm Agulhas Current, which aids in the southwards dispersal of tropical marine organisms (Heydorn 1978; Blaber 1981; Turpie *et al.* 2000). This dispersal occurs mainly during the austral summer, when sea surface temperatures are in the region of 24 °C and the Agulhas Current is more defined than during winter (Heydorn 1978), suggesting that it is facilitated both by water temperature

and current strength. During the Pleistocene, sea-surface temperatures along South Africa's east coast were up to 4 °C cooler and the Agulhas Current was considerably weaker (Lindesay 1998). Moreover, although the Agulhas Current is deflected away from the coast in the Eastern Cape region because the continental shelf widens, transport of tropical species towards the south coast is possible as eddies may transport water from the Agulhas Current towards the shoreline (Branch & Branch 1992). In contrast, during much of the Pleistocene, the Agulhas Current was deflected eastwards just south of Madagascar (Lindesay 1998). However, a founder event is conceivable during the Eemian interglacial period (127 000–122 000 years ago), a short warm phase within the Pleistocene during which sea surface temperatures along the south coast were higher than present-day temperatures (southwest coast: +3.8 °C; southeast coast: +0.9 °C; Crowley & North 1991). The fossil record indicates that much of the south coast's Eemian coastal marine fauna included species currently restricted to warmer Indian Ocean currents off the Eastern Cape Province, KwaZulu Natal Province, or even Mozambique (Martin 1962; Davies 1971). This suggests that the tropical or subtropical marine seahorses that gave rise to the Knysna population may also have been considerably more abundant in the region during the Eemian interglacial than they are today. On the basis of the results of the first method to calculate θ and g, a generation time of 1–2 years and a mutation rate of 3.6%/Myr or slightly higher, the possibility of a founder event during the Eemian interglacial is well supported (Fig. 5b). The age of the population suggests that apart from being geographically isolated from its sister species and living in a habitat that is likely to be inhospitable to other seahorses because of unstable physical and chemical conditions, *H. capensis* may also be phylogenetically distinct, and the high conservation status of this species thus seems justified. Preliminary results based on control region sequences of closely related marine seahorses confirm this (Teske *et al.* unpublished data).

As there is no suitable habitat along the south coast of South Africa, and *H. capensis* is highly specialized to survive in estuaries, physiological adaptation to the estuarine environment must have taken place in one of the three estuaries. If one assumes that a founder event took place during the Eemian interglacial, then environmental and biotic conditions characterizing the three estuaries must have been similar to present conditions, as sea surface temperatures were similar, and the sea level was only 4 m higher than it is today (Gribnitz & Kent 1989). The Knysna Estuary, with its high haplotypic diversity, seems the most likely location for this event to have taken place. First, the marine-dominated lower reaches of this estuary provide an optimal transition zone for the gradual adaptation of marine species to the estuarine environment. Second, the large size of the estuary reduces the detrimental effects of freshwater floods and can also support a larger population of seahorses. This improves the prospect for an estuarine population to establish itself and ensures its long-term survival. Third, because of its permanently open mouth, the estuary is readily accessible to seahorses migrating along the coastline. The notion that the Knysna population is the oldest is also supported by the fact that a rare estuarine goby, *Pandaka silvana*, is endemic to this estuary (Penrith & Penrith 1972). In contrast, no endemics have been found in the other two estuaries. The evolutionary history of *P. silvana* may be similar to that of *H. capensis*, because like the majority of seahorses, all other species in the genus *Pandaka* are exclusively tropical and marine.

Convergence of the Knysna, Keurbooms and Swartvlei rivers lower on the continental shelf during the colder period between the Eemian and present interglacial, and possibly a vicariance event brought about by a subsequent rise in sea levels, may eventually have resulted in three extant populations. The lower genetic diversity of the population residing in the Swartvlei Estuary can be explained by subsequent loss of haplotypes due to genetic bottlenecks. However, this scenario does not explain why the population in the Keurbooms Estuary has a significantly higher genetic diversity. Preliminary survey data suggest that seahorse densities are approximately equal in the Keurbooms and Swartvlei estuaries (Lockyear, personal communication), and the small sizes of these two estuaries suggest that their populations are thus likely to be similar in size and considerably smaller than the Knysna population. An additional explanation for the observed haplotype pattern is the small-scale migration of seahorses between the estuaries. Assuming that the source population resides in the Knysna Estuary, it is more likely that seahorses that have been flushed out of this system end up in the Keurbooms Estuary rather than the Swartvlei Estuary, because the prevailing coastal current flows eastwards (Branch & Branch 1992). The fact that site 3 in the Knysna Estuary was inhabited exclusively by juveniles and subadults, and that no population substructure was found within this estuary, suggests that juveniles and adults may be spatially separated. Dispersal within the estuary is probably accomplished passively through tidal currents, and it is likely that some juvenile seahorses may be flushed out to the sea before finding suitable habitat. This may provide the two smaller estuaries with an infrequent but continuous input of new colonists.

Although the Keurbooms population is characterized by high genetic diversity, the different ratio of haplotypes to that in the Knysna population (and particularly the low abundance of the common haplotype 1), suggests that this population may nevertheless undergo fluctuations in population size. The estuary experiences floods of substantial magnitude (Duvenage & Morant 1984), which are likely to

be more detrimental to the fauna than floods of similar magnitude in the Knysna Estuary, on account of the Keurbooms Estuary's smaller size. For that reason, the Keurbooms population may be dependent on gene flow from the Knysna population in order to maintain its high genetic diversity. Lastly, it cannot be ruled out that presence of seahorses in the two smaller estuaries is the result of recent introductions by humans, as suggested by Kok (1981). This is particularly plausible in the case of the Swartvlei population: although anecdotal evidence suggests that the species has been present in the Swartvlei Estuary for several decades, this system was not included in the list of estuaries inhabited by *H. capensis* in 1986 (Dawson 1986).

Population size and conservation implications

The high number of rare haplotypes in the Knysna assemblage suggests that this population is sufficiently large to tolerate floods without being at risk of undergoing a genetic bottleneck. This notion is supported by the preliminary results of survey work. The census population size of adult seahorses has been estimated to be ≈ 60 000 (Lockyear & Teske unpublished data). The population has an even sex ratio and > 90% of the males were found to be pregnant during the breeding season (Lockyear & Teske personal observation). This suggests that the effective female population size (N_f) may be close to 30 000. Behavioural work on the Australian seahorse *H. whitei* has shown that these form long-term faithful pair bonds (Vincent & Sadler 1995). Although it is not known whether this is also the case in *H. capensis*, the fact that male seahorses do not collect eggs from more than one female at a time (Vincent 1995), suggests that there is no justification for incorporating reproductive skew into calculations of N_f.

Methods to calculate effective population sizes based on genetic data yield higher values than the survey results (in all cases $\mu = 0.036$/Myr and generation time = 1.5 years were used to calculate N_f from θ, and standard deviations are given). Approaches based on pairwise comparisons of the number of segregating sites or the number of nucleotide differences, arrived at values approximately two to three times as high as N_f calculated from census estimates (N_f [from θ_S] = 90 740 ± 35 185; N_f [from θ_π] = 68 519 ± 46 296). Even higher values of N_f were calculated using FLUCTUATE. When jointly estimating θ and g, effective female population sizes of $2.4 \times 10^6 \pm 0.4 \times 10^6$ (first method) and $3.8 \times 10^6 \pm 1.2 \times 10^6$ (second method) were determined. The considerable differences between the estimates of N_f calculated using different methods suggests that these should be interpreted cautiously because of uncertainties about input variables and the use of a single neutral marker. Given the caveat that the survey prediction is accurate, the fact that N_f values calculated using genetic methods were higher may be an indication that the population size of *H. capensis* in the Knysna Estuary has decreased considerably during the species' short evolutionary history. However, even a drastic decline in population numbers is unlikely to be detected using the genetic methods employed here, because it is unlikely to result in a loss of rare haplotypes. Unless severe and for a prolonged period (which is not applicable in the case of a freshwater flood), it is also unlikely to have an effect on the proportional abundance of each haplotype within the population. The results of the simulation program ALLELOCIDE illustrate this point. A 1% chance of losing the first rare haplotype in the Knysna population (i.e. haplotype/s lost in 1 of 100 runs, Fig. 5) exists when population numbers have been reduced to 500 individuals, and the number of runs in which haplotypes are lost then increases rapidly as population size is reduced further. If we assume a present effective female population size of 30 000 individuals, then the removal of up to 98% of adult females would not result in the loss of any haplotypes. This conclusion is, however, based on the premise that all of the rare haplotypes present in the population were represented in the sample, which is highly unlikely. The actual minimum number of females remaining in the population in order for haplotypes to be lost may thus be slightly above 500 individuals. Although one also needs to consider that an Allee effect (Allee 1938) resulting from the low population density is likely to further increase the population size at which haplotypes are lost, the fact that adult seahorses were often found in aggregations (see Introduction) suggests that many of the individuals that have survived a freshwater flood may be confined to relatively small areas.

The difference between theoretical and observed effective female population sizes may be an indication that anthropogenic pressures during the past decades could already have had a significant negative impact on the Knysna population. The present rate at which construction developments and other human activities are increasing along the estuary is all the more alarming in the light of these findings. The resulting habitat degradation may make recovery of the population after a naturally occurring disaster such as a freshwater flood increasingly difficult. Whichever scenario resulted in the observed mtDNA population structure in the two smaller estuaries (i.e. vicariance followed by possible genetic bottlenecks in the Swartvlei Estuary vs. migration resulting in recent founder events, or a combination of the two), from a ecological perspective the Knysna Estuary has the greatest potential to ensure long-term survival of the species on account of the large size of the population and the less detrimental effects of freshwater floods. In addition, our mtDNA data also indicated high mtDNA diversity in this estuary and it is thus imperative that conservation efforts for this population be prioritized. Although the three populations of

H. capensis constitute individual management units, this conclusion was reached mainly on the basis of differences in haplotype frequencies among the populations rather than a large proportion of private haplotypes. It cannot be ruled out that these differences in haplotype frequencies may be an artefact of small sample sizes or unbalanced cohort sampling in the two smaller estuaries. Because of the absence of distinct monophyletic clades of haplotypes unique to individual populations and the generally good support for the migration hypothesis, there is at this stage little reason to discourage the translocation of seahorses among the different estuaries, but only if this should become necessary. The lack of population structure within the Knysna Estuary suggests that areas temporarily affected by habitat degradation due to construction developments in the vicinity are likely to become repopulated on their own once they have reverted to their original state. Translocations in the Knysna Estuary are feasible in cases in which the degradation of formerly populated areas is expected or documented, in which case seahorses could be collected and released in more pristine parts of the estuary.

Acknowledgements

The authors thank Marlene Gunter (Port Nolloth Sea Farms) and Ian Russell (SANP Research Station in Rondevlei) for providing tissue samples. Jackie Lockyear, Jim Morel and Mark Noffke (Seahorse Research Group, Knysna) helped with the sampling. The program 'ALLELOCIDE — a Visual Basic simulation program to detect loss of haplotypes in a population that has undergone a catastrophic reduction in population size' was written by Zeen Weight. Bettine Jansen van Vuuren, Krystal Tolley, Jackie Lockyear and three anonymous reviewers are thanked for comments on an earlier version of the manuscript. This research was supported by the National Research Foundation and the University of Stellenbosch.

References

Allee WC (1938) *The Social Life of Animals*. Norton, New York.

Anonymous (1978) Hidrouliese studie van die Swartvlei estuarium: effek van bestaande en voorgestelde brue op die waterbeweging, I. Ingeneurs verslag. In: *Opdrag Van die Department van beplanning en die omgewing Stellenbosch*. WNNR Verslag C/SEA 7805/1.

Avise JC (1989) A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution*, **4**, 279–281.

Ayvazian SG, Johnson MS, McGlashan DJ (1994) High levels of genetic subdivision of marine and estuarine populations of the estuarine catfish *Cnidoglanis macrocephalus* (Plotosidae) in southwestern Australia. *Marine Biology*, **118**, 25–31.

Bermingham E, Avise JC (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics*, **113**, 939–965.

Blaber SJM (1981) The zoogeographic affinities of estuarine fishes in south-east Africa. *South African Journal of Science*, **77**, 305–307.

Bowen BW, Bass AL, Rocha LA, Grant WS, Robertson DR (2001) Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. *Evolution*, **55**, 1029–1039.

Bower D, Crawford R (1981) A black south-easter. *Custos*, **10**, 20–21.

Branch G, Branch M (1992) *The Living Shores of Southern Africa*. Struik, Cape Town.

Brown WM, George M, Jr Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the USA*, **76**, 1967–1971.

Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L (2001) Holartic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution*, **55**, 573–586.

Chenoweth SF, Hughes JM, Keenan CP, Lavery S (1998) Concordance between dispersal and mitochondrial gene flow: isolation by distance in a tropical teleost, *Lates calcarifer* (Australian barramundi). *Heredity*, **80**, 187–197.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.

Crowley TJ, North GR (1991) *Paleoclimatology — Oxford Monographs on Geology and Geophysics, No. 18*. Oxford University Press, New York.

Davies O (1971) Pleistocene shorelines in the southern and south-eastern Cape Province (Part 1). *Annals Natal Museum*, **20**, 403–442.

Dawson CE (1986) Family no. 145: Syngnathidae. In: *Smith's Sea Fishes* (eds Smith MM, Heemstra PC), pp. 445–458. Macmillan South Africa, Johannesburg.

Day JH, Blaber SJM, Wallace JH (1981) Estuarine fishes. In: *Estuarine Ecology with Particular Reference to Southern Africa* (ed. Day JH), pp. 197–221. Balkema, Cape Town.

Donaldson KA, Wilson RR Jr (1999) Amphi-panamic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA control region of fishes. *Molecular Phylogenetics and Evolution*, **13**, 208–213.

Duvenage IR, Morant PD (1984) *Estuaries of the Cape. Part II: Synopses of available information on individual systems*. Report no. 31: Keurbooms/Bitou System (CMS 19), Piesang (CMS 18). CSIR Research Report 430. Stellenbosch, South Africa.

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

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

Fu Y-X, Li W-H (1993) Statistical tests of neutral mutations. *Genetics*, **133**, 693–709.

Geldenhuys ND (1979) *Guidelines for the future utilisation of knysna estuary*. Unpublished report, Department Environmental Planning and Energy, Pretoria.

Gold JR, Richardson LR (1998) Mitochondrial DNA diversification and population structure in fishes from the Gulf of Mexico and Western Atlantic. *Journal of Heredity*, **89**, 404–414.

Grewe PM, Billington N, Hebert PDN (1990) Phylogenetic relationships among *Salvelinus* as inferred from mitochondrial DNA divergence. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**, 984–991.

Gribnitz K-H, Kent LE (1989) Marine and fluviatile terraces and colluvial fans reflecting Quarternary climate changes, Gough Island, South Atlantic. *South African Journal of Antarctic Research*, **13**, 2–9.

Grindley JR (1985) *Estuaries of the Cape; Part II: Synopses of available information on individual systems*. Report no. 30: Knysna (CMS 13). CSIR research report 429. Stellenbosch, South Africa.

Gyllensten U (1985) The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous and freshwater species. *Journal of Fish Biology*, **26**, 691–699.

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

Heydorn AEF (1978) Ecology of the Agulhas Current region: an assessment of biological responses to environmental parameters in the south-west Indian Ocean. *Transactions of the Royal Society of South Africa*, **43**, 151–190.

Hilton-Tyler C (2000) *2000 IUCN Red List of Threatened Species*. IUCN, Gland, Switzerland.

Kok HM (1981) Knysna seahorse distribution poses a problem. Part of an article by MM Smith: The seahorse of the Knysna lagoon. *African Wildlife*, **35**, 9.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, **149**, 429–434.

Kvarnemo C, Moore GI, Jones AG, Nelson WS, Avise JC (2000) Monogamous pair bonds and mate switching in the Western Australian seahorse *Hippocampus subelongatus*. *Journal of Evolutionary Biology*, **13**, 882–888.

Largier JL, Attwood C, Harcourt-Baldwin JL (2000) The hydrographic character of the Knysna Estuary. *Transactions of the Royal Society of South Africa*, **55**, 107–122.

Lindesay J (1998) Past climates in Africa. In: *Climates of Southern Continents: present, past and future* (eds Hobbs JE, Lindesay JA, Bridgman HA), pp. 161–206. John Wiley & Sons Ltd., New York.

Liptrot MRM (1978) Community metabolism and phosphorous dynamics in a seasonally closed South African estuary. MSc Thesis, Rhodes University, Grahamstown, South Africa.

Lockyear JF (1997) Studies on captive breeding of the Kysna seahorse, *Hippocampus capensis*. *Aquarium Sciences and Conservation*, **1**, 129–136.

Luikart G, Allendorf FW, Cornuet J-M, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238–247.

Mamuris Z, Stamatis C, Triantaphyllidis C (1999) Intraspecific genetic variation of striped red mullet (*Mullus surmuletus* L.) in the Mediterranean Sea assessed by allozyme and random amplified polymorphic DNA (RAPD) analyses. *Heredity*, **83**, 30–38.

Martin ARH (1962) Evidence relating to the Quaternary history of the Wilderness Lakes. *Transactions of the Proceedings of the Geological Society of South Africa*, **65**, 19–42.

Moritz C (1994) Defining 'evolutionary significant units' for conservation. *Trends in Ecology and Evolution*, **9**, 373–375.

Nesbø CL, Rueness EK, Iverson SA, Skagen DW, Jakobsen KS (2000) Phylogeography and population history of Atlantic mackeral (*Scomber scombrus* L.): a genealogical approach reveals genetic structuring among the eastern Atlantic stocks. *Proceedings of the Royal Society of London, Series B*, **267**, 281–292.

Noble RG, Hemens J (1978) *Inland water ecosystems in South Africa — A review of research needs*. South African Natural Science Programmes Report 34. Pretoria, South Africa.

O'Brien SJ (1994) Genetic and phylogenetic analyses of endangered species. *Annual Review of Genetics*, **28**, 467–489.

Penrith MJ, Penrith M-L (1972) Redescription of *Pandaka silvana* (Barnard) (Pisces: Gobiidae). *Annals of the South African Museum*, **60**, 105–108.

Pitman WV (1981) *Surface water resources in South Africa*, Vol. IV. *Drainage regions EGHJKL the Western Cape. Part 1 (Text) Part 2 (appendices)*. Hydrological Research Unit Report no. 13/81. Johannesburg, South Africa.

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

Planes S, Doherty PJ, Bernardi G (2001) Strong genetic divergence among populations of marine fish with limited dispersal, *Acanthochromis polyacanthus*, within the Great Barrier Reef and Cora Sea. *Evolution*, **55**, 2263–2273.

Pogson GH, Mesa KA, Boutilier RG (1995) Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics*, **139**, 375–385.

Rozas J, Rozas R (1999) DNASP, Version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–168.

Russell IA (1994) Mass mortality of marine and estuarine fish in the Swartvlei and Wilderness Lake systems, Southern Cape. *South African Journal of Aquatic Sciences*, **20**, 93–96.

Sambrook J, Fritsch E, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN, *Version 2000: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

Schumann EH (2000) Oceanic exchanges and temperature variability in the Knysna Estuary. *Transactions of the Royal Society of South Africa*, **55**, 123–128.

Skelton PH (1987) South African Red Data Book — fishes. *South African National Scientific Programmes Report*, **137**, 93–95.

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

Swofford DL (2002) PAUP: *Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0. Beta 10*. Sinauer Associates, Sunderland, MA.

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

Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Thomas WK, Withler RE, Beckenbach AT (1986) Mitochondrial DNA analysis of Pacific salmonid evolution. *Canadian Journal of Zoology*, **64**, 1059–1064.

Toeffie Z (2000) A preliminary examination of the morphology and genetic structure within populations of the Knysna seahorse, *Hippocampus capensis* (Pisces: Syngnathidae), from two South African estuaries. MSc Thesis, University of the Western Cape.

Toerin DK (1979) *The geology of the Oudtshoorn area*. Explanation to sheet 3322. Department of mines, Pretoria. Government Printer.

Turpie JK, Beckley LE, Katua SM (2000) Biogeography and the selection of priority areas for conservation of South African coastal fishes. *Biological Conservation*, **92**, 59–72.

Vincent ACJ (1995) A role for daily greetings in maintaining seahorse pair bonds. *Animal Behaviour*, **49**, 258–260.

Vincent ACJ, Sadler LM (1995) Faithful pair bonds in wild seahorses, *Hippocampus whitei*. *Animal Behaviour*, **50**, 1557–1569.

Waters JM, Burridge CP (1999) Extreme intraspecific mitochondrial DNA sequence divergence in *Galaxias maculatus* (Osteichthyes: Galaxiidae), one of the world's most widespread freshwater fish. *Molecular Phylogenetics and Evolution*, **11**, 1–12.

Waters JM, Craw D, Youngson JH, Wallis GP (2001) Genes meet geology: fish phylogeographic pattern reflects ancient, rather than modern, drainage connections. *Evolution*, **55**, 1844–1851.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

Whitfield AK (1995) Threatened fishes of the world: *Hippocampus capensis* Boulenger, 1990 (Syngnathidae). *Environmental Biology of Fishes*, **44**, 362.

Whitfield AK, Allanson BR, Heinecken TJE (1983) *Estuaries of the Cape. Part II: Synopses of available information on individual systems*. Report no. 22: Swartvlei (CMS 11), CSIR Research Report 421. Stellenbosch, South Africa.

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

This conservation genetic study on the Knysna seahorse reflects one of the research focuses of the evolutionary genomics group at Stellenbosch University, South Africa. The present paper is a component of P.R. Teske's Ph.D. study on seahorse population genetics and molecular systematics. P.R. Teske is primarily interested in the ecology and population structure of estuarine fishes and invertebrates, in particular estuarine endemics. M.I. Cherry has a keen interest in mating systems, and most of his work has focused on sexual selection and on brood parasitism. C.A. Matthee is mainly interested in molecular systematics and evolutionary population genetics.

ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Molecular Phylogenetics and Evolution xxx (2003) xxx–xxx

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

The evolutionary history of seahorses (Syngnathidae: *Hippocampus*): molecular data suggest a West Pacific origin and two invasions of the Atlantic Ocean

Peter R. Teske,* Michael I. Cherry, and Conrad A. Matthee

Molecular Laboratory, Zoology Department, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

Received 2 May 2003; revised 19 May 2003

Abstract

Sequence data derived from four markers (the nuclear RP1 and Aldolase and the mitochondrial 16S rRNA and cytochrome *b* genes) were used to determine the phylogenetic relationships among 32 species belonging to the genus *Hippocampus*. There were marked differences in the rate of evolution among these gene fragments, with Aldolase evolving the slowest and the mtDNA cytochrome *b* gene the fastest. The RP1 gene recovered the highest number of nodes supported by >70% bootstrap values from parsimony analysis and >95% posterior probabilities from Bayesian inference. The combined analysis based on 2317 nucleotides resulted in the most robust phylogeny. A distinct phylogenetic split was identified between the pygmy seahorse, *Hippocampus bargibanti*, and a clade including all other species. Three species from the western Pacific Ocean included in our study, namely *H. bargibanti*, *Hippocampus breviceps*, and *Hippocampus abdominalis* occupy basal positions in the phylogeny. This and the high species richness in the region suggest that the genus probably originated in this region. There is also fairly strong molecular support for the remaining species being subdivided into three main evolutionary lineages: two West Pacific clades and a clade of species present in both the Indo-Pacific and the Atlantic Ocean. The phylogeny obtained herein suggests at least two independent colonization events of the Atlantic Ocean, once before the closure of the Tethyan seaway, and once afterwards.

1. Introduction

Seahorses belong to the Syngnathidae, a teleost family whose oldest fossils date back to the Eocene (Lutetian: 52 mya; Patterson, 1993). The family also includes the pygmy pipehorses (grouped with seahorses in the subfamily Hippocampinae), pipehorses and seadragons (Solegnathinae), flag-tail pipefishes (Doryrhamphinae), and pipefishes (Syngnathinae; Kuiter, 2000). The monophyly of seahorses is supported by a number of synapomorphic morphological characters distinguishing them from most other Syngnathids. These characters include a prehensile tail, the absence of a caudal fin, the position of the head at a right angle to the trunk, a brood pouch sealed along the midline (except for a small anterior opening), and a raised dorsal fin base (Fritzsche, 1980). Seahorses (genus *Hippocampus*) and possibly also pygmy pipehorses (genera *Amphelikturus*, *Acentronura*, and *Idiotropiscis*), are phylogenetically most closely associated with pipefishes of the genus *Syngnathus* (Wilson et al., 2001).

The world's tropical marine faunas can be divided into those associated with an Atlantic Ocean biome (including the Caribbean and Mediterranean), and those associated with an Indo-Pacific biome (Rosen, 1988). It has been suggested that this pattern arose after the closure of the Tethyan seaway, a tectonic event that resulted from the convergence of the African and Eurasian plates during the late Oligocene and Miocene (Rosen, 1988). Seahorses are found throughout the tropical and temperate regions of both the Atlantic and Indo-Pacific biomes, but their origin and evolutionary

* Corresponding author.
E-mail addresses: ptl@sun.ac.za, ptl@acad.sun.ac.za (P.R. Teske),

1055-7903/03/$ - see front matter © 2003 Published by Elsevier Science (USA).
doi:10.1016/S1055-7903(03)00214-8

history are not well understood. In a study based on cytochrome *b* sequences, Casey (1999) concluded that the genus *Hippocampus* probably evolved in the Atlantic biome. An Atlantic origin is also supported by the fact that most species of the closely related pipefish genus *Syngnathus* are associated with the Atlantic biome (Kuiter, 2000), as well as the fact that to date the only known seahorse fossils have been found in Italy (Sorbini, 1988). On the other hand, it is interesting to note that the majority of seahorse species are found in the Indo-West Pacific region (>27 species, Lourie et al., 1999). This pattern is not unique to seahorses—the majority of tropical marine taxa have their greatest concentration of species within the East Indies triangle formed by the Philippines, the Malay Peninsula, and New Guinea (Briggs, 1999). The present-day marine fauna in the Indo-West Pacific is characterized by comparatively recent genera (Newman et al., 1976; Stehli and Wells, 1971) and a large proportion of apomorphic species (Fricke, 1988; Menon, 1977; Ricklefs and Latham, 1993; Specht, 1981). Although these characteristics may suggest that the high species richness in the Indo-West Pacific is a result of recent speciation or colonization, several authors suggested that the region is a centre of origination and radiation of various marine taxa (Briggs, 1999; Lessios et al., 2001; Rosen, 1984), which might include the seahorses.

Irrespective of the origin of the genus, the circumglobal distribution of seahorses reflects major dispersal events. It is possible that some tropical shore species have been able to migrate around the Cape of Good Hope to establish themselves in the Atlantic Ocean, but there is no evidence for such dispersal events in the opposite direction (Briggs, 1995). Migration events from the Atlantic Ocean towards the Indo-West Pacific via the Central American Seaway prior to cessation of gene flow due to the rising of the Isthmus of Panama (3.1–3.5 mya; Coates and Obando, 1996; Collins, 1996; Duque-Caro, 1990a,b; Keigwin, 1982) are theoretically possible, but the expanse of the Pacific Ocean has been shown to be a formidable barrier to dispersal (Ekman, 1953; Rohde and Hayward, 2000).

Apart from the uncertain evolutionary history, the exact species boundaries of many seahorses are obscure. Morphology-based taxonomic methods have shown to be problematic. More than 100 species of seahorses have been described (Eschmeyer, 1998), but a recent attempt by Lourie et al. (1999) at revising the genus accepts only about 32 valid species names. These controversies seem to be mainly due to convergence of morphological characteristics: since seahorses avoid predators by means of camouflage, it seems reasonable to assume that many morphological characters are under strong selection pressure. Genetic methods have great potential to both resolve disputed taxonomic issues and to infer phylogenetic relationships among different species (Arnaud et al., 1999; Bowen et al., 2001; Burridge and White, 2000; Colborn et al., 2001; Grant and Leslie, 2001; McMillan and Palumbi, 1995; Muss et al., 2001).

With the exception of an unpublished study using cytochrome *b* sequences of 22 species of seahorses (Casey, 1999) and a number of additional sequences (mitochondrial cytochrome *b*, 12S rRNA and 16S rRNA) used to investigate the placement of the genus among other Syngnathids (Wilson et al., 2001), genetic data useful for *Hippocampus* phylogeny reconstruction are lacking. Our preliminary analyses of the cytochrome *b* data available on GenBank indicated that although the gene contributed signal towards the tips of the trees (reflecting recent divergence events), the data were not able to resolve the deeper nodes with high confidence. In the present paper we extended these sequence data and used more slowly evolving mitochondrial 16S rRNA sequences, as well as two nuclear DNA gene fragments (the first intron of the S7 ribosomal protein and a section of the Aldolase gene) to construct a phylogeny for seahorses. By using four genes and three independent evolutionary markers we attempted to infer a robust evolutionary tree addressing both the recent and older evolutionary events.

2. Materials and methods

Most of samples used in this study were provided by Project Seahorse (University of British Columbia, Vancouver) and were preserved by drying. Additional ethanol preserved samples, comprising fin clips, muscle, operculum or internal organs were obtained from various other sources (Table 1). The total sample size consisted of 51 individuals from 32 species, following the classification system of Lourie et al. (1999). Two of the species sequenced were not yet included in Lourie et al. (1999), namely *Hippocampus procerus* (Kuiter, 2001) and *Hippocampus queenslandicus* (Horne, 2001). A pipefish of the genus *Syngnathus* (*S. temminckii* in the case of nuclear data and *S. acus* in the case of mitochondrial data) was used as an outgroup taxon.

2.1. PCR and sequencing

Tissue samples were cut into small pieces in order to improve digestion and were then subjected to proteinase K and SDS digestion at 37 °C followed by phenol/chloroform extraction procedures (Sambrook et al., 1989). The polymerase chain reaction (PCR) was used to amplify the first intron (RP1) of the S7 ribosomal protein (primers published by Chow and Hazama, 1998), a portion of the Aldolase gene containing both coding and non-coding regions (forward primer: 5′-TGTGCCCAG TATAAGAAGGATGG-3′; reverse primer: 5′-CCCAT CAGGGAGAATTTCAGGCTCCACAA-3′) and mi-

Table 1
List of specimens sequenced in this study, including code used on phylogenetic trees, region of origin, and individual or institution that provided the samples

Species name	Code	Collection locality	Collector/Source	Gene fragments sequenced		
				Ald	RP1	16S rRNA
Hippocampus abdominalis	abdoAU	SE Australia	(Project Seahorse)	•	•	
H. abdominalis	abdoNZ	New Zealand	(Project Seahorse)	•	•	
H. algiricus	algiBN	Benin	Z. Sohou	•	•	•
H. angustus	anguAU	W Australia	(Project Seahorse)	•		
H. barbouri	barbID1	Indonesia	(Project Seahorse)	•	•	
H. barbouri	barbID2	Indonesia	A. Tuwo		•	
H. bargibanti	bargID1	Indonesia	S. Lourie		•	
H. bargibanti	bargID2	Indonesia	M. Erdmann			•
H. borboniensis	borbTZ	Tanzania	J. Schulz	•		
H. borboniensis	borbMG	Madagascar	(Project Seahorse)	•		
H. breviceps	brevAU1	SE Australia	(Project Seahorse)	•	•	
*H. breviceps**	brevAU2	SE Australia	(Australian Museum)	•		•
H. camelopardalis	cameTZ	Mozambique	M. Cherry	•	•	•
H. capensis	capeZA1	S Africa	P. Teske	•	•	•
H. capensis	capeZA2	S Africa	P. Teske		•	
H. comes	comeVN	Vietnam	A. Vincent	•		•
H. comes	comePH	Philippines	N. Perante	•	•	•
H. comes	come??	(captive)	M. Gunter			
H. coronatus	coroJP	Japan	C. Kawamura	•	•	•
H. erectus	erecUS	Florida	J. Campsen	•	•	
H. fisheri	fishUS	Hawai'i	(Project Seahorse)	•	•	
H. fuscus	fuscEG	Egypt	H. Gabr		•	•
*H. fuscus?**	fuscSL	Sri Lanka?	(Aquarium trade)		•	
H. fuscus	fusc??	(captive)	(Project Seahorse)	•	•	
H. guttulatus	guttIT	Italy	(Project Seahorse)	•	•	
*H. guttulatus**	guttPG	Portugal	J. Curtis	•		•
*H. hippocampus**	hippPG	Portugal	J. Curtis	•	•	•
H. histrix	histVN	Vietnam	Hoang		•	
H. ingens	ingePE	Peru	(Project Seahorse)	•	•	•
H. ingens	ingeMX	Mexico	J. Baum		•	
H. kelloggi	kellIN	India?	(Project Seahorse)	•	•	
H. kelloggi	kellVN	Vietnam	(Project Seahorse)	•	•	•
H. kuda	kudaZA	S Africa	(Sea World Durban)	•	•	•
H. kuda	kudaPH	Philippines	M. Santos	•	•	•
H. kuda?	kudaIN	India?	(Project Seahorse)		•	
H. mohnikei	mohnVN	Vietnam	L.-S. Feng		•	
H. mohnikei	mohnJP1	Japan	C. Kawamura		•	•
H. mohnikei	mohnJP2	Japan	T. Mukai	•	•	
*H. procerus**	procAU	Australia?	(Aquarium trade)		•	
H. queenslandicus	queeAU	NE Australia	P. Southgate	•	•	•
H. reidi	reidMX	Mexico	J. Baum	•	•	•
*H. reidi**	reidBR	Brazil (captive)	(Aquarium trade)		•	
*H. sindonis**	sindJP	Japan	T. Mukai	•	•	•
H. spinosissimus	spinPH	Philippines	(Project Seahorse)	•	•	•
H. subelongatus	subeAU	SW Australia	A. Jones	•	•	•
H. trimaculatus	trimID	Indonesia	(Project Seahorse)	•	•	
H. trimaculatus	trimVN	Vietnam	(Project Seahorse)	•		
H. trimaculatus	trimHK	Hong Kong	(Project Seahorse)			•
*H. whitei**	whitAU1	SE Australia	(Australian Museum)	•	•	•
*H. whitei**	whitAU2	SE Australia	(Australian Museum)		•	
H. zosterae	zostUS	USA	(Project Seahorse)	•		
Syngnathus temminckii	*Syngnathus*	S Africa	P. Teske	•	•	

Identifications of specimens were based on Lourie et al. (1999). Alternative species names (based on Kuiter, 2000) are *H. bleekeri* for Australian *H. abdominalis*, *H. elongatus* for *H. subelongatus*, and *H.* cf. *reidi* for Brazilian *H. reidi*. *H procerus* and *H. queenslandicus* are recently described species not included in Lourie et al. (1999). Specimens used to obtain sequence data for a particular gene fragment are marked with a dot. Specimens not seen by the senior author or Sara Lourie (tissue samples only) are marked with asterisks.

tochondrial 16S rRNA using universal primers (Palumbi, 1996). In some species, tandem repetitive series in the RP1 intron resulted in sequencing difficulties. In order to sequence the remaining portions of the fragment, two internal primers were designed, forward primer S7RPEX3F (5′-TGGTGGAGTWGCAGTGA-3′) and reverse primer S7RPEX4R (5′-ACAAACAACAG ACYRGTAA-3′).

Each 50 μl PCR contained approximately 200 ng/ml of DNA, 0.2 μM of each dNTP reaction buffer (100 mM NaCl, 0.1 mM EDTA, and 20 mM Tris–HCl, pH 8.0), variable $MgCl_2$ (2 mM for RP1, 2.5 mM for Aldolase, or 2 mM for 16S rRNA), 0.4 μM of each primer, respectively, and 1 unit of thermostable polymerase (Southern Cross Biotechnology). The PCR profile consisted of an initial denaturation step (5 min at 94 °C), followed by 35–40 cycles of denaturation (30 s at 94 °C), annealing for 1 min (at 60 °C for RP1 and Aldolase and 50 °C for 16S rRNA), and extension (1 min at 72 °C), and a final extension step (10 min at 72 °C). PCR products were then purified using a QIAquick PCR purification kit (Qiagen), cycle-sequenced using BigDye sequencing kit (Applied Biosystems), and analysed on a 3100 AB automated sequencer.

2.2. Alignment and characterizations of gene fragments

When sequences were characterized by large length variation (which was the case in RP1 and 16S sequences), POY software (Gladstein and Wheeler, 1997) was used to establish character homologies. In each case, one of the implied alignments (computed a posteriori for each of the equally parsimonious trees inferred) was used as an input matrix for further analyses. Length differences in the Aldolase sequences were rare and in this case, ClustalX (Thompson et al., 1997) using the default parameters was used to align sequences. 16S rRNA sequences of three additional seahorse species and one pipefish, as well as all of the cytochrome *b* sequences used in this study, were downloaded from GenBank (16S rRNA Accession Nos. AF355013 [*Hippocampus abdominalis*], AF354999 [*Hippocampus barbouri*], AF355007 [*Hippocampus erectus*], AF354991 [*Syngnathus acus*], Wilson et al., 2001; cytochrome *b* Accession Nos. AF192679–AF192686, Casey, 1999; AF356040 [*S. acus*], Wilson et al., 2001). Despite the fact that these sequences originated from different specimens, there was little reason to assume inconsistencies with regard to identification. Most of the samples used in this study and in the study by Casey (1999) had been identified by Lourie (pers. commun.), or were identified based on the criteria in Lourie et al. (1999).

To ensure authenticity, several representatives from each species were included whenever possible, and these were analysed at different times. However, the DNA of many of the samples used in this study was of poor quality and did not amplify readily during PCR. For this reason, some specimens included in the RP1 phylogeny are not represented in the Aldolase phylogeny, and vice versa. Mitochondrial 16S rRNA was sequenced only in species for which both Aldolase and RP1 sequences were available. Alignment gaps in Aldolase, RP1, and 16S rRNA sequences were treated as missing characters. The phylogenetic information of indels that consisted of two or more consecutive base-pairs, had clearly defined alignment borders, and were present in at least two different species was incorporated into phylogenetic analyses by coding them as fifth characters. Properties of the four gene fragments used in this study and combinations thereof were compared by determining nucleotide frequencies, maximum uncorrected *p*-distances among specimens, the relative proportion of informative sites and empirical transition:transversion ratios using the ML search option in PAUP* version 4.0 beta 10 (Swofford, 2002). For consistency, the data base used consisted of selected ingroup taxa for which sequences were available for all four gene fragments.

2.3. Phylogenetic analyses

As phylogenetic signal emerges better through the interaction of all data (Baker and DeSalle, 1997; Buckley et al., 2002; Cognato and Vogler, 2001; Murrell et al., 2001), topologies recovered from combined data are generally better resolved than those based on individual partitions (Baker and DeSalle, 1997; Buckley et al., 2002; Gatesy et al., 2002; Matthee et al., 2001; Murphy et al., 2001). In order to maximize the descriptive and explanatory power of the evidence, the four partitions used in this study were thus initially combined into a supermatrix. As mentioned above, many samples were of poor quality and did not amplify, resulting in the introduction of missing data. Wiens (1998), however, suggested that unless the proportion of missing data is very large, addition of incomplete data sets is more likely to improve phylogenetic accuracy than reduce it. In order to determine the effect of missing data and incomplete taxon sampling on the phylogenies, we used several different approaches of combining the data. First, we compiled a combined data set that included a single representative from each species for which sequence data was available for at least one of the four data partitions (referred to as data base 'combined I'). Our second combined analysis included all the specimens sequenced for this study (Table 1) in one large combined data set (data base 'combined II'). Lastly, we analysed six data partitions to screen for topological congruence: (1) combined nuclear data; (2) combined mitochondrial data; (3–6) each of the four data partitions.

Phylogenetic relationships among taxa were investigated using maximum parsimony (MP) and Bayesian inference (BI). Maximum parsimony tree searches were performed in PAUP* using the heuristic search option with 10 random-addition sequences and tree-bisection–reconnection branch swapping. The reliability of nodes was assessed using 1000 nonparametric bootstrap replicates (Felsenstein, 1985) and in each case the heuristic search was limited to a maximum of 10,000 saved trees. In order to verify that no shorter trees are likely to be found using an alternative method, trees were also searched using the 'parsimony ratchet' in NONA v. 2.0 (Goloboff, 1998) in combination with WINCLADA (Nixon, 1999–2002). Strict consensus topologies of MP trees of individual partitions and combinations thereof were compared to each other in a pairwise fashion using the Shimodaira–Hasegawa test (SH test, Shimodaira and Hasegawa, 1999) as implemented in PAUP*. For consistency, the data base used included 20 ingroup species for which sequence data were available from all four partitions, as well as pipefish sequences.

The program MODELTEST version 3.06 (Posada and Crandall, 1998) was used to conduct hierarchical likelihood ratio tests to determine the most appropriate substitution model for each data partition based on the Akaike Information Criterion (AIC; Akaike, 1973). Bayesian analyses were performed using MRBAYES v3.0B (Huelsenbeck and Ronquist, 2001). The MCMC process was set for four chains to run simultaneously for 1,000,000 generations, with trees being sampled every 50 generations for a total of 20,000 trees in the initial sample. Maximum likelihood parameters in MRBAYES were specified for each partition according to the most appropriate evolutionary model identified by MODELTEST. To ensure that analyses were not trapped on local optima, each independent run was repeated up to five times, and the posterior probabilities for individual clades from separate analyses were compared for congruence (sensu Huelsenbeck and Imenov, 2002). Graphic examination of variation in maximum likelihood scores identified that in all cases, 'burnin' was complete prior to the 20,000th generation. Hence, the first 400 trees were discarded, and the posterior probability of the phylogeny and its branches was determined from the remaining 19,600 trees. Consensus trees were constructed using the sumpt option in MRBAYES.

All trees were rooted using outgroups and tree topologies obtained with MP and BI were compared using the SH test. In order to determine the effect of using a different outgroup on tree topologies, several modified data sets were analysed using MP only. Topologies derived from data sets utilizing different outgroups were also compared using SH tests. To obtain comparable data sets, taxa not represented in both data bases were removed from trees.

3. Results

3.1. Characterization of the four gene fragments

Primers for Aldolase amplified a single fragment 267–268 bp in length and this region included 127 exon and 140–141 intron characters. The RP1 primers amplified intron sequence that varied from 658 bp in *Hippocampus kelloggi* to 678 bp in *Hippocampus camelopardalis*. The fragment also included 30 bp of flanking sequence from exon 1 and 65 bp from exon 2. In some cases, two peaks equal in intensity were present at a single position on the chromatograph, indicating heterozygosity. The presence of duplicate copies is unlikely, as no intra-individual variation was found in the exon portions of these sequences, and no multiple bands were amplified. All the heterogeneous sites were coded with IUB ambiguity codes.

Total length of the aligned RP1 intron sequences was 706 bp. Sequencing of the complete RP1 intron was problematic in the three species *Hippocampus comes*, *Hippocampus subelongatus*, and *Hippocampus sindonis* due to the presence of several AT-rich arrays. This problem was partly resolved by designing internal primers but unfortunately the final data sets lack approximately 62 (*H. comes*), 184 (*H. subelongatus*), and 176 (*H. sindonis*) nucleotides, respectively. The length of mitochondrial 16S rRNA fragments ranged from 520 bp (*Hippocampus bargibanti*) to 527 bp (*H. comes*).

A total of 38 indels was found among aligned ingroup sequences in RP1, the largest one being 22 bp in length. However, as most of these indels overlapped, were present in only a single species, or were a single base-pair in length, only five indels were coded as characters. Among the Aldolase sequences, a single indel one base-pair in length each was found in *Hippocampus capensis* and *Hippocampus breviceps*, whereas the pipefish sequence contained three indels. The aligned 16S rRNA sequences contained 15 indels. As all of the indels in Aldolase and 16S rRNA sequences either overlapped or were only a single base-pair in length, they were not coded as characters. We pruned the edges of all sequences by removing characters close to the primers to reduce ambiguity. The remaining lengths of each of the aligned gene fragments sequenced were 640 characters (RP1, exon portions were removed entirely as they were invariable), 188 characters (Aldolase), and 464 characters (16S rRNA). Cytochrome *b* sequences downloaded from GenBank were reduced to a total length of 1020 nucleotides. All sequences generated in this study have been deposited in GenBank (Accession Nos. AY277286–AY277374).

Aldolase is the most conserved gene fragment, and the cytochrome *b* gene evolves most rapidly (Table 2). The number of transitions and transversions was approximately equal in the two nuclear gene fragments,

Table 2
Comparisons of partitions and combinations thereof using 20 ingroup species for which sequences are available for all four molecular markers

Gene fragments	Type	Total sites	Informative sites	Nucleotide frequencies				t_i/t_v	Maximum
				%A	%C	%G	%T		*p*-Distance
Aldolase	nDNA	188	11 (6%)	23.8	23.6	22.1	30.5	1.08	0.04
RP1	nDNA	645	58 (9%)	24.5	18.3	24.3	32.9	0.92	0.09
16S rRNA	mtDNA	464	62 (13%)	29.5	24.4	19.9	26.2	2.55	0.09
Cytochrome *b*	mtDNA	1020	308 (30%)	23.7	27.4	15.5	33.5	6.70	0.19
nDNA combined	nDNA	833	69 (8%)	24.4	21.8	23.2	30.6	0.94	0.09
mtDNA combined	mtDNA	1484	362 (24%)	30.6	25.5	13.6	30.4	5.60	0.16
Combined	Both	2317	431 (19%)	27.5	24.2	16.6	31.7	3.79	0.15

Table 3
Uncorrected *p*-distances among the two outgroup taxa *Syngnathus temminckii* and *Hippocampus bargibanti*, among outgroup and ingroup taxa, and maximum values found among ingroup taxa

Gene fragment	Type	*Syngnathus* vs. *H. bargibanti*		*Syngnathus* vs. ingroup		*H. bargibanti* vs. ingroup		Max. value within ingroup	
		A	B	A	B	A	B	A	B
Aldolase	nDNA	–	–	0.37–0.39	0.37–0.39	–	–	0.06	0.05
RP1	nDNA	0.37	0.38	0.31–0.35	0.31–0.35	0.14–0.20	0.17–0.21	0.10	0.09
16S rRNA	mtDNA	0.15	0.14	0.13–0.16	0.12–0.16	0.11–0.14	0.12–0.14	0.09	0.09
Cytochrome *b*	mtDNA	–	–	0.21–0.24	–	–	–	0.19	–

Note that sequences of *H. bargibanti* were not available for Aldolase and cytochrome *b*. To determine the impact of indels on genetic distances, Aldolase, RP1, and 16S sequences were analysed both with and without sections containing indels (A and B, respectively).

whereas mitochondrial gene fragments were characterized by high t_i/t_v ratios.

Uncorrected sequence divergence values among ingroup and outgroup taxa were considerably larger in the case of nuclear markers than in the case of mitochondrial markers (Table 3). For example, divergence values between the pipefish and the most distant ingroup species ranged from 0.16 in 16S rRNA (1.3 times the maximum value found in the ingroup) to 0.39 in Aldolase (6.5 times the maximum value found in the ingroup). This indicates that most sites in the nuclear gene fragments are free to vary, whereas real distances for the mtDNA data are underestimated. As most indels were coded as missing data, their effect on *p*-distances was negligible.

3.2. Phylogenetic reconstructions

Initially, phylogenetic trees constructed in PAUP* using maximum parsimony were rooted using a pipefish (genus *Syngnathus*) as an outgroup taxon. However, due to the considerable sequence divergences between most seahorses and pipefishes (for example up to 35% in RP1 and up to 39% in Aldolase, Table 3) it was considered necessary to explore the utility of a basal ingroup species as an additional outgroup taxon. Evidence from the RP1, 16S rRNA and combined sequences (using both outgroup rooting with a pipefish sequence and midpoint rooting) indicated a strongly supported basal position of the pygmy seahorses (*H. bargibanti*) within the phylogeny. Hence, the pygmy seahorse was used as an additional outgroup taxon whenever sequence data for this species were available.

Maximum parsimony analyses of the four separate data sets using the heuristic search function in PAUP* yielded from 3 (Aldolase and cytochrome *b*) to 706 (RP1) equally parsimonious trees (Table 4). In the case of combined analyses, the highest number of trees was found with the 'combined II' data base. The shortest trees found using the parsimony ratchet in NONA were of equal length in all cases. SH tests revealed that most of the MP tree topologies constructed for each partition and combinations thereof using consensus data bases that included 20 ingroup species and pipefish sequences were not significantly different from each other. An exception was the topology of the Aldolase tree, which differed significantly (at $\alpha = 0.05$) from the RP1, 16S, cytochrome *b*, combined mtDNA and combined nDNA tree topologies. Despite differences in topology, resolution and the amount of homoplasy, no well-supported

Table 4
Results of ILD tests comparing phylogenetic information from the four genetic markers

	Aldolase	RP1	16S rRNA	Cytochrome *b*
Aldolase		6	8	23
RP1	0.04		17	20
16S rRNA	0.57	0.01		21
Cytochrome *b*	0.99	0.70	0.82	

Above diagonal: number of steps; below diagonal: *P* values.

nodes were found that were in conflict among different partitions. Topologies based on BI were not significantly different from their MP counterparts in three of the individual partitions (Aldolase, RP1, and 16S), but they differed significantly in the case of cytochrome *b*, as well as in all analyses based on combined data.

The highest number of nodes was recovered using the data set including all four gene fragments (combined I). At least four monophyletic clades were retrieved irrespective whether cytochrome *b* was included or not, or whether several specimens were included from each species (Figs. 1a and b). Apart from clade 1, which was always placed basal in the phylogeny, the exact branching patterns among clades differ when taxa/data were included and excluded. In order to determine whether the utility of more closely related outgroup species resulted in significantly different phylogenies, the seahorses of clade 1 were used as an alternative outgroup. Pairwise SH tests on MP trees constructed using different outgroups (i.e., *Syngnathus*/*H. bargibanti* vs. *H. breviceps*/*H. abdominalis*) revealed a significant topological difference in the case of combined mitochondrial data only; in contrast, trees constructed exclusively with nuclear data (both individual partitions and combined data) had identical topologies (i.e., $P = 1.0$, Table 5).

Because the inclusion of additional taxa and cytochrome *b* data did not alter the results of this study, we based our conclusions on a consensus phylogeny (Fig. 2), which is based on congruence among the results obtained from all the partitioned and combined analyses and also congruence among the different phylogenetic methods (Table 6). In this consensus approach we argue that nodes A, B, D, F, G, H, and L were generally well supported by at least one of the markers/combined data sets and overall nodal support was weakest for nodes C, E, and I (Table 7).

As mentioned above, *H. bargibanti* was always placed basal in the phylogeny, and the south-west Pacific seahorses *H. breviceps* and *H. abdominalis* comprised the next basal clade (supported by node B). The data also suggest that the remaining species can be placed into three monophyletic lineages. The first of these (clade 2) mainly comprises species from the south-western Pacific, an exception being the more widespread *Hippocampus histrix* (Pacific basin to East Africa, Lourie et al., 1999). Note that all Australian seahorses have been considered Pacific Ocean species. Faunas even in Western Australia are often genetically more closely associated with those of the Pacific rather than the Indian Ocean (Berquist and Kelly-Borges, 1995; McMillan and Palumbi, 1995;

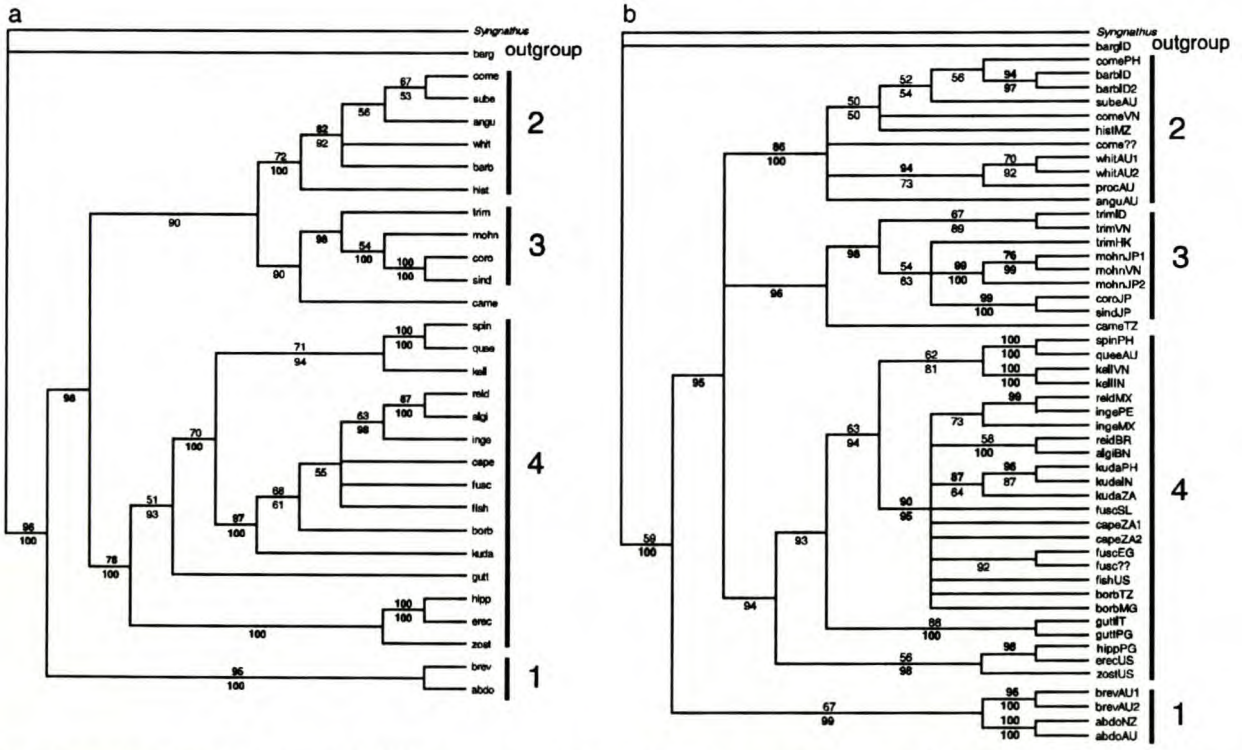

Fig. 1. Cladograms of combined analyses: (a) the supermatrix used is a combination of sequences from up to four partitions (Aldolase, RP1, 16S rRNA, and cytochrome *b*) and contains a single representative from each species; (b) the supermatrix used is a combination of sequences from all the specimens sequenced in this study and comprises a maximum of three partitions (Aldolase, RP1, and 16S rRNA). Nodal support is indicated by bootstrap values (above branches) and posterior probabilities (below branches). High nodal support (bootstrap values >75% and posterior probabilities >95%) is shown in boldface. The numbers 1–4 have been assigned to major clades.

Table 5
Tree statistics from parsimony analyses of individual and combined data partitions

	Number of				Tree length	RI
	Ingroup specimens	Characters	Informative characters	Equally parsimonious trees		
Aldolase[a]	34	188	21	3	97	0.95
RP1[b]	40	645	148	706	501	0.87
16S rRNA[2]	25	464	79	8	293	0.67
Cytochrome *b*[a]	23	1020	338	3	1471	0.54
Combined nDNA[b]	21	833	131	1	597	0.75
Combined mtDNA[b]	28	1484	415	3	1864	0.46
Combined I[b]	28	2317	536	218	2460	0.52
Combined II[b*]	49	1297	242	2588	883	0.80

[a] *Syngnathus* used as outgroup species.
[b] *Syngnathus* and *Hippocampus bargibanti* used as outgroup species.

Table 6
Results of Shimodaira–Hasegawa tests for pair-wise comparisons of consensus tree topologies constructed with data bases from four genes and combinations thereof using maximum parsimony (MP) and Bayesian inference (BI)

Data base	Method	$-\ln L$	$\Delta - \ln L$	*P*
Aldolase	MP	674	16	0.68
	BI	690		
RP1	MP	3371	15	0.06
	BI	3355		
16S rRNA	MP	2277	20	0.15
	BI	2298		
Cyto-chrome *b*	MP	8214	109	<0.01
	BI	8323		
nDNA combined	MP	3969	28	0.02
	BI	3941		
mtDNA combined	MP	11,200	111	<0.01
	BI	11,088		
Combined I	MP	14,327	181	<0.01
	BI	14,508		
Combined II	MP	6364	79	
	BI	6444		

Pipefish and pygmy seahorse (if available) sequences were included. Results were obtained using RELL optimisation and 10,000 boostrap replicates.

Williams and Benzie, 1998) because upwelling west of the Australian coast constitutes a biogeographic barrier (Fleminger, 1986; Wells et al., 1994). The second assemblage (clade 3) comprises species most of which are confined to the north-western Pacific. Only *Hippocampus trimaculatus* is more widespread (west Pacific to eastern Indian Ocean; Lourie et al., 1999). The third and last assemblage (clade 4) includes species that together are distributed circumglobally. The species comprising subclades 4a and b are found exclusively in the Indo-Pacific and those in subclades 4c and d are endemic to the Atlantic Ocean. The monophyletic lineage defined by node L (clades 4b and c) is interesting in that several of its species are the sole representatives of the genus *Hippocampus* in their home region, including *H. ingens* on the American west coast, *H. algiricus* in West Africa, and *H. capensis* in estuaries located on the south coast of South Africa. The phylogenetic placement of two species remains uncertain. First, the East African species *Hippocampus camelopardalis* was sometimes associated with the species of clade 3 (node E), but this association was well-supported in a single case only. Second, the Mediterranean species *Hippocampus guttulatus* was grouped with the species of subclades 4a–c (node I) in several analyses, but this association was never strongly supported. On one of the MP trees, this species was instead grouped with subclade 4d (combined mtDNA, 70% bootstrap support).

The monophyly of most species represented by more than one sample could not be challenged: although some specimens were closely associated with seahorses other than their conspecifics (Fig. 1b), such clades tended to be weakly supported (e.g., the three specimens of *H. comes* among clade 2 seahorses and the three specimens of *H. trimaculatus* among clade 3 seahorses). An exception was the West Atlantic species *Hippocampus reidi*: the specimen from the Gulf of Mexico was closely associated with *H. ingens* from the East Pacific, whereas the Brazilian specimen had a sister taxon relationship with the West African *H. algiricus*.

4. Discussion

4.1. Comparison of nuclear and mitochondrial markers

Even though nuclear genes were difficult to sequence in individual cases, they generally had a greater potential to recover the deeper nodes in the phylogeny. This is particularly true for the larger RP1 intron. The considerable amount of missing characters in some of the RP1

Table 7
Summary of branch support for nodes A–N in Fig. 2 using four molecular markers and combinations thereof

Data set		Nodes													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
Aldolase	MP	–	–	–	**76**	–	–	–	–	–	–	69	**87**	61	54
	BI	–	–	–	**100**	–	–	–	–	–	–	**97**	**98**	**100**	–
	MP[a]	?	–	–	**85**	–	–	–	–	–	–	67	**86**	–	56
RP1	MP	**100**	**91**	–	**80**	72	**86**	**89**	**85**	–	–	–	**77**	–	?
	BI	89	89	–	**100**	54	**100**	**100**	89	–	–	–	89	–	?
	MP[a]	?	**100**	–	**83**	65	**81**	**88**	**88**	–	–	–	**81**	–	?
16S rRNA	MP	–	–	–	–	–	–	–	–	–	55	–	–	**76**	?
	BI	–	–	–	–	–	–	–	–	–	–	–	–	**94**	?
	MP[a]	?	**82**	–	–	–	–	–	–	56	**75**	–	**90**	**81**	?
Cytochrome *b*	MP	–	–	51	**77**	–	–	–	**88**	–	55	63	**100**	**92**	?
	BI	–	–	–	–	–	–	–	**100**	–	–	–	**100**	–	?
	MP[b]	?	**100**	–	**72**	–	–	51	**97**	–	–	58	**100**	–	?
Combined nDNA	MP	**98**	**85**	–	**92**	69	**82**	**87**	68	–	–	54	**89**	–	67
	BI	**95**	93	–	**100**	71	**97**	**99**	**95**	66	–	76	**100**	–	**95**
	MP[1]	?	**97**	–	**95**	57	**80**	86	62	–	–	–	**92**	54	62
Combined mtDNA	MP	**85**	–	–	56	–	–	–	**80**	–	72	58	**100**	–	?
	BI	62	–	–	55	–	87	–	77	–	**97**	64	**99**	–	?
	MP[c]	?	**100**	–	**100**	–	–	55	**98**	–	71	59	**100**	–	?
Combined I	MP	**96**	–	–	72	–	–	54	**78**	51	70	71	**97**	63	–
	BI	**100**	**98**	**90**	**100**	90	**98**	**100**	**100**	93	**100**	94	**100**	**98**	**100**
	MP[b]	?	**100**	–	**90**	–	–	68	**97**	–	64	69	**96**	61	–
Combined II	MP	59	–	–	**86**	–	–	–	–	–	63	62	**90**	–	56
	BI	**100**	**95**	–	**100**	**96**	**98**	–	94	93	94	81	**95**	–	**98**
	MP[2]	?	**82**	–	**92**	–	–	–	–	52	71	65	**92**	–	62

First number: Bootstrap support from parsimony analysis using a pipefish (genus *Syngnathus*) and, if available, a pygmy seahorse (*Hippocampus bargibanti*) as outgroup species. Second number: posterior probabilities from Bayesian inference using the same data sets and outgroup. Third number: Bootstrap support from parsimony analysis using *H. abdominalis* and *H. breviceps* as outgroup species. The result of SH tests comparing the topologies of MP trees constructed using this alternative outgroup with topologies using the *Syngnathus*/*H. bargibanti* outgroup are indicated by superscript numbers (see legend below table). Boostrap and posterior probability values below 50% are indicated as "–". Strongly supported nodes (bootstrap values $\geqslant 75\%$ and posterior probabilities $\geqslant 95\%$) are shown in boldface. Roman numerals for combined sequences refer to: (I) supermatrix contained a single individual from each species and a maximum of four molecular markers each; (II) supermatrix contained all specimens sequenced for this study and a maximum of three molecular marker each (cytochrome *b* excluded). If a clade could not be recovered because a particular sample was not represented in the data matrix (e.g., *H. zosterae* in clade N), this was indicated with a question mark.

[a] The topology of the strict MP tree constructed with the *H. abdominalis*/*H. breviceps* outgroup was identical to the topology of the corresponding MP tree constructed with the *Syngnathus*/*H. bargibanti* outgroup ($P = 1.0$).

[b] The topology was different, but not significantly so.

[c] The topology was significantly different ($P < 0.05$).

sequences seems to have had little effect on phylogenetic placement of the affected species. The well-supported sister taxon relationship of *H. sindonis* with *Hippocampus coronatus*, for example, was confirmed by the Aldolase and 16S phylogenies. In addition, the sister taxon relationship of *H. comes* and *H. subelongatus*, which both contained a high proportion of missing characters in their RP1 sequences, was confirmed by both the cytochrome *b* and the 16S phylogenies.

A particularly striking difference between nuclear and mitochondrial sequences was the genetic distance between ingroup and outgroup species. The comparatively small genetic distances among mitochondrial sequences of ingroup and outgroup species indicates that the occurrence of homoplasies becomes considerably greater when comparing more distantly related species, which was confirmed by the fact that saturation plots constructed using both data sets indicated saturation (not shown). A saturation plateau of about 20% pairwise differentiation has been suggested for ribosomal DNA by Ortí and Meyer (1997). In case of the nuclear fragments used in this study, this level of differentiation is surpassed when comparing sequences of 'true' seahorses, pygmy seahorses, and pipefishes. However, the low number of homoplasies and lack of saturation among distantly related taxa (saturation plots not shown) suggests that the pipefish nevertheless performed satisfactorily as an outgroup species. This was confirmed by the fact that MP

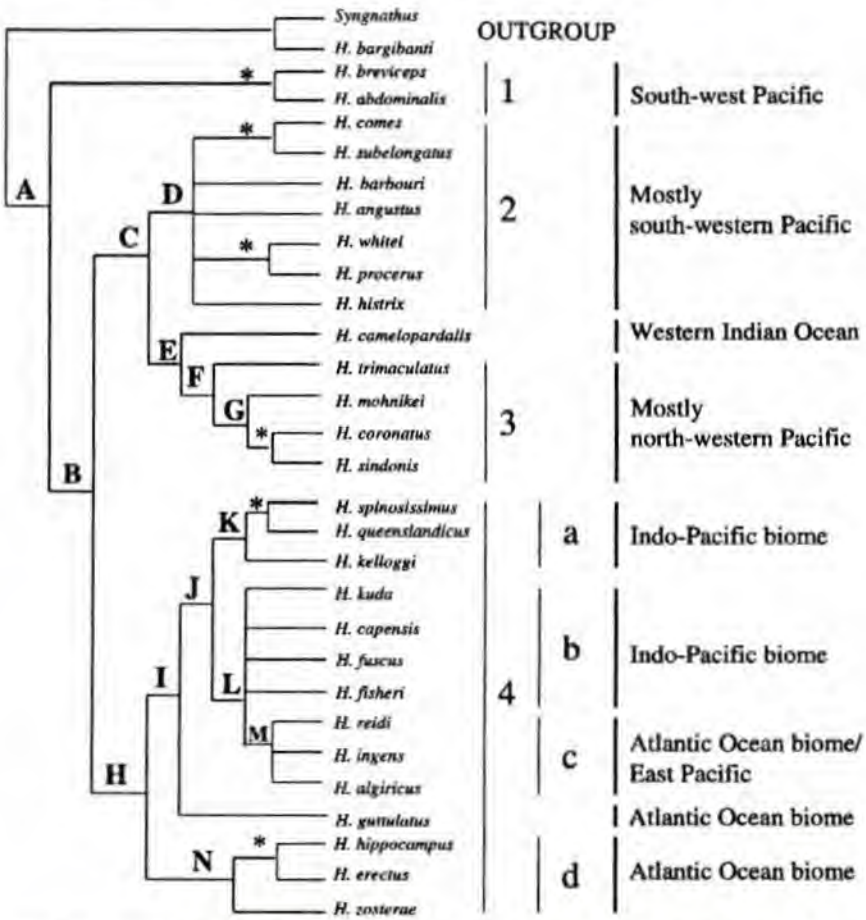

Fig. 2. Synthesis of phylogenetic information of MP and BI trees constructed using combined and individual partitions. Nodes of interest are labelled with letters A–N and correspond to those given in Table 6. Additional terminal nodes that were recovered in all analyses are marked with asterisks. The monophyly of clades 1–4 is supported by high bootstrap values and/or posterior probability values of the nodes defining them (Table 6). Grouping of seahorses in clade 4 into four subclades was based primarily on their associations with specific geographic regions.

analyses of nuclear data sets using clade 1 as alternative outgroup resulted in trees with identical topologies, and bootstrap support for individual nodes on these trees differed only slightly. In contrast, in the case of mitochondrial data, using clade 1 as outgroup resulted in different MP tree topologies (significant in one case) and increased bootstrap support for some nodes.

4.2. Indo-pacific origin of seahorses

The large genetic distance of the pygmy seahorse, *H. bargibanti*, to all other seahorses based on RP1 and 16S rRNA sequences suggests an ancient divergence of this group from the main clade of seahorses. *H. bargibanti* is widely distributed throughout the western Pacific, but the fact that this species is highly adapted to parasitise a certain species of *Muricella* gorgonian corals (Kuiter, 2000) suggests that it is unlikely to disperse readily beyond the region where this species occurs. Among the species associated with the main clade of seahorses, the most basal positions are occupied by *H. breviceps* and *H. abdominalis*. Both species are associated with the Australian continent, suggesting that this may be the region from which seahorses originated. An Australian or south-west Pacific origin of seahorses is also supported by the distributions of the three possible sister genera of the genus *Hippocampus*. Pygmy pipehorses of the genus *Amphelikturus* are restricted to the Atlantic biome, the genus *Acentronura* is widely distributed throughout the Indo-Pacific, and all known specimens of *Idiotropiscis* have been found in Australian waters. Among these three genera, the species of the genus *Idiotropiscis*, and particularly a recently discovered species from southern New South Wales, are most seahorse-like in appearance (Kuiter, 2000).

4.3. Biogeography and evolutionary history

Although some nodes of the phylogeny presented in this paper were not supported by high bootstrap and

posterior probabilities values, several novel insights can be proposed. It seems that subsequent to the origin of the genus *Hippocampus* in the Indo-Pacific biome, the main clade of seahorses split into three major lineages. Two of these remained in the Indo-Pacific (clades 2 and 3) and can be divided into a mostly south-western and a mostly north-western Pacific group. As node C was only weakly supported, it cannot be concluded whether these two clades are more closely associated with each other than either of them is with clade 4. There was some support for an association of the East African species *H. camelopardalis* with clade 3, but the results remain inconclusive. The basal placement of this species (node E) suggests that it is part of a lineage that became geographically separated from west-Pacific seahorses early during the evolutionary history of the genus.

In contrast to the first two assemblages, most of whose species remained in the West Pacific, the evolutionary history of the third major assemblage (clade 4) was characterized by dispersal events on a global scale. It is likely that this group became genetically distinct because it established itself in the Atlantic biome, a scenario that is supported by the fact that the basal position of subclade 4d within this group was well supported on most trees and the fact that the next derived species, *H. guttulatus*, is also associated with the Atlantic biome. The Caribbean has been a center of origin for many of the species found throughout the tropical Atlantic (Briggs, 1974), and it is possible that seahorses from this region then re-colonized the Indo-Pacific biome, giving rise to subclades 4a–c. Fig. 3a depicts this scenario. Alternatively, the group defined by node J (subclades 4a–c) remained in the Indo-Pacific and was the source of a maximum of three colonization events of the Atlantic biome (Fig. 3b). Additional scenarios set between these two extremes are equally plausible, and cannot be resolved in the absence of a more robust phylogeny and/or fossil data.

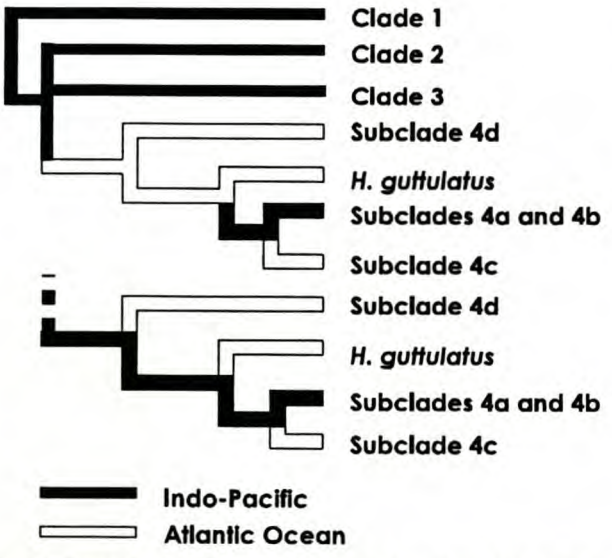

Fig. 3. Two alternative hypotheses regarding the history of colonization of the Atlantic biome by members of the genus *Hippocampus*. In both cases, it is assumed that *H. guttulatus* is more closely associated with seahorses in subclades 4a and 4b than it is with those of subclade 4d.

Due to the highly fragmented fossil record of seahorses, it is as yet difficult to confidently date divergence events within the genus *Hippocampus*. Much more fossil data are available for pipefishes than for seahorses (Fritzsche, 1980), and in future studies, it may be appropriate to calibrate a molecular clock for a combination of sequences from different Syngnathid genera. However, two well-documented vicariance events, namely the closures of the Tethyan and Central American seaways, can nevertheless be used to put the divergences among some of the ingroup clades into a temporal perspective. The divergence of subclade 4d (exclusively present in the Atlantic biome) and its nearest Indo-Pacific sister taxa is likely to have coincided with the closure of the Tethyan seaway which once connected the Atlantic biome with the Indian Ocean. Estimates for a final closure of the connection range from approximately 14 mya (Hsü and Bernoulli, 1978; Vrielynk et al., 1997) to 6.7 mya (Sonnenfeld, 1985). To our knowledge, the oldest fossilized seahorses have been found in Italy (Sorbini, 1988) and were identified as *Hippocampus ramulosus* (a synonym for *H. guttulatus*). The deposits containing these specimens have been dated as being from the upper Miocene, which confirms the presence of these seahorses in the western Tethys (today's Mediterranean) close to the time when the Tethyan seaway closed. We propose that the species of subclade 4d (*Hippocampus hippocampus*, *H. erectus*, and *Hippocampus zosterae*) as well as *H. guttulatus* represent descendents of a western Tethyan/Atlantic/Caribbean lineage, whereas subclades 4a (*H. kelloggi*, *Hippocampus spinosissimus*, and *Hippocampus queenslandicus*), 4b (*Hippocampus kuda*, *Hippocampus borboniensis*, *Hippocampus fisheri*, *Hippocampus fuscus*, and *H. capensis*), and 4c (*H. algiricus*, *H. reidi*, and *H. ingens*) are descendants of an eastern Tethyan/Indo-Pacific lineage.

The complications associated with the phylogenetic placement of *H. guttulatus* make a reconstruction of the early history of seahorses in the Atlantic Ocean difficult, but two hypotheses are likely, depending on whether this species is more closely associated with subclades 4a–c or with subclade 4d. In the first scenario, the split between subclade 4d and the remainder of the group preceded the closure of the Tethyan seaway. East Atlantic seahorses may already have been distinct from West Atlantic seahorses, and their distribution may have extended into the Indian Ocean or beyond. This differentiation into two major groups is most likely to be the result of isolation by distance (due to the expanse of the

Atlantic Ocean), which resulted in speciation. If, alternatively, *H. guttulatus* is more closely associated with subclade 4d than it is with subclades 4a–c, differentiation of Atlantic seahorses into a European and an American lineage is equally likely to have taken place after the closure of the Tethyan seaway. Whichever scenario is correct, in both cases, the split between the lineage including *H. guttulatus* and subclades 4a–c provides a calibration point for this important vicariance event to be used in future studies. The other European species, *H. hippocampus*, is presently distributed throughout the range occupied by *H. guttulatus* (Lourie et al., 1999), but its close association with the two American species in subclade 4d, and the small genetic distances among them, suggests that this species may have recently diverged from an American ancestor. Such a colonization event may have taken place due to an intensification of the Gulf Stream's current velocity as a result of the gradual rising of the Isthmus of Panama, which culminated 3.8 million years ago (Kaneps, 1979; Keller and Barron, 1983).

Note that although subclade 4c includes species present in the Atlantic Ocean, their derived position on all phylogenetic trees suggests that they colonized this region more recently. This second invasion of the Atlantic Ocean may have occurred either in a westward direction via southern Africa, or in an eastward direction via the Central American Seaway. Our results are inconclusive in this regard, and additional data are required to establish whether the three species in this clade are more closely associated with Pacific species or with Indian Ocean species. Divergence of the Indo-Pacific and Atlantic species defined by node L (subclades 4b and 4c, respectively) occurred prior to the closure of the Central American Seaway, because the close association between the east Pacific *H. ingens* and the west Atlantic *H. reidi* suggests that these two species diverged from a common ancestor due to the rising of the Panamanian Isthmus between 4.6 mya (reorganization of ocean circulation; Haug and Tiedemann, 1998) and 3.1–3.5 mya (cessation of gene flow; Coates and Obando, 1996; Collins, 1996; Duque-Caro, 1990a,b; Keigwin, 1982).

The distribution of species (or species complexes) on both sides of the tropical Atlantic Ocean has been attributed to either vicariance (spreading of the Atlantic basin 65–20 mya and rise of the Panamanian isthmus; Rosen, 1975) or recent dispersal events with the Caribbean as a center of origin (Briggs, 1974). Both of the Atlantic seahorse clades have representatives on both sides of the Atlantic Ocean: in subclade 4c, *H. reidi* and *H. ingens* occur in the west (West Atlantic and East Pacific, respectively), and *H. algiricus* in the east (West Africa); in subclade 4d, *H. erectus* and *H. zosterae* are West Atlantic species and *H. hippocampus* is an East Atlantic species. In both cases, genetic distances between members of these pairs are minimal. This is particularly striking in case of the Aldolase sequences, which are identical for all three members of each pair. The fact that East Atlantic and West Atlantic/East Pacific lineages are likely to have diverged after the closure of the Tethyan seaway suggests that the dispersal hypothesis seems more appropriate to explain present-day distribution patterns of the geminate seahorse species.

Acknowledgments

We thank the following individuals and organizations for providing tissue samples: Sara Lourie, Janelle Curtis, Adam Jones, Andre Bok, Howaida Gabr, Mark Erdmann, Ambo Tuwo, Zacharie Sohou, Leo Smith, Paul Southgate, Cathi Lehn, Takahiko Mukai, Melchor Santos, Project Seahorse, the California Academy of Sciences, the American Museum of Natural History, the Australian Museum, the Smithsonian Institution, and Sea World Durban. Sara Lourie, Wei-Jen Chen, Michael Cunningham, Rudie Kuiter, and Steven Cumbaa are thanked for providing valuable advice. The manuscript was greatly improved by the comments of Giacomo Bernardi. This study was supported by the National Research Foundation and by the University of Stellenbosch.

References

Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281.

Arnaud, S., Bonhomme, F., Borsa, P., 1999. Mitochondrial DNA analysis of the genetic relationships among populations of scad mackerel (*Decapterus macarellus*, *D. macrosoma*, and *D. russelli*) in South-East Asia. Mar. Biol. 135, 699–707.

Baker, R.H., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian *Drosophila*. Syst. Biol. 46, 654–673.

Berquist, P.R., Kelly-Borges, M., 1995. Systematics and biogeography of the genus *Ianthella* (Desmospongidae; Verongida; Ianthellidae) in the South Pacific. Beagle, Rec. Northern Terr. Mus. Arts Sci. 12, 151–176.

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S., Robertson, D.R., 2001. Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. Evolution 55, 1029–1039.

Briggs, J.C., 1974. Marine Zoogeography. McGraw-Hill, New York.

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

Briggs, J.C., 1999. Coincident biogeographic patterns: Indo-West Pacific Ocean. Evolution 53, 326–335.

Buckley, T.R., Arensburger, P., Simon, C., Chambers, G.K., 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. Syst. Biol. 51, 4–18.

Burridge, C.P., White, R.W.G., 2000. Molecular phylogeny of the antitropical subgenus *Goniistius* (Perciformes: Cheilodactylidae: *Cheilodactylus*): evidence for multiple transequatorial divergences and non-monophyly. Biol. J. Linn. Soc. 70, 435–458.

Casey, S.P., 1999. A phylogenetic study of seahorses using the cytochrome *b* gene of mitochondrial DNA. Unpublished Ph.D.

thesis, Institute of Zoology, Regents Park, London, NW1 4RY, United Kingdom.

Chow, S., Hazama, K., 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. Mol. Ecol. 7, 1255–1256.

Coates, A.G., Obando, J.A., 1996. The geologic evolution of the Central American isthmus. In: Jackson, J.B.C., Budd, A.F., Coates, A.G. (Eds.), Evolution and Environment in Tropical America. University of Chicago Press, Chicago, pp. 21–56.

Cognato, A.I., Vogler, A.P., 2001. Exploring data interaction and nucleotide alignment in a multiple gene analysis of Ips (Coleoptera: Scolytinae). Syst. Biol. 50, 758–780.

Colborn, J., Crabtree, R.E., Shaklee, J.B., Pfeiler, E., Bowen, B.W., 2001. The evolutionary enigma of bonefishes (*Albula* spp.): cryptic species and ancient separations in a globally distributed shorefish. Evolution 55, 807–820.

Collins, L.S., 1996. Environmental changes in Caribbean shallow waters relative to the closing tropical American seaway. In: Jackson, J.B.C., Budd, A.F., Coates, A.G. (Eds.), Evolution and Environment in Tropical America. University of Chicago Press, Chicago, pp. 130–167.

Duque-Caro, H., 1990a. The Choco Block in the northwestern corner of South America: structural, tectonostratigraphic, and paleographic implications. J. S. Am. Earth Sci. 3, 71–84.

Duque-Caro, H., 1990b. Neogene stratigraphy, palaeoceanography, and palaeobiology in northwestern South America and the evolution of the Panama seaway. Palaeogr. Palaeoclimatol. Palaeoecol. 777, 203–234.

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

Eschmeyer, W.N., 1998. In: Catalog of Fishes, vols. 1–3. California Academy of Sciences, San Francisco.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Fleminger, A., 1986. The Pleistocene equatorial barrier between the Indian and Pacific Oceans and a likely cause for Wallace's Line. UNESCO Technical Paper in Marine Science 49, pp. 84–97.

Fricke, R., 1988. Systematik und historische Zoogeographie der Callionymidae (Teleostei) des Indischen Ozeans, 2 volumes. Ph.D. dissertation, Albert-Ludwigs-Universität, Freiburg im Breisgau, Germany.

Fritzsche, R.A., 1980. Revision of the eastern Pacific Syngnathidae (Pisces: Syngnathiformes), including both recent and fossil forms. Proc. Cal. Acad. Sci. 42, 181–227.

Gatesy, J., Matthee, C.A., DeSalle, R., Hayashi, C., 2002. Resolution of a supertree/supermatrix paradox. Syst. Biol. 51, 652–664.

Gladstein, D.S., Wheeler, W.C., 1997. POY. The optimization of alignment characters. Program and Documentation. American Museum of Natural History, New York, NY. Available from ftp.amnh.org/pub/molecular/poy.

Goloboff, P., 1998. NONA ver. 2. Published by the author. Tucumán, Argentina.

Grant, W.S., Leslie, R.W., 2001. Inter-ocean dispersal is an important mechanism in the zoogeography of hake (Pisces: *Merluccius* spp.). J. Biogeogr. 28, 699–721.

Haug, H.H., Tiedemann, R., 1998. Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation. Nature 393, 673–676.

Horne, M.L., 2001. A new seahorse species (Syngnathidae: *Hippocampus*) from the Great Barrier Reef. Rec. Austr. Mus. 53, 243–246.

Hsü, K.J., Bernoulli, D., 1978. Genesis of the Tethys and the Mediterranean. Initial Report DSDP, pp. 943–949.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Kaneps, A.G., 1979. Gulf Stream: velocity fluctuations during the late Cenozoic. Science 204, 297–301.

Keigwin, L.D., 1982. Isotopic paleooceanography of the Caribbean and east Pacific: role of the Panama uplift in late Neogene time. Science 217, 350–352.

Keller, G., Barron, J.A., 1983. Paleooceanographic implications of Miocene deep-sea hiatures. Geol. Soc. Am. Bull. 97, 590–613.

Kuiter, R.H., 2000. Seahorses, Pipefishes and their Relatives—A Comprehensive Guide to Syngnathiformes. TMC Publishing, Chorleywood, UK.

Kuiter, R.H., 2001. Revision of the Australian seahorses of the genus *Hippocampus* (Syngnathiformes: Syngnathidae) with a description of nine new species. Rec. Austr. Mus. 53, 293–340.

Lessios, H.A., Kessing, B.D., Pearse, J.S., 2001. Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. Evolution 55, 955–975.

Lourie, S.A., Vincent, A.C.J., Hall, H. J., 1999. Seahorses: an identification guide to the world's species and their conservation. Project Seahorse, London, UK.

McMillan, W.O., Palumbi, S.R., 1995. Concordant evolutionary patterns among Indo-West Pacific butterflyfishes. Proc. R. Soc. Lond. B 260, 229–236.

Matthee, C.A., Burzlaff, J.D., Taylor, J.F., Davis, S.K., 2001. Mining the mammalian genome for Artiodactyl systematics. Syst. Biol. 50, 367–390.

Menon, A.G.K., 1977. A systematic monograph of the tongue soles of the genus *Cynoglossus* Hamilton-Buchanan (Pisces: Cynoglossidae). Smithson. Contrib. Zool. 283, 1–129.

Murrell, A., Campbell, N.J., Barker, S.C., 2001. A total-evidence phylogeny of ticks provides insights into the evolution of life cycles and biogeography. Mol. Phylogenet. Evol. 21, 244–258.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.

Muss, A., Robertson, D.R., Stepien, C.A., Wirtz, P., Bowen, B.W., 2001. Phylogeography of *Ophioblennius*: the role of ocean currents and geography in reef fish evolution. Evolution 55, 561–572.

Newman, W.A., Jumars, P.A., Ross, A., 1976. Diversity trends in coral-inhabiting barnacles (Cirripedia, Pyrgomatinae). Micronesica 12, 69–82.

Nixon, K.C., 1999–2002. WinClada ver. 1.0000. Published by the author, Ithaca, NY. Available from http://www.cladistics.com.

Ortí, G., Meyer, A., 1997. The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. Syst. Biol. 46, 75–100.

Palumbi, S.R., 1996. Nucleid acids II: The polymerase chain reaction. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MD, pp. 205–247.

Patterson, C., 1993. Chapter 36. Osteichthyes: Teleostei. In: Brenton, M.J. (Ed.), The Fossil Record. Chapman & Hall, London, England, pp. 1–145.

Posada, D., Crandall, K.A., 1998. MODELTEST: Testing the model of DNA substitution. Bioinformatics 14, 817–818.

Ricklefs, R.E., Latham, R.E., 1993. Global patterns of diversity in mangrove floras. In: Ricklefs, R.E., Schluter, D. (Eds.), Species Diversity in Ecological Communities. University of Chicago Press, Chicago, IL, pp. 215–229.

Rohde, K., Hayward, C.J., 2000. Oceanic barriers as indicated by scombrid fishes and their parasites. Int. J. Parasitol. 30, 579–583.

Rosen, B.R., 1975. A vicariance model of Caribbean biogeography. Syst. Zool. 24, 431–464.

Rosen, B.R., 1984. Reef coral biogeography and climate through the late Cainozoic: just islands in the sun or a critical pattern of islands. In: Brenchley, P.J. (Ed.), Fossils and Climate. Wiley, Chichester, pp. 201–262.

Rosen, B.R., 1988. Progress, problems and patterns in the biogeography of reef corals and other tropical marine organisms. Helgoländer Meeresunters 42, 269–301.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning—A Laboratory Manual, second ed. Cold Spring Harbor Laboratory Press, New York.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Specht, R.L., 1981. Biogeography of halophytic angiosperms (salt-marsh, mangrove and sea-grass). In: Keast, A. (Ed.), Ecological Biogeography of Australia. W. Junk, The Hague, The Netherlands, pp. 557–589.

Sonnenfeld, P., 1985. Models of upper Miocene evaporite genesis in the Mediterranean region. In: Stanley, D.J., Wezel, F.C. (Eds.), Geological Evolution of the Mediterranean Basin. Springer, Heidelberg, pp. 323–346.

Sorbini, L., 1988. Biogeography and climatology of Pliocene and Messinian fossil fish of Eastern Central Italy. Boll. Mus. Civ. Storia Nat. Verona 14, 1–85.

Stehli, F.G., Wells, J.W., 1971. Diversity and age patterns in hermatypic corals. Syst. Zool. 20, 115–126.

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 beta 10. Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J, Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Vrielynk, B., Odin, G.S., Dercourt, J., 1997. Reconstruction of Mediterranean Late Cenozoic hydrography by means of carbon isotope analyses. In: Meulenkamp, J.E. (Ed.), Reconstruction of Marine Paleoenvironments. Utrecht Mpal. Bull, Utrecht, pp. 25–47.

Wells, P.E., Wells, G.M., Cali, J., Chivas, A., 1994. Response of deep-sea benthic foraminifera to late Quaternary climate changes, southeast Indian Ocean, offshore Western Australia. Mar. Micropal. 24, 185–229.

Wiens, J.J., 1998. Combining data sets with different phylogenetic histories. Syst. Biol. 47, 568–581.

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

Wilson, A.B., Vincent, A., Ahnesjö, I., Meyer, A., 2001. Male pregnancy in seahorses and pipefishes (Family Syngnathidae): rapid diversification of paternal brood pouch morphology inferred from a molecular phylogeny. J. Hered. 92, 159–166.