The Molecular Systematics of Southern African Testudinidae

Richard Joseph Varhol

Thesis submitted in fulfilment of the requirements for the degree of
Master of Science (Medicine)
in the department of Chemical Pathology in the Faculty of Medicine
at the University of Cape Town

February 1998

DECLARATION

I, Richard Joseph Varhol, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise), and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other University.

I empower the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

_____

SIGNATURE

_____

DATE

Contents

Acknowledgements

I would like to extend my gratitude to the following people:

Professor **Eric Harley**, for his friendship, patience and guidance throughout this study and **Dr. Michael Cherry** for initiating this study, his support, advice, and for organising the collection of blood samples from Mlawula and Mkomazi Nature Reserves.

The Foundation for Research and Development for financial support

Lynne Doyle for her encouragement, support, patience and assistance.

Dr. Pim Rebholtz, for his friendship, insight, computer assistance and for those meaningful and in-depth discussions

Dr. Russell Burke (University of Michigan), for generously donating a copy of Charles Crumly's Ph.D. thesis.

Melissa Stander, **Dr. Tony Marinaki**, and **Dr. Colleen O'Ryan** for their extraordinary effort in proof reading rough drafts of this thesis.

In addition I would like to thank the following people who assisted me in finding and allowing me take blood samples of their tortoises: **Tamara Harris-Smith** (Reptile Curator of Tygerberg Zoo, Cape Town), **Dr. Ernst Baard**, **Atherton** and **Riki de Villiers** (Cape Nature Conservation, Jonkershoek), **Neil Van Heyningen** and his assistant **Soekie** (Kleinplasie Reptile World, Worcester), **Bill Branch** (Port Elizabeth Museum), **Dr. Clive McDowell** (Cape Town), and **John Visser** (Cape Town).

Abstract

Sixteen of the world's 42 species of land tortoises occur in Africa, 10 of which are endemic to southern Africa. South Africa itself, which occupies 0.8% of the earth's total land mass, has the highest tortoise biodiversity in the world, with 13 species. This is the first study to use molecular techniques to investigate the evolutionary history of this group, which displays an unusually high level of speciation on the continent.

Four hundred and fifty base pairs of mtDNA cytochrome *b* sequence were obtained, using direct PCR-based sequencing, from 32 individual tortoise blood samples, comprising 13 different species from 6 genera. PAUP 3.1.1, and MEGA were used to infer a phylogeny using Chrysemys scripta elegans (an Emydid) an outgroup.

Both phenetic and cladistic methods generated similar results. With the exception of *Malacochersus*, both morphological and molecular work show largely congruent results. When intra-specific relationships, using the molecular results, were compared to the existing morphological data, *Psammobates* was the only genus with a consistent topology.

Proposals for the re-evaluation of *Homopus, Kinixys* and *Geochelone* have been made. Suggestions, based on molecular results, include the distinction between *Chersobius* and *Homopus* (Hewitt 1937), incorporating *Malacochersus tornieri* into *Kinixys*, and the elevation of *Geochelone pardalis pardalis* and *G.p. babcocki* to species level. Sequencing a further nine individuals within *Homopus areolatus* showed a higher than expected sequence variation, suggesting a distinct population structure and possibly cryptic species.

List of Abbreviations

A	adenine
ATP	adenosine triphosphate
bp	base pair
C	cytosine
°C	degrees Celsius
cm	centimeter
D-loop	displacement loop
DNA	deoxyribonucleic acid
EDTA	ethylene diamine tetraacetic acid
e.g.	exempli gratia (for example)
et al.	et alia (and others)
Fig.	figure
G	guanine
g	gram
min	minutes
mL	mililiter
mt DNA	mitochondrial DNA
PCR	polymerase chain reaction
RNA	ribonucleic acid
s	second
T	thyamine
Taq	DNA polymerase derived from *Thermus aquaticus*
TEMED	N,N,N',N',-tetramethylethylenediamine
Ti	transition
TRIS	Tris-hydroxymethyl-aminomethane
Tv	transversion
µCi	microcurie
µg	microgram
µL	microliter
U	units
V	volts
W	watts
w/v	weight : volume ratio

1.0 Introduction

The order Testudines (Linnaeus 1758) which includes tortoises, terrapins and turtles has also been referred to as Chelonia and Testudinata (Branch 1988a, Carroll 1988, Romer 1966). However, according to Laurin and Reisz (1995) "The nomen Testudines is by far the oldest and should be used instead of the other nomina.". The order is very diverse consisting of 14 families, (Table 1.1) of which only one (Testudinidae) is completely terrestrial. The most speciose of the Testudines are the freshwater semi-terrestrial Emydidae.

Land tortoises are considered to have evolved from an ancestral emydid stock (Meylan and Auffenberg 1986) and by the end of the Cretaceous period (36-65 million years ago [MYA]), were fully established and formed a separate group from their semi-terrestrial predecessors (Auffenberg 1974). Members of the Testudinidae are currently found on all land masses (with the exception of Antarctica, Australia, New Zealand and Greenland), including islands such as Madagascar, Mauritius, and the Galapagos. Fossil data confirm that tortoises were once more diverse and had a wider distribution than those of the present day (Crumly 1984, Laurin and Reisz 1995). With their preference for warmer climates, tortoises may be reliable indicators of the prehistoric climates in which their ancestors were found (Holman 1971).

Table 1.1. Summary of Testudine diversity (adapted from Iverson 1992)

Family	Number of Genera	Number of species
Chelidae	10	40
Pelomedusidae	5	25
Carettochelyidae	1	1
Cheloniidae	5	7
Chelydridae	2	2
Dermatemydidae	1	1
Dermochelidae	1	1
Emydidae	33	94
Batagurinae	23	59
Emydinae	10	35
Kinosternidae	3	22
Playsternidae	1	1
Testudinidae	**11**	**40**
Trionychidae	14	23
Total	87	257

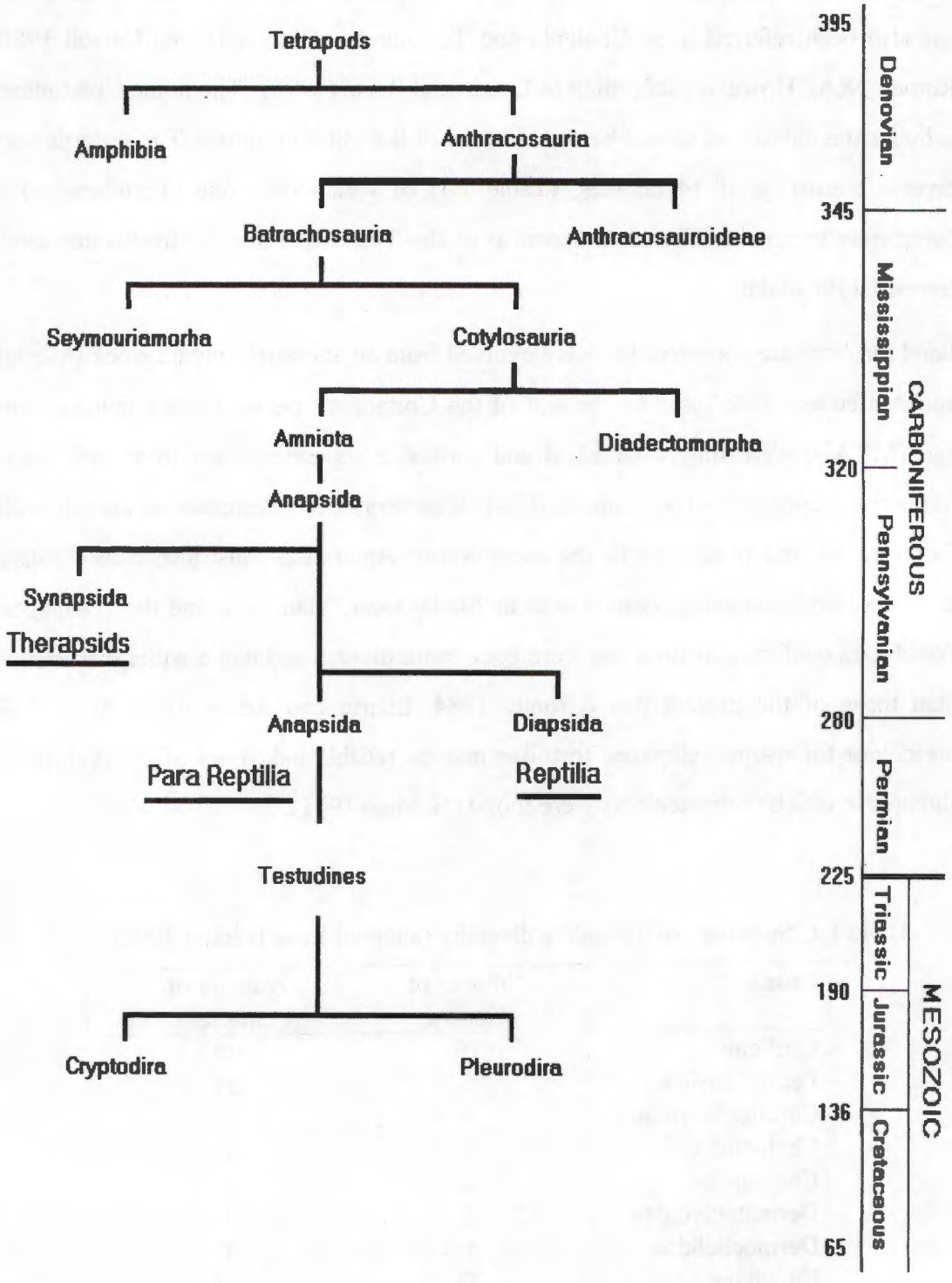


Figure 1.1. - Diagrammatic representation of the evolutionary events prior to the evolution of Testudines (Crumly 1984, Carroll 1988)

Apart from being geographically widespread, the family Testudinidae inhabits a wide variety of environments from the deserts of south central Africa (*Psammobates*) and western North America (*Xerobates* and *Gopherus*) to areas of lush vegetation in Asia (*Indotestudo* and *Manouria*), south eastern Africa (*Kinixys*) and South African coastal regions (*Chersina*) (Crumly 1984, Lehrer 1990).

Southern Africa has a wide range of ecosystems ranging from the arid environs of the Namib Desert to the Mediterranean regions of the Western Cape province and Mpumalanga. Tortoise adaptation to these varying habitats is related to their unique structural and possibly genetic diversity. Occupying 0.8% of the earth's total land mass, South Africa has the largest diversity of land tortoises in the world (Branch 1988a, Greig and Burdett 1976). It contains 13 of the 40 (33%) presently recognised tortoise species world wide, and 5 out of the 11 (45%) accepted genera (Iverson 1992, IUCN 1993).

The great diversity of land tortoises within southern Africa has become an intriguing evolutionary enigma. Construction of phylogenetic trees to gain insight to the sequence of evolutionary radiation of the Afro-Testudinidae has been based predominantly on qualitative methods using paleontological and morphological data (Auffenberg 1974, Baard 1990, Baard 1993, Boycott 1986, Branch 1992, Broadley 1962, Broadley 1993, Hewitt 1937, Loveridge and Williams 1957). Difficulties in constructing phylogenies using these techniques arise from complications such as differentiating between genuine inherited characters (synapomorphies - shared derived characters), convergence (similar characters which evolved independently in different species) (Auffenberg 1974) and the possibility that some of the variation of unique characters between certain taxa (autapomorphies) may have arisen due to specific environmental selective pressure (e.g. the flattened shape of *Malacochersus*) (Hillis 1987).

Recently, new molecular methodologies have been developed and employed to assist in reconstructing phylogenies independent of morphology to complement the existing trees. These molecular techniques allow for a database to be constructed from the characterisation of macro-molecules such as proteins, nuclear DNA or organelle (mitochondrial) DNA.

In this dissertation, I attempt to infer a molecular phylogeny using mitochondrial DNA of the six genera representing the southern African Testudinidae. Both intra- and inter-specific relationships will be investigated. An attempt at estimating the time of divergence from the last common ancestor through the calibration of the molecular clock will also be made, in order to provide a fuller understanding of the diversity of southern African Testudinidae.

1.1. Evolutionary History of the Testudine Order

The evolutionary history of the land tortoise is somewhat cryptic, even though there is an abundance of fossil evidence. Comparisons of both shell and skull fossil morphology between Testudines and primitive Anapsida reveals many evolutionary discrepancies that have led to much debate as to which amniote group evolved into the modern day Testudine (Carroll 1988). These differences have also been responsible for the uncertainty in sister group relationships amongst both the living and extinct reptiles (Kemp 1980, Laurin and Reisz 1995). Recently, Lee (1996) has proposed that the extinct pareiasaurs, which were large herbivorous anapsid reptiles, are the sister group of the Testudines. Other studies, using mitochondrial tRNA sequence analysis, have also tried to resolve the organisation of the reptilian order (Kumazawa and Nishida 1995).

There is little doubt as to the basal relationships of the Cotylosauria (stem reptiles) to that of the rest of the reptiles. The suborder Captorhinomorpha, which separated from the anthracosaurs in the late Carboniferous (Fig. 1.1), gave rise to the reptilian class, which subsequently became the dominant vertebrate group during the Mesozoic era (Bellairs 1975, Carroll 1988, Gauthier *et al.* 1988, Laurin and Reisz 1995, Romer 1966,).

Prior to the "Age Of The Great Reptiles" during the Mesozoic, the Cotylosauria gave rise to the Amniota from which collectively, reptiles, birds and mammals are assumed to have originated (Fig. 1.1). The amniotes evolved as a monophyletic group in the Lower Carboniferous, but by the Upper Carboniferous, had diverged into three distinct lineages: Therapsida (mammal-like reptiles), Diapsida (birds and reptiles) and Anapsida, a group which Laurin and Reisz (1995) refer to as Parareptilia, and which is effectively equivalent to the Testudines (Carroll 1988, Gauthier *et al.* 1988, Lee 1996).

Amniota are characterised according to the number of temporal openings. Primitive amniotes that have a solid skull (*i.e.* no temporal openings) are referred to as anapsids, and form the ancestral stock of primitive stem reptiles (Fig. 1.2A) (Bellairs 1975, Carroll 1988, Zugg 1993). The first to branch off from the anapsids were the synapsids, also referred to as therapsids, or "mammal-like" reptiles (Bellairs 1975, Carroll 1988, Romer 1966), whose skulls had a single pair of temporal openings (Fig. 1.2B). The subsequent introduction of an additional pair of openings in the cheek region of the skull, gave rise to diapsids (Fig. 1.2D), the second group to evolve from the anapsids, and which consequently became the most dominant reptilian subclass of the Mesozoic (Romer 1966).

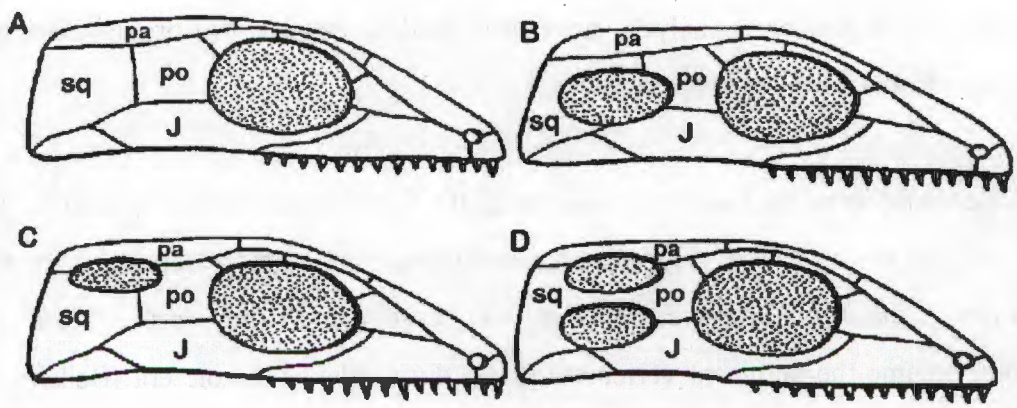


Figure 1.2 - Diagrammatic side views of early amniotic skulls depicting the pattern of temporal openings. A, no opening (anapsid); B, lower opening joining the postorbital and squamosal bones (synapsid); C, an upper opening (eurapsid); D, both openings present (diapsid). J - jugal; pa - parietal; po - postorbital; sq squamosal [After Romer 1966].

By the end of the Triassic all the major reptilian groups were present. Reptiles multiplied and radiated extensively from the Jurassic until the end of the Cretaceous, when many of the reptilian orders became extinct. The demise of the reptiles at the onset of the Cenozoic, marked the beginning of the mammalian radiation (Carroll 1988).

The reptilian amniotes which survived the mass extinction include the Lepidosauria (lizards, snakes, tuataras), Archosauria and the Para-reptilia. The Testudines, which have a unique dermal shell and anapsid skull structure, are the only representatives of the Para-reptilia (Fig. 1.1) (Gauthier *et al.* 1988, Laurin and Reisz 1995, Lee 1996).

1.2. Testudine Evolution

The Testudines, which share anapsid features with the extinct cotylosaurs, are the only extant amniotes representing the anapsids (Carroll 1988, Zugg 1993). It is plausible that due to their characteristic substantial body armour, the Testudines managed to survive independently with little change since the Triassic (Carroll 1988, Crumly 1984, Zugg 1993).

The earliest Testudine fossils are represented by *Proganochelys* in the deposits of Germany's Upper Triassic (Fig. 1.3) (Crumly 1984). Both the ribs and vertebrae were joined to the dermal bones (carapace) and the pelvis was fused to the dermal bones (plastron). Even though a horny beak had already been developed, the palate still retained teeth (Carroll 1988, Zugg 1993).

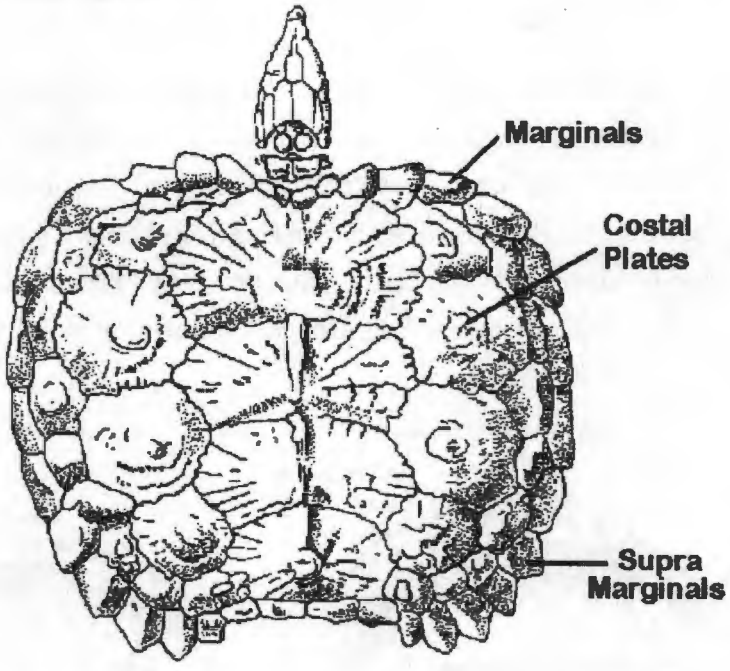


Figure 1.3 - Dorsal view of the carapace of the Triassic *Proganochelys*, showing the pattern of the horny epidermal, marginal scutes [adapted from Carroll 1988, Romer 1966)

This primitive Testudine stock may have been ancestral to the more advanced Amphichelydia (Romer 1966). Lacking teeth and dominating the Jurassic and Cretaceous periods, the Amphichelydia are thought to be ancestral to both extant suborders, Pleurodira and Cryptodira.

The (Greek) nomenclature Pleurodira and Cryptodira, refer to the method of how the Testudine neck and head are retracted into the body. The former does so by lateral flexure of the cervical vertebrae (Fig. 1.4A), and the latter retracts the neck vertically and posterialy into the body, forming an "S" shaped curve in the vertebrae (Fig. 1.4B and Fig. 1.5) (Boycott & Bourquin 1988, Carroll 1988, Romer 1966, Zugg 1993).

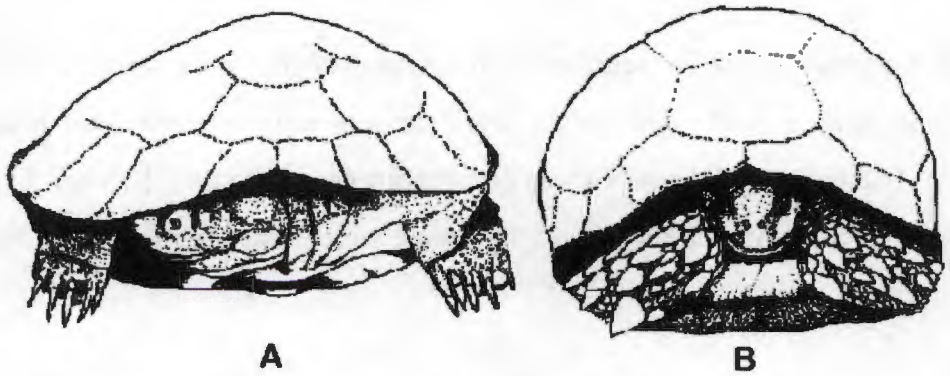


Figure 1.4 - Anterior view of the two methods of head retraction into the shell commonly used by extant Testudines. A, The head and neck are folded sideways under the shell, common to freshwater species (Pleurodirans). B, Both the head and neck are withdrawn into the shell. A feature displayed by both marine and terrestrial species (Cryptodirans) [Reproduced from Boycott and Bourquin 1988]

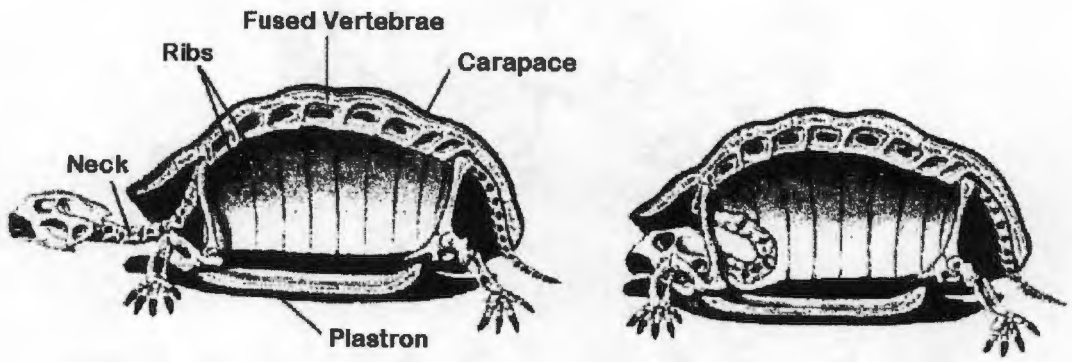


Figure 1.5 - Lateral view of a tortoise shell and skeleton depicting the fusion of the carapace to the vertebrae and ribs, and the mechanism by which the flexible neck is retracted into the body [Adapted from Hickman *et al.* 1993]

By the late Jurassic, the shell was common to all Testudines. However, minor modifications have taken place since then. These include:

1) Loss of extra marginal scutes (refer to Fig. 1.3)
2) Loss of mesoplastra (transitional plastral components) which are absent in all modern cryptodirans
3) Joining, in Pleurodirans, of the pelvic girdle to the plastron and carapace

It has been suggested (Romer 1966), that during the Upper Cretaceous of North America, terrestrial cryptodirans went back to the sea to escape their predators. This evasive action may have subsequently given rise to the reduction of the carapace and the development of webbed feet, both of which are common features shared by all sea turtles of the Chelonioidea super-family.

As the Jurassic period provided favourable conditions for world wide pleurodiran radiation to occur, the current restriction of pleurodirans to Australia, New Guinea and South America (Zugg 1993), may provide an indication as to the nature of the climate and habitat of the late Cretaceous. Presently, two families represent the pleurodirans, Chelidae and Pelomedusidae (Fig. 1.6A) that are further subdivided into seven genera (not shown).

The Cryptodira are considered to be more advanced than the pleurodirans because of the sophisticated method of head retraction (Fig. 1.4B and Fig. 1.5) (Romer 1966). When compared to pleurodirans, the cryptodirans are more diverse and have a greater geographical distribution (Zugg 1993).

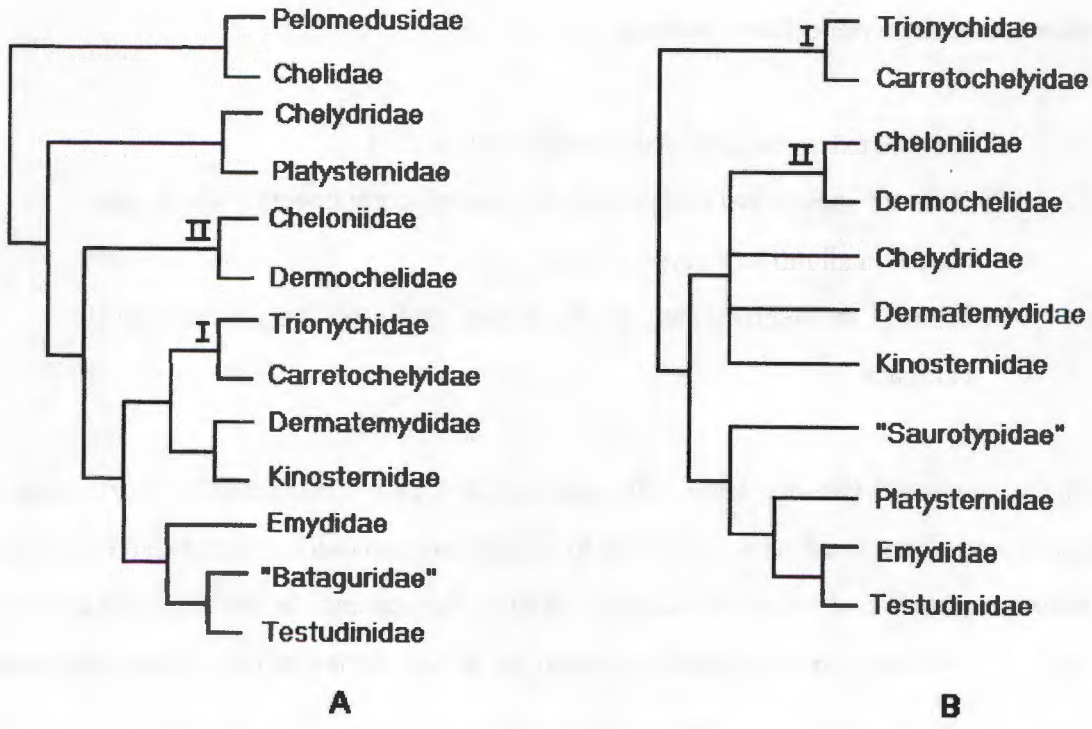


Figure 1.6 - The phylogenetic representations of the cryptodiran suborder based on; (A) morphological data (including two pleurodiran representatives) (Gaffney and Meylan 1988), and (B) cytogenetic data (Bickham and Carr 1983), each showing the distinct groupings of the soft shelled aquatic turtles (I) and the marine turtles (II).

Cryptodirans can be divided into three super-families: the sea turtles, represented by the Cheloniidae and Dermochelyidae referred to as Chelonioidea (Bickham and Carr 1983, Carroll 1988, Dutton *et al.* 1996, Gaffney 1975); the Trionychoidea including the Kinosternidae, Carretochelydae, Dermatemydidae and Trionychidae families (soft-shelled turtles); and the Testudinoidae, representing the Emydidae, Testudinidae and Chelydridae families (land tortoises and most fresh water terrapins) (Carroll 1988, Gaffney 1975).

Of the three super families, Chelonioidea and Trionychoidea are monophyletic, whereas the Testudinoidea, lacking unique features, are said only to be representatives of cryptodiran evolution instead of being a specific taxonomic group (Gaffney 1975, Carroll 1988). The Chelonioidea predate the Testudinoidea in the fossil record by appearing first in the Upper Jurassic, while the latter appear in the Eocene, and are typically found throughout Tertiary deposits (Carroll 1988).

Two proposed phylogenies of the families of recent Testudines are based both on molecular cytogenetics (karyotyping) (Bickham and Carr 1983) and morphological (Gaffney and Meylan 1988) techniques (Fig. 1.6A and 1.6B). The two phylogenies are in some agreement in that they propose the Testudinidae and the Emydidae as having evolved most recently. The phylogeny based on morphological data (Fig. 1.6A) which places the pleurodiran suborder as basal to the cryptodirans, specifies the "Bataguridae" as a potentially paraphyletic group. Gaffney and Meylan raised the Batagurinae from sub-family level to family status but admit that the family may be problematic in terms of low confidence of the proposed sister group relationships. Similarly, the "Saurotypidae" (Fig. 1.6B), have been raised to family status on the basis of chromosomal differences from that of the Kinosternidae (2n = 54 vs. 56). The overall topologies of both phylogenies are very different. The phylogeny based on morphological data resolves the Trionichoidea into a monophyletic super-family, whereas this super-family is unresolved using cytogenetic data. Even though the trees are incongruent there is some support for the independent sister group relationships of (I) soft-shelled aquatic turtles, and (II) marine turtles.

1.3. The Family Testudinidae

The earliest representation of the Testudinidae appears in the Eocene sediments of North America, Asia and Europe. The oldest known fossil is *Manouria majusculus* from the North American deposits (Auffenberg 1974, Crumly 1984, Estes and Hutchinson 1980). Although the origin of this specimen is not yet known, Crumly (1984) speculates that *Manouria*, which is represented by five different fossilised species, may have migrated from Asia to North America by way of the mid-continental contraction of the North American sea-way during the Eocene.

By the end of the Eocene and the start of Oligocene, testudinids had differentiated quite substantially from one another as they radiated to North America, Asia, Europe and Africa (Crumly 1984). Differentiation in the African tortoises is less apparent, as there are no known fossil records of testudinids from the Oligocene (Auffenberg 1974, Crumly 1984, Loveridge and Williams 1957, Meylan and Auffenberg 1986).

It was hypothesised by Auffenberg (1974) that during the Tertiary, both temperate and tropical biomes occurred together, allowing tortoises to disperse widely. The subsequent decrease in temperature and increase in moisture during the glacial periods of the Quaternary, forced tortoises occupying the northern latitudes to migrate south. These glacial periods, however, were not the main reasons for the gradual extinction of land tortoises. It is thought that many of the larger land tortoises became extinct as a result of the harsh dry and cold climate from the last ice age (Auffenberg 1974, Meylan and Auffenberg 1986). By adopting a burrowing habit, some taxa may have managed to survive these harsh climates. However, the burrowing habit has been lost by many modern testudinids and only *Gopherus*, *Testudo* (Auffenberg 1974) and *Xerobates* (Nichols 1989) of northern latitudes are known to have retained it (Auffenberg 1974).

1.3.1 The Monophyly Of the Family Testudinidae

Although the definition of the super-family Testudinoidea has been very problematic, the families Testudinidae and Chelydridae, housed within this super-family, are both well defined and assumed to be monophyletic (Benton 1990a & b, Crumly 1984).

Two morphological studies by Crumly (1984) and Gaffney and Meylan (1988) shown in Figure 1.7A and 1.7B respectively depict similar relationships between the different genera of the Testudinidae. The former phylogeny is based on cranial osteology, whereas the latter is a compilation of independent studies based on skeletal morphology, including the previously mentioned analysis by Crumly (1984).

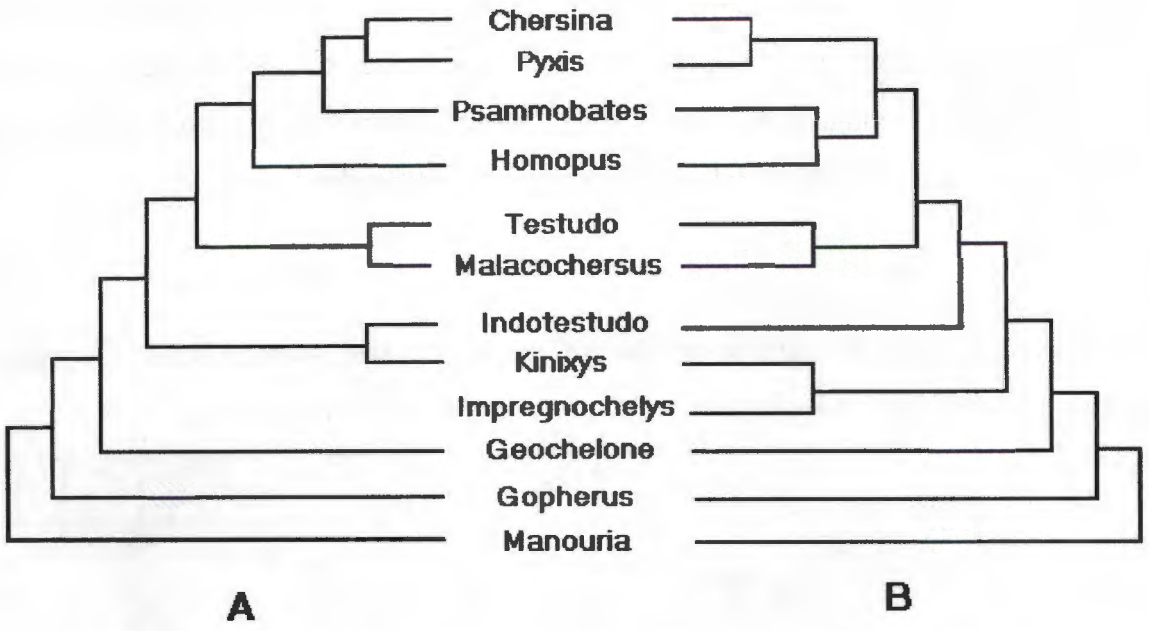


Figure 1.7. Two phylogenies of the family Testudinidae based on morphological data taken from (A) Crumly 1984, and (B) Gaffney and Meylan 1988.

1.4. Review of the Southern African Tortoises: Classification and Distribution

1.4.1. *Malacochersus*

Of the six genera found in southern Africa, *Malacochersus* is unique in its characteristic flat "pancake"-like appearance, as well as its relationship amongst the other genera within the Testudinidae. This monotypic genus, which is restricted to the rocky environs of Kenya and Tanzania, was first described by Tornieri in 1896 as *Cinicys belliana* (*Kinixys belliana*) based on the assumption that the flexible, flat carapace was related to a pathological defect of an adult *Kinixys* (Loveridge and Williams 1957, and references therein).

Seven years later Siebenrock (1904) re-described it as *Testudo tornieri* after reanalysing the Tornieri specimen (Loveridge and Williams 1957, and references therein). Siebenrock agreed with his colleague that the malleable shell may have been attributed to arrested development, however Siebenrock also noticed many similarities relating to *Testudo*. It wasn't until 1929 that Lindholm, after the rearrangement of the *Testudo* genus, proposed the generic name *Malacochersus* (Loveridge and Williams 1957, and references therein).

This monotypic genus is the flattest tortoise in the world. Its flat flexible carapacial characteristics have been thought to be primitive, resembling soft-shelled members of the Emydidae family (Loveridge and Williams 1957). These authors further postulate from these presumed primitive features that Malacochersus evolved independently from an ancestral stock of Testudines.

Malacochersus is typically found in narrow rock crevices in south-central Kenya and central Tanzania (Loveridge and Williams 1957, Klemens and Moll 1995) (Fig. 1.9A).

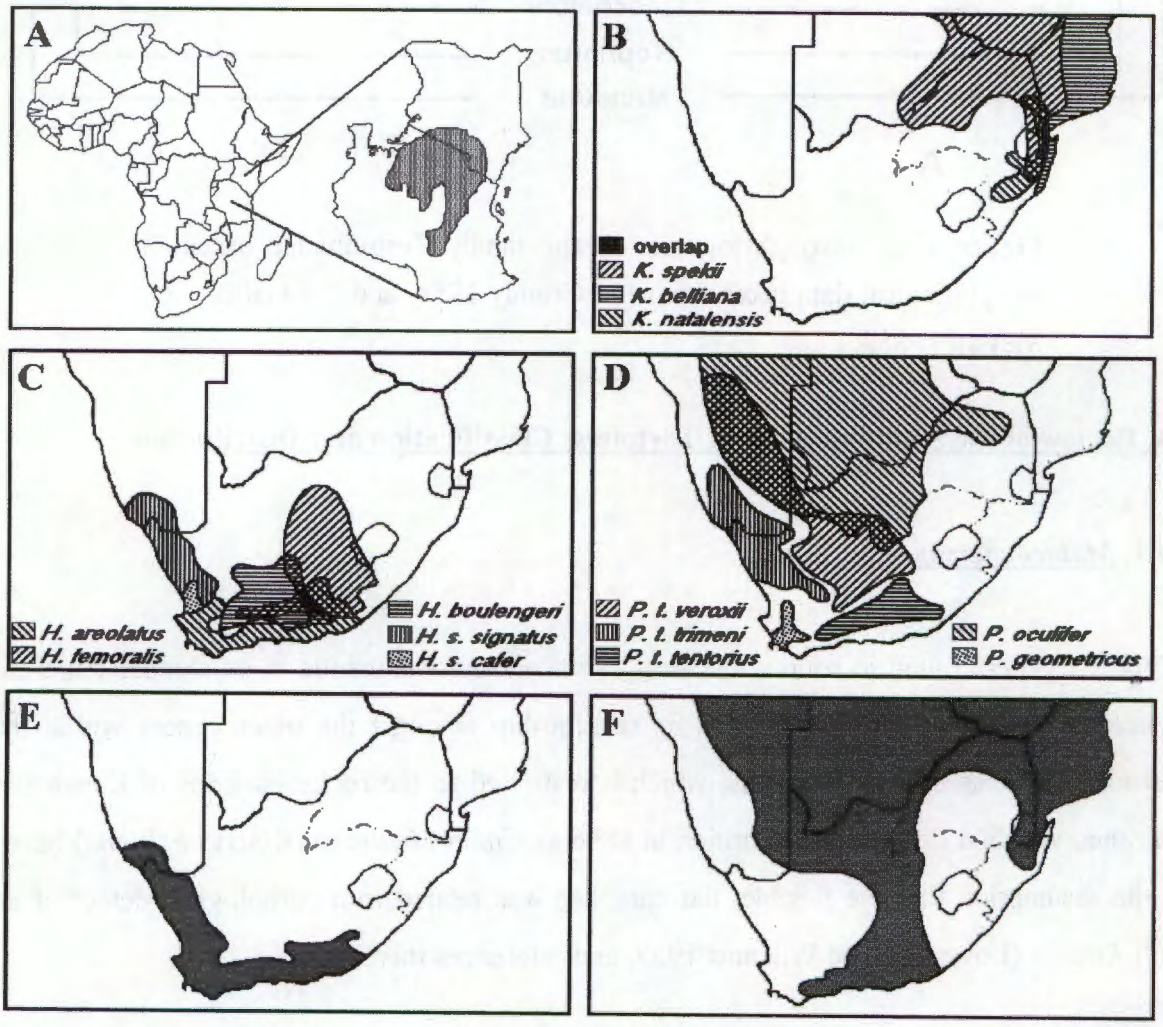


Figure 1.9. - Distribution of southern African Testudines. (A) *Malacochersus* (taken from Klemens and Moll 1995), (B) *Kinixys* (taken from Boycott and Bourquin 1988), (C) *Homopus* (taken from Boycott and Bourquin 1988), (D) *Psammobates* (taken from Boycott and Bourquin 1988), (C) *Chersina* (taken from Branch 1988a), and (F) *Geochelone* (taken from Branch 1988a).

1.4.2. *Kinixys*

First described by Bell in 1827 (Loveridge and Williams 1957), the adult *Kinixys* has a distinguishing hinged shell. Hinged between the costal and marginal bones, the carapace posterior, by movement of the vertebrae, is allowed to come into contact with the fixed plastron, providing ample protection of the hind legs (Loveridge and Williams 1957)

This hinge however is not a character which should be used solely for generic identification, as it is only found in adults as a secondary state, and may sometimes fail to develop (Loveridge and Williams 1957), resulting in a misleading specimen.

In addition to the potentially misleading hinge feature, initial species definition within *Kinixys* was further complicated by the lack of consistent morphological features (Hewitt 1931). This author treated all the southern African *Kinixys* representatives as distinct species, suggesting however, that perhaps *K. belliana, K. lobatsiana*, and *K. spekii*, might eventually be grouped as either *K. belliana* (domed) or *K. spekii* (flat), depending on their carapacial height. In 1935 Hewitt subsequently described *K. natalensis* as an additional species. This however was not generally accepted due to the paucity of distinctive features (Loveridge and Williams 1957) until 1981, when Broadley reinstated *K. natalensis* to full species status, based on carapacial length and height (Broadley 1981, 1993). To substantiate both Hewitt's (1935) and Broadley's (1981) claim, Boycott and Jacobsen (1988) confirmed the validity of *K. natalensis,* through quantitative morphological findings.

In reviewing former *Kinixys* classifications, Loveridge and Williams (1957) recognised only *K. belliana* as being a distinct species because the previously proposed species and subspecies lacked distinctive morphological characters as well as having a wide ranging savannah type habitat (Loveridge and Williams 1957, Broadley 1993).

From 1957 the genus remained monotypic until Broadley (1981) revised the genus *Kinixys* by resurrecting *K. natalensis* to full species status, making it sympatric to K. *belliana* and renewing *K. spekii* as a subspecies of *K. belliana* (*K. belliana spekii*). *K. b. belliana* was restricted to coastal plains of north-eastern South Africa and its other subspecies inhabited the interior. *Kinixys natalensis* was later confirmed as a full species (Boycott and Jacobsen 1988,

Broadley 1993). After scrupulous examination of more *Kinixys* specimens, Broadley suggested that *K. spekii* and *K. lobatsiana* both be considered as full species (Broadley 1989, 1992, 1993).

The relationships within *Kinixys* has been based on two major characters; one being the poorly developed carapacial hinge and the second being a primitive gular lip (Fig. 1.8). Taking both of these characters into account, Broadley (1993) hypothesised that *K. natalensis* is the most primitive of the four species (Fig 1.9B). *Kinixys belliana* and *K. lobatsiana* seem to be most recent while *K. spekii* is thought to be the second most primitive after *K. natalensis* (Broadley 1989, 1993).

1.4.3. ***Homopus***

Homopus is distinguishable from the other southern African genera by a flattened carapace in which the shields have a large central areola, and by having a wide gular shield. The first reference to this genus was by Dumeril and Bibron (1835), even though the *areolatus* and *signatus* species were individually identified by Thunberg (1787) and Schoepf (1782) as *Testudo areolatus* and *Testudo signatus* respectively (Loveridge and Williams 1957, and references therein).

Originally comprising four species, *Homopus areolatus*, *H. femoralis* (the largest of the four species [Greig and Burdett 1976]), *H. signatus* (the smallest tortoise in the world [Bayoff 1995]) and *H. boulengeri*, Hewitt, in 1931, noticed two discernible groupings and proclaimed that *areolatus* and *femoralis* should retain their *Homopus* distinction, and a new genus, *Pseudomopus*, be utilised for the remaining two species.

The former genus has four claws on the front and rear limbs (a feature unique to this genera alone, since other species in the family Testudinidae have five claws on their front and four on their rear feet), and a small inguinal shield which does not reach the femoral (Hewitt 1931). Creating this new genus made *Pseudomopus* a closer relative to *Psammobates* than to *Homopus* proper, or to *Kinixys* (Hewitt 1931). "*Pseudomopus, Homopus* and *Kinixys* are derived from a common ancestor, which is essentially a simplified *Pseudomopus*...." (Hewitt

1931). Six years later, Hewitt (1937) renamed the genus *"Chersobius"* (Boycott 1986, Loveridge and Williams 1957).

After Hewitt's (1931, 1937) convincing argument for the dichotomous separation of *Homopus*, it was subsequently rejected by Loveridge and Williams (1957). They refused to acknowledge the proposed generic division for fear of complicating the genus even further, even though they admit appreciable morphological skull differences between *boulengeri* and *areolatus*. Instead Loveridge and Williams (1957) contemplated the species within *Homopus* as being similarly related, with two identifiable species assemblages.

"We regard the species of *Homopus* as forming a single closely-knit series within which are recognisable two species groups: one embracing *signatus* and *boulengeri*; the relatively primitive members...the other, comprising the more specialised members - *femoralis* and *areolatus*." (Loveridge and Williams 1957)

Distribution of the genus is very heterogeneous with *H. areolatus*, *H. femoralis* and *H. boulengeri being* endemic to South Africa, whereas *H. signatus* is endemic to both the Western Cape, Northern Cape and southern Namibia (Fig. 1.9C) (Loveridge and Williams 1957, Greig and Burdett 1976, Boycott 1986, Boycott and Bourquin 1988, Branch 1988a). Ecologically, both *signatus* and *boulengeri* are found in similar habitats (Greig and Burdett 1976) of rocky outcrops and regions of low rain-fall. *Homopus signatus*, occurring predominantly in the Northern and Western Cape and *H. boulengeri*, found to be restricted to elevations of 2000 to 3000 meters of the Great Karoo, are allopatric. The two species *areolatus* and *femoralis* occur together in areas of high rain-fall (Boycott 1986). *Homopus areolatus* is predominantly a coastal species thriving on the indigenous fynbos, yet it can occur inland at elevations of 900 meters (Loveridge and Williams 1957, Greig and Burdett 1976, Boycott 1986), where it overlaps with *H. femoralis*, which is restricted to high plateau regions of over 1800 meters (Greig and Burdett 1976).

It is in this region of overlap that *Homopus areolatus* appears to display morphological differences from that of other *H. areolatus*. *H. areolatus* which occur in coastal areas have an olive-green carapace coloration, where as the same species which occur inland appear to have a carapace which is less vaulted with a uniform rustic brown coloration resembling that of *H.*

boulengeri and *H. femoralis* found in the same region (Ernst Baard pers. comm.). Other differences include small buttock tubercles and evidence of split supracaudals, both of which are atypical for *H. areolatus* (Ernst Baard pers. comm.)

Even though the geographic distribution of *H. femoralis* and *H. areolatus* is sympatric (over-lapping), they are detached from one another by habitat and altitude. An analogous situation occurs with the sympatric distribution of *H. boulengeri* and *H. femoralis* (Boycott 1986), where *H. femoralis* is confined to grassveld of high altitudes, and *H. boulengeri* is restricted to rocky outcrops at low altitudes (Boycott 1986).

1.4.4. *Psammobates*

Recognising three species, Loveridge and Williams (1957) mention that there were 33 possible names of which Hewitt (1937) was responsible for 48% of these citations, all due to taxonomic error, mostly dealing with *Psammobates tentorious*. "No group of tortoises has fostered the making of so many specific names as has *Psammobates*." (Loveridge and Williams 1957).

First described by Fitzinger (1835) as *Testudo geometrica*, *Psammobates* is said to be, according to Loveridge and Williams, ".... emydine-like in maintaining as a normal condition the hexagonal neural pattern that *Geochelone*, *Testudo* and *Gopherus* - all three independently - abandoned (except as an occasional variant) 30 or more million years ago."

Currently there are three species constituting the *Psammobates* genus. They are: *P. oculifer*, *P. geometricus*, both being monotypic species, and *P. tentorius*, which is subdivided further into three subspecies: *P. tentorius verroxii*, *P. t. trimeni* and *P. t. tentorius* (Auffenberg 1974, Baard and Mouton 1993, Boycott and Bourquin 1988, Branch 1988a, Branch *et al.* 1995, Greig and Burdett 1976, Loveridge and Williams 1957). The species *P. oculifer* (Auffenberg 1974, Baard and Mouton 1993, Boycott and Bourquin 1988, Branch 1988a, Branch *et al.* 1995, Greig and Burdett 1976, Loveridge and Williams 1957) has also been referred to as *P. oculiferus* (Iverson 1992), and *P. oculifera* (IUCN 1993), however since *P. oculifer* is used more frequently, it was chosen to be used throughout this dissertation.

The species within *Psammobates* can be separated into two groups, based on nuchal size and plastral pattern. Originally suggested by Siebenrock (1904), *P. geometricus* and *P. oculifer*, which are allopatric, form one group (Loveridge and Williams 1957, references therein). This is distinct from the second complex, *P. tentorius*, which features a plastral pattern over the entire plastral area (Baard and Mouton 1993, Greig and Burdett 1976, Loveridge and Williams 1957). A later study by Baard (1990) enabled the confirmation of the sister grouping of *P. geometricus* with *P. oculifer* (Baard 1990, Baard and Mouton 1993).

The distribution of the "sand-loving" *Psammobates* is quite peculiar with respect to the sister species *P. geometricus* and *P. oculifer* in that the latter two are allopatric, separated from each other by the *tentorius* complex (Fig. 1.9D) (Greig and Burdett 1976, Boycott and Bourquin 1988, Branch 1988a, Baard and Mouton 1993). Baard and Mouton (1993) speculate that the retreating ancestors of *P. geometricus* and *P. oculifer* may have been separated by the radiation of *P. tentorius*.

Currently the status of the geometric tortoise (*P. geometricus*) according to the IUCN/SSC is "vulnerable" (Groombridge 1982) and according to South Africa's Red Data Book: Reptiles and Amphibians (Branch 1988b), it is placed in the "endangered" category. At risk due to depletion of habitat from encroaching farmers, *P. geometricus* is endemic to the Western Cape (Baard and Mouton 1993, Baard 1993, Boycott and Bourquin 1988, Branch 1988b, Greig and Burdett 1976).

Psammobates oculifer is a wide ranging endemic species (Branch 1988a). Although it occupies scrub desert and savannah regions of central-southern Africa (never occurring south of the Orange River), it occupies a region of a higher rainfall than *P. tentorius* whose distribution is varied, from the dry arid environment of the west coast and the Karoo to the semi-desert of the Little and Great Karoo (Branch 1988a, Boycott and Bourquin 1988, Greig and Burdett 1976).

1.4.5. *Chersina*

The monotypic angulate tortoise (*Chersina angulata*), is the only African tortoise with a large undivided gular scute (Hewitt 1937). However, tortoise species from Mauritius and Madagascar within the genus *Geochelone* also share this plastral feature (Loveridge and Williams 1957). A feature, which sets Chersina apart from Geochelone, is the thickening of the gular region, a feature also found in *Kinixys* (Loveridge and Williams 1957).

First described by Gray (1825), the species is restricted to the coastal regions of South Africa from East London to the mouth of the Orange River (Fig. 1.9E) (Loveridge and Williams 1957, Greig and Burdett 1976, Boycott and Bourquin 1988).

1.4.6. *Geochelone*

The largest tortoise in southern Africa is the mountain, or leopard tortoise (*Geochelone pardalis*). This paraphyletic genus (Crumly 1982, Crumly 1984) is wide ranging, and can be found on the Galapagos Islands, Madagascar, islands in the Indian Ocean, South America and Africa (Loveridge and Williams 1957). As *Geochelone* has the largest distribution, its great diversity has lead to subdivision into 13 subgenera, of which seven are extinct and six are extant, and each are considered to be monophyletic (Auffenberg 1974, Crumly 1982). The species G. *pardalis* in Africa inhabits varied environments such as arid savannah regions to valley bushveld, from the Sudan in Northern Africa all the way to the Southern Cape (Branch 1988a). Not found historically in the South Western Cape, it has been suggested that they have been introduced there (Branch 1988a). There has been confusion and disagreement with regards to the status of *Geochelone* in Africa. Loveridge and Williams (1957) recognised two subspecies: *G. pardalis pardalis* and *G. p. babcocki*. Reservations have been expressed towards Loveridge's (1935) subspecific establishment of the former, because of the lack of strong morphological characters (Boycott and Bourquin 1988), and because of their sympatric distribution (Greig and Burdett 1976, Branch *et al.* and ref. therein 1995). However, a few herpetologists (Broadley 1989, Visser pers. comm.) still acknowledge the occurrence of a larger and flatter *G. p. pardalis* restricted to the Western Cape and Namibia, and a dwarfed race *G. p. babcocki* which occurs in eastern portion southern Africa.

1.5. Fossil Evidence

Land tortoises are currently distributed world-wide inhabiting the warmer climates of every continent except Antarctica and Australia (Auffenberg 1974, Crumly 1984, Meylan and Auffenberg 1986, Shaffer *et al.* 1997, Zugg 1993). According to the fossil record, land tortoises were larger than extant Testudines and had a greater geographical range than their present Gondwanaland distribution, and their extinction began in the early Cenozoic (Auffenberg 1974, Crumly 1984). Approximately half of these fossilised forms are associated with the genus *Geochelone*, whose fossil history dates back to the Eocene (Auffenberg 1974, Crumly 1984).

The first appearance of the family Testudinidae in Africa was described by Andrews (1906) who identified three fossil specimens from the Fayum of Egypt dating to the Eocene. Auffenberg (1974) allocated these three to a subgenus grouping of *Geochelone*. After considerable re-evaluation of the preceding descriptions, Crumly (1984) noted that the *Testudo ammon* specimen identified by Andrews, was in fact more comparable to the genus *Indotestudo* than to *Geochelone*, ignoring Auffenberg's (1974) classification and Hewitt's (1937) suggestion that *T. ammon* was a distant relative of *Chersina*. The type specimen agreed on by both Andrews (1906) and Crumly (1984) was *Testudo beadnelli*, which is thought to be a distant ancestor of *Kinixys*.

The southern African tortoise fossil record is fairly sparse, although there have been a few accounts signifying differentiation from their relatives of north Africa. In southern Africa *Geochelone* has been identified from the Miocene deposits of the Namib Desert, by fossil species such as *Geochelone namaquensis* (Stromer 1926) and *Geochelone stromeri* (Meylan and Auffenberg 1986). The latter has also been characterised from the Langebaanweg deposits of the South Western Cape, South Africa (Cooper and Broadley 1990). Even more recently, dating from the Pleistocene, a *Geochelone pardalis* has been excavated from the Makapansgat lime works in the Transvaal (Broadley 1962).

In the Pliocene phosphate deposits of Langebaanweg, fossils of *Chersina* have been recorded (Meylan and Auffenberg 1986 and references therein). The earliest fossil representative of

Chersina, found near Arrisdrift, Namibia, dates back to the early Miocene (Meylan and Auffenberg 1986).

Although Afro-Testudine fossil evidence is scarce (Table 1.2), it may be possible, using the fact that tortoises are animals limited to a slow rate of locomotion and inhibited geographically by physical boundaries, to shed some light on times of origins of southern African Testudinidae.

Table 1.2 African Testudinidae fossil record (Compiled from Auffenberg 1974, Cooper and Broadley 1990*, Meylan and Auffenberg 1986)

Genus	species	locality	Era
Geochelone	*ammon*	The Fayum, Egypt	Late Eocene
Geochelone	*beadnelli*	The Fayum, Egypt	Late Eocene
Geochelone	*namaquensis*	Namib Desert, Namibia	Miocene
Geochelone	*stromeri*	Namib Desert, Namibia	Miocene
Geochelone	*laetoliensis*	Laetoli, Tanzania	Pliocene
Geochelone	*crasa*	Kanapoi, Kenya	Pliocene
Geochelone	*pardalis*	Makapansgat Limeworks, Transvaal South Africa	Mid Pleistocene
Kinixys	*erosa*	Songor Hill, Kenya	Miocene
Chersina		Arrisdrift, Namibia	Miocene
Chersina		Langebaanveg, South Africa	Pliocene
Homopus	*fenestratus* sp. nov.	Albany, South Africa	Neogene* (Uncertain horizon)
Impregnochelys	*pachytectis*	Russinga Isl, Lake Vict., Kenya	Lower Miocene

1.5.1. Paleontological Limitations

Even though there is a substantial amount of fossilised Testudine shells, little is known about the order's evolutionary or fossil history (Carroll 1988, Crumly 1984), due to the lack of paleontological skull, limb and girdle evidence (Auffenberg 1974). The difficulty of classifying the paleontological Testudinidae into distinct generic groups, can therefore be rationalised by the following reasons (Auffenberg 1974):

- Evolution has mostly occurred in parallel
- Many of the genera are closely related
- Natural and convergent characters are difficult to distinguish
- There are many combinations of primitive features
- Advanced characters have evolved independently
- Advanced characters are not synchronised with geological time.

1.5.2. Molecular versus Morphological Data

Construction of phylogenetic trees using morphological data relies on the accurate description of physical, sometimes including behavioural, characters which form the basis of cladistic approaches (Hillis 1987). Unrecognised homoplasy (similar characters which are not inherited), or lack of sufficient numbers of such characters can limit the value of this approach for phylogenetic systematics. Extensive autapomorphic changes in a taxon adapting to a specific environmental niche may cause difficulties in classification (Hillis 1987). In the present context, for example, the extreme flattened shape of the *Malacochersus* in its adaptation to living in rock crevices makes classification difficult.

Molecular characters, especially DNA sequencing data have the advantage that they represent objective, unweighted, and unordered characters (Avise *et al.* 1987, Felsenstein 1988, Graybeal 1994, Nei 1987). Large data sets can be accumulated, assisted by accessioning data via the internet or related sequence databanks such as GenBank, which are generated and archived by other researchers. In addition to producing large data sets, material from study taxa can be obtained non-destructively from small amounts of blood or tissue. Advances have

also been made so as to utilise preserved museum specimens (e.g. flies in amber - Grimaldi 1996, Service 1996 and fossilised Woolly Mammoth - Yang *et al.* 1996). The disadvantages of molecular data are: unrecognisable homoplasies which can be extensive, choice of an inappropriate gene which can either have too little or too much change for a particular taxon study group, and expense.

Frequently morphological and molecular data sets give congruent results, but where this is not the case, molecular data sets can provide evidence, (such as lower than expected divergence rates), that the taxa involved need to be studied further. This can be resolved by reanalysing the data or by additional sequencing studies (e.g. using a different gene) (Graybeal 1994).

In the case of the family Testudinidae there are problems where morphological studies have not resolved all of the problems completely. These include whether *Homopus* is monophyletic (Should *Chersobius* [Hewitt 1931, 1937] be resurrected to genus level?), and the relative affinities of the included genera and the status of the subspecies (Are some subspecies cryptic species *i.e.* subspecies within *Homopus areolatus*?)

2.0 Mitochondrial DNA And Molecular Approaches Used In Systematics

During the last sixteen years, analysis of mitochondrial DNA (mtDNA) has been utilised to generate matriarchal phylogenies of many animal species (Avise 1994). Early analysis of this type, predominantly involved restriction fragment or site comparisons of the complete mtDNA genome, employing gel electrophoretic techniques (Avise *et al.* 1992, Avise 1992, Bowen *et al.* 1991, Bowen *et al.* 1992, Meylan *et al.* 1990, Nichols 1989). With the advent of polymerase chain reaction (PCR), segments of mtDNA have been amplified and sequenced, and are appropriate for answering many phylogenetic questions (Kocher *et al.* 1989).

2.1. Mitochondrial DNA

Mitochondrial DNA is a covalently closed circular molecule which has become widely used for both evolutionary systematics and population genetics. The main reason for mtDNA's popularity in genealogical studies is because of its haploid character, maternal inheritance, and lack of recombination. This is useful when studying colonisation, founder events and sympatric populations (Avise *et al.* 1987, Harrison 1989). MtDNA lineages of sympatric interbreeding populations are thought to remain isolated from one another, and any structural similarities would have most likely have arisen maternally (Avise 1992). MtDNA generally evolves at a fast rate and displays extensive polymorphisms within species (Brown *et al.* 1979, Rand 1994, Zhang and Hewitt 1996). Other characteristics which make mtDNA so versatile are: its stable, double stranded, circular structure (Fig. 2.1); its relative small size of approximately 15000 - 17000 nucleotide base pairs (bp) which allows restriction sites to be scored with relative ease, and its presence in multiple copies per cell. It is assumed that mtDNA's genetic variation is due almost entirely to genetic drift there-by acting as a neutral molecular marker (Avise 1992, Kimura 1983).

A review by Ballard and Kreitman (1995) however, questions the neutrality of the mitochondrial genome. They consider high heterogeneity and replacement polymorphisms found in *Drosophila*, rodents and humans to be directly correlated with mitochondrial recombination. Their theory runs counter to the neutral theory, implying that selection takes

over from random genetic drift, which is caused by random mutations in gene frequency resulting in the "average" genetic make-up of a population to diverge from the original starting composition (Quicke 1993).

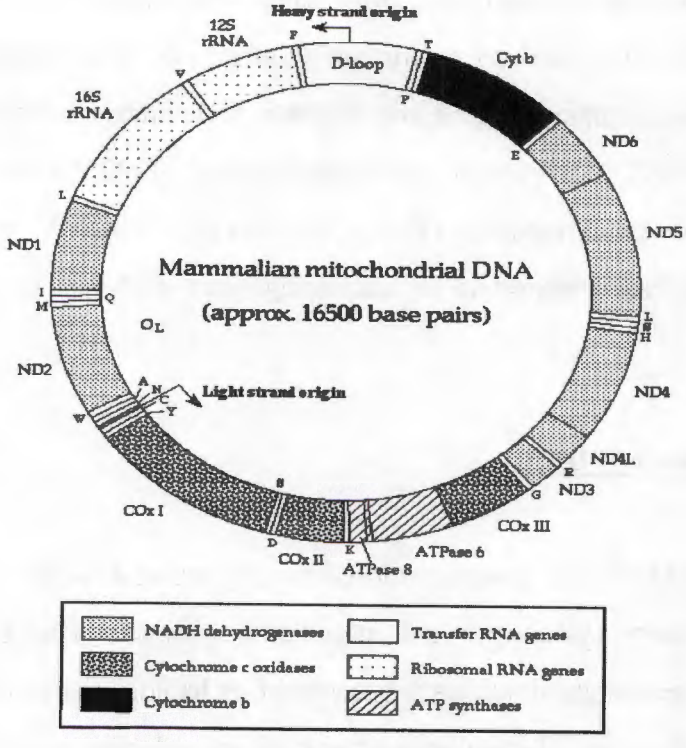


Figure 2.1. - Diagram of mammalian MtDNA [Adapted from Avise and Lansman 1983]

The animal mtDNA genome typically contains one rapidly evolving, non-coding variable region (D-loop) which controls mtDNA replication and tRNA transcription, 22 genes that code for transfer RNA (tRNA), 13 protein coding genes and 2 ribosomal (RNA) coding genes.

2.2. Choosing The Right Gene

Estimating phylogenetic utility of certain genes requires integrating divergence times and relationships amongst the taxa studied. This information can be obtained from paleontological and morphological data (Carroll 1988, Romer 1966). The amount of sequence divergence can be used to estimate the rate of evolution of a specific gene. However, factors such as the proportion of sites free to change, the nucleotide substitution

rate at a site, and the type of substitution allowed at that particular site all contribute differently to the utility of a gene (Fitch 1986). With these varying components that establish gene utility, many different combinations can occur that represent the rate of gene evolution. Three potential patterns are shown below (Fig. 2.2).

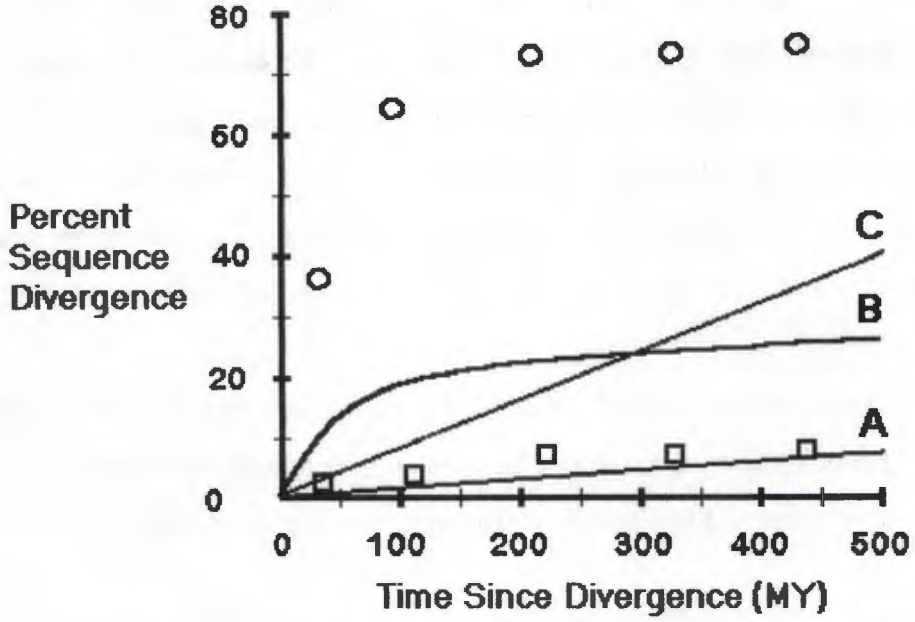


Figure 2.2. - Divergence times of three hypothetical DNA sequences plotted against overall percent sequence divergence (Adapted from Graybeal 1994). Line A, characterises a gene with all of its sites changing very slowly. Line B, depicts a gene which is evolving rapidly, but only at a small proportion of sites, and line C, represents a gene evolving at a moderate rate at a large proportion of sites. B is also further divided into silent (circles) and replacement (squares) mutations.

The first pattern (A) is characteristic of a gene that accumulates little sequence divergence with time (slowly evolving). Accordingly, few phylogenetically informative sites are detected and a gene of this nature would be appropriate for detecting divergences as old as 500 MY (Graybeal 1994). Genes that code for amino acids in mtDNA are expected to be representative of pattern (A) and are used effectively for resolution of deeper divergences (Kocher *et al.* 1989, Meyer and Wilson 1990). Amino acids are thought to be selected for more than at the nucleotide level, relying on the individual three codon positions for substitution (Adachi *et al.* 1993). Saturation between distantly related taxa will not be apparent for amino acids as it is for nucleotides (Kocher *et al.* 1989).

2.4. Techniques Used To Generate Molecular Data

There are currently many techniques to produce molecular data (e.g. isozyme electrophoresis, restriction fragment length polymorphisms and microsatellites to name a few), and each, whether it is old or new, has its merits as long it is applied appropriately. The methods used in this dissertation, polymerase chain reaction (PCR) and sequencing, have proved to be appropriate for this study since only a small amount of blood (100 µl) could be obtained without stressing the tortoise.

2.4.1. Polymerase Chain Reaction

The polymerase chain reaction (PCR) since its conception (Mullis 1990) has become a widely used technique in molecular biology. It is inexpensive, quick and generates multiple copies, in an exponential form, from a minute source of starting DNA via in vitro amplification (Gyllensten *et al.* 1992).

The reaction pivots around the DNA polymerase (*Thermus aquaticus*) which conducts the complementary synthesis of the previously heat denatured template DNA strand from the 5' to the 3' end. An oligonucleotide primer attaches to its complementary sequence on the template DNA and DNA polymerase then synthesises the complementary nucleotide sequences. Excess amounts of primer and deoxynucleotides are required so that the whole cycle (denaturing of the DNA, annealing of the primer to the template DNA, and extension of the primer by the polymerase), can be repeated many times to give an exponential increase in product (Gyllensten *et al.* 1992).

In PCR reactions, contamination of the template DNA with DNA from DNA from another source is a potential threat (Hillis *et al.* 1990). However, the template genomic DNA is in most cases much more concentrated than any potential foreign DNA. The chance of amplifying contaminating DNA is minimal unless the oligonucleotide primers have a higher affinity for the contaminating DNA, or the DNA from the species of interest is old, or damaged or degraded.

In pattern (B), mutations are accumulated rapidly, followed by a constant mutation rate. This hypothetical example, is similar to actual results found by Brown *et al.* (1979) when rapid evolution between higher primates was detected to be linear (increased sequence divergence in the first part of line B) followed by random multiple hits, representing the flattened part of line B (saturation). Genes that code for proteins provide additional differences between silent and replacement mutations both suggested to evolve at differing rates (Nei 1987). Mutations that replace amino acids occur less frequently than silent mutations which are abundant from the fact that synonymous substitutions occur at the third codon position without changing the amino acid. Resolution of recent divergences (100 MY) would be suitable for a gene such as the one portrayed in line B.

A gene that would be most likely to resolve distant divergences is the one representing line C. This gene evolves at a moderate rate accumulating continuous sequence divergence over time. More differences will be found in C than in A over the same time period.

Preliminary sequencing results on this data set, as well as published studies on marine turtles (Bowen *et al.* 1993, Laurent *et al.* 1993), emydids (Lamb *et al.* 1994) and North American gopher tortoises (Lamb and Lydeard 1994), suggested that cytochrome *b* had a sequence divergence profile similar to line B in figure 2.2 and had an appropriate amount of variation to make it suitable for this study.

2.3. Cytochrome *b*

Cytochrome *b* is the gene most widely used to date for inferring phylogenies (Meyer 1994). Its popularity stems largely from the development of the so called universal primers established for all classes of vertebrates (Kocher *et al.* 1989). With the diversity and availability of such primers, cytochrome *b* soon gained wide acceptance as an "industrial standard" for inferring vertebrate phylogenies (Avise *et al.* 1994). The rate of third codon (silent) substitution in cytochrome *b* (refer to circles associated with line B in figure 2.2.) is comparable to other genes in the mtDNA, though the rate of amino acid substitution is slow (refer to squares associated with line B in figure 2.2.) (Graybeal 1993, Meyer 1994).

2.4.2. Nucleic Acid Sequencing

Together with the appropriate gene selection and PCR optimisation, sequencing of the amplified target DNA has become a routine method of providing data for molecular phylogenetic analysis. In contrast to other sequencing techniques (e.g. Maxam and Gilbert 1977, 1980), the Sanger dideoxy method (Sanger *et al.* 1977) has become widely used because unlike in the previous methods an a priori knowledge of the sequence is not needed.

The Sanger technique uses DNA template strands (often amplified via PCR), a DNA polymerase, an oligonucleotide primer, the four deoxyribonucleotide triphosphates and smaller quantities of one of the dideoxynucleotide triphosphates (ddNTP's) to establish DNA sequences. The reaction is initiated by denaturing the double stranded DNA. A primer, which is complementary to the single stranded, target DNA, is then annealed to the template. The sample is further separated into four sub-samples into which the deoxynucleotides are added (*i.e.* dGTP, dATP [radio actively labelled], dTTP, and dCTP). To these mixtures the DNA polymerase and the respective dideoxynucleotide (ddNTP) is added.

The primer instigates the extension process where-by the polymerase attaches the deoxynucleotides. Similarly to the normal nucleotides, the ddNTP is added to the 3' end of the growing complementary chain, but once the polymerase attaches it to the end of the fragment it will terminate the extension process, because the ddNTP lacks a 3' hydroxyl group (Hillis *et al.* 1990). A series of radio-actively labelled, length varying primer extensions are generated, and electrophoretically separated according to their size on a denaturing polyacrylamide gel. The gel is then autoradiographed and the separated gel fragments displayed on the developed plate.

Researchers, (Allard *et al.* 1994 [*Chelonia mydas*], Norman *et al.* 1994 [*Chelonia mydas*], Lahanas *et al.* 1994 [*Chelonia mydas*], Walker *et al.* 1995 [*Sternotherus minor*]), investigating the control region of mtDNA and cytochrome *b* (Bowen *et al.* 1989, 1993 [Chelonioidea], Laurent *et al.* 1993 [*Caretta caretta*], Lamb *et al.* 1994 [*Graptemys*], Lamb and Lydeard 1994 [Gopher tortoises]), have used sequencing techniques and applied them to Testudine systematics for identification of ancestral lineages, construction of molecular

phylogenies, estimating the evolutionary diversity of specified genes and ascertainment of the process and patterns of macroevolution (*i.e.* phylogeography and migratory homing patterns through generation of phylogenies.).

3.0 Molecular Evolution

3.1. Molecular Clock

The basis of the molecular clock theory constitutes correlating the nucleotide substitution rate with time, the calibration of which involves paleontological or comparative molecular data. If the molecular divergence is known and can be calibrated, evolutionary events such as the time when two species diverged, can be estimated (Lewin 1988).

Some of the controversy surrounding the molecular clock, stems from theoretical and experimental estimations of complex DNA mutations, while some also extends from scepticism of the implication of unique molecular clock inferences when compared with classical morphological and paleontological interpretations - *i.e.*, primates (Brown *et al.* 1979) and turtles (Avise *et al.* 1992). Opponents of the molecular clock argue that molecular divergence correlated with time, is complicated by many sources of error leading to misleading results; the greatest of which are, accurately calibrating the paleontological record and identifying the age of the last common ancestor of the extant taxa (Carlson *et al.* 1978). Sequence divergence rates are also known to vary between groups of organisms (Li 1993, Martin and Palumbi 1993, Vawter and Brown 1986).

3.2. Concept of the Molecular Clock

There are three dominating features of the molecular clock: the locus effect, the lineage effect and the locus within lineage interaction effect.

1) **Locus effect**: This applies specifically to internal differences between two or more molecular clocks. For example, a nuclear gene coding for a protein consisting (in a theoretical extreme example) entirely of methionine (AUG) would be anticipated to have a slower synonymous substitution rate than a protein consisting entirely of leucine, since leucine is coded for by six synonymous codons. The methionine codon can not have synonymous substitutions. In addition, a protein made up entirely of cysteine would be expected to have a lower nonsynonymous substitution rate than a protein consisting entirely of proline because there are more alternative

amino acids that are similar in chemical and physical properties to proline than cysteine (Ridley 1993, Xia 1995 pers. com., Voet and Voet 1990).

2) **Lineage effect**: A phenomenon which refers to the same clock for protein coding genes, and is concerned with evolution at different rates for different evolutionary lineages. A lineage effect can be caused by factors such as population size, generation times, mutation rate or by any factor that changes the mutation - selection balance. If the lineage effect is substantial enough then it can compromise the validity of the molecular clock concept (Ridley 1993, Hillis and Moritz 1990).

3) **Locus within lineage effect**: The possibility that gene A evolves slower than gene B in one lineage but faster in a second. This may indicate that genes within the same lineage could evolve at different rates (Xia 1995 pers. com.).

3.2.2. Rate of Sequence Divergence

It has been postulated that functionally important proteins or genes evolve at slower rates than less important ones (Zuckerkandl 1962). Functioning predominantly as a template for conserved proteins of the oxidative phosphorylation pathway, the mtDNA genome with its highly conserved organisation, size and function (Brown *et al.* 1982), might be expected to contain a conserved primary sequence; accumulating little change over time. This however is not the case. Brown *et al.* (1979) observed mtDNA divergence of four species of higher primates, compared their results to an evolutionary time scale and calculated the substitution rate of vertebrate mtDNA to be approximately 1% every million years (Avise 1994). When compared to single copy nuclear DNA, the rate of mtDNA evolution was found to be about ten times faster. Mitochondrial DNA of Rodents (Gu and Li 1993) and *Xenopus* (Carr *et al.* 1987) has also been found to accumulate more change than nuclear DNA.

Using sequence data from cytochrome b and cytochrome oxidase I of 13 species of shark, Martin *et al.* (1992) tested the validity of the molecular clock by comparing their results to that of primates and ungulates. They concluded that the substitution rate in both genes is slower in sharks than in mammals by seven to eight times. Analogous findings were obtained by observing slow divergence rates (eight times lower) in Chelonians, using restriction site data, when compared with

nucleotide divergence rates in other vertebrate groups (Avise *et al.* 1992). In *Drosophila* (Powell *et al.* 1986) and echinoids (Vawter and Brown 1986), the rates of substitution between mtDNA and nDNA are very similar, indicating that a general molecular clock hypothesis may be incorrect (Rand 1994).

3.2.3. Possible Factors Influencing Rates of Nucleotide Substitutions

There have been many suggestions to account for the varying rates of base substitution in DNA including body size (Martin and Palumbi 1993), generation time (Vawter and Brown 1986), and metabolic rate (Martin and Palumbi 1993, Rand 1994), all of which are considered to contribute collectively towards evolutionary rate.

Related variables such as life span, population size, generation time and weight specific metabolic rate, have been regarded as being attributable to the varying rates of DNA evolution between diverging taxonomic groups (Martin and Palumbi 1993). The lack of support for the relationships between cellular characteristics and molecular rate of evolution, prompted Martin and Palumbi (1993) to compare various animals' body size and their respective sequence divergence. Compiling previous results they detected a trend that the evolutionary rate of mtDNA was inversely proportional to body size. (*i.e.* smaller animals accumulate DNA mutations at a faster rate than larger ones.)

Similar effects on the evolutionary rate have also been related to the life span, or generation time of certain species (Wu and Li 1985, Li *et al.* 1987, Li 1993). However the theory that DNA substitutions accumulate at slower rates in animals with shorter generation times, has been questioned by studies on the mutation rates of both mitochondrial (Ishida *et al.* 1995, Kumazawa and Nishida 1995) and nuclear DNA (Shlotterer *et al.* 1991). An increased accumulation of mutations may occur as a result of a higher metabolic rate of certain taxa (Rand 1994).

The rate of sequence divergence or mutation rate can be explained by differences in specific metabolic rates (Martin and Palumbi 1993). It is thought that the mutation rate may be affected by oxygen-free radicals which are generated by mitochondria as a by product of processing reducing equivalents for ATP production (Weindurch 1996). Mitochondrial DNA is more directly

susceptible to oxidative damage from free radicals than nuclear DNA, since it lacks a protective protein barrier (Richter *et al.* 1988, Weindurch 1996); and this could contribute to the reason why nuclear DNA accumulates less change over time than mtDNA (Brown *et al.* 1982). Larger animals with a lower metabolic rate should in theory accumulate fewer mutations since there are fewer protons pumped across cellular membranes decreasing oxidative damage (Rand 1994).

The evolutionary rate difference between ectotherms and endotherms may also have a correlation with metabolic rate. A four to five fold decrease in the rate of molecular evolution in poikilotherms, compared with homeotherms, has been noted in two separate studies using four different mtDNA genes, (Thomas and Bechenbach 1989) and the amino acid sequence from the complete mtDNA genome (Adachi *et al.* 1993). The authors in the latter study observed a graded evolutionary rate ranging from slow to rapid mutation rate as: fish < amphibians < birds < mammals.

3.3. Random Drift and Stochasticity

The irregularity of accumulating nucleotide mutations is a consequence of stochasticity and genetic drift, and is a major source of error affecting sequence divergence (Hillis and Moritz 1990). Predicting the estimated time of divergence of two related taxa can prove to be complex since the substitution rate is not constant with time. Randomness can be minimised by comparing larger data sets because of the inverse relationship between random error and the amount of data used.

3.4. Calibration of the Molecular Clock

Molecular clocks are calibrated by dividing the average estimate of the age of the last common ancestor with the average measure of molecular divergence. Frequently the only error which is taken into account is that obtained from sequence divergence resulting from stochasticity and genetic drift, however this error is insignificant when compared to other sources of error such as the accuracy of paleontological data (Carlson *et al.* 1978, Hillis and Moritz 1990). This point is particularly relevant when it comes to Testudine fossil record as this data is very incomplete.

3.5. Conclusion

Although there is controversy surrounding the molecular clock hypothesis, evidence suggests that a molecular clock can be applied to a particular gene within a particular group. To construct a relatively accurate calibration, certain criteria should be met, such as: only neutral mutations should be used (*i.e.* mutations which are neither more nor less advantageous than those they previously replaced), utilisation of a large data set to avoid stochastic error, comparison of homologous regions and use, if possible, of an absolute time calibration.

4.0 Approaches For Phylogenetic Inference

Inferring the evolutionary history of a species as well as the individual genes within that species, (which is not necessarily the same thing), has become an important scientific discipline incorporating complex statistical computations. It is through the generation of viable phylogenies that an understanding of evolutionary mechanisms and species history can be achieved. Calibration of these phylogenies with respect to evolutionary time based on morphological data is problematic because phenotypic changes are sporadic and irregular. With the discovery of the relative constancy in rate of base substitutions in DNA, molecular data were shown to display a more uniform pattern of evolutionary change than morphological or physiological data (Nei 1987).

Phylogeny construction using the parsimony method is based on phylogenetically informative characters (Felsenstein 1988, Li and Graur 1991, Quicke 1993). These assume that ancestral characters are inherited, and that through time these characters undergo changes creating an imprint in the organisms evolutionary history. Characters are most informative when their evolutionary rate is appropriate for the divergence times of interest. If they evolve quickly, homoplasy, which includes false resemblance through convergences, reversals and parallelisms, may conceal the phylogenetic signal (Moritz *et al.* 1987); if they evolve slowly then they may be uninformative, or have accumulated too little change to provide for an adequate sized data set.

Molecular data can be used to generate phylogenetic trees in many ways, although the two most commonly used approaches are distance methods (phenetics) and cladistic methods. The former approach generates a distance matrix by measuring the similarity or difference between pairs of taxa, from which a phylogeny may be constructed using a number of different algorithms. This approach is used when the evolutionary history of the taxa is unknown, and may not necessarily represent true evolutionary relationships (Nei 1987), since similarity alone is not a reliable indicator of evolutionary mechanisms (Stewart 1993).

Cladistic methods such as maximum parsimony, transform morphological and/or molecular data into character states which in the simplest cases may be comparable to a binary system (e.g. either presence or absence). Based on these phylogenetically informative characters, a tree which is explained by the least number of steps is

constructed. Cladistic analyses focus on the evolutionary pathways and ancestral relationships. Caveats such as differing patterns of evolution and unequal frequencies of homoplasies in distinct characters, should also be considered as they may be responsible for generating an incorrect phylogeny (Quicke 1993). Some authors (Stewart 1993) suggest that parsimony has a wider range of applications. Proponents of distance methods argue conversely that distance looks at the data as a whole, rather than isolating only the informative sites of e.g. a DNA sequence as does cladistics (Kumar *et al.* 1993). Since the debate as to which method (*i.e.* Cladistics vs. Phenetics) is better than the other has a history which is sometimes acrimonious, this dissertation will attempt, as much as possible, to treat both objectively.

In the following section both distance and cladistic methods will be discussed along with their respective parameters and character states which have a bearing on the phylogenetic outcome. In many instances more than one tree is generated and the problem of choosing the optimal tree and its respective robustness is also addressed.

4.1. Distance Methods

All distance methods require the initial generation of a distance matrix based on pairwise similarity. This transformation of sequence data to a phylogeny is carried out in two steps. First, the sequences are reduced to distance values for all the taxa in a pairwise manner, and secondly, phylogenetic analysis is performed on these distance comparisons. Since multiple events have a greater probability of occurring at a single position because of increasing distance, (Swofford and Olsen 1990), the generation of a distance matrix may be used to postulate the nature of the evolutionary pattern.

The number of hidden events or multiple hits which occur along the branch leading to a node, can be evaluated by determining the evolutionary distance which is based on the number of substitutions per position. Some of the algorithms used in this dissertation for determining evolutionary distance are discussed below.

4.1.1. Estimation of Evolutionary Distance : Unseen Events

There are many different methods of transforming the aligned homologous sequence data into a distance matrix. Algorithms are used to calculate the data and use different ways to compensate for multiple hits or unseen events. Correcting for potentially unseen events provides a better estimation of evolutionary distance and in the process, phylogenies can be generated with more precision and accuracy. Here, only P-distance and Kimuras' 2-parameter correction are discussed.

P-distance

The simplest method of comparing two sequences is through proportion or percentage similarity, referred to as p-distance (Swofford and Olsen 1990, Kumar *et al.* 1993). In this model all nucleotide positions and substitutions are considered equal. The proportion of differing sites between two aligned sequences is calculated by dividing the total number of dissimilar sites or fractional differences (Nd) by the amount of nucleotides (N) compared.

$$p=Nd/N$$

$$\text{Variance:} \qquad V(p) = [p(1-p)]/N$$

When p is small (<0.1), the number of nucleotide site substitutions is approximately equal to the p-distance (Kumar *et al.* 1993). The constructed phylogeny is in most cases the same as when the more complex methods below are used, as long as the difference between pairs of taxa is small (Swofford and Olsen 1990). The p-distance estimation retrieves the correct topology more often than the Kimura 2-parameter and Jukes Cantor models when the rate of substitutions between lineages (evolutionary rate) is similar and the amount is small. This is largely due to the low variance of this method. Lowering the variance of the other two algorithms is usually done by obtaining more sequence data. As the data size increases the similarities between sequences also increases, thereby lowering the variance (Kumar *et al.* 1993, Weir 1990). Since this method is very basic, it is not appropriate for estimating divergence times between two species because all substitutions are treated equally without giving any consideration to the difference in frequency of transitions (purine ↔ purine or pyrimidine ↔ pyrimidine) or transversions

(purine ↔ pyrimidine) and doesn't take into account multiple hits such as those characters not arising from inheritance.

Kimura 2-Parameter

This distance estimation method allows for the rate of transitions and transversions to occur independently from the four nucleotides, which themselves are assumed to be substituted at equal rates (Swofford and Olsen 1990, Kumar *et al.* 1993).

$$d = -1/2 \ln[(1-2p-Q)v1-2Q]$$

$$p = U_p/N \quad \text{(for transitions)}$$
$$Q = U_Q/N \quad \text{(for transversions)}$$
$$N = M+U_p+U_Q$$

where p represents the proportion of transitions in the sequence, U_p refers to the amount of transitions and N relates to the length of the two sequences compared. M represents the number of shared nucleotides.

DNA Sequences That Code For Proteins

Sequence data are more informative than protein data in most instances, because there is more data to compare and it provides opportunities to identify nucleotide substitutions that may not have resulted in a protein change. There are however factors to consider when using sequence data that code for proteins.

In protein coding genes, silent mutations/substitutions are more common than replacement substitutions (Swofford and Olsen 1990), resulting in a rapidly randomised third codon position that could lead to misinformative phylogenetic information. Nucleotide variation at the third codon position also tends to be similar between certain species. This suggests that different lineages have unique selective forces, which could result in misleading phylogenies based on convergent sequences (Swofford and Olsen 1990). The following approaches can be used when using protein coding sequence data:

1. A phylogeny can be constructed by converting the sequence data into protein data

2. The third codon position tends to vary randomly when sequence divergence increases and becomes uninformative; in which case, only the first two coding positions may be used.
3. If the divergence between two sequences is small, the p - distance can be used directly. However, the implicit assumption that all sites are treated equally will be violated, resulting in an unresolved phylogenetic tree (Swofford and Olsen 1990).

4.1.2. Neighbor-Joining

The neighbor-joining method takes the second category of distance strategies into account and generates an additive phylogeny (Saitou and Nei 1987). Additive trees are based on additive distances, which can be utilised to generate an unrooted tree. Unlike the UPGMA's cluster analysis, neighbor-joining does not assume the data to be ultrametric, and does not infer a molecular clock (*i.e.* not all lineages evolved at equal rates) (Swofford and Olsen 1990). In this method the distance matrix is modified so that the division of each nodal pair is calibrated according to their average distance from all other nodes, which in effect normalises the divergence of each taxon for its average evolutionary rate (Swofford and Olsen 1990).

Neighbor-joining converts an additive distance matrix into an ultrametric matrix and infers a phylogeny using cluster analysis methods. This method is not based on taxon clusters as is UPGMA, but rather on the nodes linking the least distant pair of nodes. By linking two nodes together, their common ancestral node is added to the tree, and the two linked (terminal nodes) along with their branches are removed. This ancestral node becomes a terminal node on a diminished tree. Again, the algorithm selects two terminal nodes that are similar, which it then replaces by a single ancestral node. This process continues until a single branch combining two nodes remains (Swofford and Olsen 1990). This algorithm is widely used for construction of dendograms from evolutionary sequence divergence values (Hillis 1995).

4.1.3. General Drawbacks of Distance Methods

The major concern over the use of distance matrix comparisons is the potential loss of valuable information when reducing the sequence data to the distance matrix (Li and Graur 1991). Distance also overlooks the differentiation between overall similarity and true evolutionary association (Stewart 1993).

4.2. Cladistics

Cladistic methods were developed to circumvent problems associated with the older phenetic models (Li and Graur 1991, Stewart 1993). The accumulation of morphological data is poorly correlated with time (Hillis 1987). Therefore a new method for phylogenetic reconstruction was devised, which looked at discrete character sets and ignored time. A tree generated in this manner allows only the topology to be obtained, unlike distance methods, which emphasise branch length measurements enabling calibration using a molecular clock. Cladistic methods rely on the identification of shared character states or synapomorphies (Fig. 4.1).

To construct a phylogeny using a cladistic approach, character states are compared and phylogenetically informative characters (constituting potential synapomorphies) identified. Algorithms such as PAUP and Hennig 86 have been devised and implemented to construct the most parsimonious (shortest) phylogenies (i.e. the trees with the least number of changes). To polarise the different character states, an outgroup is generally used to identify the states which are likely to be plesiomorphic (ancestral) or apomorphic (derived).

Similar character states can be classified into three categories:

- shared derived (synapomorphy) Fig. 4.1
- shared ancestral (symplesiomorphy) Fig. 4.1
- homoplasies - characters which are similar but derived from separate evolutionary events (parallelism, convergence, and reversal) Fig. 4.2

Shared derived characters (synapomorphies) occur along the branches of the most recently diverged lineages of common descent and arise from the development of new traits or modification (Fig. 4.1).

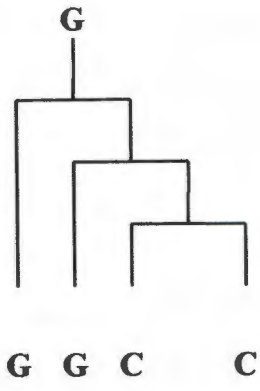


Figure 4.1. - Shared ancestral and derived characters. The shared ancestral state is represented by G. The ancestral G, relates to a shared ancestral character (symplesiomorphy) retained from the common ancestor.

Homoplasies are similar characters which did not result from inheritance (Hillis and Moritz 1990, Quike 1993, Stewart 1993) and occur independently by either parallelism, convergence or reversal (Fig 4.2).

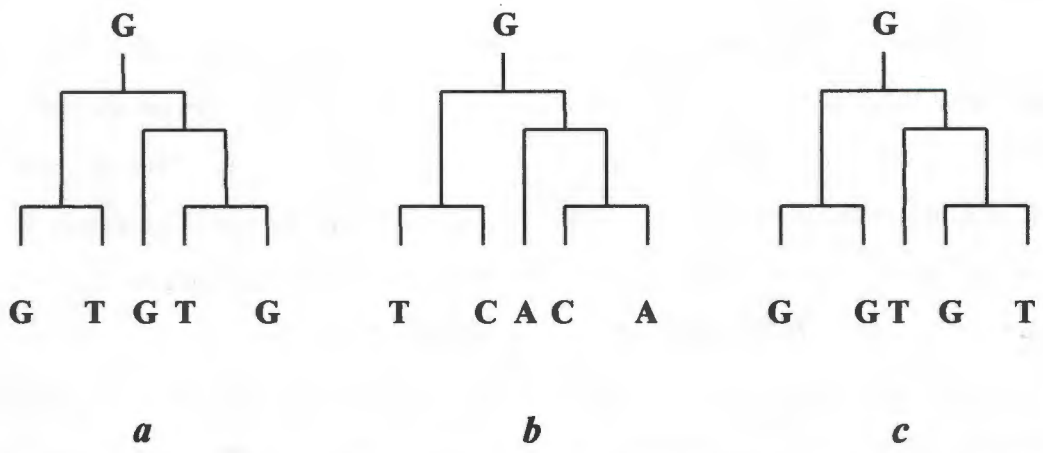


Figure 4.2. - Types of Homoplasy. Parallel mutations (*a*) arise when different lineages share the same character state. Here an ancestral G has mutated to T in two separate taxa. Convergence (*b*) occurs when two different ancestral character states (G and A) mutate to the same (C) state. Reversal or back mutation (*c*) results when the G mutates to a T and then back to its ancestral state G.

4.2.1. Parsimony

The principle of parsimony states that the simplest explanation consistent with a given data set should be chosen over more complex interpretations (Swofford and Olsen 1990). For molecular data, it has been used to generate phylogenies for amino acid (Eck and Dayoff 1966) and nucleotide sequences (Fitch 1971, 1977). The principle of maximum parsimony is to choose a tree that requires the least amount of mutational changes.

To construct a phylogeny using this method, a topology is first considered for a group of taxa. The minimum number of substitutions explaining the evolutionary changes (tree length) are counted and another topology is tried. This process continues for all possible topologies, or if the number of these to be examined is too great to be accomplished in a reasonable time then heuristic approaches which find some but not necessarily all of the shortest trees can be used. The tree, (or trees) with the lowest tree length is deemed to be most parsimonious (Nei 1987).

When using nucleotide data, not all polymorphic sites are informative because the observed change may have resulted from a single mutation in the branch leading to the terminal taxa (autapomorphy). Informative sites arise when there are at least two different types of nucleotides, each represented at least twice (Nei 1987).

The principle of parsimony has been criticised as not always providing a consistent estimate of the true tree. A method is said to consistent if it converges on the correct tree when the amount of nucleotide data is large or becomes effectively infinite (Kumar *et al.* 1993, Hillis 1995). If the average number of substitutions per site (Felsenstein 1978) is large, allowing many parallel or back mutations to occur (homoplasies), then errors occur unless the rate of substitution is constant (Kumar *et al.* 1993). Even if the substitution rate is constant, parsimony may generate the wrong tree if the number of nucleotide substitutions is small and there are many homoplasies (Huelsenbeck *et al.* 1996).

A disadvantage of parsimony, when compared to distance methods is that if both the number of taxa examined and the number of substitutions is small, most of the mutational changes at a site occur in only one taxon (*i.e.* autapomorphic) and these are

uninformative for constructing a parsimonious phylogeny (Nei 1987). Distance measures can use all the available sites where there is change, including autapomorphies. Studying the evolutionary relationships of apes and humans, Saitou and Nei (1987) found that the probability for obtaining the correct tree was higher for distance than for parsimony. Authors such as Saitou and Imanishi (1989) and Tateno *et al.* (1994) also found that parsimony performed less effectively than distance methods. These authors have generated renewed interest in the use of distance methods for molecular data, although it may be that the methods or assumptions used in evaluating the nucleotide sequence may have biased their results (Huelsenbeck 1995).

4.2.2. Rooting Trees

Most programs that contain parsimony algorithms (e.g. PAUP, MEGA and Hennig 86) generate unrooted trees unless specified otherwise. To distinguish which character states are apomorphic (derived) or plesiomorphic (ancestral) an appropriate outgroup must be chosen (Maddison *et al.* 1984). The underlying assumption or rule for outgroup analysis is that if a character within the ingroup has more than two states, the state which occurs in another group (outgroup) is assumed to be plesiomorphic (Maddison *et al.* 1984). It is important to assertain the character polarity because monophyletic groups can only be determined by apomorphies, and the inheritance of an evolutionary character can be traced only if the historical sequence of the character stated can be inferred (Clark and Curran 1986).

4.3. Testing Phylogenetic Reliability Using Resampling Methods

Random errors (homoplasies) may sometimes cause two taxa that have diverged early in evolutionary history, to be considered close relatives. Such occurrences are common if the data set is small, and often result in erroneous phylogenies (Swofford and Olsen 1990). The best way to minimise the effect of random errors is to have an infinite amount of data. Since this is not possible, finite data sets should be as large as possible and should be examined using methods that estimate the reliability of the generated phylogenies.

Methods such as the bootstrap and the jackknife estimate the variability which is associated with the sample distribution (Felsenstein 1988). They function by a repeated resampling of the data set in order to estimate the form of the sample distribution. The difference between bootstrapping and jackknifing, is that the former resamples the data randomly, with replacement, allowing some points to be sampled once, more than once, or not at all. The latter resamples randomly, but each time it resamples, the original data set is decreased by one (without replacement). With each replication, a specific statistic is calculated and can be related to a chosen confidence limit (Swofford and Olsen 1990). Felsenstein (1985) defines the statistic of interest by the presence or absence of a monophyletic group (which is chosen prior to the resampling), in each replication. By inferring such a statistic, characters can be weighted according the number of times they appear per sampling. The frequency with which they appear is a measure of confidence.

Misinterpretation of accuracy and precision has prompted some systematists to use large numbers of bootstrap replicates (Hedges 1992), with the intention of reducing the estimate bias, when in fact this high number makes the estimates biased with more precision (Hillis 1995). Performing many bootstrap replicates does not increase the accuracy of the phylogeny. If the tree is incorrect to start with, increasing the bootstrap replicates will not correct the tree.

4.4. Conclusion

This brief review of the various methods which estimate phylogenetic relatedness shows that a number of methods can be used, each having their specific assumptions, limits and merits. In some cases certain phylogenetic questions can be answered with one method and not the other. With the increasing processing power of the latest computers, techniques that were computationally cumbersome are now becoming more feasible. With advancing technology and powerful new algorithms, the generation of robust phylogenies should be become increasingly refined.

5.0 Materials and Methods

5.1. Sample Sources

Blood samples were acquired from 34 individuals (3 outgroups and 31 ingroups), representing each of the currently accepted southern African species (Table 5.1) using a non-destructive blood withdrawal method. The *Chersina angulata* sample was post mortem material from the posterior marginal of the carapace. Nine different specimens of *Homopus areolatus* were studied in order to provide information on the extent of within species sequence variation and because Ernst Baard (pers. comm.) had noticed morphological differences within the species in the Sutherland region of the Little Karoo (refer to 1.4.3.)

5.1.1. Blood Sampling

Prior to sampling, the tortoise limb was cleaned with 70% ethanol. Blood samples were obtained from individual tortoises using a 2 ml Promex syringe and a Promex needle varying in size from 0.45 x 16 mm LB to 0.7 x 32 mm LB, depending on the size of the individual.

The thick skin rarely allows one to see the vessels associated with the limbs therefore the sampling is done blindly. Experience enabled one to eventually obtain blood from a vein under a lose skin fold in the fore or hind limb with relative ease on most occasions; 10 to 100 µl being considered a sufficient blood sample. After sampling, the tortoise was observed for 15 minutes to ensure there were no ill effects from the procedure. Prior to blood extraction, the interior of the syringe and needle were moistened with heparin solution, so as to minimise clotting within the syringe. Once the blood had been obtained, it was expelled into a 1.5 ml microfuge tube containing lysis buffer (Appendix A), appropriately labelled and kept as cool as field conditions allowed.

Table 5.1. - Taxa localities that were used in this study (bracketed numbers coincide with numbered localities in Fig. 5.1 and Fig. 5.2)

Taxa	Reference Name	Locality
Chersina angulata	Cangul (1)	Jonkershoek (Stellenbosch)
Geochelone pardalis babcocki	Gpardb (2)	Mlawula Nature Reserve (Swaziland)
Geochelone pardalis pardalis	Gpardp1	Unknown (South African Museum)
Geochelone pardalis pardalis	Gpardp2 (3)	Victoria West (South Africa)
Psammobates geometricus	Pgeo1/2 (4)	Elandsberg Nature Reserve (Stellenbosch - Wellington)
Psammobates oculifer	Poculifr (5)	Windhoek (Namibia)
Psammobates tentorious veroxii	Ptentv (6)	Aus Townlands (Namibia)
Psammobates tentorious trimenii	Ptentt	Unknown (Tygerberg Zoo South Africa)
Homopus areolatus	HaHerm1/2 (7)	Hermanus
	Hareolat	Unknown (Tygerberg Zoo)
	HaEland (8)	Elandsberg Nature Reserve
	HaMUfema (9)	Müller Farm (Sutherland Region)
	HaMUmale (9)	Müller Farm (Sutherland Region)
	HaSUfema (10)	Driefontein (Sutherland Region)
	HaMTmale (11)	Matjiesfontein
	HaMTfema (11)	Matjiesfontein
	HaDeHoop (12)	DeHoop
Homopus femoralis	Hfem	Unknown (Tygerberg Zoo)
Homopus signatus signatus	Hsigsig	Unknown (Tygerberg Zoo)
Homopus signatus cafer	Hsigcaf	Unknown (Tygerberg Zoo)
Homopus boulengeri	Hboul (13)	Aus Townlands (Namibia)
Kinixys belliana	Kbelbel	Unknown (Tygerberg Zoo)
Kinixys spekii	Kspekii (14)	Mlawula Nature Reserve (Swaziland)
Kinixys natalensis	Knat (15)	Mlawula Nature Reserve (Swaziland)
Kinixys lobatsiana	Klobatsi	Unknown (Tygerberg Zoo)
Malacochersus tornieri	Mtorner	Mkomazi Game Reserve (Tanzania)
Testudo marginata	Tmargin	Unknown (Tygerberg Zoo)
Testudo hermanni	Thermani	Unknown (Tygerberg Zoo)
**Chrysemys scripta elegans*	CHscrip1/2	Unknown (Worcester Snake Park)
**Emydura australis*	Eaustral	Unknown (Worcester Snake Park)
**Pelomedusa subrufa*	PELsub (16)	Cape Peninsula (South Africa)

* outgroup representatives

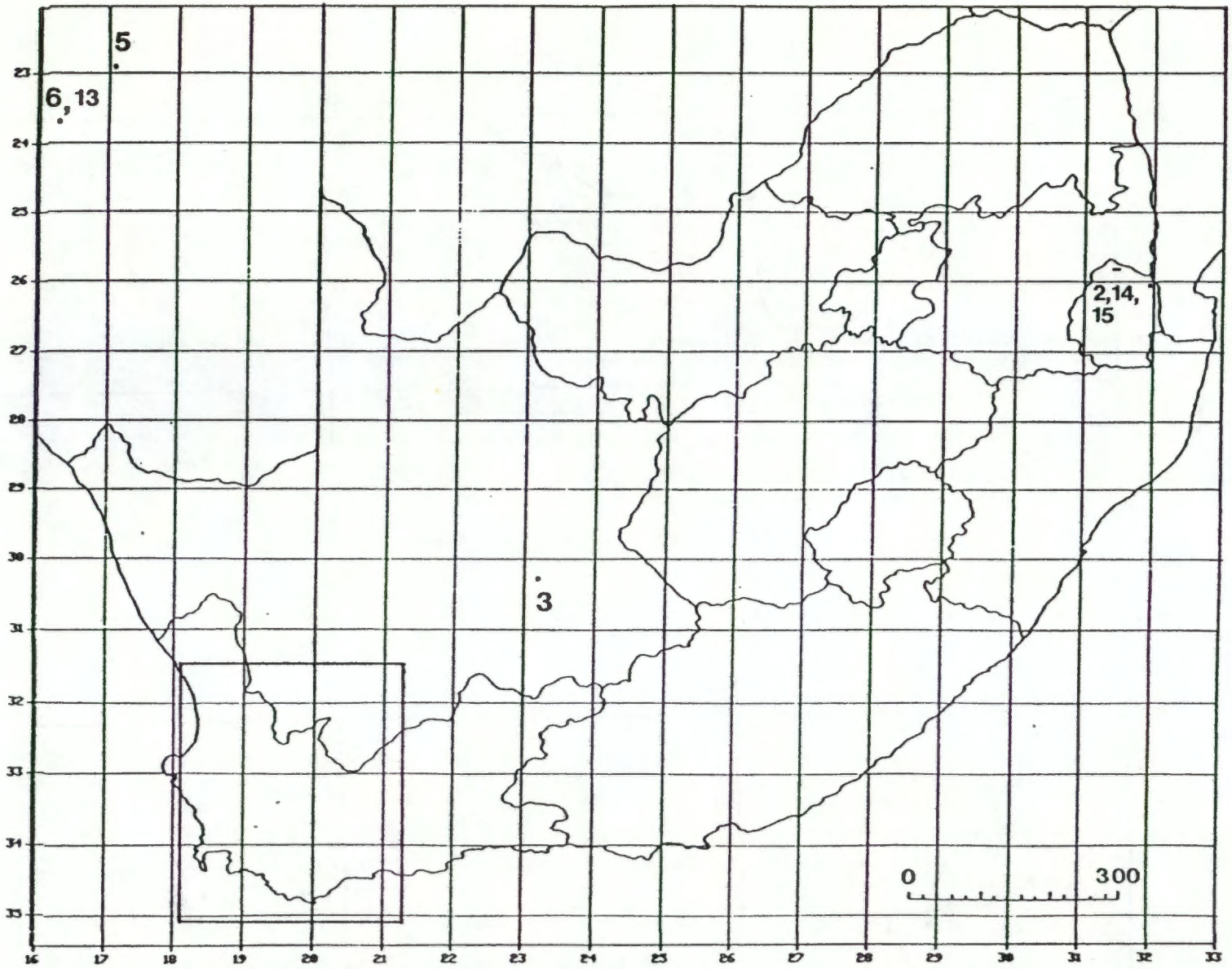


Figure 5.1. - Distribution map of samples obtained in this study. Boxed region is enlarged in accompanying Fig. 5.2. (refer to Table 5.1 for taxa representation).

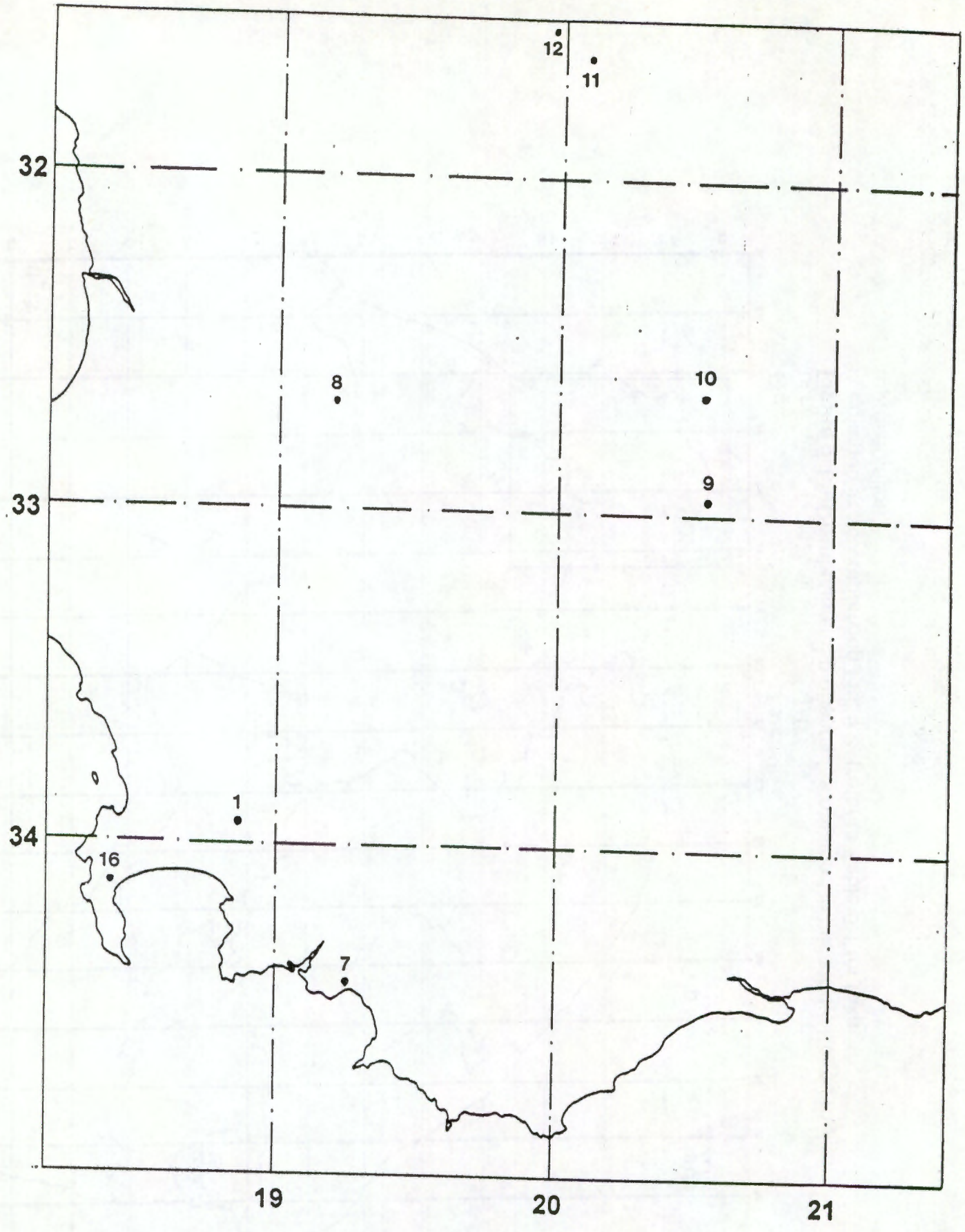


Figure 5.2. - Distribution map of samples obtained in this study (refer to Table 5.1 for taxa representation).

5.2. DNA Extraction From Nucleated Reptilian Blood

In the laboratory, 15 µl of 10 mg/ml of proteinase K (Appendix A) was added to the lysed blood mixture, and incubated at 37°C over night. The following day, a final concentration of 3 M NaCl was added to the solution to precipitate out residual proteins, and centrifuged for 15 minutes at 1700 g and 20°C, using a Sigma 2MK centrifuge. The supernatant containing the genomic DNA (both mtDNA and nuclear DNA) was saved and to it was added twice the volume of 100% ethanol. After vortexing briefly, the precipitated DNA, was centrifuged for 10 minutes at 1700g and 20°C. Without disturbing the pellet, the ethanol was taken off and 70% ethanol was added to wash the pellet of contaminating salt. After centrifuging for 5 minutes at 1700g and 20°C and taking off the 70% ethanol, the pellet was air dried under vacuum for 10 minutes and reconstituted in TE (Appendix A) to a final volume of 100 to 200 µl depending on the size of the pellet. This pellet should consist of total cell DNA which contains both mitochondrial and nuclear DNA.

5.2.2. DNA Extraction From Carapacial Material

After the posterior marginal of the carapace was excised, using scissors flamed with ethanol, it was cut into small pieces with a sterile dissecting knife and placed into a clean mortar and pestle.

Liquid nitrogen was poured onto the sample and an attempt was made to pulverise the fragmented carapace to obtain greater surface area. The powdered carapace was then placed into a 1.5 ml microfuge tube with 500 µl of extraction buffer (Appendix A). 75 µl of 10 mg/ml proteinase K were added, and the sample incubated at 37°C overnight. The following day 60 µl of 1 mg/ml RNase was added and incubated for an additional 30 minutes at 37°C.

An equal volume of Phenol/TE pH 7.5, was added to remove any residual proteins and vortexed briefly prior to centrifugation at 15 000 rpm for 15 minutes. The result was a heterogeneous mixture containing two phases. The top phase contained DNA, while the dark bottom layer consisted of the phenol and any haemoglobin and proteins that may have

been present. The supernatant was saved, and the sample was phenol-extracted a second time.

To purify the sample further, the supernatant was vortexed briefly with an equal volume of phenol/chloroform (1:1) and centrifuged at 15 000 rpm for 5 minutes. Again the supernatant was saved and an equal volume of chloroform : isoamyl alcohol (24:1) was added to remove the residual phenol. This centrifugation step took 10 minutes at 15 000 rpm.

The top layer was taken off carefully, avoiding any organic phase contamination and twice the volume of 100% ethanol was added and mixed gently until the DNA came out of solution. A pellet was formed by centrifuging again at 15 000 rpm for 5 minutes, after which the ethanol was removed and replaced with a 70% ethanol wash to eliminate residual salt and chloroform. After centrifuging, the ethanol was poured off and the pellet was allowed to dry overnight. 100-200 µl of TE. (depending on the size of the pellet), was then added to dissolve the DNA pellet. The DNA concentration, in both procedures, was assessed qualitatively by running 10 µl of the final sample volume on a 2% agarose gel.

5.2.3. Extraction Controls

Regardless of the DNA extraction method used, an extraction control was carried out by substituting the sample with filtered, distilled and autoclaved water. The control was subjected to the same conditions and procedures as the individual DNA containing samples.

5.3. Agarose Gel Electrophoresis

DNA samples were separated according to size using 2% (w/v) horizontal agarose (SIGMA Agarose Type II - Medium EEO) gels, in 1 x TAE (Appendix A) containing 1µl of 10 mg/ml EtBr (Sigma Chemicals Company). Electrophoresis was conducted at 110 volts in 1 X TAE buffer in either a Hoefer Scientific Instruments (HSI) Model HE 33 - Minnie•the•GelCicle™ or (HSI) model HE 99 - Dedicated Design Line, horizontal

submarine agarose gel apparatus. The latter was used prior to DNA sequencing to ensure complete band separation. DNA samples that were to be checked (10 µl), were mixed with bromophenol blue loading dye (2 µl) in a microtitre plate prior to loading.

5.4. Amplification Of DNA Using The Polymerase Chain Reaction (PCR)

A 450 base pair segment representing the 5' end of the cytochrome *b* gene was obtained for all 30 specimens (Appendix B). A series of available mammalian primers were tested initially to see which set was compatible with tortoise DNA. The primer combination of L14724 (L) and H15756 (H5) (Stanley *et al.* 1994) was found to amplify a fragment of 1032 bp for all taxa in the study except *Homopus areolatus* and *Homopus femoralis*. For these two specimens a shorter fragment of 668 bp was amplified using the former L primer with H15392 (H3) (Stanley *et al.* 1994) (see Fig. 6.1 in Results).

5.4.1. Primer Design

Initially six Testudine species were successfully amplified and sequenced with the L14724 (L) and H15392 (H3) primers. All of the sequences were aligned using the DNA and Protein Sequence Analysis program, DAPSA (Harley 1997), and a unique tortoise specific internal primer was then designed close to the end of the readable section of the sequence (Table 5.2.). Seven different species were used to design the primers to confirm that the region was conserved.

These new tortoise specific primers enabled further reading within the 450 bp region of the cytochrome *b* gene in either the 5' or 3' direction. Additional primers had to be designed for both *Geochelone pardalis* and *Chrysemys scripta elegans* (Tables 5.3. and 5.4. respectively) since the other primers were not able to generate readable sequence data.

Table 5.5. Summary of amplification and sequencing primers used throughout the study.

Primer Name	Primer Sequence
L14724(L) •	5'-TGA TAT GGG AAA CCA TCG TTG-3'
L14848(L21)*	5'-GGA TCA CTA CTA GGC-3'
H15263(H2) *•	5'-GAA TCG TGT TAG GGT GGC-3'
H15392(H3) •	5'-GGA TGG AAT GGG ATT TTA TC-3'
H15756(H5) •	5'-TAC TGG TTG TCC TCC GAT TC-3'
L15055(GpL31)*	5'-TAC TAT GGC TCT-3'
H14943(GpH31)*	5'-AGT AGC CCA TAT-3'
TERL2*	5'-ATT TTT CCT AGC-3'
TERH2*	5'-GAG GCG CTA CAG TTA TCA CCA AC-3'

Primers marked as (*) were used only for internal sequencing. Those marked as (•) signify primers designed by Stanley *et al.* 1994.

Table 5.2. - Sequence of seven Testudine species that were used to design the internal sequencing primer L14848. All of these sequences were obtained using the L14724 (Stanley *et al.* 1994) primer, except the *Caretta caretta* which was downloaded from GenBank (accession #L12720, Bowen *et al.* 1993). Dots represent nucleotides that are identical to the top sequence.

Position 14848↓

Chers. ang.	TGA	TGA	AAC	TTC	GGA	TCA	CTA	CTA	GGC	ATC	TGC	TTA	ATC
Caretta car.	...	...	...	...	...	...	...	...	...	.C.	..T	C..	GCA
Chrys. scrip.e	...	..G	..T	..T	...	...	...	T..	...	..T	...	C.G	...
Testudo marg	...	...	...	...	...	...	...	...	...	...	...	C..	.C.
Homo. fem.	...	...	...	..T	...	...	T..	..G	...	.C.	...	C..	...
Psam. geo.	...	..G	..T	..T	...	...	...	T..	...	..T	...	C.G	...
Geo. pard.	...	...	...	..T	...	...	...	...	...	...	..T	C..	...

Primer L14848 → GGA TCA CTA CTA GGC

Table 5.3. - Internal sequencing primers GPLH31 and GPL31, designed specifically for *Geochelone pardalis*. Dots represent nucleotides that are identical to the top sequence.

Position 14943↓

Chers. ang.	AGC	ATT	CTC	ATC	AGT	AGC	CCA	TAT	TAC	CCG	AGA	TGT
Kin. natalen	...	...	...	...	...	...	...	...	C..	...	...	...
Homo. fem	...	...	...	...	T..	...	...	...	...	...	...	...
Psam. geo	...	...	...	...	...	.A.	...	...	C..	...	...	...
Geo. pard. b.	...	...	...	...	...	...	...	...	C..	...	...	...
Geo. pard. p.	...	...	...	...	...	...	...	...	C..	...	...	C..
Geo. pard. p	...	...	...	...	...	...	...	...	C..	...	...	C..

Primer GPH31 → AGT AGC CCA TAT

Position 15055↓

Chers. ang.	GGC	CGA	GGA	CTT	TAT	TAC	GGC	ACC	TAC	ATA	TAC	AAA
Kin. natalen	...	...	...	...	..C	..T	...	T..	...	C.T	...	...
Homo. fem	..T	...	...	...	..C	...	..T	T..	...	C..	...	...
Psam. geo	...	...	...	...	...	...	...	T..	...	C..	..T	...
Geo. pard. b	...	...	...	...	..C	..T	...	T..	..T	C..	...	...
Geo. pard. p	..T	.A.	...	...	...	...	...	T.T	..T	C..	...	...
Geo. pard. p.	...	...	...	...	..C	...	...	T.T	..T	C..	...	...

Primer GPL31 → TAC TAT GGC TCT

Table 5.4. - Internal sequencing primers TERH2 and TERL2, designed specifically for *Chrysemys scripta elegans*. Dots represent nucleotides that are identical to the top sequence.

Position 15170↓

Chers, ang.	TATCATTCT	GAGGCGCCACAGTCATCACCAAC	CTACTCTCAG
Geo. pard.	.....T.	.....T.....T.....	.....T....
Emydura aus.	.....T.	.....T.....T.....	.....T....
Pelo. sub.	.....T.	.....T.....T.....	.....T....
Chrys. scrip.ele 1	.....T.	.....T.....T.....	.....T....
Chrys. scrip. ele 2	.....T.	.....T.....T.....	.....T....

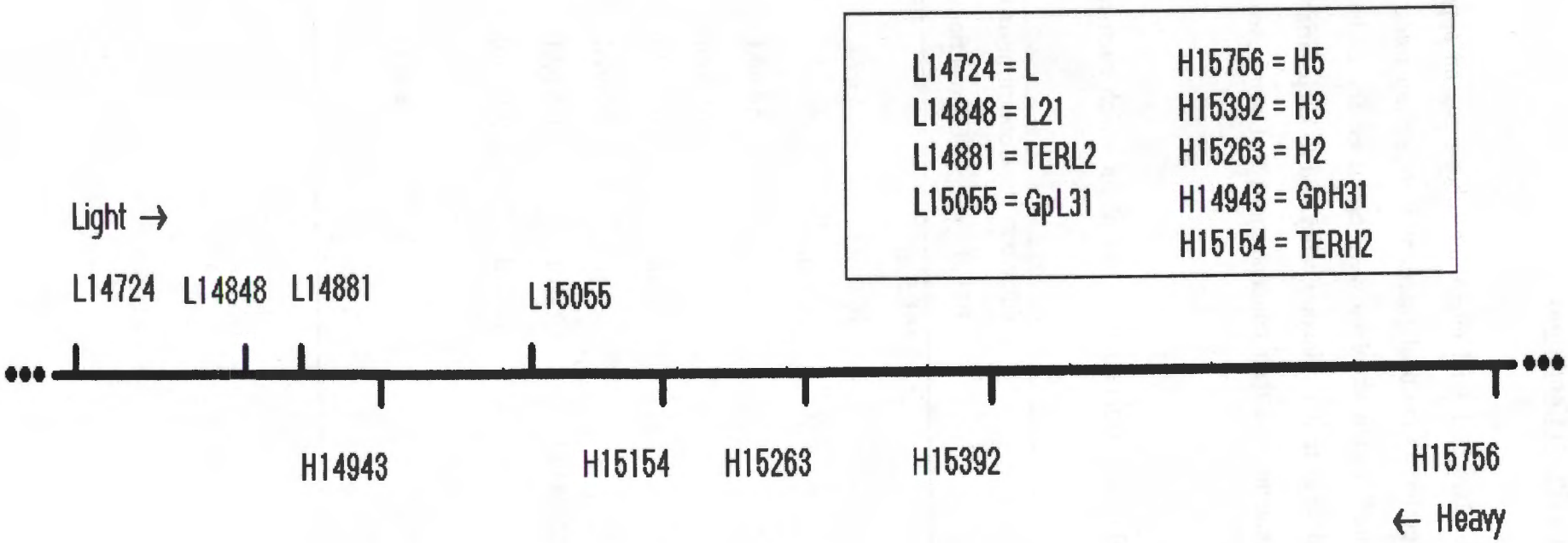
Primer TERH2 → GAGGCGCTACAGTTATCACCAAC

Position 14893↓

Chers. ang.	ATC ATT ACC GGA	ATC TTT CTA TCA	ATA CAC TAC TCA
Geo. pard.	... .T ...	..T ... ..G G..	... ...
Emydura aus.	..T ... ..T ...	..T ... ... G..	... ...
Pelo. sub.	..T ... ..T ...	... ..C ..G G..	... ...
Chrys. scrip. ele 1	..T	..T ..C ... G..	... ...
Chrys. scrip. ele 2	..C	..T ..C ... G..	..T ..T ... ...

Primer TERL2 → ATT TTC CTA GCA

Before a designed primer was ordered, it was tested rigorously using the Oligo primer program (version 3.4, National Biosciences Inc., Plymouth MN, USA) to identify optimal conditions. Note also that as these primers were used for internal sequencing, with the exception to TERH2, they could be shorter than those used for PCR amplification. The primers used in this study are shown in Table 5.5. and diagrammatically in Fig. 5.3.


L14724 = L
L14848 = L21
L14881 = TERL2
L15055 = GpL31
H15756 = H5
H15392 = H3
H15263 = H2
H14943 = GpH31
H15154 = TERH2
Light →
L14724
L14848
L14881
L15055
H14943
H15154
H15263
H15392
H15756
← Heavy

5.4.2. Polymerase Chain Reaction (PCR) Conditions

PCR amplification reactions took place in 1.5 ml microfuge tubes using an appropriate TECHNE PHC-2 PCR thermal cycler. Occasional reactions were set up using 0.6 ml microfuge tubes, but once the amplification variables were optimised for each specific taxon, a standard format was used so as to avoid inconsistency and to keep variables at a minimum. Table 5.6. shows the reagents and their concentrations that were used during amplification.

Table 5.6. - The total volume for each reaction was 100 µl, in which each reaction contained the following:

Reagent	Volume per 100 µl	Concentration in PCR reactions
H_2O (filtered and distilled)	61.5 µl	
$MgCl_2$ (25 mM)	16.0 µl	4 mM
10 x reaction buffer	10.0 µl	
(KCl 500mM)		50 mM
(TRIS - HCl 100mM)		10 mM
dNTP (5 mM of each dNTP)	4.0 µl	0.2 mM
Primer 1 (L) (25 ρM/µl)	2.0 µl	0.5 mM
Primer 2 (H3, H5 or TERH2 H) (25 ρM/µl)	2.0 µl	0.5 mM
TAQ DNA polymerase 5 U/µl (Advanced Biotechnologies, Surrey, UK)	0.5 µl	2.5 U/µl
DNA template (0.1 of genomic stock)		4 µl
Mineral Oil	2 µl	

5.4.3. PCR Optimisation

A preliminary amplification was performed using the above reagents on four dilutions of the extracted and purified genomic DNA (1/5, 1/10, 1/100, 1/1000), using the described conditions in table 5.7. Controls such as, an extraction blank (see above), and a PCR blank, in which genomic DNA was substituted with filtered, distilled, autoclaved water, were also amplified.

Table 5.7. Amplification conditions for most PCR reactions in this study

1. 40 sec. at 94°C (denaturation of double stranded DNA)
2. 60 sec. at 40°C (annealing of the primers to the single stranded DNA template)
3. 90 sec. at 72°C (PCR product extension)

These first three timed reactions were repeated 30 times followed by:

4. 10 min. at 72°C (final extension for all PCR products)

To optimise PCR conditions; amplification products were viewed and analysed with an ethidium bromide stained checking gel. In most instances DNA dilutions of 1/5 and 1/10 worked best. After identifying the appropriate DNA dilution, optimisation of the salt concentration was assessed by using a series of $MgCl_2$ titrations, ranging from 2 mM to 6 mM. In almost all cases, 4 mM was suitable to produce amplification products with the optimal DNA dilution.

If in either of these conditions multiple bands were observed, the initial annealing temperature of 40°C was increased by increments of 5 degrees, until a single band was produced. The conditions that generated the brightest, single band were repeated an additional four times to ensure ample target DNA was present during purification, concentration and sequencing.

Usually five DNA amplification reactions were pooled and precipitated using ammonium acetate and isopropanol. To the 500 µl total volume was added 250 µl of 8M ammonium

acetate and 750 µl of 100 % isopropanol. The microfuge tube with its contents was vortexed vigorously and then put into the -80°C deep freeze overnight. The following day the sample was centrifuged at 15 000 rpm for 20 minutes in a cool rotor, after which the supernatant was taken off and a 200 µl aliquot of 100 % ethanol was gently poured on top of the pellet. Again the sample was centrifuged at 15 000 rpm for 20 minutes after which the ethanol was taken off carefully and the pellet allowed to air dry in a vacuum for 5 minutes. When the pellet had dried, 20 µl of TE was added and the DNA allowed to dissolve in a 37°C water bath for 20 minutes.

Any residual primers or nucleotides, which may have influenced the sequencing reactions, were electrophoretically separated on a 2%(w/v) agarose gel using the conventional 200 X 150 mm electrophoretic apparatus. Prior to loading, a 5 µl aliquot of loading buffer was added to the sample. When the DNA had migrated approximately 4/5 of the length of the gel, the band was cut out from the gel with a flamed scalpel blade and put into a new, clearly labelled microfuge tube. Caution was exercised as to the amount of time the DNA was exposed to the ultraviolet light, since UV light destroys DNA (Morin *et al.* 1994, Mathew 1991).

5.5. DNA Purification From an Agarose Gel Slice

Recovery of the DNA from the extricated gel slice, was accomplished by using the Wizard™ PCR Prep (PROMEGA) purification system according to manufacturers instructions.

5.6. Sequencing the Purified Amplified DNA Fragment

Prior to sequencing, 2 µl of the purified DNA was run on an ethidium bromide stained, 2% (w/v) agarose minigel, to see if there was a sufficient amount of DNA present. Again this was done qualitatively and a bright glowing band signified ample DNA. If the band was very faint, a repeat of the amplification process was required.

The recovered amplification products were then sequenced in both directions, obtaining a consensus overlap of about 100 bp's. Sequencing was conducted as described in the

USB™ DNA Sequencing Kit (Amersham). All sequencing was done on single stranded DNA using a modified dideoxy sequencing technique described by Sanger (1977).
In a 1.5 ml microfuge tube the following was aliquoted, making up the annealing mixture:

7 µl of purified DNA
2 µl Reaction Buffer
2 µl Primer (25pM/µl) (*)
1 µl Detergent (see Appendix)

(*) Only L, L21, H2, GPL31, GPH31 TERL2 and TERH2 primers were used for sequencing (see above for sequence).

This mixture was vortexed, centrifuged for 15 sec. and then heated for 3 minutes in a TECHNE Dri-Block® DB.1M heating block at 94°C (to denature both the primer and template DNA) and then snap cooled in a slurry of ethanol and dry ice, (allowing the primer, because of its small size, to bind to the template quickly and efficiently). To prevent the sample from thawing, which may have resulted in non specific binding, the microfuge tube was quickly placed in a -20°C freezer.

In four separately labelled microfuge tubes, 2 µl of each dideoxy termination mixtures (ddGTP, ddATP, ddTTP, ddCTP) was added, along with 0.5 µl of detergent and then placed in a 37°C water bath until required.

The final reagents, making up the labelling reaction were aliquoted out into a single clean 1.5 ml microfuge tube and labelled appropriately.

2 µl Label mix (1:4 [Label mix : H_20])
1.75 µl Detergent
1 µl DTT (Dithiothreitol)
1 µl [S^{35}] ATP (10 µCi/µl)
2 µl Sequenase Polymerase (1:7 [enzyme : enzyme dilution buffer])

[On occasion 1 µl of Mn was used in addition to the above reagents, when sequences were required close to the primer.]

The label mix contains 7.5 µM of dGTP, dCTP and dTTP. ATP is obtained from the radio actively labelled S^{35}. The mixture was vortexed, centrifuged for 10 seconds and placed on ice, while in a separate 0.6 µl microfuge tube the enzyme dilution was carried out, and also placed on ice.

Having aliquoted out and treated all the three reactions (annealing, labelling and terminating) appropriately, the final timed sequencing reaction then took place. The following steps were taken during sequencing.

1. The annealing reaction was taken from the -20°C freezer and centrifuged for 10 seconds to incorporate any condensation that may have occurred during the snap cooling phase.

2. 5.75 µl of the label mix was added to the bottom of the microfuge tube, followed by the addition of 2 µl of the diluted enzyme. This was then mixed thoroughly by drawing up the sample into the pipette tip and expelling it a few times in succession, and then placed into a 12°C water bath for 3 minutes. The polymerase synthesises DNA complementary to the template from the 3'OH end of the bound primer. Incubation at this temperature (12 °C) enables one to obtain sequence data closer to the primer because the enzyme activity is slowed down.

3. After the 3 minutes had expired, 3.5 µl of this reaction was aliquoted to each of the pre-warmed termination mixes, centrifuged and incubated for an additional 3 minutes at 37°C. Polymerisation comes to a stop when the ddNTP becomes incorporated into the sequence, since the 3' end of the extending DNA now lacks a 3' hydroxyl group.

4. Following the termination reaction, 4 µl of Stop Solution were added, and the mixture stored at -20°C.

Each of the four microfuge tubes contained various fragment lengths of complementary DNA template all of which were radioactively labelled. These fragments were separated according to size, electrophoretically, on a polyacrylamide gel.

5.7. Separation of the Sequenced DNA

To separate out the amplified and sequenced DNA, a 6% polyacrylamide gel was poured from a 47% acrylamide stock solution (see Appendix A) between two glass plates (large - 370 mm x 441 mm and small - 370 mm x 410). Once the plates had been assembled and poured, using 75 µl TEMED, 350 µl of 20% ammonium persulphate (APS), and 100 ml of 6% acrylamide, the gel was allowed to set overnight.

To obtain as much readable sequence as possible, the sample was loaded twice after denaturing at 80 °C for 90 seconds. To read farther away from the primer, (long run), the sample was initially run until the light blue Xylene Cyanol FF dye migrated 3/4 of the gel length (approximately 2 hours). A subsequent loading (short run) was conducted in the same fashion in separate loading lanes, allowing sequences to be read closer to the primer. Both long and short runs were subjected to 80 watts (~2000 volts).

Once the dye of the short run had reached the 3/4 mark, the gel was disassembled and submerged in a fixing solution (see Appendix A) for 45 - 60 minutes. The gel was carefully taken off the glass plate with Whatman® chromatography paper covered with cling wrap and dried on a pre- warmed (80°C) HSI Model SE 1160 Drygel Sr. Slab gel drier, for approximately 30 minutes.

5.8. Autoradiography

The dried gel was screened with a Geiger counter to measure, qualitatively, the intensity of the label incorporation in order to give an indication as to how long the gel should be exposed to the autoradiograph film. Tape was used to affix the dried gel to the autoradiograph cassette. In the dark room X-ray film (Amersham Hyperfilm™ - MP X-ray film) was placed on top of the gel and the cassette sealed. After exposure, autoradiographs were developed, fixed and dried.

5.9. Analysing DNA Sequencing Results

The generated sequence data was aligned using the DNA And Protein Sequence Analysis program (DAPSA) version 3.8 (Harley 1997). Phylogenetic relationships of the aligned sequences were achieved by using both cladistic and distance methods. Maximum parsimony trees, distance dendograms and their associated confidence limits (bootstrap), were constructed using PAUP version 3.1.1. (Swofford 1993) and HENNIG '86 (Farris 1986). Additional distance dendograms were constructed using MEGA version 1 (Kumar *et al.* 1993) and the beta version of PAUP 4 D.XX (Swofford 1996) to see if there was a difference between PAUP 3.1.1. and MEGA. Since no discernible variations were noticed, PAUP 4D.XX was used throughout this study to maintain consistency. There was no difficulty in the alignment since the amplified fragment that was sequenced was part of the cytochrome *b* protein coding region, and there were no insertion and/or deletion (indel) variations detected.

Nucleotide sequences were subsequently converted to their corresponding amino acid sequences which served as a proof reading measure, ensuring that the nucleotide alignment was correct. For both nucleotide and amino acid sequence data refer to Appendix B1 and B2 respectively.

6.0 Results

6.1.1 Mitochondrial Amplification and Sequencing

All primer combinations successfully amplified a DNA fragment of 1032 bp from representatives of the 10 genera studied. A single sized product was generated for all primer combinations except in the case of *Homopus areolatus* where amplification with this primer set was unsuccessful. This was overcome by amplifying a smaller 668 bp product using primers, L14724 and H15392 (Stanley *et al.* 1994) shown in Fig. 6.1.

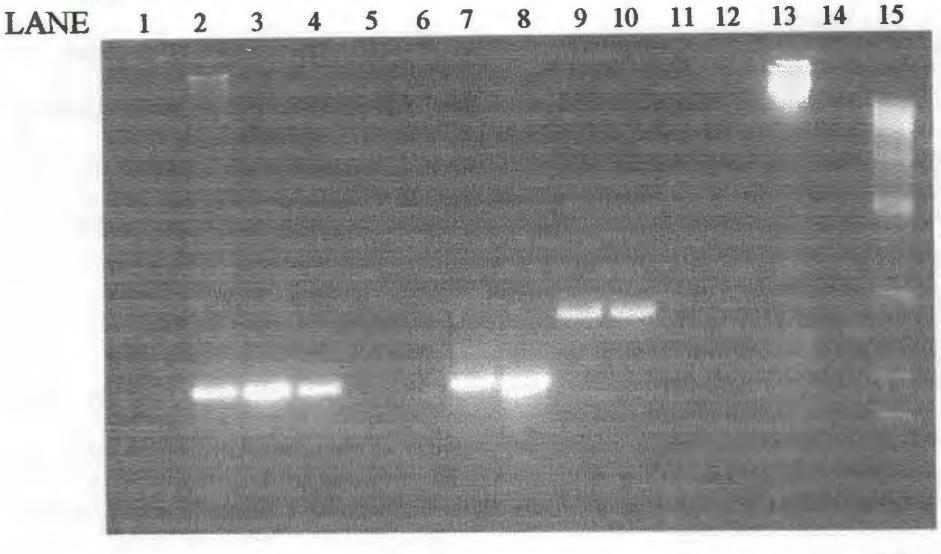


Figure 6.1. Comparison of PCR amplifications using different primer sets and $MgCl_2$ dilutions. The amplifications all contained a 1/10 dilution of the extracted genomic DNA, and 4mM of $MgCl_2$, with the exception of lanes 2 to 4, where a dilution series of 1mM, 3mM and 2mM are used respectively in association with a *Homopus areolatus* DNA specimen using primers L14724 and H15392.

Lane 1 shows the amplification blank. Lanes 5 to 8 show that the H15392 primer works only for *H.femoralis* and *H.areolatus* (lanes 7 and 8), and not for *H.signatus* or *H.boulengeri* (lanes 5 and 6). When the H15756 primer is used, both *H.signatus* and *H.boulengeri* (lanes 9 and 10) are amplified however *H.signatus* and *H.areolatus* in lanes 11 and 12 are not responsive to the primer. Genomic DNA of the *H.areolatus* and its respective extraction blank are pictured in lanes 13 and 14. The marker, (λ

DNA cut with DRA I) is shown in lane 15. Amplifications were performed according to the protocol described in 5.4.2.

Sequencing of the amplified fragments was successfully carried out using specifically designed Testudine primers (see 5.4.1). Clean sequences were obtained in most cases (i.e. no stuttering or overlapping sequences).

6.1.2. Sequence Variation

A 450 bp fragment representing the 5' end of the cytochrome *b* gene was obtained for all 34 individuals. Within the sequenced fragment, a total of 170 variable and 138 phylogenetically informative sites were observed amongst all taxa. The amount of variation in the first, second and third codon positions observed in the different genera in the study was found to be 14 (9%), 12 (8%), 37 (25%) respectively for the Testudinidae and 16 (11%), 15 (10%) and 39 (26%) representing the whole data set. No insertions or deletions were recorded.

When the pattern of nucleotide variation amongst different codon positions was assessed (Fig. 6.2.), it was found to be generally similar to that of mammals (Irwin *et al.* 1991) and amphibians (Graybeal 1993). A bias in the second and third codon positions was noticed, with the former having a high preponderance of T, whereas the latter showed a very low G content. The first position varied less than the third, and also showed less bias. Most of the nucleotide sequence variation occurred at the third position (Fig. 6.3), since most of the mutations here are silent (i.e. do not affect the encoded amino acid). The second position contained the least amount of change, because a single change at this site will in all cases lead to an amino acid change (Appendix 3), thereby altering the protein sequence (Muse 1996).

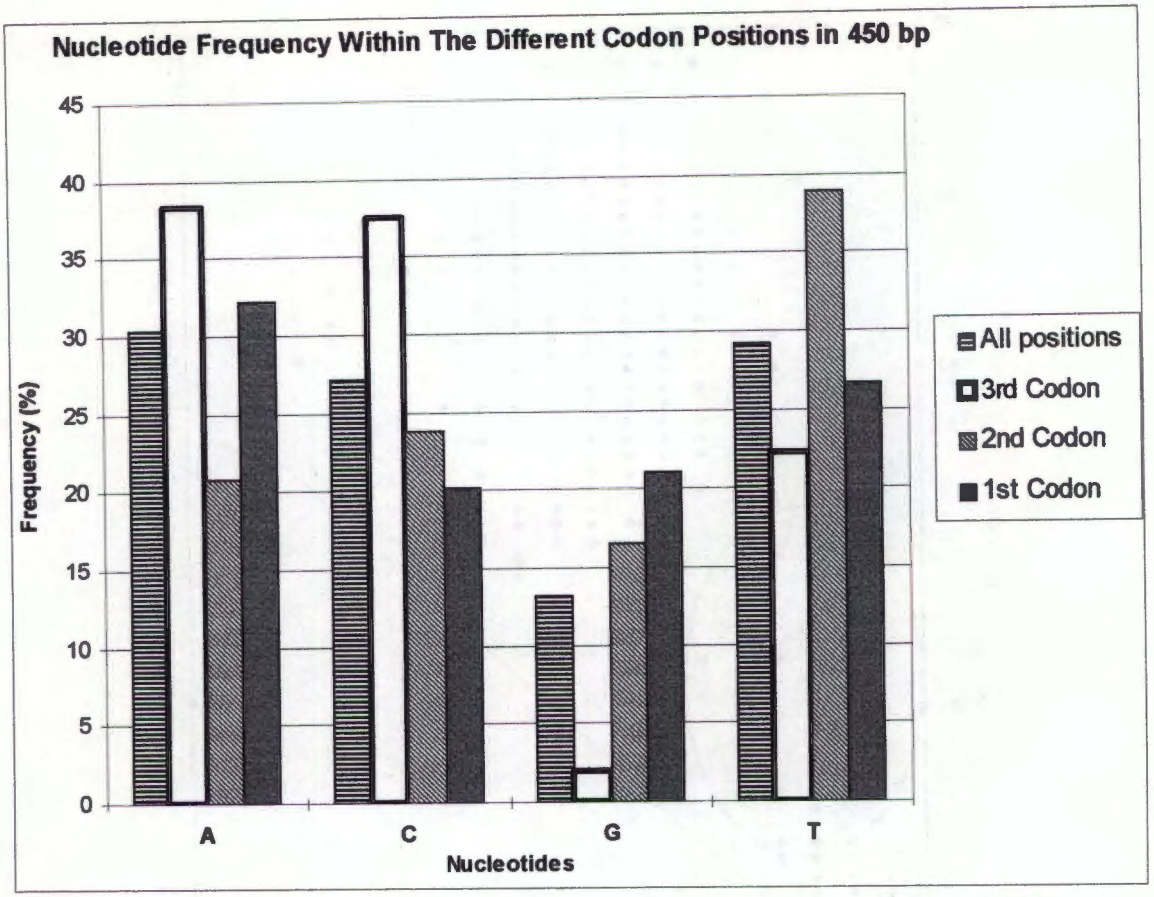


Figure 6.2. - Nucleotide frequencies at the three codon positions in 450 bp incorporating all 34 taxa in the study.

Using the complete data set, it was found that the transition/transversion ratio ranged between 2 and 14, with three comparisons out of 1122 pair of taxa, (*Hareolat-HaSUfem*, *Hareolat-HaEland* and *CHsrip1- Chsrip2*) which were in the range of 0 to 1. Transition bias was evident in pair-wise comparisons of closely-related species. Deeply divergent taxa indicated less bias towards transitions. This is to be expected, because saturation increases with time (Fig. 6.3.). A transition/transversion ratio greater than two (Fig. 6.3 - straight line with a slope of two), along with a pair-wise similarity of between 5 and 30% suggested the use of Kimura's 2-parameter method be used for distance analysis (Kumar *et al.* 1993).

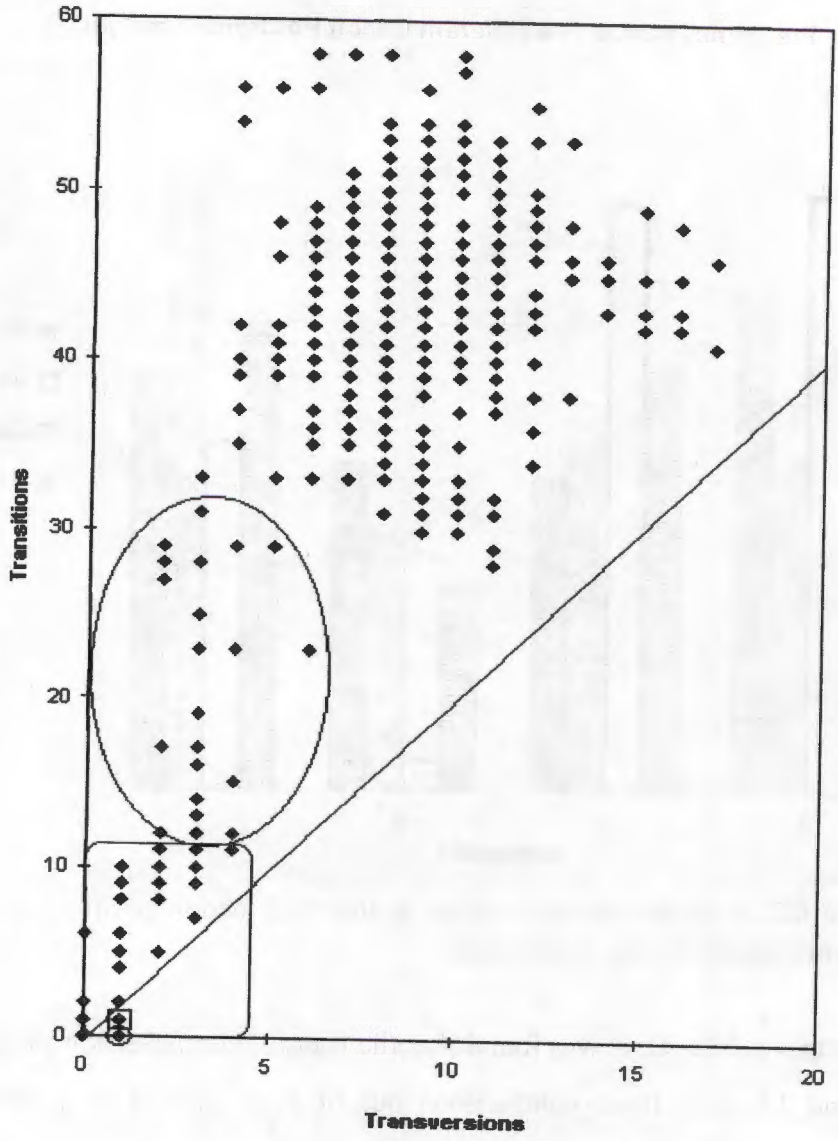


Figure 6.3. - Number of transversions plotted against number of transitions for all pairwise comparisons of cytochrome b sequences from 34 individuals within the Testudine order. Points with less than 12 transitions (bottom rectangle) represent within *Homopus areolatus* comparisons. The data point enclosed in both a box and circle, directly below the solid line, represent within-species (intraspecific) comparison of *Chrysemys scripta elegans* and *Psammobates geometricus* respectively. The cluster within the oval represents comparisons between species of the same genera (interspecific). Points remaining are those comparisons amongst the rest of the taxa in this study. The solid line (slope=2) depicts where the transition/transversion ratio is equal to two, corresponding with the Kimura 2 parameter requirement.

Pairwise sequence divergence estimates ranged from 0.0 to 15.1% for the whole data set (see Appendix 3). Within all substructure classifications, (i.e. subspecies, species and genera), higher degrees of sequence divergence than anticipated were found at each level. Intra-specific subspecies variation (Table 6.1) was observed to range from 1.6 to 6.7% for all currently accepted subspecies. *Homopus areolatus* showed a pairwise sequence divergence which ranged from 0 to 3.6% (Table 6.1). Grouping the individuals into different clades on the basis of both the parsimony and neighbor joining trees (Figs. 6.6a & 6.6b), found that the individual designated *HaSUfem* was between 2.2 and 3.6% different from all other individuals. Although not as detailed as the *Homopus areolatus* study, two individuals from both *Psammobates geometricus* and *Chrysemys scripta elegans* were found to have a 0.2% and 0.9% within-species difference, respectively. To obtain a more accurate rate of intra-specific sequence divergence more individuals of the same species will need to be sequenced.

Table 6.1. - Pairwise sequence divergence (%) between subspecies within a species, and between individuals of a single species with no recognised subspecies.

Category	Sequence Divergence(%) (p-distance)
Intra - specific subspecies variation	
Homopus signatus (signatus/cafer)	1.6
Psammobates tentorius (tentorius/verroxii)	6.4
Geochelone pardalis (pardalis/babcocki)	4.7 - 6.7
Intra - specific population variation	
Homopus areolatus (10 individuals)	0 - 3.6
Within Western Clade (3 individuals)	0.2 - 0.4
Within South Central Clade (6 individuals)	0 - 1.6
remaining individual (Sutherland region)	N.A.
Psammobates geometricus (2 individuals)	0.2
Chrysemys scripta elegans (2 individuals)	0.9
Geochelone pardalis pardalis (2 individuals)	2.0

An intra-generic comparison revealed a range of sequence divergence values from 1.1 % (*Kinixys spekii* vs. *K. natalensis*) to 14% (*Homopus signatus cafer* vs. *H. areolatus*-DeHoop) (Table 6.2). A similar study of North American Gopher tortoises, based on 352 bp of cytochrome *b* (Lamb and Lydeard 1994), found the range within *Gopherus* to be between 3.1 and 7.4%.

Table 6.2. - Sequence divergence estimates (%) within genera.

Genus	Range (%)	Number of Individuals sequenced
Homopus	6.7-14.0	14
Kinixys	1.1-7.6	4
Psammobates	4.2-9.8	5
Geochelone	4.7-6.7	2
Testudo	10.2	2

Sequence divergence values between different genera featured in this study were found to be lowest between *Malacochersus tornieri* vs. *Kinixys lobatsiana* and *Testudo hermani* vs. *Psammobates geometricus*, both of which shared a value of 6% (Table 6.3). The genera that displayed the highest sequence difference (15.1%) were *Malacochersus tornieri* vs. *Psammobates tentorius verroxii*. In comparison with other sequence data, Lamb and Lydeard (1994) showed inter-generic differences between North American Testudines to be in a narrower range (*Manouria* vs. *Testudo* [9.5%] to *Gopherus* vs. Geochelone [11.9%]), than the southern African genera.

Table 6.3. - Inter-generic range differences (%) between genera of the same family (Testudinae). The highlighted numbers represent the upper and lower values found.

	1	2	3	4	5	6
1 *Chersina*	-					
2 *Homopus*	6.9-11.1	-				
3 *Kinixys*	9.8-11.8	10.7-14	-			
4 *Malacochersus*	12.4	11.8-14.9	**6**-10.7	-		
5 *Psammobates*	10.4-14.9	9.1-14.7	12.2-14.7	13.6-**15.1**	-	
6 *Testudo*	9.1-12	9.6-13.8	12-14.4	12.7-14.2	**6**-14.7	-
7 *Geochelone*	8.9-10	9.1-11.6	10.7-12.2	12.9-13.6	9.3-13.6	9.1-11.6

A comparison between the generic members of the Testudinidae family with a representative of the sister family, Emydidae (Table 6.4) showed values of between 8.9 and 14.4% for *Geochelone pardalis pardalis* vs. *Chrysemys scripta elegans* and *Psammobates geometricus* vs. *C. s. elegans* respectively. Representatives of the Pleurodiran suborder were also compared to the Testudines (Table 6.4). It was observed that *G. p. pardalis* shared a sequence divergence of 2.7% with both *Pelomedusa subrufa* and *Emydura australis*. The largest difference was 12.7%

between *P.subrufa* vs. *Kinixys lobatsiana* and *P. subrufa* and *M. Tornieri*. Other intra-familial comparisons between the Emydidae the Chelidae and Pelomedusidae show that the Emydid representative differs by a similar degree from both of the Pleurodiran suborder individuals, whereas the families within the Pleurodira have a sequence divergence of 2.7%.

Table 6.4. - Range differences (%) between genera within the Testudinidae and individuals of a different family # (Emydidae) and sub-order * (Pleurodira). The high lighted numbers signify upper and lower boundaries of the range. Differences were obtained from pair wise comparisons (Appendix 3).

	Chrysemys scripta#	*Pelomedusa subrufa**	*Emydura australis**
1 *Chersina*	12.17	9.6	10
2 *Homopus*	12.7-14	8.7-11.3	8.9-11.3
3 *Kinixys*	12.9-14.2	11.3-**12.7**	10.7-12.4
4 *Malacochersus*	14.2-14.2	**12.7**	13.3
5 *Psammobates*	12-**14.4**	10.2-12.4	8.9-11.3
6 *Testudo*	11.8-12.4	8.7-9.8	10.2
7 *Geochelone*	**8.9**-9.6	**2.7**-4.4	**2.7**-4.2
Chrysemys scripta#	-		
*Pelomedusa subrufa**	8.9-**9.3**	-	
*Emydura australis**	8.7-9.1	**2.7**	-

An increased similarity, or decreased sequence divergence, between individuals of two suborders (i.e. *P. subrufa* and *E. australis* - Pleurodira with *Geochelone* - Cryptodira) is highly unlikely and suggests that the informative third codon position may be saturated thus giving the impression of close relationships. For this reason the data was analysed to see which codon position was more phylogenetically informative (Fig. 6.5).

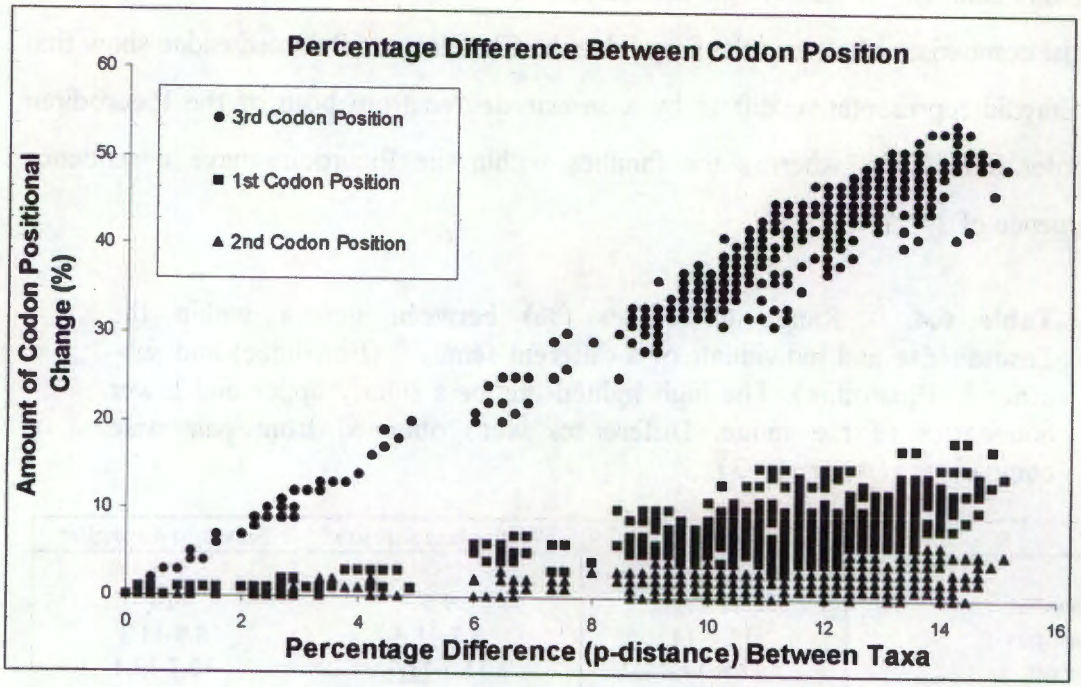


Figure 6.5. - Percentages of 1st, 2nd and 3rd codon position changes for all pair-wise comparisons for all sequenced taxa.

Independent analysis of the three codon positions depicts the third position as the most informative. But as the taxa become more divergent (i.e. increase in p-distance), the amount of positional change increases, becoming stochastic. This increased stochasticity eventually leads to saturation by which point the third position becomes non-informative (Brown *et al.* 1979). The other two positions are not as informative as the first codon position. Since this doesn't explain the close relationships between the two suborders (Cryptodiras and Pleurodirans), a possible explanation could be that the samples may have been contaminated, and thus were subsequently omitted from phylogenetic analysis.

6.2. Phylogenetic Analysis

Phylogenetic analysis of the data is subdivided into three sections. The first concentrates on resolving the *Homopus areolatus* phylogeny using *H. femoralis* as the outgroup, utilising all 450 bp (Fig. 6.6a-c). A second analysis incorporates the findings from the previous phylogeny with sequenced (450 bp) Afro-Testudo representatives of each available genus, species and subspecies (Fig. 6.7a-c). An Emydid, *Chrysemys*

scripta elegans was used as an outgroup based on previous *a priori* assumptions (Fig. 1.6 and 5.10). The final comparison (Fig. 6.8a-c) incorporates sequence data from a previous analysis by Lamb and Lydeard (1994) of North American Gopher tortoises, again using *C. s. elegans* as the outgroup. The maximum sequence overlap for the 34 individuals of the third study was 324 bp.

6.2.1. *Homopus areolatus*

The phylogenetic analysis of the *Homopus areolatus* species was based on the *a priori* assumption that *H. femoralis* and *H. areolatus* are sister species (see Fig. 1.7), and utilised 18 phylogenetically informative characters found within this group. Both phylogenies constructed by parsimony and phenetic methods show substructure within the *H. areolatus* clade. In the neighbor joining analysis (Fig. 6.6a), three distinct groupings can be seen: I (one individual from Elandsberg, Matjiesfontein and an unknown origin), II (an individual from the Driefontein region), and III (2 individuals from Hermanus, 2 individuals from Kruispad, one from De Hoop and Matjiesfontein). A similar organisation and separation of the three groups was noticed when other outgroups (i.e. *Chersina angulata, Geochelone pardalis* and *Kinixys belliana*) were chosen. When parsimony analysis was conducted on this clade, the substructuring of the three groups was not as distinct as with the previous analysis. Only group I came out as an isolated cluster whereas the individual from Driefontein was grouped with the remaining tortoises. When the other outgroups were applied to the analysis, the three groupings became evident once again. The weak bootstrap values for both neighbor joining and parsimony (Fig 6.6c) are possibly a reflection of groups II and III being incorporated as one in the parsimony analysis. The strongest bootstrap support was observed for group I as its isolation from the main group was consistent in both methods. Based on the long branch length and the results from when the other outgroups were used, the *H. areolatus* specimen from Driefontein may potentially represent a third grouping. It was therefore incorporated into the following analyses along with a randomly chosen individual representing groups I and III.

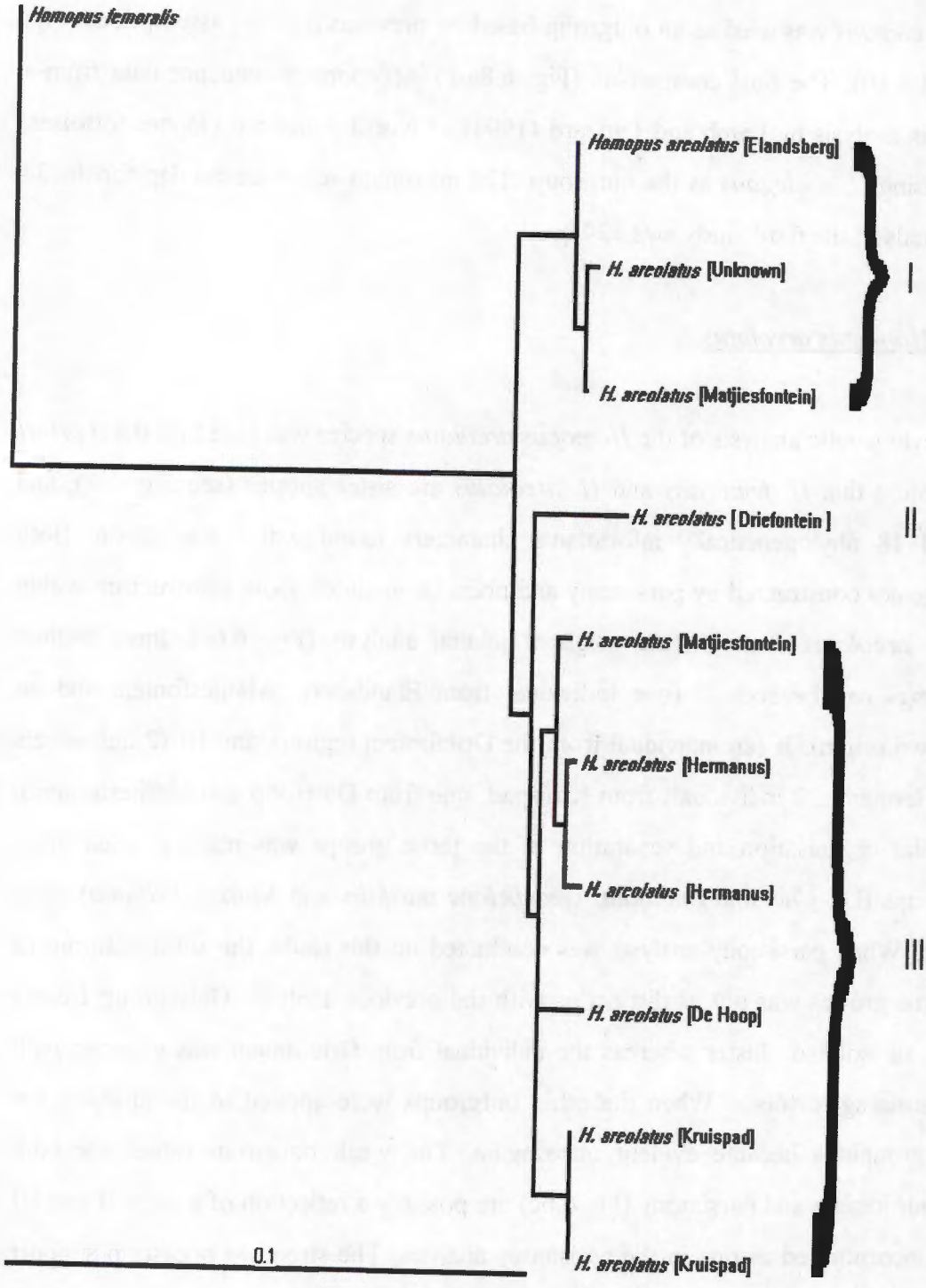


Figure 6.6a - Neighbor joining tree for the *Homopus areolatus* clade

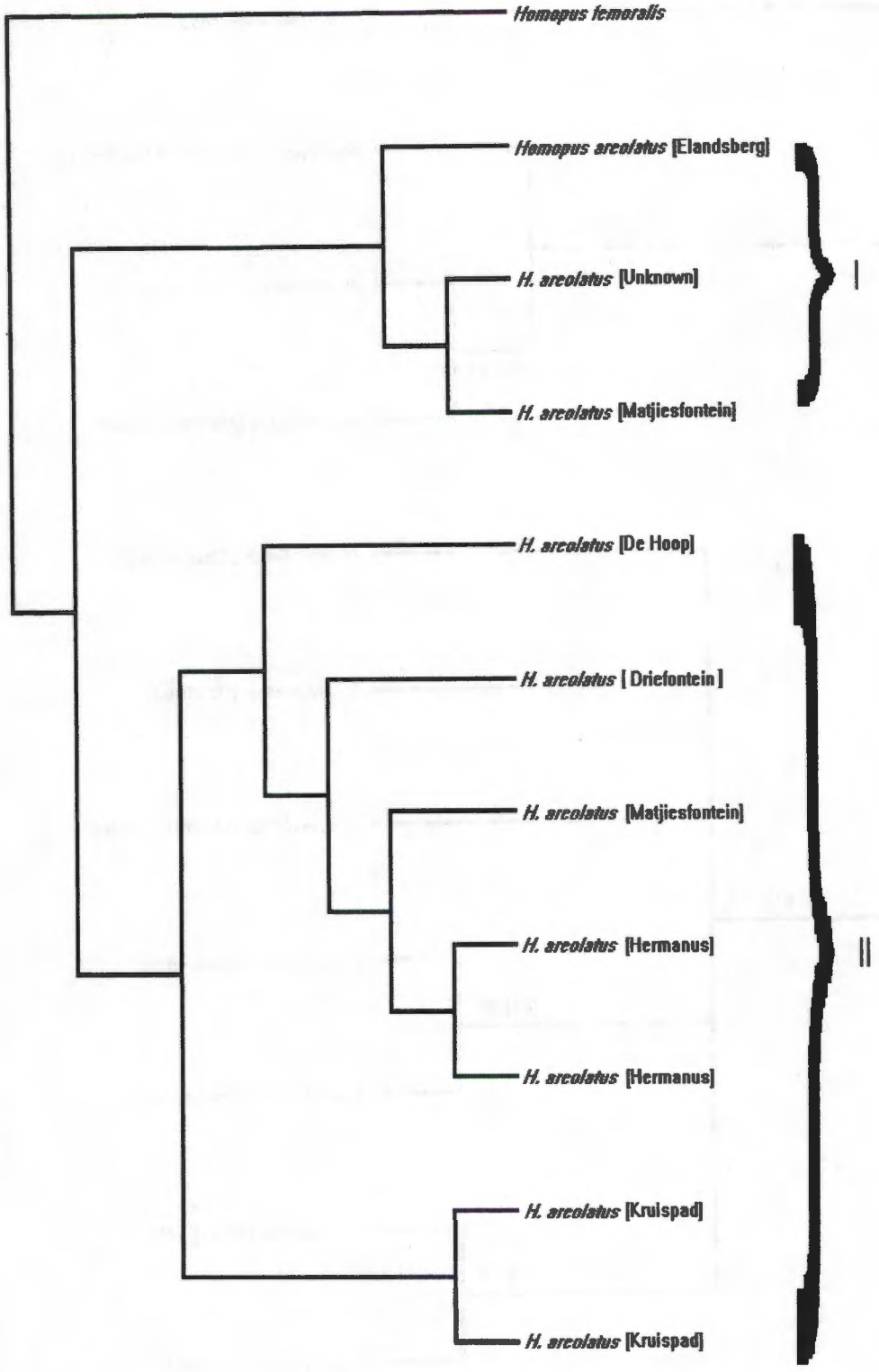


Figure 6.6b - Parsimony consensus tree for the *Homopus areolatus* clade of 3 trees. Length = 30, ci = 63, ri = 73

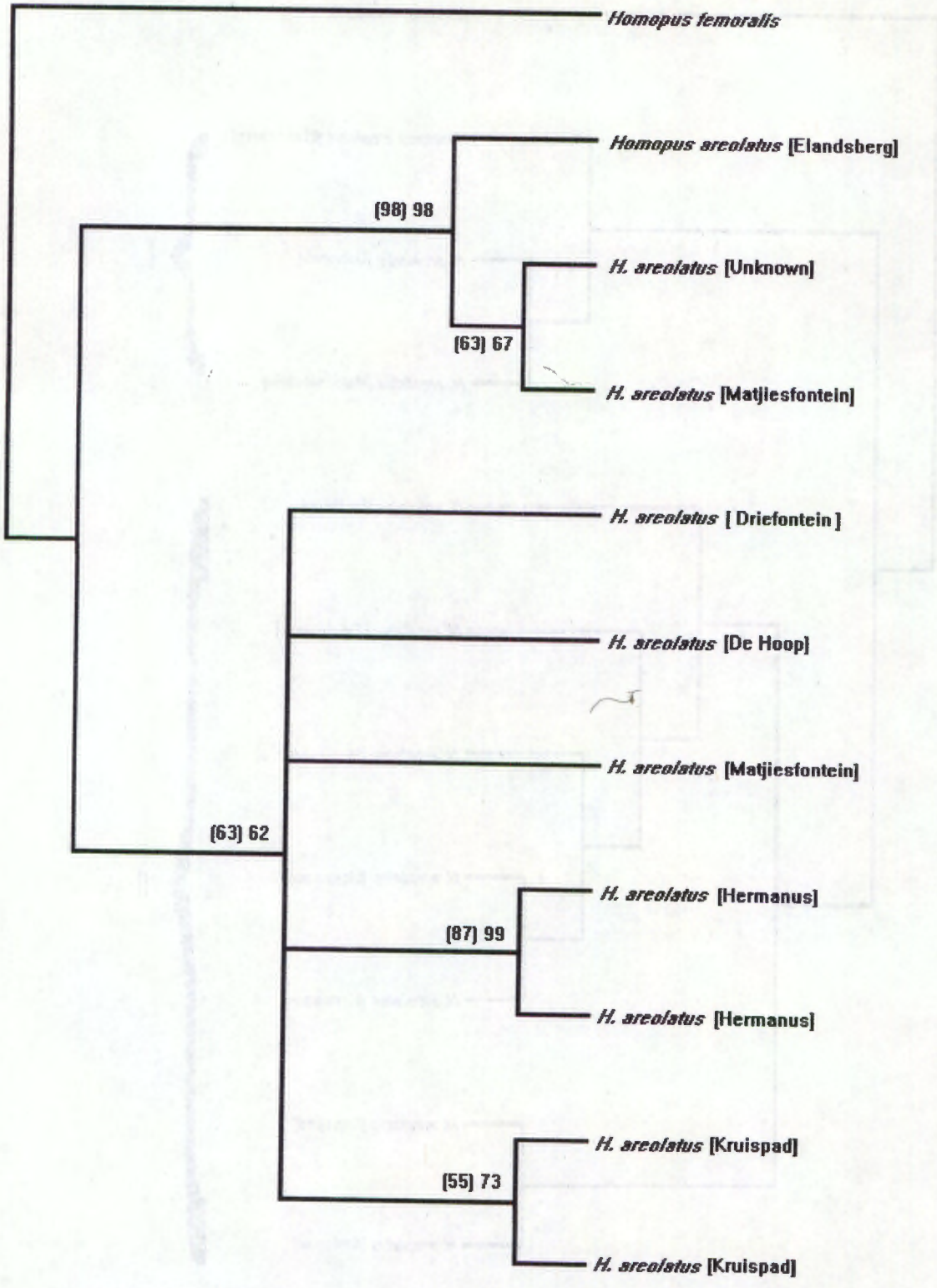


Figure 6.6c - Parsimony and neighbor joining tree using 1000 bootstrap replicates for the *Homopus areolatus* clade with bootstrap values for neighbor joining (unbracketed) and for parsimony (bracketed).

6.2.2. South African Testudines

Phylogenetic analysis comparing all southern African Testudinid representatives, rooted with *Chrysemys scripta elegans,* contained 118 phylogenetically informative characters, and placed the *Geochelone* representatives as the most deeply diverging group relative to the remaining taxa (Fig 6.7a-c).

All of the clades, with the exception of *Homopus* which showed unexpected paraphyly, clustered together with high bootstrap values. The sister species relationship thought to have been between *H. femoralis* and *H. areolatus* is only seen in the neighbor joining analysis, but due to the short branch length, becomes unresolved when bootstrapped. The only consistency observed within the *Homopus* genus is that *H. signatus* and its subspecies together with *H. boulengeri*, clustered together with *Chersina angulata* as the sister group for that particular clade.

The *Psammobates* clade was observed to be phylogenetically robust, with good bootstrap support for the topology. Within the *Psammobates* clade, *tentorious* was placed as a sister group to *oculifer* and *geometricus*. The monophylicity of subspecies within *tentorious* showed a relatively strong bootstrap estimate with phenetic analysis (86), when compared to a value of 55 using cladistic methods.

An unexpected result occurred when *Malacochersus* was observed to be included within the *Kinixys* clade (Fig. 6.7a-c). Support for a monophyletic *Malacochersus/Kinixys* clade was high and it appeared to be the sister group to *K. lobatsiana* with weaker support.

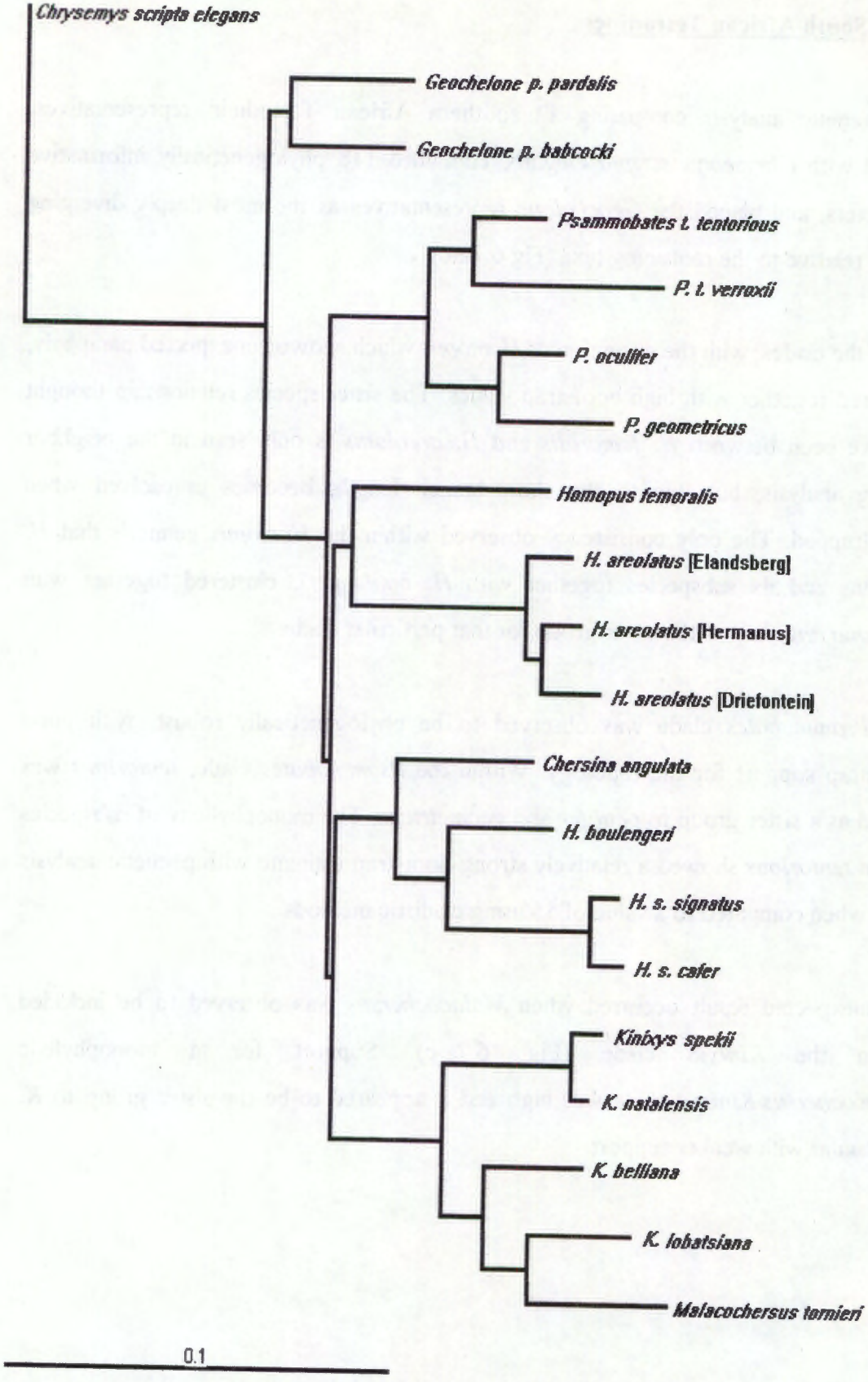


Figure 6.7a - Neighbor joining tree for the southern African Testudinidae

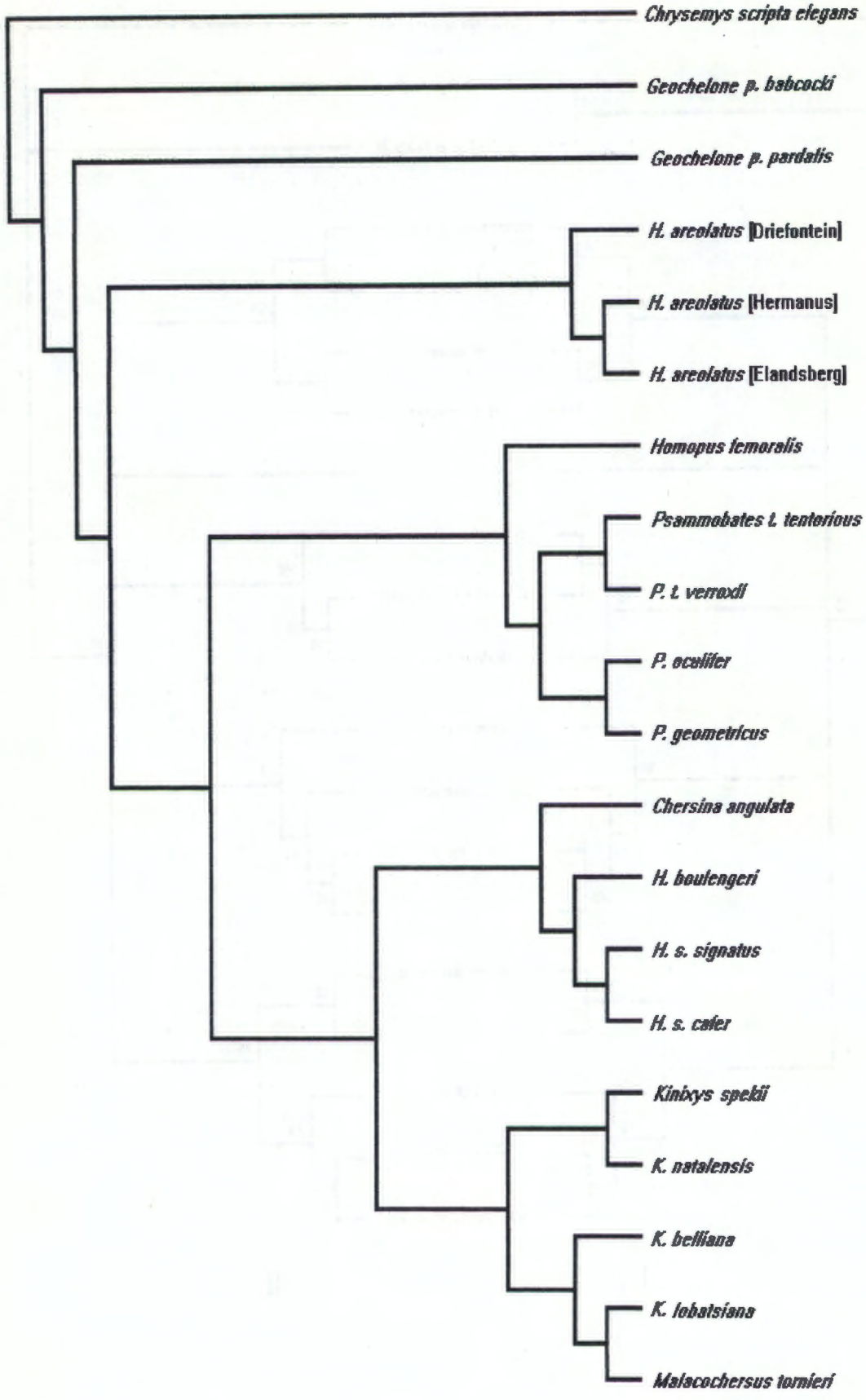


Figure 6.7b - Parsimony consensus tree of two most parsimonious trees, for the southern African Testudinidae. Length = 333, ci = 42, ri = 61

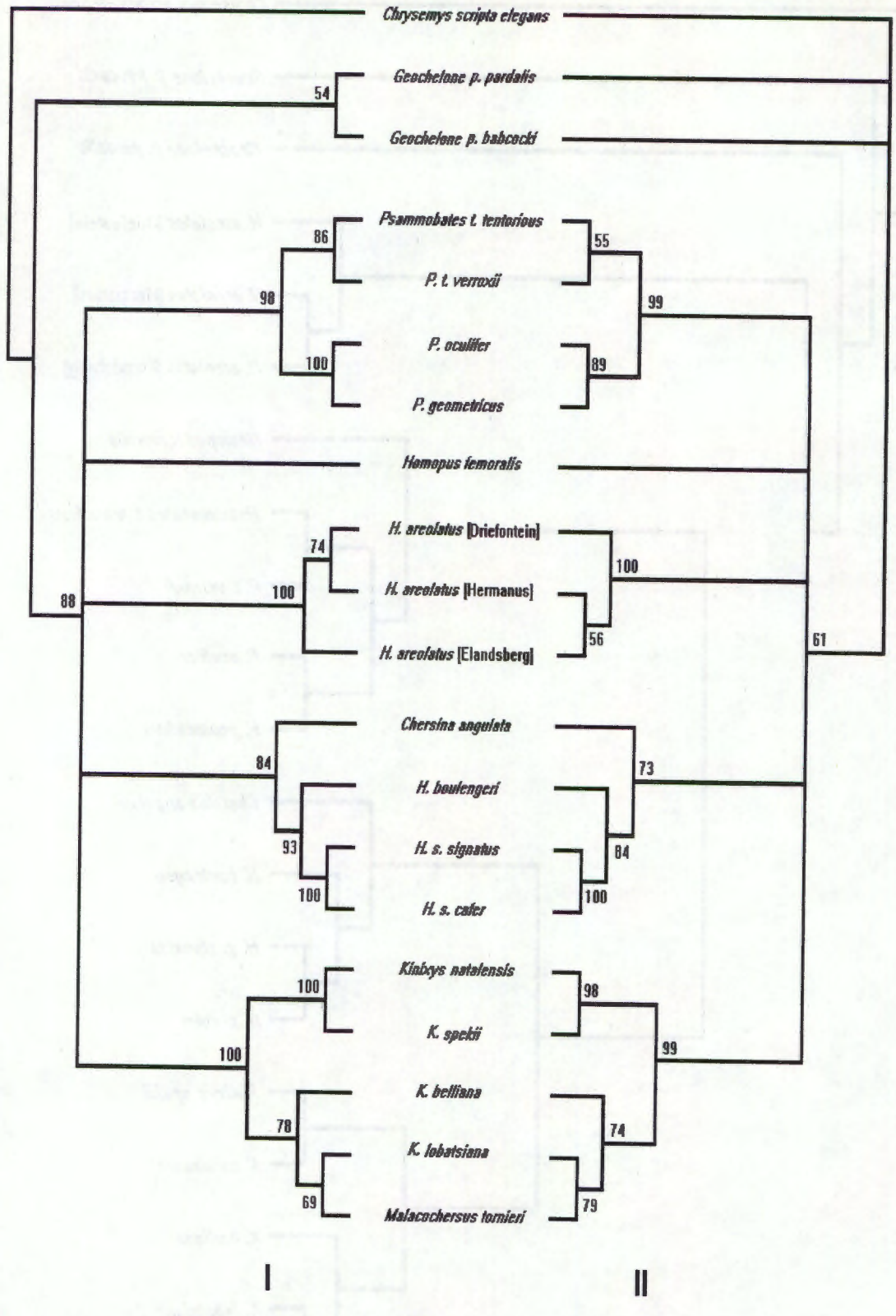


Figure 6.7c - Neighbor joining (I) and parsimony (II) trees with bootstrap values (1000 replicates) for the southern African Testudinidae

6.2.3. Phylogenetic Analysis Of All Availably Sequenced Testudines

North American gopher tortoise sequences, obtained from Lamb and Lydeard (1994) were compared to a 324 bp overlapping sequence of southern African Testudines to establish the relationship between these two groups (Fig 6.8a-c). Rooting the phylogeny with *C. s. elegans*, consistently showed the Burmese mountain tortoise (*Manouria emys*) as being basal to the robust *Gopherus* clade. The addition of three *Testudo* representatives showed that this genus was not monophyletic. *Testudo horsefieldi* and *T. marginata* clustered together whereas *T. hermanni* was found deep in the southern African clade. This observed paraphyly must be considered with caution since *T. hermanni* may have been incorrectly identified at the Tygerberg Zoo.

Paraphyly was also seen within *Geochelone*. Both African representatives appear generically and specifically paraphyletic, with *G. p. babcocki* clustering with *Psammobates* and *G. p. pardalis* grouping with the *H. areolatus* clade (Fig. 6.8 a & b). Poor support resulted in *Geochelone* being unresolved after bootstrapping.

Similar generic structuring, as was seen in the previous results (6.2.2.), was also observed. Subtle differences can be seen in the *H. areolatus*, and the *Kinixys* clades, between the two bootstrapped phylogenies

Comparing the phylogenies generated by the distance method, it can be seen that the branch lengths for the larger Testudinidae family phylogeny(Fig. 6.8a) are ten times longer than for either of the phylogenies (Figs. 6.6a & 6.7a) generated for the southern African Testudinidae. This difference in branch length may suggest that this comparison (the larger Testudinidae family - Fig 6.8a) may be beyond the utility of the cytochrome *b* gene to adequately resolve the phylogenetic relationships. However, since there were 9 out of the 12 (75%) Testudine representatives with overlapping sequence available a phylogeny was inferred to identify any possible generic structure and compare it to the existing morphological data.

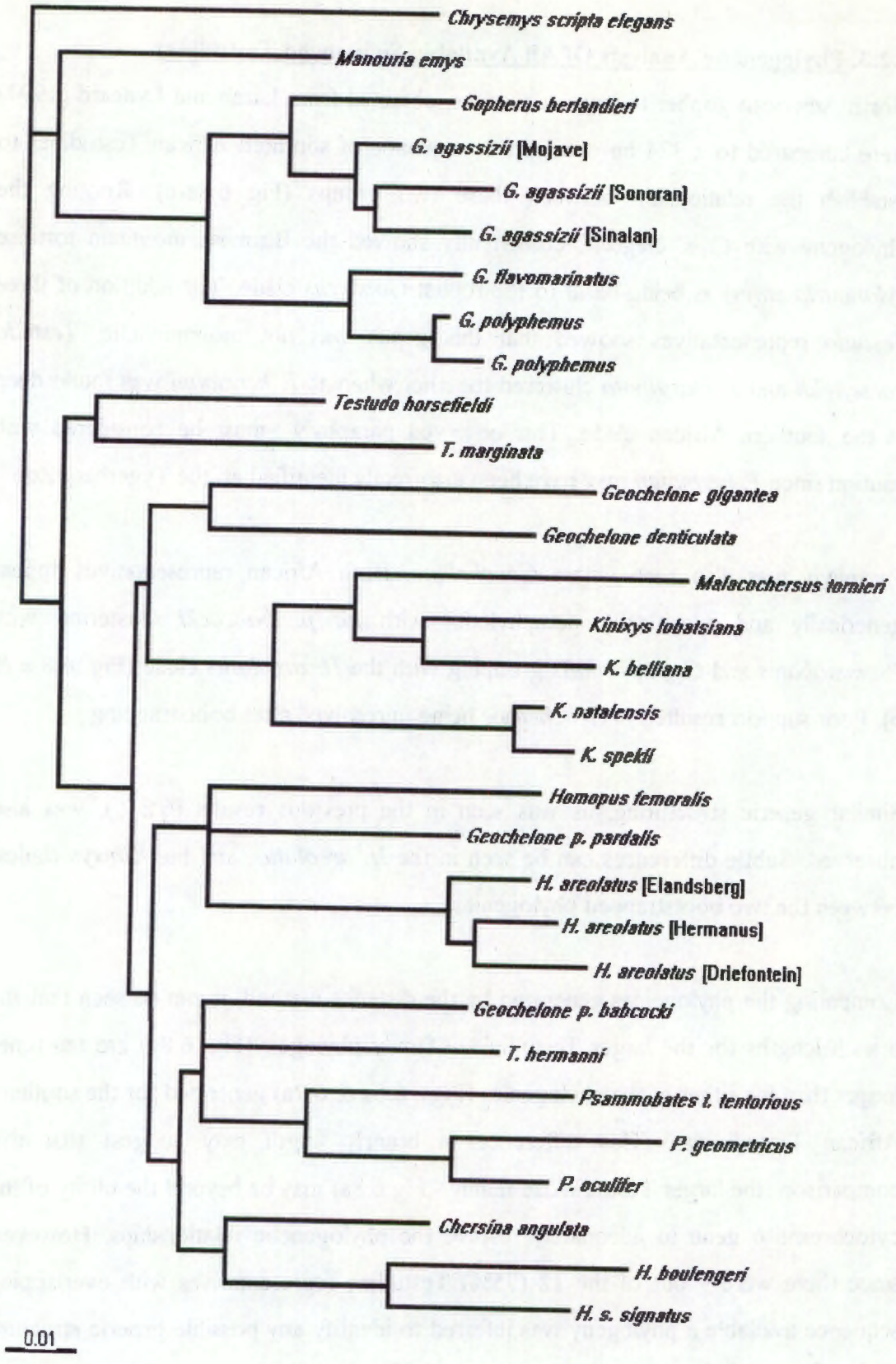


Figure 6.8a - Neighbor joining tree for all members within the Testudinidae family for which cytochrome *b* sequences are available.

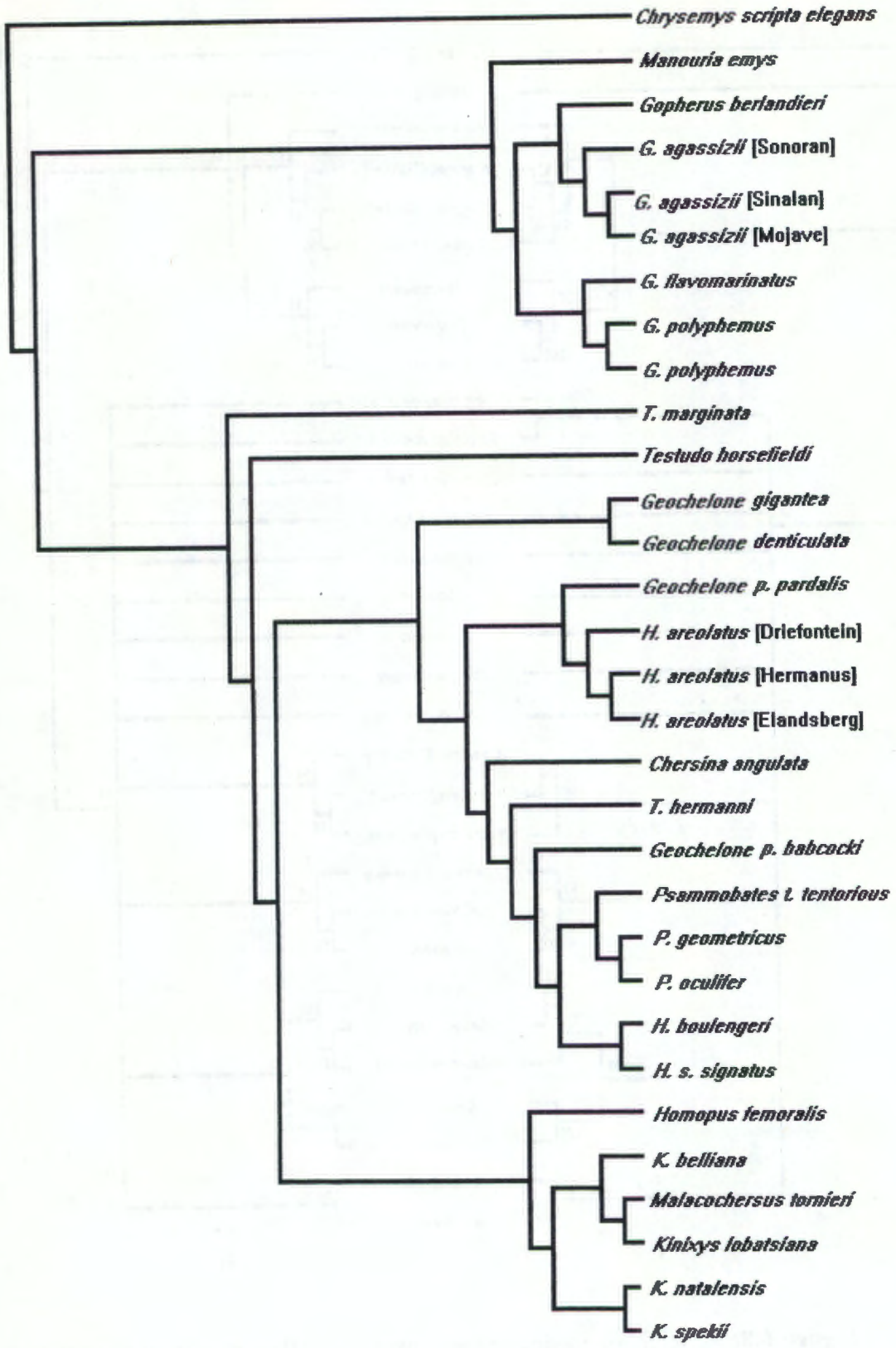


Figure 6.8b - Parsimony consensus tree of the six most parsimonious trees, for all members within the Testudinidae family which had cytochrome *b* sequences available. Length = 369, ci = 33, ri = 59

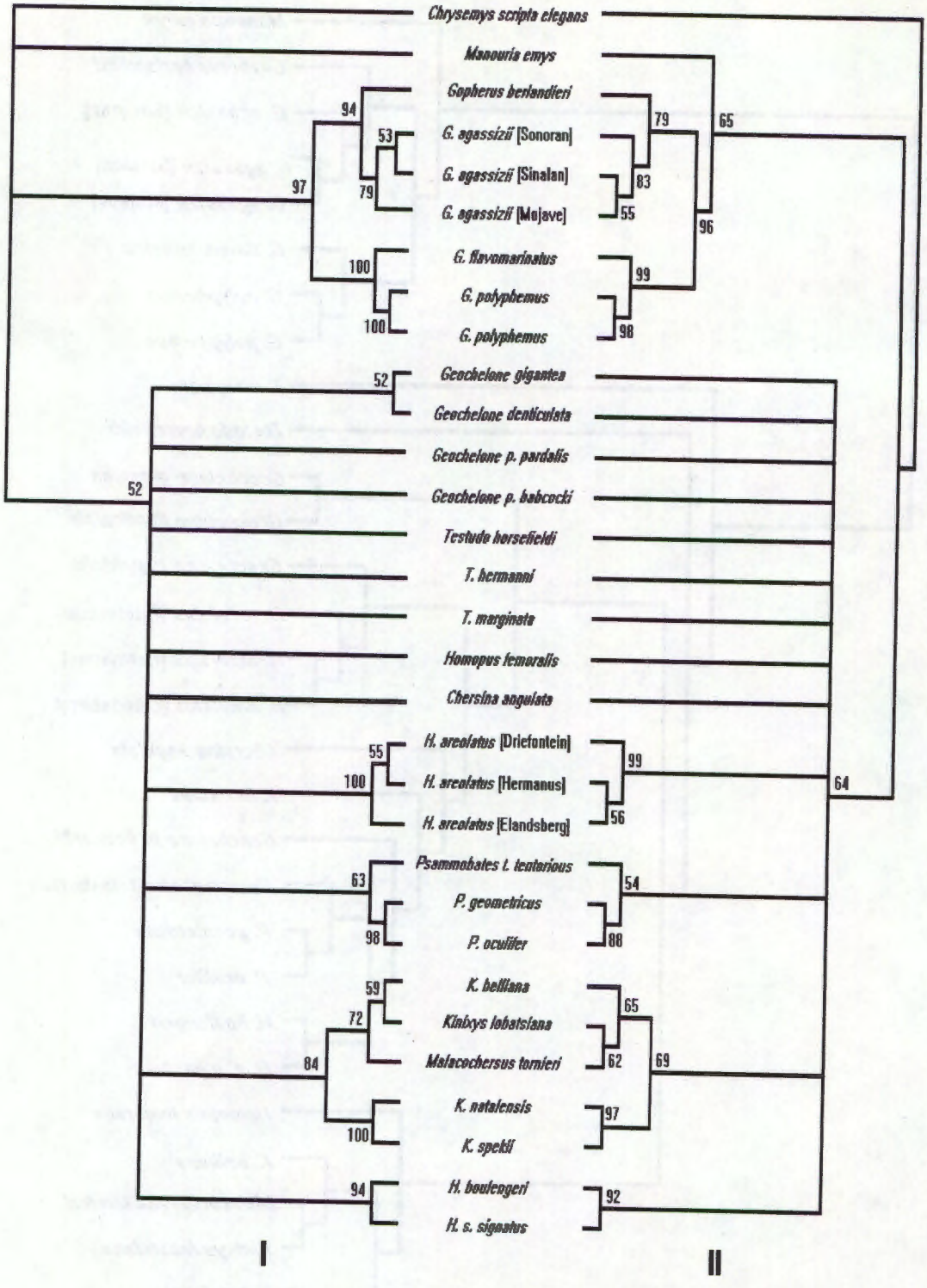


Figure 6.8c - Neighbor joining (I) and parsimony (II) trees with bootstrap values (1000 replicates) for all members within the Testudinidae for which cytochrome *b* sequences were available

7.0 DISCUSSION

Many authors have addressed the various properties of mtDNA sequencing and its applications in phylogenetic reconstruction (Avise *et al.* 1987, Graybeal 1993, 1994, Hillis 1987, Harrison 1989, Kocher *et al.* 1989, Zhang and Hewitt 1996). Although there are limitations, if properly used, mtDNA sequencing can be a powerful tool for inferring phylogenies, especially in conjunction with other data sets such as morphological data. It is instructive to compare phylogenies created by these different data sets because if they consistently show the same result it means that the tree generated with these data is likely to be the "correct" tree. Incorporation of several data sets, if available, is better than relying on only one. On its own, each data set has its limitations and will have a limited value (Huelsenbeck *et al.* 1996). This will be apparent in the case of the southern African Testudinidae, where no molecular data have been available, and a limited amount of morphological data have been documented.

One possible limitation of DNA sequencing is saturation of the third codon position in protein coding genes. Since there is a very low guanine content at the third position (Fig. 6.2), the majority of sites contain the other three nucleotides (Irwin *et al.* 1991). Therefore, in theory the third codon position should become saturated when a pairwise comparison between two divergent taxa at these positions approaches 66% sequence divergence. The probability of two random, sequences sharing identical nucleotides at the third codon position, is 0.33% under these conditions (Irwin *et al.* 1991). If this amount of divergence is found it may be better to use a different gene which is more conserved.

The origin of the Testudines dates back to the Mesozoic era (65-225 MYA), therefore it was important to determine whether cytochrome *b* fell within phylogeneticaly informative boundaries. Since phylogenetic information obtained from cytochrome *b* has been shown to be informative at certain divergence levels (Ballard and Kreitman 1995, Graybeal 1993, Meyer 1994), an initial "pilot" study was conducted to asses the informativeness of the cytochrome *b* gene within the Testudinidae. The study included

five representatives of the Testudinidae family: *Chersina angulata*, *Homopus femoralis*, *Psammobates geometricus*, *Geochelone pardalis pardalis*, *Testudo marginata*, an Emydid, *Chrysemys scripta elegans*, and one individual from the Carettochelyidae family, *Caretta caretta*. It was found that there was adequate resolution between the various lineages and that cytochrome *b* was an appropriate gene for this study. This was evident from the interspecific comparisons where the individuals within the Testudinidae family showed a sequence variation between 9.1 - 13.5% (refer to Appendix C). When the Emydid and the Carretochelid (not shown) were compared with the Testudinidae representatives, the upper limits of the range increased to 14.2% for the Emydid and 17.9% for the Carretochelid. This increase in sequence divergence was expected, as these individuals are representatives from different families. From this preliminary analysis, it was found that the sequence divergence for cytochrome *b* appeared to be significant enough to resolve the different genera within the Testudinidae

Many methods of phylogenetic reconstruction have been proposed and tested by numerous authors to establish which method generates trees that agree with current evolutionary theories (Cummings *et al.* 1995, Felsenstein 1988, Hillis *et al.* 1994, Hillis 1995, Miyamoto and Fitch 1995, Mooers 1995, Neyman 1971, Nunn 1995, Schluter 1995, Schmidt 1995, Tajima 1993, Yang *et al.* 1994, Yang and Wang 1995). With this debate still continuing, both parsimony (cladistic) and distance (phenetic) based methods were used in this study. In the latter case the neighbor joining method was used with the Kimura two parameter correction for multiple hits. A correlation between cladistic and phenetic techniques can be observed in the bootstrap analysis. Nodes that are strongly supported with neighbor joining show similar values with the bootstrapped parsimony analysis. Differences between the two arise when a node is weakly supported (Fig. 6.7c).

7.1. A Molecular Phylogeny Of The Southern African Tortoises

Relationships of Testudinidae within southern Africa have until now been inferred on the basis of morphological classification only. This discussion will attempt to provide comparisons between the results of molecular sequence data with results from morphological data, and then to present possible scenarios of southern African Cryptodiran evolution.

7.1.1. *Homopus*

The rejection of the paraphyletic status of Homopus (Hewitt 1931, 1937) by Loveridge and Williams (1957) has instigated little, if any debate over the status of *Homopus* for nearly four decades. Strong morphological characters such as the equal number of claws on the anterior and posterior limbs, as well as a small inguinal shield, which clearly identifies *H. femoralis* and *H areolatus*, were rejected by Loveridge and Williams as characters for elevating *H. signatus* and *H. boulengeri* to a separate genus *Chersobius*. The molecular sequencing data are, however, consistent with Hewitt's (1931, 1937) proposed grouping and subsequent separation of *Homopus* into two genera (*Chersobius* and *Homopus*). The phylogenies show strong support for the division of *Homopus* into two different genera. Separation can also be correlated with morphological differences (i.e. number of claws in each species). *H. signatus* and *H. boulengeri* which posses identical claw arrangement (five and four claws on their anterior and posterior limbs respectively, which are different from the equal number found *H. femoralis* and *H. areolatus*) are shown to group together in the molecular analysis with *Chersina angulata* in a separate clade from the other *Homopus* species.

The relatively strong support for *C. angulata* being grouped with *H. boulengeri* and *H. signatus*, in both cladistic and distance analyses, comes as an unexpected result because morphologically *Chersina* and *Homopus* are very different. The former has a highly vaulted, convexed shell, and the latter features a more flattened carapace (Hewitt 1931, 1937, Loveridge and Williams 1957). A pairwise difference of 6.9% indicated that although at the lower range of generic differences (Table 6.3), *Chersina* is more closely related to *Homopus* (*Chersobius*) than to any other of the studied

genera. Biogeographically, *Chersina* has a predominantly allopatric distribution, in terms of habitat, from *H. boulengeri* and *H. femoralis* which occur in the dryer regions of central South Africa (Greig and Burdett 1976). On the southern coasts of South Africa, where the rainfall is 600 to 700 mm per annum, *Chersina* is largely sympatric with respect to *H. areolatus and H. signatus.* More investigation is needed, such as increasing the sample size, using representative geographic sampling, and analysing another gene before an evolutionary relationship between *Chersina* and *Homopus* (*Chersobius*) is confirmed.

Another unexpected finding within *Homopus*, is the isolation *of H. femoralis*. Standing on its own, *H. femoralis* remains independent from other clades in the bootstrapped phylogenies. The southern African Testudine phylogram (Fig. 6.7a) however is the only phylogeny that depicts *H. femoralis* basal to the *H. areolatus* clade, but since the branch length is very short, this resolution is lost in the bootstrap analysis (Fig. 6.7c). Parsimony differed from neighbor joining in that *H.femoralis* was placed basal to the *H. areolatus* and *Psammobates* clades (Fig. 6.7 a&b). A weak bootstrap value of less than 50 for the parsimony topology resulted in the subsequent collapse and lack of resolution after bootstrap analysis (Fig. 6.7c).

7.1.1.1. ***Homopus areolatus***

In order to investigate the amount of intraspecific sequence variation within the family Testudinidae for cytochrome *b*, a number of individuals of one species were chosen. This intraspecific study which involved ten *H. areolatus* individuals, revealed substantial substructure within a highly supported monophyletic clade. This substructure is also evident in the pairwise comparisons between individuals of this species (Table 6.1). A group of three tortoises representing separate populations (Elandsberg Nature Reserve, near Wellington; Farm: Matjiesfontein, near Sutherland; Unknown: Tygerberg Zoo) shows the least sequence divergence out of the entire study sample of *H. areolatus*. The close relationship between these three representatives, along with a long branch length supporting this node, is further supported by the high bootstrap values (Fig. 6.6c). Prior to bootstrap analysis a larger clade consisting of six tortoises representing four distinct localities could be seen in the neighbor joining

analysis (Fig. 6.6a). However a clade of seven tortoises was observed in the parsimony analysis (Fig. 6.6b). It is interesting that the cladistic method placed the individual from Driefontein (near Sutherland) between tortoises from De Hoop and Matjiesfontein, even though the sequence divergence is greater between De Hoop - Driefontein (2.7%) and Matjiesfontein - Driefontein (2.2%) than between De Hoop and Matjiesfontein (1.1%) (see Appendix C). This grouping of the Driefontein individual with the six remaining tortoises, using parsimony analysis, was unique only to this phylogeny, when *H. femoralis* was the specified outgroup. When other outgroups were used (i.e. *Chersina, Geochelone* and *Kinixys*) the three groupings were very distinct. The only difference being the order of the groups.

Sequence variation within this clade may be directly related to the number of autapomorphies rather than the number of shared derived characters, as can be seen in the collapsed branches subsequent to bootstrap analysis. The short branch lengths in the phylogram (Fig. 6.6a), resulted in the tree losing much of its topology after bootstrapping, indicating recently diverged individuals.

It is quite surprising to see such a range in sequence divergence between individuals of the same monotypic species. Although no molecular work has previously been done on southern African tortoises, and very little sequencing of the cytochrome *b* gene has been conducted on other members of the family Testudinidae, it is difficult to predict what the expected range of sequence variance should be. The reason for this is that each gene and each species (or genus, or family) has a different rate of evolutionary change, and correlation by extrapolation can result in erroneous results (Adachi *et al.* 1993, Ballard and Kreitman 1995, Graybeal 1993, 1994, Meyer 1994). Sequence divergences between two individuals of the same species from both the Testudinidae and Emydidae families have shown low variation (see Table 6.1), suggesting that intraspecific variation for this gene is generally less than 1%. The data for *H. areolatus* show that, in this species, intraspecific differences go up to 3.6% , which is an even greater value than the intergeneric differences between *Kinixys* and *Geochelone*. This suggests that *H. areolatus* may contain three possible cryptic species.

From the phylogram (Fig. 6.6a), the long branch length connecting the *H. areolatus* clade to *H. femoralis* suggests that the species has had a relatively long independent evolution. The shorter branch lengths at the terminal ends of the clade indicate recent mutational events, resulting in the observed sub-structure. Expectations of relating the generated phylogram to biogeographic population patterns were thwarted by the relatively large sequence divergence (2.4%) observed between two individuals obtained from the same locality at Matjeisfontein farm in the Sutherland region (Ha MT Female). There are four possible explanations for this large intra-population difference. Since tortoises are easily transported by human intervention, individuals perceived to be from a particular geographic locality might in fact have been translocated from another area (Branch *et al.* 1995, Greig and Burdett 1976). This could certainly explain the differences between the two Matjeisfontein farm individuals, which were collected within meters of one another (DeVilliers and Baard pers. comm.).

Another possibility for the non-correlating regional sequence divergence between individuals may be that tortoises travelled across geographical and vegetative barriers, deserting their native population, and invading another separate and distinct population. Although not much is known about the migratory behaviour, it seems highly unlikely that *H. areolatus*, which favours low lying moist coastal regions as well as low foothills in the Cape Fold Mountains, would be able to tolerate habitats to which it is unaccustomed (Boycott and Bourquin 1988, Boycott 1986, Loveridge and Williams 1957). The other possibility is that in any large stable interbreeding population mitochondrial DNA polymorphisms will accumulate and may remain as stable morphs with fluctuating frequency in the population which take a long time before genetic drift results in their loss (Benirschke 1985). Therefore morphs with a long period of evolutionary separation can thus coexist. Bottle necks will promote their loss, so the existence of many divergent morphs, as here, in a population, if not due to man-made translocations, imply a long lasting panmictic population of a large and stable size.

7.1.2. *Kinixys*

A physical characteristic found in *Kinixys*, is its posterior carapacial hinge (Broadley 1981, 1989, 1992, 1993, Greig and Burdett 1976, Hewitt 1931, Loveridge and Williams 1957). As one would expect, such a feature would on morphological grounds, render *Kinixys* monophyletic (Hewitt 1931). It was therefore an unexpected observation to find *Malacochersus* deep within the *Kinixys* clade. The sequencing data and resulting phylogenies show strong bootstrap support (Fig. 6.7c) for the inclusion of *Malacochersus,* making *Kinixys* paraphyletic. Pairwise sequence comparisons also show that *Malacochersus* shares most of its sequence similarity with *Kinixys.* Although the molecular data is conclusive in showing that *Malacochersus* shares close affinities with *Kinixys*, it would be useful to search for additional confirmatory evidence for example, ontogenetic studies.

Malacochersus has previously been designated a separate clade on morphological grounds (Gaffney and Meylan 1988, Loveridge and Williams 1957) but the present study indicates that it belongs within the *Kinixys* clade. In this regard, it is interesting to note that the sister species to *Malacochersus* (*K. lobatsiana*) has been reported as having the least vaulted carapace (flat appearance) of the genus. The most plausible scenario is that the hinge of *Malacochersus* has been lost secondarily, as an adaptive response to living in crevices.

The topology of this paraphyletic clade is also robust, indicated by the high bootstrap support. The molecular and morphological (Broadley 1992) phylogenies, excluding *Malacochersus,* agree as to the grouping of the species. Morphological data suggest that *K. spekii* is a sister species to both *K. lobatsiana* and *K. belliana* and that *K. natalensis* is basal to the whole clade, as it's physical characteristics are the most primitive. Resolution of the evolutionary order of these tortoises using molecular sequencing is weakly supported and would be difficult to address since there are two unique clades. However the long branch length observed in the phylogram (Fig. 6.7a), does suggest that all the species within the clade containing *Malacochersus,* have diverged a long time ago and evolved independently. A recent divergence seen in the *K. spekii- K. natalensis* cluster is depicted by short branch lengths.

It is interesting to note that Broadley (1981) reinstated *K. natalensis* Hewitt to a full species after Loveridge and Williams (1957) grouped all previously described species and subspecies together as *K. belliana.* Eleven years later, Broadley (1992) elevated *K. b. spekii* to specific level and also reinstated *K. lobatsiana* Power as full species. A sequence divergence estimate of 1.1% between *K. natalensis* and *K. spekii* is considerably lower than the differences between other intra-generic comparisons (see Table 6.2), and falls within subspecies limits (see Table 6.1). This low difference suggests that designation of both *K. spekii* and *K. natalensis* as separate species should be considered more closely.

7.1.3. *Psammobates*

Support for monophyly of the *Psammobates* corroborates previous morphological findings (Baard and Mouton 1993, Crumly 1984, Gaffney and Meylan 1988 Greig and Burdett 1976, Loveridge and Williams 1957). The organisation of *Psammobates* into two separate groups has been suggested by previous authors (Baard and Mouton 1993, Duerden 1907, Greig and Burdett 1976). From the morphological characters, *P. geometricus* and *P. oculifer* form one group, while the *P. tentorius* complex constitutes the other. Sequencing results are consistent with these associations, showing strong topological support within the monophyletic clade.

The two sister species, *P. geometricus* and *P. oculifer* are geographically separated by *P. tentorius*, which has an area of sympatry with the latter species. Allopatric speciation has been invoked to as one of the explanations for *P. geometricus* and *P. oculifer* being sister species. Baard (1993) suggested that *P. tentorius* might have taken over the area left void by a receding *geometricus-oculifer* ancestral species.

It is also plausible that an ancestral *P. tentorius* may have once occupied a large area as a single species population. With the subsequent fluctuations of sea levels (Miller and Fairbanks 1985) and the gathering influence of the Benguela current, peripheral isolates of the ancestral population may have formed. It is thought that coastal temperatures of western South Africa would have been stabilised, as a result of the

cold Benguela current, generating increased fog and subsequent drier areas than at present (Mouton and Oelofsen 1988). Adaptation to these resulting adverse coastal climatic conditions may have instigated enough change to render a separate species (*geometricus*). Similarly, *oculifer* may have adapted to its environmental conditions of higher rainfall in south-central Africa as opposed to *tentorius* which prefers dry arid environments (Greig and Burdett 1976).

7.1.4. *Geochelone*

There is little doubt that *Geochelone* diverged from the ancestral stock of Testudinidae relatively early in time because of its many primitive characteristics (Crumly 1982, Crumly 1984, Gaffney and Meylan 1988). Sequence data are consistent with the morphological data and places the genus as a basal cluster with respect to the other tortoises in the study. The placement of the southern African representatives of the *Geochelone* subgenus has been controversial in terms of morphological designation and subsequent species recognition. The data here support the hypothesis that *Geochelone pardalis* consists of two separate subspecies: *G. p. pardalis* (found only in the Western Cape and Namibia) and *G. p. babcocki* (found in south-eastern Africa) (Loveridge 1935). Positive identification of both subspecific individuals used in this study was made by D. G. Broadley and J. D. Visser, respectively (pers. comm.). A third individual (*G. p. pardalis* 2) was acquired as a result of a road fatality in the Western Cape Province. While it was impossible to identify the specimen to subspecific level, sequence divergence comparisons between the unknown *Geochelone* and the two positively identified individuals, indicated that the unknown one was more closely related to *G. p. pardalis.* A pairwise sequence homology of 98%, supported this association, but on the other hand, such a sequence divergence (2%) is high for two individuals of the same subspecies (see Table 6.1). However, since *Geochelone* is thought to have diverged early (Eocene) from the ancestral stock of Testudinidae (Crumly 1982), it is plausible that after a rapid speciation event (Carroll 1988, Crumly 1984), and having evolved independently, few morphological changes have occurred along the *G. p. pardalis* lineage. The high value of 4.7 - 6.7% for the *babcocki/pardalis* sequence divergence (Table 6.1) would be much more consistent

with an inter-specific relationship, which would have important implications for any captive breeding programmes for *Geochelone pardalis*, therefore more individuals of *G. pardalis* should be examined before this claim can be substantiated.

7.2. Inter-familial Relationships

Previous morphological comparisons of representative families within the Testudinidae, conducted by Gaffney and Meylan (1988), generally agreed with Crumly's findings (1984) (see Fig. 1.7). Recent molecular sequence data, which focused on the phylogeny of gopher tortoises (Lamb and Lydeard 1994), were used together with the representatives sequenced in this study. This allowed for a molecular comparison involving similar taxonomic representatives as in previous morphological analyses. With the exception of *Pyxis*, *Indotestudo*, and *Impregnochelys*, for which sequence data were unavailable, 324 base pairs for each generic representative of the Testudinidae were used phylogenetically.

Results are consistent with previous hypotheses (Carroll 1988, Crumly 1984) and indicate that there may have been a rapid period of speciation, followed by a period of adaptive morphological radiation. The phylogram (Fig. 6.8a) depicts many of the generic divergences (short branch lengths) occurring early in time. The proposed timing of these events is thought to have occurred during the Eocene (Carroll 1988, Crumly 1984). Following speciation, the generic representatives seem to have evolved independently over a long period of time, which can be observed in the long branch lengths. After bootstrap analysis, the shorter branches became unresolved, although, some of the clades were still highly supported. Similar results were obtained for parsimony bootstrap analysis.

Comparisons between the resulting molecular phylogeny and the phylogeny obtained using morphological data were difficult to make, in that the molecular topology was very different from the morphological one. The two phylogenies agree only on a few taxonomic arrangements, one of which is the recent divergence of *Homopus* (*Chersobius*) and *Psammobates*. The other is the basal placement of *Gopherus*.

The poor resolution within the Testudinidae has previously been attributed to the lack of good resolution when using cytochrome *b* sequences (Lamb and Lydeard 1994). Possible ways to overcome the lack of resolution would be to use more sequencing data (i.e. longer sequence fragments), or alternatively to sequence genes which are known to evolve slowly (Graybeal 1993, 1994, Lamb and Lydeard 1994).

7.3. The Utility of Cytochrome *b* and Its Evolutionary Rate *in* Southern African Tortoises

When the rate of evolution is appropriate for the desired divergence times, the data can give good resolution. If the characters evolve at slower or faster rates then anticipated, they may respectively provide an insufficient number of phylogenetically informative sites or contain too much homoplasy as a consequence of saturation, especially at third codon positions (Brown *et al.* 1979, Graybeal 1994). To test the efficiency of cytochrome *b*, the amount of change per codon position was plotted against the p-distance (Fig. 6.5). Most of the changes occurred at the third codon position, as was observed in figure 6.2. With increasing p-distance, the taxa became more divergent and less related. Extensive homoplasy resulting from saturation at third codon positions could explain the poor resolution observed in the deeply diverged taxa. Close affinities observed in the pairwise comparisons between *Geochelone* and the outgroups could be a result of such a homoplasic effect. Comparisons between closely related taxa also show that transitions occur more frequently than do transversions (Fig. 6.3). For pairwise comparisons between more divergent taxa, this effect is less noticeable, and would imply multiple changes at these sites. This is consistent with the work of Graybeal (1993).

The slow evolutionary rate of mtDNA in turtles, proposed by Bowen *et al.* (1989), has received support from different RFLP assays using various representatives within the Cryptodiran suborder (Avise *et al.* 1992, Bowen *et al.* 1994). Additional evidence from comparisons of cytochrome *b* in the Testudinidae (Lamb and Lydeard 1994), Emydidae (Lamb et al. 1994), and marine turtles (Bowen *et al.* 1993), has shown that the rate of evolution for Cryptodirans is approximately 0.2 - 0.4% per MY; which is seven times slower than the proposed rate of 2.0% per MY for mammals (Brown *et al.*

1979). Previous calibrations have been made using convincing fossil evidence as well as accurately dated geological events. The paucity of fossil evidence for African Cryptodira (Auffenberg 1974, Crumly 1984, Meylan and Auffenberg 1986), makes calibration of the molecular clock difficult, and as a result, unreliable. Assuming the clock is evolving at the same rate in African Cryptodirans as in the Testudine taxa in previous studies, and that the rate of evolution is linear up until 20 MYA (Bowen *et al.* 1993), the phylogenetic relationships of these taxa, with respective inter-generic divergences of less than 8% can be evaluated reliably using cytochrome *b*. Of the seven genera sequenced in this study, only *Kinixys* and *Geochelone* were able to be calibrated tentatively. Taking the upper divergence limit for *Kinixys* and dividing it by the proposed evolutionary rate of 0.4% substitutions per MY (Avise *et al.* 1992), it can be calculated that this genus dates back at least 19 million years. This result confirms the generic designation in the *Kinixys erosa* fossil (see Table 1.2), which had been dated to the Miocene epoch (Auffenberg 1974). The *Geochelone* calibration dates back 17 million years (mid-Miocene) suggesting that the ancestor of the present day *Geochelone* may be a representative of the ancient *G. namaquensis* or *G. stromeri pardalis* (see Table 1.2). The *G. pardalis* fossil which was dated to the mid-Pleistocene (Broadley 1962), may be a recent representative of this ancient lineage.

8.0 Conclusion

PCR-based direct sequencing has shown to be a valuable molecular technique for inferring phyolgenies, identifying unknown taxa, and assisting in calibration of fossil data.

When the molecular results were compared with the morphological findings, only *Psammobates* was consistent with previous morphological studies. Other proposed relationships based on morphology were not supported using molecular data. There was strong support for the inclusion of *Malacochersus* into the *Kinixys* clade, as well as the close association of *Chersina* with the *Chersobius* clade. The molecular data agree with Hewitt's (1937) proposal of separating *Homopus* into two genera, thereby supporting the use of *Chersobius*. *Homopus* individuals with five claws on their posterior and four on their anterior limbs (*Chersobius*) were found to cluster with *Chersina*, from which *H. boulengeri* and *H. signatus* seem to be much more distantly related. The true *Homopus*, with an identical claw assemblage, appeared paraphyletic and clustered independently of each other.

Within *H. areolatus*, population substructure was also observed. Sequencing ten individuals from different populations showed a considerable and unexpected degree of sequence variation. Positive identification of all of these individuals placed them unconditionally into the *H. areolatus* species, but with no discernible morphological differences between them. The results indicate that there may be a possibility of cryptic speciation with some individuals deserving elevation to either species or subspecies status. An in-depth study incorporating molecular and morphological studies, with careful documentation of the precise localities of the sequenced individuals will be necessary in order to develop this interesting finding further.

Resolution of the generic relationships, between the six proposed southern African Testudinidae representatives, was ascertained even though both phenetic and cladistic phylogenies were inconsistent. Both methods agree with the basal placement of *Geochelone* to the rest of the study group and the relatively recent divergence of *Kinixys* and the *Chersobius* clade (including *Chersina*). This level of resolution was

subsequently lost after bootstrap analysis because the support for the structuring was very weak. A lack of generic definition may be related to rapid speciation proposed during the Eocene (Carroll 1988, Crumly 1984). Evolving from a common ancestor, diversification may have occurred in a relatively short time span, allowing for very little genetic change to accumulate in the cytochrome *b* gene during this period.

The PCR technique was also found to be an appropriate powerful tool for forensic identification of unknown taxa. An individual tortoise, obtained from an accident scene, could not be identified beyond the subspecific level *Geochelone pardalis*. Having sequenced the two positively identified proposed subspecific varieties, this unknown tortoise was matched to *G. p. pardalis*.

The issue of whether the base substitution rate is similar to the rest of the Testudinidae still remains unresolved. The possible inaccuracy of the molecular clock (mainly due to the stochasticity of mutational events and calibration difficulties) and the paucity of the fossil record for African Cryptodira mitigates against a definitive conclusion being made.

Appendix A: Buffers and Reagents

TE (Tris / EDTA) 1M pH 8.0

1 mM EDTA pH 8.0
10 mM Tris/HCl

Make up to 500 ml with H_2O and autoclave

TAE (Tris / Acidic Acid / EDTA) 50 X

242 g Tris/HCl
57.1 ml Glacical Acetic Acid
100 ml .05 M EDTA pH 8.0

Make up to 1 l with H_2O and autoclave. Dilute 1:50 with H_2O prior to use and store at room temperature.

Detergent (used in sequencing)

0.05% Tween 20
0.05% MP40

TBE (Tris / Boric Acid / EDTA) 10 X

216 g Tris/HCl
110 g Boric Acid
80 ml 0.5 mM EDTA pH 8.0

Make up to 1 l with H_2O and autoclave. Dilute 1:10 with H_2O proir to use and store at room temperature.

Stock Acrylamide Solution (47%) Bis (2.5%) 20:1

250 g Acrylamide
12.5 g Bis

Make up to 500 ml with H_2O and store in dark container at 4°C

Working Acrylamide Stock (6%)

63.8 mL 47% Stock Acrylamide Solution
50 mL T.B.E 10 X
240 g Urea

Make up to 500 ml with H_2O. Dissolve by heating the mixture slowly on an hot plate and filter through. Store in a dark bottle at 4°C.

Proteinase K

10 mg/ml stock solution. Made up with distillled H_2O

λ DNA Marker

40 µl λ DNA stock (0.49 µg/µl)
4 µl 10 X buffer M (supplied with λ DNA)
4 µl Dra1 stock (10µg/µl)

Incubate for 60 minutes at 37°C. Load 4.8 µl per lane

Extraction Buffer

100 mM Tris pH 7.5
150 mM NaCl
20 mM EDTA
10% w/v surose

Make up to 1 l with H_2O and autoclave

Lysis Buffer

TNE (Tris / NaCl / EDTA)

50 mM Tris/HCl
100 mM NaCl
5 mM EDTA pH 7.5

3 ml TNE
300 µl 1 M Tris/HCl pH 8.0
80 µl 25% SDS

The solution was made in a small beaker and aliquoted out into eppendorfs, each receiving 300 - 500 µl.

Bromophenol Blue Loading Buffer

0.25% Bromophenol Blue
40% w/v Sucrose

Stored at room temperature.

Ethidium Bromide (10 mg/ml stock solution)

For 100 ml: 1g Ethidium Bromide
100 ml H_2O

Dissolve the Ethidium Bromide and store in a container wrapped in aluminium foil at 4°C.

Fixer for fixing acyrlamide gels

10 % Methanol
10% Acetic Acid
10 ml of Glycerol

Appendix B1: Sequence data (450 bp) for all taxa studied

```
                                                                                                                    60
Cangul      ATAATTAAAA TCATTAACAA CTCATTCATC GATCTGCCAA GCCCCTCCAA CATCTCTGCC
Hsigsig     ATAATTAAAA TTATTAATAA TTCATTCATC GATCTACCAA GCCCCTCTAA CATCTCTGCT
Hsigcaf     ATAATTAAAA TTATTAATAA TTCATTCATC GATCTACCAA GCCCCTCTAA CATCTCTGCT
Hboul       ATAATTAAAA TTATTAACAA TTCATTCATC GATCTGCCAA GCCCCTCTAA CATCTCTGCC
Hfem        ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCAAA CATCTCTGCC
Hareolat    ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaHermI     ATAATTAAAA TTATCAACAG CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaHermII    ATAATTAAAA TTATCAACAG CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaDeHoop    ATAATTAAAA TTATCAACAG CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaMUmale    ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaMUfema    ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaMTmale    ATAATTAAAA TTATCAACAG CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaMTfema    ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaEland     ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
HaDriefn    ATAATTAAAA TTATCAACAG CTCATTCATT GATTTACCAA GCCCCTCCAA CATCTCTGCC
Kbelbel     ATAATTAAAA TCATCAACAA CTCGTTCATT GACCTGCCAA GTCCTTCTAA CATCTCTACC
Klobatsi    ATAATTAAAA TCATCAACAA CTCATTCATC GACCTACCAA GTCCATCCAA CATCTCTACC
Kspeki      ATAATTAAAA TCATCAACAA TTCATTCATC GACCTGCCAA GTCCATCCAA CATCTCTGCC
Knat        ATAATTAAAA TCATCAACAA TTCATTCATC GACCTGCCAA GTCCATCCAA CATCTCTGCC
Mtorner     ATAATTAAAA TTATCAACAA CTCATTCATT GACCTACCAA GCCCCCCCAA CATCTCTGCC
Ptentt      ATAATTAAAA TTATCAACAA CTCATTCATC GATTTACCAA GCCCCTCAAA CATCTCTGCC
Ptentv      ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCTCCCAA CATCTCTGCC
Poculifr    ATAATTAAAA TTATTAACAA CTCATTCATT GATTTACCAA GCCCTTCTAA CATCTCTGCC
Pgeo1       ATAATTAAAA TTATCAACAA CTCATTCGTT GATTTACCAA GTCCTTCTAA CATCTCTGCC
Pgeo2       ATAATTAAAA TTATCAACAA CTCATTCGTT GATTTACCAA GTCCTTCTAA CATCTCTGCC
Tmargin     ATAATAAAAA TCATTAACAA CTCGTTCATT GATCTACCAA GCCCCTCTAA CATCTCTGCT
Thermeni    ATAATAAAAA TCATTAACAA CTCATTCATT GACCTACCAA GCCCTTCTAA CATCTCTGCT
Gpardp1     ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCTAA CATCTCTGCC
Gpardp2     ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCTAA CATCTCTGCC
Gpardb      ATAATCAAAA TCATCAATAA CTCATTCATT GATCTACCAA GCCCCTCCAA CATCTCTGCC
CHscrip1    ATAGTAAAAA TCATCAACAA CTCATTCATT GATCTACCAA GCCCCTCCAA CATCTCTGCT
CHscrip2    ATAGTAAAAA TCATCAACAA CTCATTCATT GATCTACCAA GCCCCTCCAA CATCTCTGCT
PELsub      ATAATTAAAA TTATCAACAA CTCATTCATT GATCTACCAA GCCCCTCTAA CATCTCTGCC
Eaustral    ATAATTAAAA TTATCAACAA CTCATTCATT GATTTACCAA GCCCCTCTAA CATCTCTGCC
CARcar      ATAATAAAAA TCATCAACAA TTCACTCATC GACTTACCAA GCCCCTCCAA CATCTCTGCA


                                                                                                                   120
Cangul      TGATGAAACT TCGGATCACT ACTAGGCATC TGCTTAATCC TACAAATCAT TACCGGAATC
Hsigsig     TGATGAAATT TCGGATCACT ACTAGGCATC TGCCTAATTC TACAAGTCAT TACTGGAATT
Hsigcaf     TGATGAAATT TCGGATCACT ACTAGGCATC TGCCTAATCC TACAAGTCAT TACTGGAATT
Hboul       TGATGAAATT TTGGGTCACT ACTAGGTATC TGCCTAATCC TGCAAATCAT TACCGGAATT
Hfem        TGATGAAACT TTGGATCATT ACTGGGCACC TGCCTAATCC TACAAATTAT CACCGGAATA
Hareolat    TGATGAAACT TTGGATCACT ACTAGGCATC TGCCTAATCC TACAAATTAC TACTGGAATC
HaHermI     TGATGAAACT TTGGATCACT ACTAGGCATC TGCCTAATCC TACAAATTAC TACTGGAATC
HaHermII    TGATGAAACT TTGGGTCACT ACTAGGCATC TGCCTAATCC TACAAATTAC TACTGGAATC
HaDeHoop    TGATGAAACT TTGGGTCACT ACTAGGTATC TGCCTAATAC TACAAATTAC TACTGGAATC
HaMUmale    TGATGAAACT TTGGGTCACT ACTAGGCATC TGCCTAATCC TACAAATTAC TACTGGAATC
HaMUfema    TGATGAAACT TTGGGTCACT ACTAGGCATC TGCCTAATCC TACAAATTAC TACTGGAATC
HaMTmale    TGATGAAACT TTGGGTCACT ACTAGGCATC TGCCTAATCC TACAAATTAC TACTGGAATC
HaMTfema    TGATGAAACT TTGGATCACT ACTAGGCATC TGCCTAATCC TACAAATTAC TACTGGAATC
HaEland     TGATGAAACT TTGGATCACT ACTAGGCATC TGCCTAATCC TACAAATTAC TACTGGAATC
HaDriefn    TGATGAAACT TTGGGTCACT ACTAGGCATC TGCCTAATAC TACAAATCAC TACTGGAATC
Kbelbel     TGATGAAACT TTGGATCACT ACTAGGCATC TGTCTAATCC TACAAATCAT TACTGGAATT
Klobatsi    TGATGAAACT TTGGATCACT ACTAGGCATC TGTCTAATCC TACAAATCAT TACTGGGATT
Kspeki      TGATGAAACT TTGGATCACT ACTAGGCATC TGCCTAATCC TACAAATTAT TACCGGAATT
Knat        TGATGAAACT TTGGATCACT ACTAGGCATC TGCCTAATCC TACAAATTAT TACCGGAATT
Mtorner     TGATGAAATC TAGGATCACT ACTAGGCACC TGCCTAATTC TACAAATTAT CACAGGAATC
Ptentt      TGATGAAATT TTGGATCATT ACTAGGCATC TGCCTAATCC TACAAATCAT TACCGGAATT
Ptentv      TGATGAAATT TTGGATCATT ACTAGGCATC TGCCTAATCC TACAAATTAT TACCGGAATT
Poculifr    TGATGAAATT TTGGATCACT ACTAGGCATT TGCCTAATCC TACAGATCAT TACTGGAATT
Pgeo1       TGATGGAATT TTGGATCACT ATTAGGCATT TGCCTGATCC TACAAATCAT TACTGGAATC
Pgeo2       TGATGGAATT TTGGATCACT ATTAGGCATT TGCCTGATCC TACAAATCAT TACTGGAATC
Tmargin     TGATGAAACT TCGGATCACT ACTAGGCATC TGCCTAACCC TACAAATTAT TACCGGAATC
Thermeni    TGATGAAATT TCGGCTCACT ACTAGGCATC TGCCTAATCC TACAAATTAT TACCGGAATC
Gpardp1     TGATGAAACT TCGGATCATT ACTAGGCATC TGCCTAATCG TACAAATCAT TACTGGAATT
Gpardp2     TGATGAAACT TCGGATCATT ACTAGGCATC TGCCTAATCG TACAAATCAT TACTGGAATC
Gpardb      TGATGAAACT TTGGATCACT ACTAGGCATC TGTCTAATCC TACAAATAAT TACTGGAATC
CHscrip1    TGATGAAACT TTGGATCGTT ACTAGGCATC TGCCTAATCC TACAAATTAT TACCGGAATT
CHscrip2    TGATGAAACT TTGGATCATT ACTAGGCATC TGCCTAATCC TACAAATCAT TACCGGAATT
PELsub      TGATGAAACT TCGGCTCACT ACTAGGCATC TGCCTAATCG TACAAATTAT TACTGGAATC
Eaustral    TGATGAAACT TCGGATCATT ACTAGGCATC TGCCTAATCG TACAAAATAT TACTGGAATT
CARcar      TGATGAAACT TCGGATCACT ACTAGCCACC TGTCTAGCAC TACAAATCAT TACCGGAATC
```

```
                                                                    180
Cangul      TTTCTATCAA TACACTACTC ACCAAACATC TCACTAGCAT TCTCATCAGT AGCCCATATT
Hsigsig     TTTCTATCAA TACACTACTC ACCAAATATC TCACTAGCAT TCTCATCAGT AGCCCATATC
Hsigcaf     TTTCTATCAA TACACTACTC ACCAAATATC TCACTAGCAT TCTCATCAGT AGCCCATATC
Hboul       TTTCTATCAA TACACTACTC ACCAAATATC TCACTAGCAT TCTCATCAGT AGCCCATATT
Hfem        TTTCTAGCAA TACACTACTC ACCAGATATC TCACTAGCAT TCTCATCTGT AGCCCATATT
Hareolat    TTCCTAGCAA TACACTACTC CCCTAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaHermI     TTCCTAGCAA TACACTACTC CCCAAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaHermII    TTCCTAGCAA TACACTACTC CCCAAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaDeHoop    TTCCTAGCAA TACACTACTC CCCAAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaMUmale    TTCCTAGCAA TACACTACTC CCCAAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaMUfema    TTCCTAGCAA TACACTACTC CCCAAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaMTmale    TTCCTAGCAA TACACTACTC CCCAAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaMTfema    TTCCTAGCAA TACACTACTC CCCTAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaEland     TTCCTAGCAA TACACTACTC CCCTAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
HaDriefn    TTCCTAGCAA TACACTACTC CCCAAATATT TCACTAGCAT TCTCATCCGT AGCCCATATC
Kbelbel     TTTCTAGCAA TACATTACTC ACCTAATATC TCACTAGCAT TCTCATCAGT AGCCCACATC
Klobatsi    TTTCTAGCAA TACACTACTC ACCAAACATC TCACTAGCAT TCTCATCAGT AGCCCACATC
Kspeki      TTTCTAGCAA TACATTACTC ACCCAACATT TCACTAGCAT TCTCATCAGT AGCCCATATC
Knat        TTTCTAGCGA TACATTACTC ACCCAACATT TCACTAGCAT TCTCATCAGT AGCCCATATC
Mtorner     CTCCTAGCAA TACACTACTC ACCTAATATC TCACTAGCAT TCTCATCAGT AGCCCACATC
Ptentt      TTTCTAGCAA TACATTACTC ATCAAACATC TCACTAGCAT TCTCATCAGT AGCCCACATC
Ptentv      TTCTTAGCAA TACATTATTC ATCAAATATC TCACTAGCAT TCTCATCAGT AACCCACATC
Poculifr    TTTCTAGCAA TACATTACTC ATCAAACATC TCACTAGCAT TCTCATCAGT AACCCATATC
Pgeo1       TTTCTAGCAA TACATTACTC ATCAAACATC TCACTAGCAT TCTCATCAGT AACCCATATC
Pgeo2       TTTCTAGCAA TACATTACTC ATCAAACATC TCACTAGCAT TCTCATCAGT AACCCATATC
Tmargin     TTCCTAGCAA TACACTACTC ACCAAACATC TCACTAGCAT TTTCATCAGT AGCCCATATC
Thermeni    TTCTTAGCAA TACACTACTC ACCAAACATC TCACTAGCAT TCTCATCAGT AGCCCATATC
Gpardp1     TTTCTGGCAA TACACTACTC ACCAAACATC TCACTAGCAT TCTCATCAGT AGCCCATATC
Gpardp2     TTTCTGGCAA TACACTACTC ACCAAACATC TCACTAGCAT TCTCATCAGT AGCCCATATC
Gpardb      TTTCTAGCAA TACACTACTC ACCAAACATC TCACTAGCAT TCTCATCAGT AGCCCATATC
CHscrip1    TTCCTAGCAA TACATTACTC ACCAGACATC TCCCACGTCC CTCCATCAGT ATCCCACATC
CHscrip2    TTCCTAGCAA TTCATTACTC ACCAGACATC TCCCACGTCC CTCCATCAGT ATCCCACATC
PELsub      TTCCTGGCAA TACACTACTC ACCAAACATC TCACTAGCAT TCTCATCAGT AGCCCATATC
Eaustral    TTTCTAGCAA TACACTACTC ACCAAACATC TCACTAGCAT TCTCATCAGT AGCCCATATC
CARcar      TTCCTAGCAA TACATTACTC ACCAGACATC TCCATAGCCT TTTCATCAAT TACCCACATC

                                                                    240
Cangul      ACCCGAGATG TACAGTACGG ATGACTTATC CGAAATATAC ATGCTAACGG AGCCTCTATC
Hsigsig     ACCCGAGACG TACAGTACGG ATGACTTATC CGAAATATAC ATACTAACGG AGCCTCCCTC
Hsigcaf     ACCCGAGACG TACAGTACGG ATGACTTATC CGAAATATAC ATACTAATGG AGCCTCCCTC
Hboul       ACCCGAGACG TACAGTATGG ATGACTTATC CGAAATATAC ATGCTAATGG AGCCTCCATC
Hfem        ACCCGAGATG TACAATACGG ATGACTTATC CGAAATATAC ATGCTAACGG GGCCTCCATC
Hareolat    ACCCGAGATG TACAATATGG ATGACTTATC CAAAACATAC ATGCCAACGG AGCCTCTATC
HaHermI     ACCCGAGATG TGCAGTATGG ATGACTTATC CAAAATATAC ATGCCAACGG AGCCTCTATC
HaHermII    ACCCGAGATG TGCAGTATGG ATGACTTATC CAAAATATAC ATGCCAACGG AGCCTCTATC
HaDeHoop    ACCCGAGATG TACAGTATGG ATGACTTATC CAAAACATAC ATGCCAACGG AGCCTCTATC
HaMUmale    ACCCGAGATG TGCAGTATGG ATGACTTATC CAAAATATAC ATGCCAACGG AGCCTCTATC
HaMUfema    ACCCGAGATG TGCAGTATGG ATGACTTATC CAAAATATAC ATGCCAACGG AGCCTCTATC
HaMTmale    ACCCGAGATG TACAGTATGG ATGACTTATC CAAAATATAC ATGCCAACGG AGCCTCTATC
HaMTfema    ACCCGAGATG TACAATATGG ATGACTTATC CAAAACATAC ATGCCAACGG AGCCTCTATC
HaEland     ACCCGAGATG TACAATATGG ATGACTTATC CAAAACATAC ATGCCAACGG AGCCTCTATC
HaDriefn    ACCCGAGATG TACAGTATGG GTGACTTATT CGAAATATAC ATGCCAATGG AGCCTCTATC
Kbelbel     ACCCGAGATG TACAATATGG ATGACTTATT CGAAATGTAC ACGCTAACGG AGCCTCTATC
Klobatsi    ACCCGAGACG TGCAATATGG ATGACTCATT CGAAATGTAC ACGCTAACGG AGCCTCTATC
Kspeki      ACCCGAGATG TACAATATGG ATGACTTATT CGAAACATAC ACGCTAACGG AGCCTCCATC
Knat        ACCCGAGATG TACAATATGG ATGACTTATT CGAAACATAC ACGCTAACGG AGCCTCCATC
Mtorner     ACCCGAGACG TGCAATATGG ATGACTCATT CGAAATGTAC ACGCTAACGG AGCCTCTATC
Ptentt      AATCGAGATG TACAGTACGG ATGACTTATC CGAAACGTAC ATGCCAACGG AGCCTCCATC
Ptentv      ACTCGAGATG TGCAATACGG ATGACTGATC CGAAACATAC ATGCCAACGG AGCCTCCATC
Poculifr    ACCCGAGATG TACAGTACGG ATGACTTATC CGAAACATAC ATGCTAACGG AGCCTCTATC
Pgeo1       ACCCGAGATG TACAGTACGG ATGACTTATC CGAAACATAC ATGCTAATGG AGCCTCTATC
Pgeo2       ACCCGAGATG TACAGTACGG ATGACTTATC CGAAACATAC ATGCTAATGG AGCCTCTATC
Tmargin     TCCCGAGATG TACAATACGG ATGACTCATC CGAAATATAC ACGCCAACGG AGCCTCCATT
Thermeni    ACCCGAGATG TACAGTACGG ATGACTTATC CGAAACATAC ATGCTAATGG AGCCTCTATC
Gpardp1     ACCCGAGACG TACAGTACGG ATGGCTTATC CGAAATATAC ATGCCAACGG AGCCTCCATC
Gpardp2     ACCCGAGACG TACAGTACGG ATGGCTTATC CGAAATATAC ATGCCAACGG AGCCTCCATA
Gpardb      ACCCGAGATG TACAGTACGG GTGACTTATC CGAAACATAC ATGCCAACGG AGCCTCCATA
CHscrip1    ACCCGAGACG TACATTACGG ATGACTCATC CGCAACATAC ACGCCAACGG AGCTTCCCTC
CHscrip2    ACCCGAGACG TACATTACGG ATGACTCATC CGCAACATAC ACGCCAACGG AGCTTCCCTC
PELsub      ACCCGAGATG TACAGTACGG ATGACTTATC CGAAACATAC ATGCCAACGG AGCCTCCATC
Eaustral    ACCCGAGATG TACAGTACGG ATGACTTATC CGAAACATAC ATGCCAACGG AGCCTCCATA
CARcar      ACCCGAGATG TACAATACGG ATGACTCATC CGCAACATGC ACGCCAACGG AGCCTCCCTA
```

						300
Cangul	TTCTTCATAT	GTATCTACCT	CCACATCGGC	CGAGGACTTT	ATTACGGCAC	CTACATATAC
Hsigsig	CTCTTTATAT	GTATTTATCT	TCACATTGGC	CGAGGACTTT	ATTACGGCTC	CTACATATAT
Hsigcaf	CTCTTTATAT	GTATTTATCT	TCACATTGGC	CGAGGACTTT	ACTACGGCTC	CTACATATAT
Hboul	CTCTTCATAT	GTATCTATCT	TCATATTGGC	CGAGGACTTT	ATTATGGCTC	CTACATATAT
Hfem	TTCTTTATAT	GTATCTACCT	TCACATTGGT	CGAGGACTTT	ACTACGGTTC	CTACCTATAC
Hareolat	TTCTTCATAT	GTATTTACCT	TCACATCGGC	CGAGGACTTT	ATTATGGTTC	CTATTTATAC
HaHermI	TTCTTCATAT	GCATCTACCT	TCACATCGGC	CGAGGACTTT	ATTACGGTTC	CTATTTATAC
HaHermII	TTCTTCATAT	GCATCTACCT	TCACATCGGC	CGAGGACTTT	ATTACGGTTC	CTATTTATAC
HaDeHoop	TTCTTCATAT	GCATCTACCT	TCACATCGGC	CGAGGACTTT	ATTACGGTTC	CTATTTATAC
HaMUmale	TTCTTCATAT	GCATCTACCT	TCACATCGGC	CGAGGACTTT	ATTACGGTTC	CTATTTATAT
HaMUfema	TTCTTCATAT	GCATCTACCT	TCACATCGGC	CGAGGACTTT	ATTACGGTTC	CTATTTATAT
HaMTmale	TTCTTCATAT	GCATCTACCT	TCACATCGGC	CGAGGACTTT	ATTACGGTTC	CTATTTATAC
HaMTfema	TTCTTCATAT	GTATTTACCT	TCACATCGGC	CGAGGACTTT	ATTATGGTTC	CTATTTATAC
HaEland	TTCTTCATAT	GTATCTACCT	TCACATCGGC	CGAGGACTTT	ATTATGGTTC	CTATTTATAC
HaDriefn	TTCTTCATAT	GTATCTACCT	TCACATCGGC	CGAGGACTTT	ATTACGGTTC	CTATTTATAC
Kbelbel	TTCTTTATAT	GTATCTACCT	ACACATTGGT	CGAGGACTTT	ATTATGGCTC	CTACCTGTAT
Klobatsi	TTCTTTATAT	GCATCTACCT	ACACATTGGT	CGAGGACTTT	ATTACGGCTC	CTACCTGTAT
Kspeki	TTCTTTATAT	GTATCTACCT	ACATATTGGT	CGAGGACTTT	ACTATGGCTC	CTACCTTTAC
Knat	TTCTTTATAT	GTATCTACCT	ACACATTGGC	CGAGGACTTT	ACTATGGCTC	CTACCTTTAC
Mtorner	TTCTTTATAT	GCATCTACCT	ACACATTGGT	CGAGGACTTT	ATTACGGCTC	CTACCTGTAT
Ptentt	TTCTTCATAT	GTATCTATCT	CCATATTGGC	CGAGGACTTT	ACTACGGTTC	CTACCTATAT
Ptentv	TTCTTTATAT	CTATTTATCT	CCATATCGGC	CGAGGACTTT	ACTATGGTTC	CTACCTATAT
Poculifr	TTCTTTATAT	GTATTTATCT	CCATATCGGC	CGAGGACTTT	ATTACGGCTC	CTATTTATAT
Pgeo1	TTCTTTATGT	GTATTTACCT	ACATATCGGC	CGAGGACTTT	ATTACGGCTC	CTACCTATAT
Pgeo2	TTCTTTATGT	GTATTTACCT	ACATATCGGC	CGAGGACTTT	ATTACGGCTC	CTACCTATAT
Tmargin	TTCTTTATAT	GTATTTACCT	TCACATCGGC	CGAGGACTTT	ACTACGGCTC	ATATCTACAT
Thermeni	TTCTTCATAT	GTATTTACCT	ACACATCGGC	CGAGGACTTT	ATTATGGCTC	CTACCTATAC
Gpardp1	TTCTTCATAT	GTATCTACCT	TCACATCGGT	CAAGGACTTT	ATTACGGCTC	TTATCTATAC
Gpardp2	TTCTTCATAT	GTATCTACCT	TCACATCGGC	CGAGGACTTT	ACTACGGCTC	TTATCTATAC
Gpardb	TTCTTTATAT	GTATTTACCT	CCACATTGGC	CGAGGACTTT	ACTATGGCTC	CTATCTATAC
CHscrip1	CTCTTCATGT	GCATTTATCT	TCACATCGGC	CGAGGACTTT	ACTACGGCTC	GTACTTATAC
CHscrip2	CTCTTCATGT	GCATTTATCT	TCACATCGGC	CGAGGACTTT	ACTACGGCTC	GTACTTATAC
PELsub	TTCTTTATAT	GTATTTACCT	TCACATTGGC	CGAGGACTTT	ATTACGGCTC	TTATCTATAC
Eaustral	TTCTTTATAT	GTATTTACCT	CCACATTGGC	CGAGGACTTT	ACTACGGCTC	CTATCTATAC
CARcar	TTTTTCATCT	GCATCTACCT	CCACATCGGA	CGAGGAATCT	ACTACGGTTC	CTATCTATAC
						360
Cangul	AAAGAAACCT	GAAACACAGG	AATTATTCTA	CTATTTCTAG	TCATAGCCAC	TGCATTCGTA
Hsigsig	AAAGAAACCT	GAAATACAGG	AATTATCCTA	CTATTACTAA	CCATAACCAC	CGCATTCATA
Hsigcaf	AAAGAAACCT	GAAATACAGG	AATTATCCTA	CTATTACTAA	CCATAACCAC	CGCATTCATA
Hboul	AAAGAAACCT	GAAATACAGG	AATTATTCTA	CTACTACTAG	TCATAGCCAC	TGCATTCGTA
Hfem	AAAGAAACCT	GAAACACAGG	AATTATCCTA	CTCCTCCTAG	TCATAGCCAC	TGCATTCATA
Hareolat	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCCTC	CGCATTTGTG
HaHermI	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCTAC	CGCATTTGTT
HaHermII	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TCATAGCTAC	CGCATTTGTT
HaDeHoop	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCTAC	CGCATTTATG
HaMUmale	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCCAC	CGCATTTATG
HaMUfema	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCCAC	CGCATTTATG
HaMTmale	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCTAC	CGCATTTGTA
HaMTfema	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCCAC	CGCATTTGTG
HaEland	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCCAC	CGCATTTGTG
HaDriefn	AAAGAAACCT	GAAACACAGG	AATTATTTTA	TTACTCCTAG	TTATAGCCTC	CGCATTTGTT
Kbelbel	AAAGAAACCT	GAAACACAGG	AATTATTCTA	CTATTCCTAG	TCATAGCTAC	TGCATTCGTA
Klobatsi	AAAGAAACCT	GAAACACAGG	AATCATTCTA	CTATTCCTAA	TTATAGCTAC	TGCATTCGTA
Kspeki	AAAGAAACCT	GAAACACTGG	AATTATTCTA	CTATTCCTAA	TTATAGCTAC	TGCATTCGTA
Knat	AAAGAAACCT	GAAACACAGG	AATTATTCTA	CTATTCCTAA	TTATAGCTAC	TGCATTCGTA
Mtorner	AAAGAAACTT	GAAACACAGG	AATCATTCTA	CTATTCCTAG	TTATAGCTAC	TGCATTCGTA
Ptentt	AAAGAAACCT	GAAATACAGG	AATCATCTTA	TTACTACTAG	TCATAGCCAC	TGCATTCATA
Ptentv	AAAGAAACCT	GAAATACAGG	AATCATCTTA	TTACTACTAT	TAATAGCCAC	TGCATTCATG
Poculifr	AAAGAAACCT	GAAATACAGG	AATCATTCTA	TTACTACTAG	TCATAGCTAC	CGCATTCATA
Pgeo1	AAAGAAACCT	GAAATACAGG	AATCATCCTA	TTACTACTGG	TCATAGCCAC	TGCATTCATA
Pgeo2	AAAGAAACCT	GAAATACAGG	AATCATCCTA	TTACTACTGG	TCATAGCCAC	TGCATTCATA
Tmargin	AAAGAAACCT	GAGACACAGG	AATCATACTA	TTACTCCTAG	TTATAGCCAC	TGCATTTATG
Thermeni	AAAGAAACCT	GAAATACAGG	AATCATCCTA	TTACTACTAG	TCATAGCCAC	TGCATTCATA
Gpardp1	AAAGAAACCT	GAAACACAGG	AATTATTCTA	CTACTCCTAG	TTATAGCTAC	TGCATTCGTA
Gpardp2	AAAGAAACCT	GAAACACAGG	AATTATCTTA	CTACTCTTAG	TTATAGCCAC	TGCATTCGTA
Gpardb	AAAGAAACCT	GAAATACAGG	AATCATCTTA	CTACTCCTAG	TCATAGCCAC	TGCATTCGTA
CHscrip1	AAAGAAACCT	GAAACACTGG	AATCGTTCTA	CTAATACTAG	TTATAGCCAC	TGCATTCGTA
CHscrip2	AAAGAAACCT	GAAACACTGG	AATCGTTCTA	CTAATACTAG	TTATAGCCAC	TGCATTCGTA
PELsub	AAAGAAACCT	GAAACACAGG	AATTATCTTA	CTACTCCTAG	TCATAGCCAC	TGCATTCGTA
Eaustral	AAAGAAACCT	GAAACACAGG	AATTATCTTA	CTACTCCTAG	TTATAGCCAC	TGCATTCGTA
CARcar	AAAGAAACCT	GAAATACCGG	AATCATCCTC	TTACTAC---	----------	----------

```
                                                                                                    420
Cangul      GGTTATGCTC TACCATGAGG CCAAATATCA TTCTGAGGCG CCACAGTCAT CACCAACCTA
Hsigsig     GGTTATGTTC TACCATGAGG CCAAATATCA TTCTGAGGCG CCACAGTCAT CACCAACCTA
Hsigcaf     GGTTATGTCC TACCATGAGG CCAAATATCA TTCTGAGGCG CCACAGTCAT CACCAACCTA
Hboul       GGTTACGTCC TACCATGAGG CCAAATATCA TTCTGAGGCG CCACAGTCAT CACCAACCTA
Hfem        GGCTATGTCC TACCATGAGG CCAAATATCA TTCTGAGGCG CTACAGTCAT CACCAACCTA
Hareolat    GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGTG CTACAGTCAT CACCAACCTA
HaHermI     GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGCG CTACAGTCAT CACCAACCTA
HaHermII    GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGCG CTACAGTCAT CACCAACCTA
HaDeHoop    GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGCG CTACAGTCAT CACCAACCTA
HaMUmale    GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGCG CTACAGTCAT CACCAACCTA
HaMUfema    GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGCG CTACAGTCAT CACCAACCTA
HaMTmale    GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGCG CTACAGTCAT CACCAACCTA
HaMTfema    GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGTG CTACAGTCAT CACCAACCTA
HaEland     GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGTG CTACAGTCAT CACCAACCTA
HaDriefn    GGTTATGTCC TGCCATGAGG TCAAATATCA TTCTGAGGCG CTACAGTCAT CACCAACCTA
Kbelbel     GGTTATGTAC TACCATGAGG CCAAATGTCA TTCTGAGGCG CTACCGTCAT CACCAACCTA
Klobatsi    GGTTACGTAC TGCCATGAGG CCAAATGTCA TTCTGAGGCG CTACCGTCAT CACCAACCTA
Kspeki      GGTTACGTAC TACCATGAGG CCAAATATCA TTCTGAGGCG CTACCGTCAT CACCAACCTA
Knat        GGTTACGTAC TACCATGAGG CCAAATATCA TTCTGAGGCG CTACCGTCAT CACCAACCTA
Mtorner     GGTTACGTAC TGCCATGAGG CCAAATGTCA TTCTGAGGCG CTACCGTCAT CACCAACCTA
Ptentt      GGCTACGTAC TACCATGAGG CCAAATATCA TTCTGAGGTG CTACAGTCAT CACCAACCTA
Ptentv      GGCTATGTAC TACCATGAAA CCAAATATCA TTCTGAGGTG CTACAGTCAT TACCAACCTA
Poculifr    GGCTACGTAC TACCATGAGG CCAAATATCA TTCTGAGGTG CTACAGTCAT TACTAACCTA
Pgeo1       GGCTACGTAC TACCATGAGG CCAAATATCA TTCTGAGGTG CTACAGTCAT TACAAACCTA
Pgeo2       GGCTACGTAC TACCATGAGG CCAAATATCA TTCTGAGGTG CTACAGTCAT TACAAACCTA
Tmargin     GGTTACGTCC TACCCTGAGG CCAAATATCA TTTTGAGGGG CCACCGTTAT CACTAATCTA
Thermeni    GGCTACGTAC TACCATGAGG CCAAATATCA TTCTGAGGTG CTACAGTCAT TACAAACCTA
Gpardp1     GGTTATGTCT TACCATGAGG ACAAATATCA TTTTGAGGCG CTACAGTTAT CACCAACCTA
Gpardp2     GGTTATGTCT TACCATGAGG ACAAATATCA TTTTGAGGCG CTACAGTTAT CACCAACCTA
Gpardb      GGTTATGTCT TACCATGAGG ACAAATATCA TTTTGAGGCG CTACAGTTAT CACCAACCTA
CHscrip1    GGTTATGTCT TACCATGAGG ACAAATATCA TTTTGAGGCG CTACAGTTAT CACCAACCTA
CHscrip2    GGTTATGTCT TACCATGAGG ACAAATATCA TTTTGAGGCG CTACAGTTAT CACCAACCTA
PELsub      GGTTATGTCT TACCATGAGG ACAAATATCA TTTTGAGGCG CTACAGTTAT CACCAACCTA
Eaustral    GGTTATGTCT TACCATGAGG ACAAATATCA TTTTGAGGCG CTACAGTTAT CACCAACCTA
CARcar      ---------- ---------- ---------- ---------- ---------- ----------

Cangul      CTCTCAGCCA TCCCATACAT CGGTGACACC
Hsigsig     CTCTCAGCCA TCCCCTACAT CGGTACCACC
Hsigcaf     CTCTCAGCTA TCCCATACAT TGGTACCACC
Hboul       CTTTCAGCCA TCCCATACAT CGGTAACACC
Hfem        CTCTCAGCCA TCCCATATGT AGGTGACACC
Hareolat    CTCTCAGCCA TCCCATATAT AGGCAACACC
HaHermI     CTCTCAGCCA TCCCATATAT AGGCAACACC
HaHermII    CTCTCAGCCA TCCCATATAT AGGCAACACC
HaDeHoop    CTCTCAGCCA TCCCATATAT AGGCAACACC
HaMUmale    CTCTCAGCCA TCCCATATAT AGGCAACACC
HaMUfema    CTCTCAGCCA TCCCATATAT AGGCAACACC
HaMTmale    CTCTCAGCCA TCCCATATAT AGGCAACACC
HaMTfema    CTCTCAGCCA TCCCATATAT AGGCAACACC
HaEland     CTCTCAGCCA TCCCATATAT AGGCAACACC
HaDriefn    CTCTCAGCCA TCCCATATAT AGGCAACACC
Kbelbel     CTTTCAGCCA CCCCATATAT TGGCAACACC
Klobatsi    CTTTCAGCCA CTCCCTATAT TGGCAACATC
Kspeki      CTCTCAGCCG CCCCATATAT TGGCAATACC
Knat        CTCTCAGCCG CCCCATATAT CGGCAATACC
Mtorner     CTTTCAGCCA TCCCCTATAT CGGTAACATC
Ptentt      CTTTCAGCCA CCCCTTACAT CGGCAACACC
Ptentv      CTTTCAGCCA CTCCTTACAT TGGCAACACC
Poculifr    CTTTCAGCCA TCCCTTACAT CGGCAACACC
Pgeo1       CTTTCAGCCA TCCCTTACAT CGGCAACACA
Pgeo2       CTTTCAGCCA TCCCTTACAT CGGCAACACC
Tmargin     CTCTCAGCCA TCCCCTACAT CGGCAACACA
Thermeni    CTTTCAGCCA TCCCTTACAT CGGCAACACC
Gpardp1     CTTTCAGCCA TTCCTTATAT CGGCAACACC
Gpardp2     CTTTCAGCCA TTCCTTATAT CGGCAACACC
Gpardb      CTTTCAGCCA TTCCTTATAT CGGCAACACA
CHscrip1    CTTTCAGCCA TTCCTTATAT CGGCAACACA
CHscrip2    CTTTCAGCCA TTCCTTATAT CGGCAACACA
PELsub      CTTTCAGCCA TTCCTTATAT CGGCAACACA
Eaustral    CTTTCAGCCA TTCCTTATAT CGGCAACACA
CARcar      ---------- ---------- ----------
```

Appendix B2: Protein sequence data for all taxa studied

```
                                                                                                                    60
Cangul      MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLSMHYSPNI SLAFSSVAHI
Hsigsig     MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQVITGI FLSMHYSPNI SLAFSSVAHI
Hsigcaf     MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQVITGI FLSMHYSPNI SLAFSSVAHI
Hboul       MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLSMHYSPNI SLAFSSVAHI
Hfem        MIKIINNSFI DLPSPSNISA WWNFGSLLGT CLILQIITGM FLAMHYSPDI SLAFSSVAHI
Hareolat    MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQITTGI FLAMHYSPNI SLAFSSVAHI
HaHerm1     MIKIINSSFI DLPSPSNISA WWNFGSLLGI CLILQITTGI FLAMHYSPNI SLAFSSVAHI
HaHerm2     MIKIINSSFI DLPSPSNISA WWNFGSLLGI CLILQITTGI FLAMHYSPNI SLAFSSVAHI
HaDeHoop    MIKIINSSFI DLPSPSNISA WWNFGSLLGI CLMLQITTGI FLAMHYSPNI SLAFSSVAHI
HaMUmale    MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQITTGI FLAMHYSPNI SLAFSSVAHI
HaMUfema    MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQITTGI FLAMHYSPNI SLAFSSVAHI
HaMTmale    MIKIINSSFI DLPSPSNISA WWNFGSLLGI CLILQITTGI FLAMHYSPNI SLAFSSVAHI
HaMTfema    MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQITTGI FLAMHYSPNI SLAFSSVAHI
HaEland     MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQITTGI FLAMHYSPNI SLAFSSVAHI
HaDriefn    MIKIINSSFI DLPSPSNISA WWNFGSLLGI CLMLQITTGI FLAMHYSPNI SLAFSSVAHI
Kbelbel     MIKIINNSFI DLPSPSNIST WWNFGSLLGI CLILQIITGI FLAMHYSPNI SLAFSSVAHI
Kspeki      MIKIINNSFI DLPSPSNIST WWNFGSLLGI CLILQIITGI FLAMHYSPNI SLAFSSVAHI
Kspekii     MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAMHYSPNI SLAFSSVAHI
Knat        MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAMHYSPNI SLAFSSVAHI
Mtorner     MIKIINNSFI DLPSPPNISA WWNLGSLLGT CLILQIITGI LLAMHYSPNI SLAFSSVAHI
Ptentt      MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAMHYSSNI SLAFSSVAHI
Ptentv      MIKIINNSFI DLPSPPNISA WWNFGSLLGI CLILQIITGI FLAMHYSSNI SLAFSSVTHI
Poculifr    MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAMHYSSNI SLAFSSVTHI
Pgeo1       MIKIINNSFV DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAMHYSSNI SLAFSSVTHI
Pgeo2       MIKIINNSFV DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAMHYSSNI SLAFSSVTHI
Tmargin     MMKIINNSFI DLPSPSNISA WWNFGSLLGI CLTLQIITGI FLAMHYSPNI SLAFSSVAHI
Thermeni    MMKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAMHYSPNI SLAFSSVAHI
Gpardp1     MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLIVQIITGI FLAMHYSPNI SLAFSSVAHI
Gpardp2     MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLIVQIITGI FLAMHYSPNI SLAFSSVAHI
Gpardb      MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQMITGI FLAMHYSPNI SLAFSSVAHI
CHscrip1    MVKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAMHYSPNI SLAFPSVAHI
CHscrip2    MMKIINNSFI DLPSPSNISA WWNFGSLLGI CLILQIITGI FLAIHYSPNI SLAFPSVSHI
PELsub      MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLIVQIITGI FLAMHYSPNI SLAFSSVAHI
Eaustral    MIKIINNSFI DLPSPSNISA WWNFGSLLGI CLIVQNITGI FLAMHYSPNI SLAFSSVAHI
CARcar      MMKIINNSLI DLPSPSNISA WWNFGSLLAT CLALQIITGI FLAMHYSPDI SMAFSSITHI
                                                                                                                   120
Cangul      TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGTYMY KETWNTGIIL LFLVMATAFV
Hsigsig     TRDVQYGWLI RNMHTNGASL LFMCIYLHIG RGLYYGSYMY KETWNTGIIL LLLTMTTAFM
Hsigcaf     TRDVQYGWLI RNMHTNGASL LFMCIYLHIG RGLYYGSYMY KETWNTGIIL LLLTMTTAFM
Hboul       TRDVQYGWLI RNMHANGASI LFMCIYLHIG RGLYYGSYMY KETWNTGIIL LLLVMATAFV
Hfem        TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
Hareolat    TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMASAFV
HaHerm1     TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
HaHerm2     TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
HaDeHoop    TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
HaMUmale    TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
HaMUfema    TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
HaMTmale    TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
HaMTfema    TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
HaEland     TRDVQYGWLI QNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
HaDriefn    TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMASAFV
Kbelbel     TRDVQYGWLI RNVHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LFLVMATAFV
Kspeki      TRDVQYGWLI RNVHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LFLIMATAFV
Kspekii     TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LFLIMATAFV
Knat        TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LFLIMATAFV
Mtorner     TRDVQYGWLI RNVHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LFLVMATAFV
Ptentt      NRDVQYGWLI RNVHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
Ptentv      TRDVQYGWLI RNMHANGASI FFMSIYLHIG RGLYYGSYLY KETWNTGIIL LLLLMATAFM
Poculifr    TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
Pgeo1       TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
Pgeo2       TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
Tmargin     SRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLH KETWDTGIML LLLVMATAFM
Thermeni    TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFM
Gpardp1     TRDVQYGWLI RNMHANGASI FFMCIYLHIG QGLYYGSYLY KETWNTGIIL LLLVMATAFV
Gpardp2     TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
Gpardb      TRDVQYGWLI RNMHANGASM FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
CHscrip1    TRDVHYGWLI RNMHANGASL LFMCIYLHIG RGLYYGSYLY KETWNTGIVL LMLVMATAFV
CHscrip2    TRDVHYGWLI RNMHANGASL LFMCIYLHIG RGLYYGSYLY KETWNTGIVL LMLVMATAFV
PELsub      TRDVQYGWLI RNMHANGASI FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
Eaustral    TRDVQYGWLI RNMHANGASM FFMCIYLHIG RGLYYGSYLY KETWNTGIIL LLLVMATAFV
CARcar      TRDVQYGWLI RNMHANGASL FFICIYLHIG RGIYYGSYLY KETWNTGIIL LL?-------
```

```
Cangul      GYALPWGQMS FWGATVITNL LSAIPYIGDT
Hsigsig     GYVLPWGQMS FWGATVITNL LSAIPYIGTT
Hsigcaf     GYVLPWGQMS FWGATVITNL LSAIPYIGTT
Hboul       GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Hfem        GYVLPWGQMS FWGATVITNL LSAIPYVGDT
Hareolat    GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaHerm1     GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaHerm2     GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaDeHoop    GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaMUmale    GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaMUfema    GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaMTmale    GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaMTfema    GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaEland     GYVLPWGQMS FWGATVITNL LSAIPYMGNT
HaDriefn    GYVLPWGQMS FWGATVITNL LSAIPYMGNT
Kbelbel     GYVLPWGQMS FWGATVITNL LSATPYIGNT
Kspeki      GYVLPWGQMS FWGATVITNL LSATPYIGNI
Kspekii     GYVLPWGQMS FWGATVITNL LSAAPYIGNT
Knat        GYVLPWGQMS FWGATVITNL LSAAPYIGNT
Mtorner     GYVLPWGQMS FWGATVITNL LSAIPYIGNI
Ptentt      GYVLPWGQMS FWGATVITNL LSATPYIGNT
Ptentv      GYVLPWNQMS FWGATVITNL LSATPYIGNT
Poculifr    GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Pgeo1       GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Pgeo2       GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Tmargin     GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Thermeni    GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Gpardp1     GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Gpardp2     GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Gpardb      GYVLPWGQMS FWGATVITNL LSAIPYIGNT
CHscrip1    GYVLPWGQMS FWGATVITNL LSAIPYIGNT
CHscrip2    GYVLPWGQMS FWGATVITNL LSAIPYIGNT
PELsub      GYVLPWGQMS FWGATVITNL LSAIPYIGNT
Eaustral    GYVLPWGQMS FWGATVITNL LSAIPYIGNT
CARcar      ---------- ---------- ----------
```

Appendix C: Pairwise distances between taxa

Below diagonal: Absolute distances, Above diagonal: Mean distances

	1	2	3	4	5	6	7	8
1 Cangul	-	0.084	0.091	0.069	0.091	0.109	0.102	0.102
2 Hsigsig	38	-	0.016	0.069	0.116	0.138	0.136	0.136
3 Hsigcaf	41	7	-	0.067	0.111	0.138	0.136	0.136
4 Hboul	31	31	30	-	0.104	0.122	0.120	0.116
5 Hfem	41	52	50	47	-	0.102	0.100	0.100
6 Hareolat	49	62	62	55	46	-	0.027	0.031
7 HaHerm1	46	61	61	54	45	12	-	0.004
8 HaHerm2	46	61	61	52	45	14	2	-
9 HaDeHoop	50	62	63	54	47	13	7	7
10 HaMUmale	47	58	58	51	44	12	6	6
11 HaMUfema	47	58	58	51	44	12	6	6
12 HaMTmale	45	60	60	51	44	12	3	3
13 HaMTfema	48	61	61	54	45	1	11	13
14 HaEland	47	62	62	53	44	2	10	12
15 HaSUfema	47	61	60	51	48	15	11	11
16 Kbelbel	46	58	58	48	50	53	54	54
17 Klobatsi	53	61	63	57	60	62	55	57
18 Kspekii	47	61	59	49	50	53	55	57
19 Knat	44	58	58	48	50	51	53	55
20 Mtorner	56	61	67	57	54	58	53	55
21 Ptentt	48	50	51	41	44	54	55	55
22 Ptentv	67	65	64	61	56	55	61	62
23 Poculifr	47	46	49	43	55	50	51	51
24 Pgeo1	53	55	56	50	56	58	61	61
25 Pgeo2	52	54	55	49	55	57	60	60
26 Tmargin	54	60	61	62	59	55	58	60
27 Thermani	41	49	50	43	52	50	55	54
28 Gpardp1	40	50	52	44	41	45	42	42
29 Gpardp2	45	49	49	47	42	44	45	45
30 Gpardb	44	52	52	50	46	44	47	47
31 CHscrip1	57	62	62	59	61	58	58	60
32 CHscrip2	57	62	62	59	63	60	60	62
33 PELsub	43	49	51	48	43	40	41	40
34 Eaustral	45	51	51	51	40	42	43	45

Pairwise distances between taxa (continued)

	9	10	11	12	13	14	15	16
1 Cangul	0.111	0.104	0.104	0.100	0.107	0.104	0.104	0.102
2 Hsigsig	0.138	0.129	0.129	0.133	0.136	0.138	0.136	0.129
3 Hsigcaf	0.140	0.129	0.129	0.133	0.136	0.138	0.133	0.129
4 Hboul	0.120	0.113	0.113	0.113	0.120	0.118	0.113	0.107
5 Hfem	0.104	0.098	0.098	0.098	0.100	0.098	0.107	0.111
6 Hareolat	0.029	0.027	0.027	0.027	0.002	0.004	0.033	0.118
7 HaHerm1	0.016	0.013	0.013	0.007	0.024	0.022	0.024	0.120
8 HaHerm2	0.016	0.013	0.013	0.007	0.029	0.027	0.024	0.120
9 HaDeHoop	-	0.016	0.016	0.011	0.027	0.024	0.027	0.129
10 HaMUmale	7	-	0.000	0.013	0.024	0.022	0.029	0.122
11 HaMUfema	7	0	-	0.013	0.024	0.022	0.029	0.122
12 HaMTmale	5	6	6	-	0.024	0.022	0.022	0.118
13 HaMTfema	12	11	11	11	-	0.002	0.036	0.116
14 HaEland	11	10	10	10	1	-	0.033	0.113
15 HaSUfema	12	13	13	10	16	15	-	0.122
16 Kbelbel	58	55	55	53	52	51	55	-
17 Klobatsi	61	56	56	56	61	60	60	22
18 Kspekii	57	58	58	54	52	51	58	28
19 Knat	55	56	56	52	50	49	56	29
20 Mtorner	58	54	54	54	57	56	59	36
21 Ptentt	55	52	52	54	53	52	56	52
22 Ptentv	62	57	57	63	54	55	66	62
23 Poculifr	51	50	50	50	49	50	54	51
24 Pgeo1	61	58	58	60	57	58	60	54
25 Pgeo2	60	57	57	59	56	57	59	53
26 Tmargin	59	54	54	58	54	55	61	64
27 Thermani	54	53	53	53	49	50	55	52
28 Gpardp1	46	43	43	41	44	43	43	48
29 Gpardp2	47	46	46	44	43	44	44	51
30 Gpardb	49	48	48	46	43	44	47	49
31 CHscrip1	60	59	59	57	57	58	63	64
32 CHscrip2	62	61	61	59	59	60	63	64
33 PELsub	42	41	41	39	39	40	43	51
34 Eaustral	45	44	44	42	41	42	46	52

Pairwise distances between taxa (continued)

	17	18	19	20	21	22	23	24
1 Cangul	0.118	0.104	0.098	0.124	0.107	0.149	0.104	0.118
2 Hsigsig	0.136	0.136	0.129	0.136	0.111	0.144	0.102	0.122
3 Hsigcaf	0.140	0.131	0.129	0.149	0.113	0.142	0.109	0.124
4 Hboul	0.127	0.109	0.107	0.127	0.091	0.136	0.096	0.111
5 Hfem	0.133	0.111	0.111	0.120	0.098	0.124	0.122	0.124
6 Hareolat	0.138	0.118	0.113	0.129	0.120	0.122	0.111	0.129
7 HaHerm1	0.122	0.122	0.118	0.118	0.122	0.136	0.113	0.136
8 HaHerm2	0.127	0.127	0.122	0.122	0.122	0.138	0.113	0.136
9 HaDeHoop	0.136	0.127	0.122	0.129	0.122	0.138	0.113	0.136
10 HaMUmale	0.124	0.129	0.124	0.120	0.116	0.127	0.111	0.129
11 HaMUfema	0.124	0.129	0.124	0.120	0.116	0.127	0.111	0.129
12 HaMTmale	0.124	0.120	0.116	0.120	0.120	0.140	0.111	0.133
13 HaMTfema	0.136	0.116	0.111	0.127	0.118	0.120	0.109	0.127
14 HaEland	0.133	0.113	0.109	0.124	0.116	0.122	0.111	0.129
15 HaSUfema	0.133	0.129	0.124	0.131	0.124	0.147	0.120	0.133
16 Kbelbel	0.049	0.062	0.064	0.080	0.116	0.138	0.113	0.120
17 Klobatsi	-	0.073	0.076	0.060	0.122	0.147	0.129	0.136
18 Kspekii	33	-	0.011	0.107	0.116	0.138	0.127	0.133
19 Knat	34	5	-	0.104	0.113	0.140	0.124	0.131
20 Mtorner	27	48	47	-	0.136	0.151	0.138	0.144
21 Ptentt	55	52	51	61	-	0.064	0.064	0.076
22 Ptentv	66	62	63	68	29	-	0.087	0.098
23 Poculifr	58	57	56	62	29	39	-	0.044
24 Pgeo1	61	60	59	65	34	44	20	-
25 Pgeo2	60	59	58	64	33	43	19	1
26 Tmargin	65	63	60	64	60	66	56	60
27 Thermani	59	54	51	57	38	44	30	28
28 Gpardp1	52	55	54	58	45	61	46	52
29 Gpardp2	55	54	51	61	42	56	45	51
30 Gpardb	53	51	48	61	42	52	46	49
31 CHscrip1	60	59	58	64	54	60	60	64
32 CHscrip2	60	61	60	66	54	61	59	63
33 PELsub	57	54	51	57	46	56	47	47
34 Eaustral	56	51	48	60	40	51	45	48

Pairwise distances between taxa (continued)

	25	26	27	28	29	30	31	32
1 Cangul	0.116	0.120	0.091	0.089	0.100	0.098	0.127	0.127
2 Hsigsig	0.120	0.133	0.109	0.111	0.109	0.116	0.138	0.138
3 Hsigcaf	0.122	0.136	0.111	0.116	0.109	0.116	0.138	0.138
4 Hboul	0.109	0.138	0.096	0.098	0.104	0.111	0.131	0.131
5 Hfem	0.122	0.131	0.116	0.091	0.093	0.102	0.136	0.140
6 Hareolat	0.127	0.122	0.111	0.100	0.098	0.098	0.129	0.133
7 HaHerm1	0.133	0.129	0.122	0.093	0.100	0.104	0.129	0.133
8 HaHerm2	0.133	0.133	0.120	0.093	0.100	0.104	0.133	0.138
9 HaDeHoop	0.133	0.131	0.120	0.102	0.104	0.109	0.133	0.138
10 HaMUmale	0.127	0.120	0.118	0.096	0.102	0.107	0.131	0.136
11 HaMUfema	0.127	0.120	0.118	0.096	0.102	0.107	0.131	0.136
12 HaMTmale	0.131	0.129	0.118	0.091	0.098	0.102	0.127	0.131
13 HaMTfema	0.124	0.120	0.109	0.098	0.096	0.096	0.127	0.131
14 HaEland	0.127	0.122	0.111	0.096	0.098	0.098	0.129	0.133
15 HaSUfema	0.131	0.136	0.122	0.096	0.098	0.104	0.140	0.140
16 Kbelbel	0.118	0.142	0.116	0.107	0.113	0.109	0.142	0.142
17 Klobatsi	0.113	0.144	0.131	0.116	0.122	0.118	0.133	0.133
18 Kspekii	0.131	0.140	0.120	0.122	0.120	0.113	0.131	0.136
19 Knat	0.129	0.133	0.113	0.120	0.113	0.107	0.129	0.133
20 Mtorner	0.142	0.142	0.127	0.129	0.136	0.136	0.142	0.147
21 Ptentt	0.073	0.133	0.084	0.100	0.093	0.093	0.120	0.120
22 Ptentv	0.096	0.147	0.098	0.136	0.124	0.116	0.133	0.136
23 Poculifr	0.042	0.124	0.067	0.102	0.100	0.102	0.133	0.131
24 Pgeo1	0.002	0.133	0.062	0.116	0.113	0.109	0.142	0.140
25 Pgeo2	-	0.136	0.060	0.113	0.111	0.107	0.144	0.142
26 Tmargin	61	-	0.102	0.116	0.113	0.109	0.118	0.122
27 Thermani	27	46	-	0.107	0.109	0.091	0.120	0.124
28 Gpardp1	51	52	48	-	0.020	0.067	0.091	0.091
29 Gpardp2	50	51	49	9	-	0.047	0.089	0.089
30 Gpardb	48	49	41	30	21	-	0.093	0.096
31 CHscrip1	65	53	54	41	40	42	-	0.009
32 CHscrip2	64	55	56	41	40	43	4	-
33 PELsub	48	44	39	17	12	20	40	42
34 Eaustral	49	46	46	19	12	18	39	41

Pairwise distances between taxa (continued)

		33	34
1	Cangul	0.096	0.100
2	Hsigsig	0.109	0.113
3	Hsigcaf	0.113	0.113
4	Hboul	0.107	0.113
5	Hfem	0.096	0.089
6	Hareolat	0.089	0.093
7	HaHerm1	0.091	0.096
8	HaHerm2	0.089	0.100
9	HaDeHoop	0.093	0.100
10	HaMUmale	0.091	0.098
11	HaMUfema	0.091	0.098
12	HaMTmale	0.087	0.093
13	HaMTfema	0.087	0.091
14	HaEland	0.089	0.093
15	HaSUfema	0.096	0.102
16	Kbelbel	0.113	0.116
17	Klobatsi	0.127	0.124
18	Kspekii	0.120	0.113
19	Knat	0.113	0.107
20	Mtorner	0.127	0.113
21	Ptentt	0.102	0.089
22	Ptentv	0.124	0.113
23	Poculifr	0.104	0.100
24	Pgeo1	0.104	0.107
25	Pgeo2	0.107	0.109
26	Tmargin	0.098	0.102
27	Thermani	0.087	0.102
28	Gpardp1	0.038	0.042
29	Gpardp2	0.027	0.027
30	Gpardb	0.044	0.040
31	CHscrip1	0.089	0.087
32	CHscrip2	0.093	0.091
33	PELsub	-	0.027
34	Eaustral	12	-

References:

Adachi J., Y. Cao, and M. Hasegawa. **1993**. Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: rapid evolution in warm blooded vertebrates. Journal of Molecular Evolution 36: 270-81.

Allard M. W., M. M. Miyamoto, K. A. Bjorndal, A. B. Bolten, and B. B. Bowen. **1994**. Support for natal homing in green turtles from mitochondrial DNA sequences. Copeia: 34-41.

Andrews C. **1906**. A descriptive catalogue of the Tertiary Vertebrata of the Fayum, Egypt. British Mus. (Nat. Hist.): 1-324.

Auffenberg W. **1974**. Checklist of fossil land tortoises (Testudinidae). Bulletin of the Florida State Museum, Biological Sciences 3: 121-251.

Avise J. C. and R. A. Lansman. **1983**. Polymorphism of mitochondrial DNA in populations of higher animals. Chap. in Evolution of Genes and Proteins. Sunderland, Mass. Sinauer. 147-64.

Avise J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. **1987**. Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecological Systematics: 489-522.

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

Avise J. C., B. B. Bowen, T. Lamb, A. B. Meylan, and E. Bermingham. **1992**. Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. Molecular Biology and Evolution 3: 457-73.

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

Baard E. H. W. **1990**. Biological aspects and conservation status of the geometric tortoise, *Psammobates geometricus* (Linneaus, 1758)(Cryptodira: Testudinidae). University of Stellenbosch.

Baard E. H. W. **1993**. Distribution and status of the geometric tortoise *Psammobates geometricus* in South Africa. Biological Conservation 63: 235-9.

Baard E. H. W. and P. le F. N. Mouton. **1993**. A hypothesis expanding the enigmatic distribution of the geometric tortoise *Psammobates geometricus*, in South Africa. Herpetological Journal: 65-7.

Ballard J. W. O. and M. Kreitman. **1995**. Is mitochondrial DNA a strictly neutral marker? Trends in Evolutionary Ecology 12: 485-8.

Bayoff N. **1995**. Observations and morphometric data on the Namaqualand speckled tortoise, *Homopus signatus signatus* (Gmelin, 1789), in South Africa. Chelonian Conservation and Biology 3: 215-20.

Bellairs A. A. **1975**. Reptiles. 4th ed. London: Hutchinson & Co, Ltd. ed. Attridge J. 1-233.

Benirschke K. **1985**. The genetic management of exotic animals. Oxford: Oxford Science Publications. eds. Hearn J. P. and J. K. Hodges. 71-87.

Benton M. J. **1990a**. Phylogeny of the major tetrapod groups: morphological data and divergence dates. Journal of Molecular Evolution 30: 409-24.

Benton M. J. **1990b**. Vertebrate Palaeontology. London: Unwin Hyman Ltd. 1-377.

Bickham J. W. and J. L. Carr. **1983**. Taxonomy and phylogeny of the higher categories of Cryptodiran turtles based on cladistic analysis of chromosomal data. Copeia 4: 918-32.

Bowen B. B., A. B. Meylan, and J. C. Avise. **1989**. An odyssey of the green sea turtle: Ascension Island revisited. Proceedings in the National Academy of Science USA: 573-6.

Bowen B. B., A. B. Meylan, and J. C. Avise. **1991**. Evolutionary distinctiveness of the endangered Kemp's ridley sea turtle. Nature: 709-11.

Bowen B. B., A. B. Meylan, C. J. Limpus, G. H. Balazs, and J. C. Avise. **1992**. Global population structure and natural history of the green turtle (*Chelonia mydas*) in terms of matriarchal phylogeny. Evolution 46: 865-81.

Bowen B. B., W. S. Nelson, and J. C. Avise. **1993**. A molecular phylogeny for marine turtles: Trait mapping, rate assessment and conservation relevance. Proceedings in the National Academy of Science USA: 5574-7.

Bowen B. B., N. Kamezaki, C. L. Limpus, G. R. Hughes, A. B. Meylan, and J. C. Avise. **1994**. Global phylogeography of the loggerhead turtle (*Caretta caretta*) ad indicated by mitochondrial DNA haplotypes. Evolution 6: 1820-8.

Boycott R. **1986**. A review of *Homopus signatus* (Schoepff) with notes on related species (Cryptodira: Testudinidae). Journal of the Herpetological Association of Africa 32: 10-6.

Boycott R. C. and O. Bourquin. **1988**. The South African tortoise book. A guide to South African tortoises, terrapins and turtles. Johannesburg: Southern Book Publishers. 1-148

Boycott R. C. and N. H. G. Jacobsen. **1988**. On the distribution, habitat and identification of *Kinixys natalensis* Hewitt, 1935 (Cryptodira: Testudinidae) in southern Africa. Durban Museum Noviates 5: 93-101.

Branch W. R. **1988**a. Field guide to the snakes and other reptiles of southern Africa. Cape Town: Struik

Branch, W. R. **1988**b. South African Red Data Book - reptiles and amphibians. S. Afr. Nat. Sci. Prog. Rpt. 1-241.

Branch W. R. **1992**. *Homopus "bergeri"* - A wrong name for a new tortoise from southern Namibia. Journal of the Herpetological Association of Africa 40: 11

Branch W. R., G. A. Benn, and A. T. Lombard. **1995.** The tortoises (Testudinidae) and terrapins (Pelomedsidae) of southern Africa: their diversity, distribution and conservation. South African Journal of Zoology 30: 91-102.

Broadley D. G. **1962**. Some fossil Chelonian fragments from Makapansgat. Nature: 791-2.

Broadley D. G. **1981**. A review of the populations of *Kinixys* (Testudinidae) occurring in south-eastern Africa. Annals of the Cape Provincial Museums (Natural History) 14: 195-216.

Broadley, D. G. **1989**. The conservation biology of tortoises. The Durrell Institute of Conservation Ecology: 48-61.

Broadley D. G. **1992**. The savannah species of *Kinixys* (Testudinidae). Journal of the Herpetological Association of Africa 40: 12-3.

Broadley D. G. **1993**. A review of the southern African species of *Kinixys bell* (Reptilia: Testudinidae). Annals of the Transvaal Museum 6: 41-52.

Brown W. M., M. George, and A. C. Wilson. **1979**. Rapid evolution of animal mitochondrial DNA. Proceedings in the National Academy of Science USA 4: 1967-71.

Brown W. M., E. M. Prager, A. Wang, and A. C. Wilson. **1982**. Mitochondrial DNA sequences of primates: tempo and mode of evolution. Journal of Molecular Evolution 18: 225-39.

Carlson S. S, A. C. Wilson, and R. D. Maxson. **1978**. Do albumin clocks run on time? Science 200: 1183-5.

Carr S. M., A. J. Brothers, and A. C. Wilson. **1987**. Evolutionary inferences from restriction maps of mitochondrial DNA from nine taxa of *Xenopus* frogs. Evolution 41: 176-90.

Carroll R. L. **1988**. Vetebrate paleontology and evolution. New York: W.H. Freeman

Clark C. and D. J. Curran. **1986**. Outgroup analysis, homoplasy and global parsimony: a response to Maddison, Donoghue, and Maddison. Systematic Zoology: 422-6.

Cooper M. R. and D. G. Broadley. **1990**. A new species of fossil *Homopus*. Studia Geologica Salmanticensia 4: 41-55.

Crumly C. R. **1982**. A cladistic analysis of *Geochelone* using cranial osteology. Journal of Herpetology 3: 215-34.

Crumly C. R. **1984**. The evolution of land tortoises (family Testudinidae). Rutgers University.

Cummings M. P., S. P. Otto, and J. Wakeley. **1995**. Sampling properties of DNA sequence data in phylogenetic analysis. Molecular Biology and Evolution 5: 814-22.

Duerden J. E. **1907**. Variations in the *Geometrica* - group of South African tortoises. Report of the South African Association for the Advancement of Science: 178-208.

Dumeril A. M. C. and G. Bibron. **1835**. Erpetologie generale ou histoire naturelle complete des reptiles. Paris: 1-680.

Dutton P. H., S. K. Davis, T. Guerra, and D. Owens. **1996**. Molecular phylogeny for marine turtles based on sequences of the ND4-leucine tRNA and control regions of mitochondrial DNA. Molecular Phylogenetics and Evolution 3: 511-21.

Eck R. V. and M. O. Dayoff. **1966**. Atlas of protein sequence and structure. Silver Springs: National Biomedical Research Foundation

Estes R. D. and J. H. Hutchinson. **1980**. Eocene lower vertebrates from Ellsemere Island, Canadian Archipeligo. Paleogeography, Palaeoclimatology, Palaeoecology: 325-47.

Farris J. S. **1986**. Synapomorphy, parsimony, and evidence. Taxon: 298-315.

Felsenstein J. **1978**. Cases in which parsimony and compatibility methods will be positively misleading. Systematic Zoology: 401-10.

Felsenstein J. **1985**. Confidence limits on phylogenies: an approach using the bootstrap. Evolution: 783-91.

Felsenstein J. **1988**. Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics 22: 521-65.

Fitch W. M. **1971**. Toward defining the course of evolution: minimal change for a specific tree topology. Systematic Zoology 20: 406-16.

Fitch W. M. **1977**. On the problem of discovering the most parsimonious tree. Am. Natur. 111: 223-57.

Fitch W. M. **1986**. An estimation of the number of invariable sites is necessary for the accurate estimation of the number of nucleotide substitutions since a common ancestor. Chap. in Evolutionary perspectives and the new genetics. New York: Alan R. Liss. 149-59.

Fitzinger L. J. F. J. **1835**. Entwurf einer systematichen anordungen der schildkroten nach den grundsatzen der naturlichen methode. Annals of the Wien Museum: 103-28.

Gaffney E. S. **1975**. A phylogeny and classification of the higher categories of turtles. Bulletin of the American Museum of Natural History 155: 387-436.

Gaffney E. S. and P. A. Meylan. **1988**. A phylogeny of turtles. Chap. in The phylogeny and classification of tetrapods, amphibians, reptiles and birds. Oxford: Clarendon Press. 157-219.

Gauthier J. A., A. G. Kluge, and T. Rowe. **1988**. Amniote phylogeny and the importance of fossils. Cladistcs 4: 105-209.

Gray J. E. **1825**. A synopsis of the genera of reptiles and amphibians with a description of some new species. Annals of the Philosophical Society: 193-217.

Graybeal B. **1993**. The phylogenetic utility of cytochrome *b*: lessons from Bufonid frogs. Molecular Phylogenetics and Evolution 3: 256-69.

Graybeal B. **1994**. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. Systematic Biology 2: 174-93.

Greig J. C. and P. D. Burdett. **1976**. Patterns in the distribution of southern African terrestrial tortoises (Cryptodira: Testudinidae). Zoologica Africana 2: 249-73.

Grimaldi D. A. **1996**. Captured in Amber. Scientific American: 71-7.

Groombidge B. **1982**. Part 1. Testudines, Crocodylia, Rhynchocephalia. Chap. in The IUCN Amphibia-Reptilia Red Data Book. Gland, Switzerland: IUCN.

Gu L. and W. H. Li. **1993**. Higher rates of amino acid substitution in rodents than in man. Molecular Phylogenetics and Evolution 1: 211-4.

Gyllensten U. B., M. Allen, and A. Josefsson. **1992**. Sequencing of *in vitro* amplified DNA. Chap. in The PCR technique: DNA sequencing. Natic MA: Eaton Publishing Co. 1-15.

Harley E.E.H. **1997**. DAPSA: DNA And Protein Sequence Analysis. Ver. 3.9. (Computer Program) Department of Chemical Pathology, University of Cape Town, Cape Town.

Harrison R. G. **1989**. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Evolutionary Ecology 1: 6-11.

Hedges S. B. **1992**. The number of replications needed for accurate estimation of the boostrap *P* value in phylogenetic studies. Molecular Biology and Evolution: 366-9.

Hewitt J. **1931**. Descriptions of some African tortoises. Annals of the Natal Museum 6: 461-506.

Hewitt J. **1935**. Some new forms of batrachians and reptiles from South Africa. Rec. Albany Mus. 4: 283-357.

Hewitt J. **1937**. A note on the relationships of the Cape genera of land tortoises. South African Journal of Science: 788-96.

Hickman C. P., L. S. Roberts, and A. Larson. **1993**. Integrated principles of Zoology. 9th ed. London: Mosby Publishers. 1-843.

Hillis D. M. **1987**. Molecular versus morphological approaches to systematics. Annual Review of Ecological Systematics 18: 23-42.

Hillis D. M., A. Larson, S. K. Davis, and F. A. Zimmer. **1990**. Nucleic Acids III: Sequencing. Chap. in Molecular Systematics. Sunderland, Massachusetts: Sinauer. 318-70.

Hillis D. M. and C. Moritz. **1990**. An overview of applications of molecular systematics. Chap. in Molecular Systematics. Sunderland, Massachusetts: Sinauer Associates, Inc. 502-15.

Hillis D. M., J. P. Huelsenbeck, and C. W. Cunningham. **1994**. Application and accuracy of molecular phylogenies. Science: 671-7.

Hillis D. M. **1995**. Approaches for assessing phylogenetic accuracy. Systematic Biology 1: 3-16.

Holman J. A. **1971**. Climatic significance of giant tortoises from Wood Mountain Formation (Upper Miocene) of Saskatchewan. Canadian Journal of Earth Science 9: 1148-51.

Huelsenbeck J. P. **1995**. The performance of phylogenetic methods in simulation. Systematic Biology 44: 17-48.

Huelsenbeck J. P., J. J. Bull, and C. W. Cunningham. **1996**. Combining data in phylogenetic analysis. Trends in Evolutionary Ecology 4: 152-8.

IUCN **1993**. Tortoises and Freshwater Turtles - An action plan for their conservation. International Union for the Conservation of Nature and Natural Resources. 1-47.

Irwin D. M., T. D. Kocher, and A. C. Wilson. **1991**. Evolution of the cytochrome *b* gene of mammals. Journal of Molecular Evolution: 128-44.

Ishida, N., Oyunsuren, T., Mashima, S., Mukoyama, H., and Saitou, N. **1995** Mitochondrial DNA sequences of various species of the genus *Equus* with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. Journal of Molecular Evolution 41: 180-188.

Iverson, J. B. **1992**. A revised checklist with distribution maps of the turtles of the world. Richmond, Indiana: Privately printed. 1-291.

Kemp T. S. **1980**. Origin of the mammal-like reptiles. Nature: 378-80.

Kimura M. **1983**. The neutral theory of molecular evolution. Cambridge, England: Cambridge University Press. 94-104

Klemens M. W. and D. Moll. **1995**. An assessment of the effects of commercial exploitation on the pancake tortoise, *Malacochersus tornieri*, in Tanzania. Chelonian Conservation and Biology 3: 197-206.

Kocher T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Paabo, F. X. Villablanca, and A. C. Wilson. **1989**. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. Proceedings in the National Academy of Science USA: 6196-200.

Kumar S., K. Tamura, and M. Nei. **1993**. MEGA: Molecular Evolutionary Genetics Analysis. Ver 1.0. (Computer Program) The Pennsylvania State University:

Kumazawa Y. and M. Nishida. **1995**. Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. Molecular Biology and Evolution 12: 759-72.

Lahanas P. L., M. M. Miyamotot, K. A. Bjorndal, and A. B. Bolten. **1994**. Molecular evolution and population genetics of Greater Caribbean green turtles (*Chelonia mydas*) as inferred from mitochondrial control region. Genetics 94: 57-67.

Lamb T. and C. Lydeard. **1994**. A molecular phylogeny of the Gopher tortoises, with comments on familial relationships within the Testudinoidea. Molecular Phylogenetics and Evolution 4: 283-91.

Lamb T., C. Lydeard, R. B. Walker, and J. W. Gibbons. **1994**. Molecular systematics of map turtles (*Graptemys*): a comparison of mitochondrial restriction site versus sequence data. Systematic Biology 4: 543-59.

Laurent L., J. Lescure, L. Excoffier, B. B. Bowen, M. Domingo, M. Hadjichristophorou, L. Kornaraki, and G. Trabuchet. **1993**. Genetic studies of relationships between Mediterranean and Atlantic populations of lggerhead turtle *Caretta carreta* with a mitochondrial marker. C. R. Academy of Science Paris 316: 1233-9.

Laurin M. and R. R. Reisz. **1995**. A reevaluation of early amniote phylogeny. Zoological Journal of the Linnean Society: 165-223.

Lee M. S. **1996**. Correlated progression and the origin of turtles. Nature 29: 812-5.

Lehrer J. **1990**. Turtles and Tortoises. New York: Mallard Press. ed. Kalman S. 1-127

Lewin, R. **1988.** Molecular clocks turn a quarter century. Science 2: 561-563.

Li W. H., M. Tanimura, and P. M. Sharp. **1987**. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. Journal of Molecular Evolution 25: 330-42.

Li W. H. and D. Graur. **1991**. Fundamentals of molecular evolution. Sunderland, Massachusetts: Sinauer

Li, W.H. **1993.** So, what about the molecular clock hypothesis? Current Opinion in Genetics and Development 3:896-901.

Linnaeus C. **1758**. Systema Naturae. Holmiae 1: 1-824.

Loveridge A. **1935**. Scientific results on an expedition to rainforest regions in Eastern Africa. New reptiles and amphibians from East Africa. Bull. Mus. Comp. Zool: 1-19.

Loveridge A. and E. E. Williams. **1957**. Revision of the African tortoises and turtles of the suborder Cryptodira. Bulletin: Museum of Comparative Zoology Harvard 6: 161-557.

Maddison W. P., M. J. Donoghue, and D. R. Maddison. **1984**. Outgroup analysis and parsimony. Systematic Zoology: 83-103.

Martin P. A., G. J. P. Naylor, and S. R. Palumbi. **1992**. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature 357: 153-5.

Martin P. A. and S. R. Palumbi. **1993**. Body size, metabolic rate, generation time and the molecular clock. Proceedings in the National Academy of Science USA: 4087-91.

Mathew C. G. **1991**. Protocols in molecular human genetics. Clifton, New Jersey: Humana Press. 1-461.

Maxam A. M. and W. Gilbert. **1977**. A new method for sequencing DNA. Proceedings in the National Academy of Science USA: 560-4.

Maxam A. M. and W. Gilbert. **1980**. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods in Enzymology: 499-559.

Meyer A. and A. C. Wilson. **1990**. Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. Journal of Molecular Evolution 31: 359-64.

Meyer A. **1994**. Shortcomings of the cytochrome *b* gene as a molecular marker. Trends in Evolutionary Ecology 8: 278-80.

Meylan P. and W. Auffenberg. **1986**. New land tortoises (Testudines: Testudinidae) from the Miocene of Africa. Zoological Journal of the Linnean Society 86: 279-307.

Meylan A. B., B. B. Bowen, and J. C. Avise. **1990**. A genetic test of the natal homing versus social facilitation models for green turtle migration. Science: 724-7.

Miller K. G. and R. G. Fairbanks. **1985**. Cainozoic record of climate and sea level. South African Journal of Science 5: 248-9.

Miyamoto M. M. and W. M. Fitch. **1995**. Testing species phylogenies and phylogenetic methods with congruence. Systematic Biology 1: 64-76.

Mooers A. O. **1995**. Tree balance and tree completeness. Evolution 2: 379-84.

Morin P. A., J. Messier, and D. S. Woodruff. **1994**. DNA extraction amplification and direct sequencing from hornbill feathers. Journal of the Science Society of Thailand: 31-41.

Mouton P. le F. N. and B. W. Oelofsen. **1988**. A model explaining patterns of geographic character variation in *Cordylus cordylus* (Reptilia; Cordylidae) in the south-western Cape, South Africa. South African Journal of Zoology 1: 20-31.

Mullis K. B. **1990**. The unusual origin of the polymerase chain reaction. Scientific American April: 104-11.

Muse S. V. **1996**. Estimating synonymous and nonsynonymous substitution rates. Molecular Biology and Evolution 1: 105-14.

Nei M. **1987**. Molecular evolutionary genetics. New York: Columbia University Press. 1-512 .

Neyman J. **1971**. Molecular studies of evolution: a source of novel statistical problems. Chap. in Statistical decision theory and related topics. New York: Academic Press. 1-27.

Nichols R. A. **1989**. Mitochondrial DNA clues to gopher tortoise dispersal. Trends in Evolutionary Ecology 7: 192-3.

Norman, J.A., Moritz, C., and Limpus, C.J. **1994.** Mitochondrial DNA control region polymorphisms: genetic markers for ecological studies of marine turtles. Molecular Ecology 3:363-373.

Nunn C. L. **1995**. A simulation test of Smith's "degrees of freedom" correction for comparative studies. American Journal of Physical Anthropology: 355-67.

Oligo 3.4. (Computer Program) National Biosciences Inc. Plymouth MN, USA.

Powell J. R., A. Caccone, G. D. Amato, and C. Yoon. **1986**. Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nulcear DNA are similar. Proceedings in the National Academy of Science USA: 9090-3.

Quicke D. L. J. **1993**. Tertiary Level Biology: Principles and Techniques of Contemporary Taxonomy. 1st ed. London: Blackie Academic & Professional. 41-285.

Rand D. M. **1994**. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. Trends in Evolutionary Ecology 4: 125-31.

Richter C., J. W. Park, and B. N. Ames. **1988**. Normal oxidative damage to mitochondrial and nuclear DNA is extensive. Proceedings in the National Academy of Science USA 85: 6465-7.

Ridley M. **1993**. Evolution. Boston: Blackwell Scientific Publications. 1-670

Romer A. S. **1966**. Vertebrate paleontology. 3rd ed. Chicago: University of Chicago Press.

Saitou N. and M. Nei. **1987**. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-25.

Saitou N. and T. Imanishi. **1989**. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum evolution, neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Molecular Biology and Evolution: 514-25.

Sanger F., S. Nicklen, and A. R. Coulson. **1977**. DNA sequencing with chain-terminating inhibitors. Proceedings in the National Academy of Science USA: 5463-7.

Schlotterer C., B. Amos, and D. Tautz. **1991**. Conservation of polymorphic simple sequence loci in cetacean species. Nature: 63-5.

Schluter D. **1995**. Uncertainty in ancient phylogenies. Nature: 108-9.

Schmidt W. **1995**. Phylogeny reconstruction for protein sequences based on amino acid properties. Journal of Molecular Evolution: 522-30.

Service R. F. **1996**. Just how old is that DNA, anyway? Science: 810.

Shaffer, B.H., Meylan, P., and McKnight, M.L. **1997**. Tests of turtle phylogeny: molecular, morphological. and paleontological approaches. Systematic Biology 46(2):235-268.

Siebenrock F.**1904**. Die sudafrikanischen Testudo-Arten der *Geometrica*-Gruppe. Sitzungsb. Akad. Wiss. Wein 113: 307-323.

Stanley H. F., M. Kadwell, and J. C. Wheeler. **1994**. Molecular evolution of the family Camelidae: a mitochondrial DNA study. Proceedings in the Royal Society of London: 1-6.

Stewart C-B. **1993**. The powers and pitfalls of parsimony. Nature: 603-7.

Stromer E. V. **1926**. Reste land - und susswasser - bewohnender Wirbeltiere aus den Diamnat - feldern Deutsch -Sudwestafrikas. Die Diamantewuste Sudwestafricas 2: 139-42.

Swofford D.L. **1993.** PAUP: Phylogenetic Analysis Using Parsimony. Ver 3.1. (Computer Program) Illinois Natural History Survey, Champaign, Illinois.

Swofford D.L. **1996.** PAUP: Phylogenetic Analysis Using Parsimony. Ver Beta. (Computer Program)

Swofford D. L. and G. J. Olsen. **1990**. Phylogeny Reconstruction. Chap. in Molecular Systematics. Sunderland, Massachusetts: Sinauer Associates, Inc. 411-501.

Tajima F. **1993**. Evolutionary relationships of DNA sequences in finite populations. Genetics: 437-60.

Tateno Y., N. Takezaki, and M. Nei. **1994**. Relative efficiencies of the maximum-likelihood, neighbor-joining and maximum-parsimony methods when substitution rate varies with site. Molecular Biology and Evolution: 261-77.

Thomas W. K. and A. T. Bechenbach. **1989**. Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. Journal of Molecular Evolution 29: 233-45.

Vawter L. and W. M. Brown. **1986**. Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science: 194-5.

Voet D. and J. G. Voet. **1990**. Biochemistry. New York: John Wiley and Sons

Walker, D., Burke, V.J., Barak, I., and Avise, J.C. **1995.** A comparison of mtDNA restriction sites vs. control region sequences in phylogeographic assessment of the musk turtle (*Sernotherus minor*). Molecular Ecology 4:365-373.

Weindurch R. **1996**. Caloric restirction and aging. Scientific American January: 32-8.

Weir B. S. **1990**. Intraspecific Differentiation. Chap. in Molecular Systematics. Sunderland, Massachusetts: Sinauer Associates, Inc. 373-410.

Wu C. I. and W. H. Li. **1985**. Evidence for higher rates of nucleotide substitutions in rodents than in man. Proceedings in the National Academy of Science USA: 1741-5.

Yang Z., N. Goldman, and A. Friday. **1994**. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Molecular Biology and Evolution 2: 316-24.

Yang Z. and T. Wang. **1995**. Mixed models analysis of DNA sequence evolution. Biometrics: 552-61.

Yang H., E. M. Goldenberg, and J. Shoshani. **1996**. Phylogenetic resolution within the Elephantidae using fossil DNA sequence from the American mastodon (*Mammut americanum*) as an outgroup. Proceedings in the National Academy of Science USA: 1190-4.

Zhang D-X and G. M. Hewitt. **1996**. Nuclear integrations: challenges for mitochondrial DNA markers. Trends in Evolutionary Ecology: 247-51.

Zuckerkandl E. and L. Pauling. **1962**. Molecular disease, evolution and genetic heterogeneity. Chap. in Horizons in Biochemistry. New York: Academic Press. 189-225.

Zugg G. R. **1993**. Herpetology - An introductory biology of amphibians and reptiles. San Diego: Academic Press. 1-456