THE BIOCHEMICAL SYSTEMATICS OF THE SOUTHERN AFRICAN FELIDAE

BY

NOMUSA Y. MDA BSc. (Med)(Hons)

A thesis presented in fulfilment of the requirements for the degree of Master of Science in the Faculty of Medicine of the University of Cape Town.

JUNE 1992

Promoted by Professor E.H. Harley M.D., Ph.D

This thesis is dedicated to my mother and aunt Ntomobomzi, aunt Sisa, Pamela and Namhla.

ACKNOWLEDGEMENT

I wish to express my sincere gratitude and thanks to Professor Harley for his expert guidance and enthusiasm for this study. I am also grateful to Dr Faadiel Essop for his excellent and endless support throughout the course of this study. I am indebted to those individuals and institutions who selflessly laboured to provide suitable source of biological material, without which this study would not have been possible. I also wish to acknowledge Dr Michael Cherry and Dr Thandi Sikakana for their constructive criticisms during the write-up of this thesis. Thanks go to Ms Nadiema Jackson for typing this thesis in such a proficient, patient and cheerful manner.

ABSTRACT

The classification of the family *Felidae* (cats) is problematical due to the conservative nature of their morphology. Some workers classify the family into as many as 20 genera (Ewer, 1973) while others divide it into three genera (Walker *et al.*, 1964). Such studies have largely been based on morphological and behavioural characters. Recently, molecular studies, namely, protein albumin immunological distances (Collier and O'Brien, 1985) and protein electrophoresis (Randi and Ragni, 1991) have been used to try and resolve the problems underlying this family.

To complement the previous studies, in the present study we use mitochondrial (mt) DNA to construct a phylogeny of eight members of the southern African *Felidae* namely, African wild cat, *Felis lybica*; domestic cat, *Felis catus*; caracal, *Caracal caracal*; European wild cat, *Felis sylvestris*; leopard, *Panthera pardus*; lion, *Panthera leo*; and cheetah, *Acynonyx jubatus*. Mitochondrial DNA (mt DNA) was utilized instead of nuclear DNA since it accumulates point mutations at a rate which is 5 to 10 times as fast as the nuclear DNA and is therefore particularly useful for studying more closely related organisms between sub-species, species and genera. Its apparent potential to be used as a tool for constructing genealogical trees and time scales makes it a method of choice in evolutionary studies. We used the restriction mapping approach to generate data for phylogenetic analysis. Restriction mapping was utilized since it gives good resolution at the species and genus level and evolutionary estimates derived from this method are considered more accurate than those obtained by methods such as the restriction fragment size

comparison. We have also attempted to develop the methodology for sequencing part of the cytochrome b region of mt DNA following polymerase chain reaction (PCR) amplification. Both cladistic and distance approaches were used for phylogenetic construction.

This study will be both of academic value and may have relevance to practical conservation management since these molecular approaches help to identify or confirm specific status especially with respect to the relationship between the domestic cat and the African and the Eropean wild cats. Furthermore, such approaches can be used at the intraspecific level to address problems in biogeography and population genetics.

Our results are in concordance with the previously determined morphological studies and albumin immunological distance studies. The restriction maps for the African wild cat and the domestic cat are identical, emphasizing their close relationship and the African origin of the domestic cat. The European wild cat showed a slight variation with the African wild cat or the domestic cat with four different restriction sites and a sequence divergence of 0.9. This suggests that the common ancestral mt DNA of these cats existed about 450 000 years ago. The lion and the leopard are monophyletic in both cladistic and distance approaches. The precise placement of caracal has yet to be resolved but it is deeply rooted in the phylogenetic analysis which would be more consistent with a separate generic status of the latter species rather than its inclusion within either *Felis* or *Panthera*. The distance analyses are consistent with the placement of the cheetah as the most distantly related species amongst the eight *Felid species* examined.

Generic and common names of eight Felid species

Generic Name	**Common Name**
Acynonyx jubatus	cheetah
Caracal caracal	caracal
Felis catus	domestic cat
Felis lybica	African wild cat
Felis nigripes	black-footed cat
Felis sylvestris	European wild cat
Panthera leo	lion
Panthera pardus	leopard

LIST OF ABBREVIATIONS

bp - Base pairs
°C - Degrees celcius
CsCl - Cesium Chloride
dATP - Deoxyadenosine triphosphate
dCTP - Deoxycytidine triphosphate
dGTP - Deoxyguanosine triphosphate
DNA - Deoxyribonuclei acid
dpm - Disintergrations per minute
dTTP - Deoxythymidine triphosphate
ddATP - Dideoxyadenosine triphosphate
ddCTP - Dideoxycytidine triphosphate
ddGTP - Dideoxyguanosine triphosphate
ddTTP - Dideoxythymidine triphosphate

Restriction enzymes:

A - Asp 718
B - Bam HI
C - Cla I
E - EcoR I
g - Bgl II
H - Hind III
l - Sal I
N - NcO I
P - Pst I
R - EcoR V
s - Sac I
S - Sac II
u - Stu I
v - Pvu II
X - Xba I
o - Xho I

TABLE OF CONTENTS

Page No.

CHAPTER 1 : GENERAL INTRODUCTION

CHAPTER 2 : GENERAL METHODS

CHAPTER 3 : MITOCHONDRIAL DNA RESTRICTION SITE ANALYSIS

CHAPTER 4 : POLYMERASE CHAIN REACTION AND MITOCHONDRIAL DNA SEQUENCING

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 : GENERAL INTRODUCTION

1.1 HISTORICAL BACKGROUND OF THE FAMILY FELIDAE

The family Felidae of the order Carnivora consists of a single sub-family (*Feline*) of extant cats and two or more sub-families covering a large number of fossil genera (Rosevear, 1974). This family is widely distributed throughout the world except for the Pacific islands, New Guinea, Australia, New Zealand, Madagascar, West Indies and Antarctica (Walker et al., 1964; Rosevear, 1974; Smithers, 1983). In Africa, the indigenous wild species are found in an area extending from the Mediteranean coast to the Cape Province in the Republic of South Africa, and in all vegetation zones including the Saharan region (Rosevear, 1974).

The phylogenetic relationships of the family *Felidae* have been problematical since the original classification of the latter into a single Linnaean genus Felis in the 19th Century. Morphological characters such as the shapes and relative proportions of different parts of the skull have been predominantly used for grouping various species (Walker *et al.*, 1964). The problem with using such characters is that members of the family *Felidae* have remarkably uniform skull dimensions (Rosevear, 1974) in spite of the great variation in sizes of animals, proportion of limbs and tail, and markings (Kitchner, 1990). This uniformity has led to a very diverse classification of the members of the family *Felidae*. Pocock (1951) classified the family into many genera, Hemmer (1978) used 15 genera, while Leyhausen (1979) grouped them into 13 genera, Walker (1964) used 6 and Nowak and Paradiso (1983) grouped them into 4 genera (Figs. 1.1 to 1.5).

Robinson (1976 a, b) looked at the haploid number of chromosomes and the position of the centromere as characters in determining

Fig. 1.1 **Classification of the family Felidae according to** Ewer (1973)

FAMILY FELIDAE

SUBFAMILY	GENUS	SPECIES	ORIGIN	COMMON NAME
FELINAE	Acinonyx	jubatus	Africa	Cheetah
	Panthera	onca	S. America	Jaguar
		pardus	Africa, Asia	Leopard
		leo	Africa	Lion
		tigris	Asia	Tiger
	Caracal	caracal	Africa	Caracal (rooicat)
	Felis	libyca	Africa	African wild cat
		sylvestris	Europe	European wild cat
		nigripes	Africa	Small spotted cat
		catus	unknown	Domestic cat
		bieti	E. Asia	Chinese desert cat
		chaus	S. Asia	Jungle cat
		margarita	Sahara, Near East	Sand cat
	Leptailurus	serval	Africa	Serval (tierboscat)
	Prionailurus	viverrina	S. Asia	Fishing cat
		bengalensis	S. & E. Asia	Leopard cat
		rubiginosa	India, Ceylon	Rusty spotted cat
	Ictailurus	planiceps	Malaya, Bornea	Flat-headed cat
	Mayailurus	iriomotensis	Taiwan	
	Octocolobus	manul	Central Asia	Palla's cat
	Pardofelis	marmorata	S. Asia	Marbled cat
		badia	Borneo	Borneo marbled cat
	Profelis	temmincki	SE Asia	Asiatic golden cat
		aurata	W. Africa	African golden cat
	Puma	concolor	America	Puma
	Leopardus	pardalis	America	Ocelot
		tigrina	America	Tiger ocelot
		weidii	S. America	Margay
		geoffroyi	S. America	Geoffroy's ocelot
	Oncifelis	guigna	S. America	Kodkod (huina)
	Lynchailurus	colocola	S. America	Pampas cat
	Oreailurus	jacobita	S. America (Andes)	Mountain cat
	Herpailurus	yagouaroundi	America	Jaguarundi
	Neofelis	nebulosa	S. Asia	Clouded leopard
	Uncia	uncia	Asia	Snow leopard
	Lynx	lynx	Europe	European lynx
		pardina	Iberian peninsula	Spanish lynx
		rufus	N. America	Bobcat

Fig. 1.2 **Classification of the Family Felidae according to Hemmer (1978)**

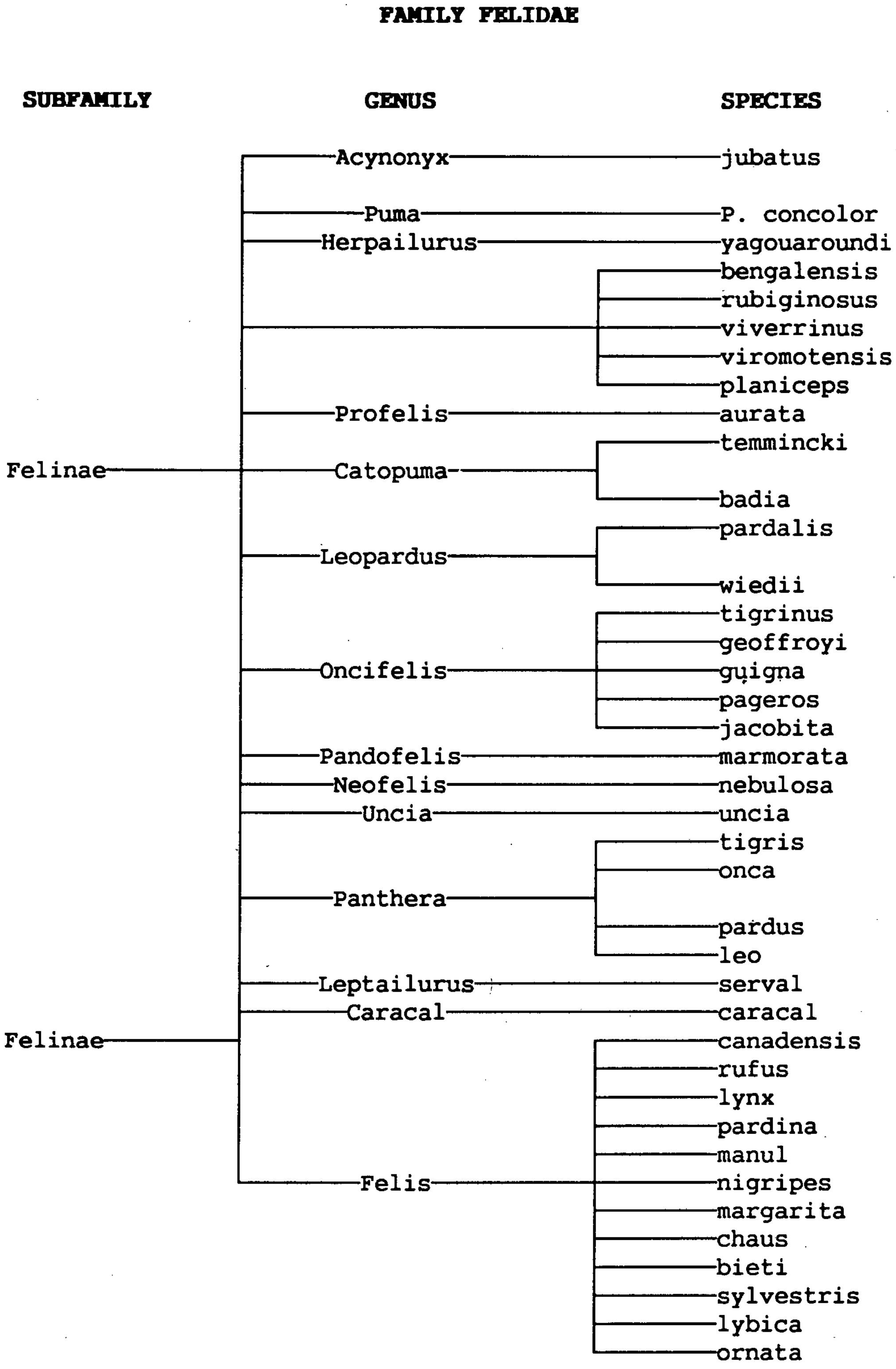


Fig. 1.3 **Classification of the Family Felidae according to Leyhausen (1979)**

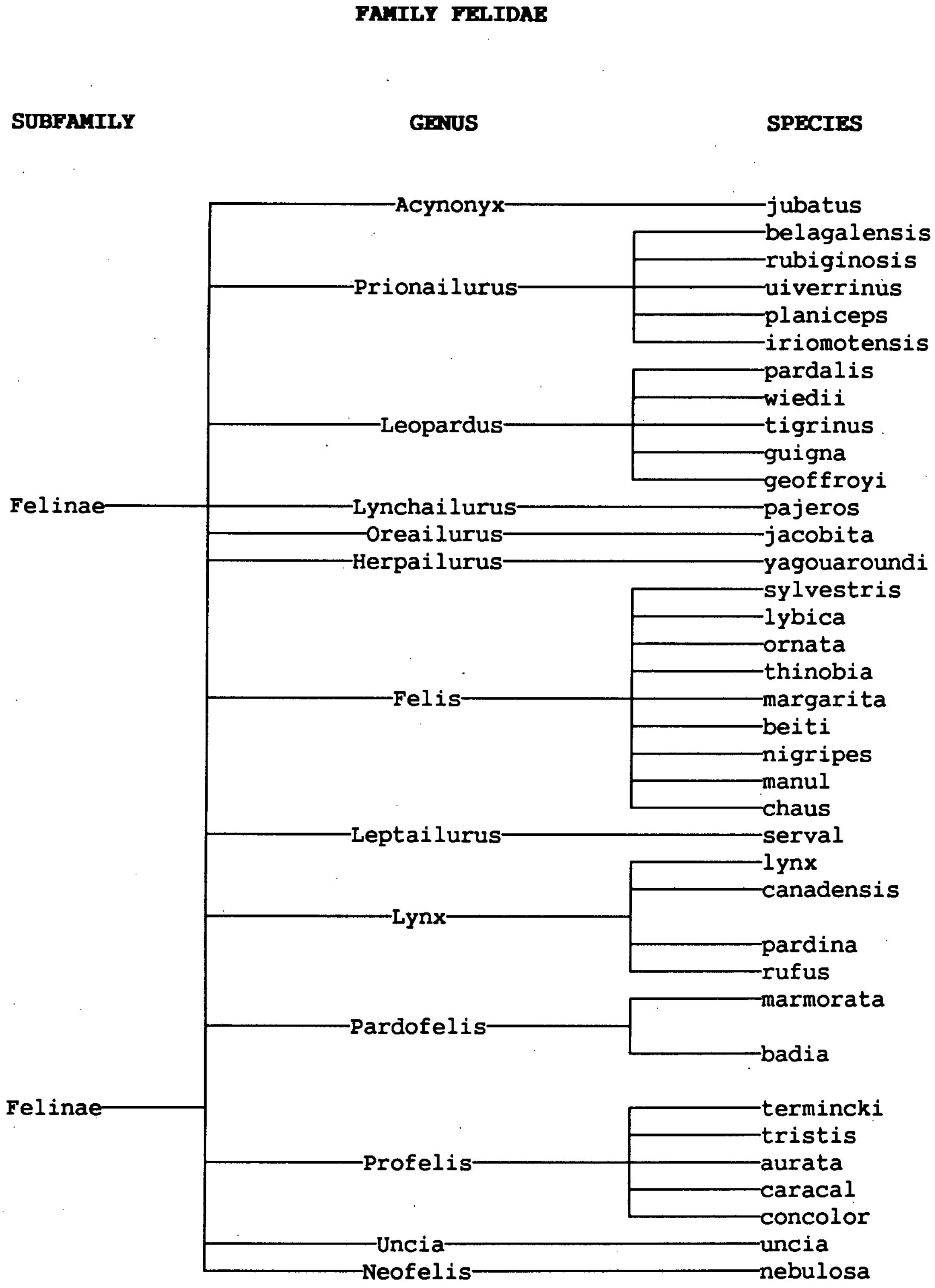
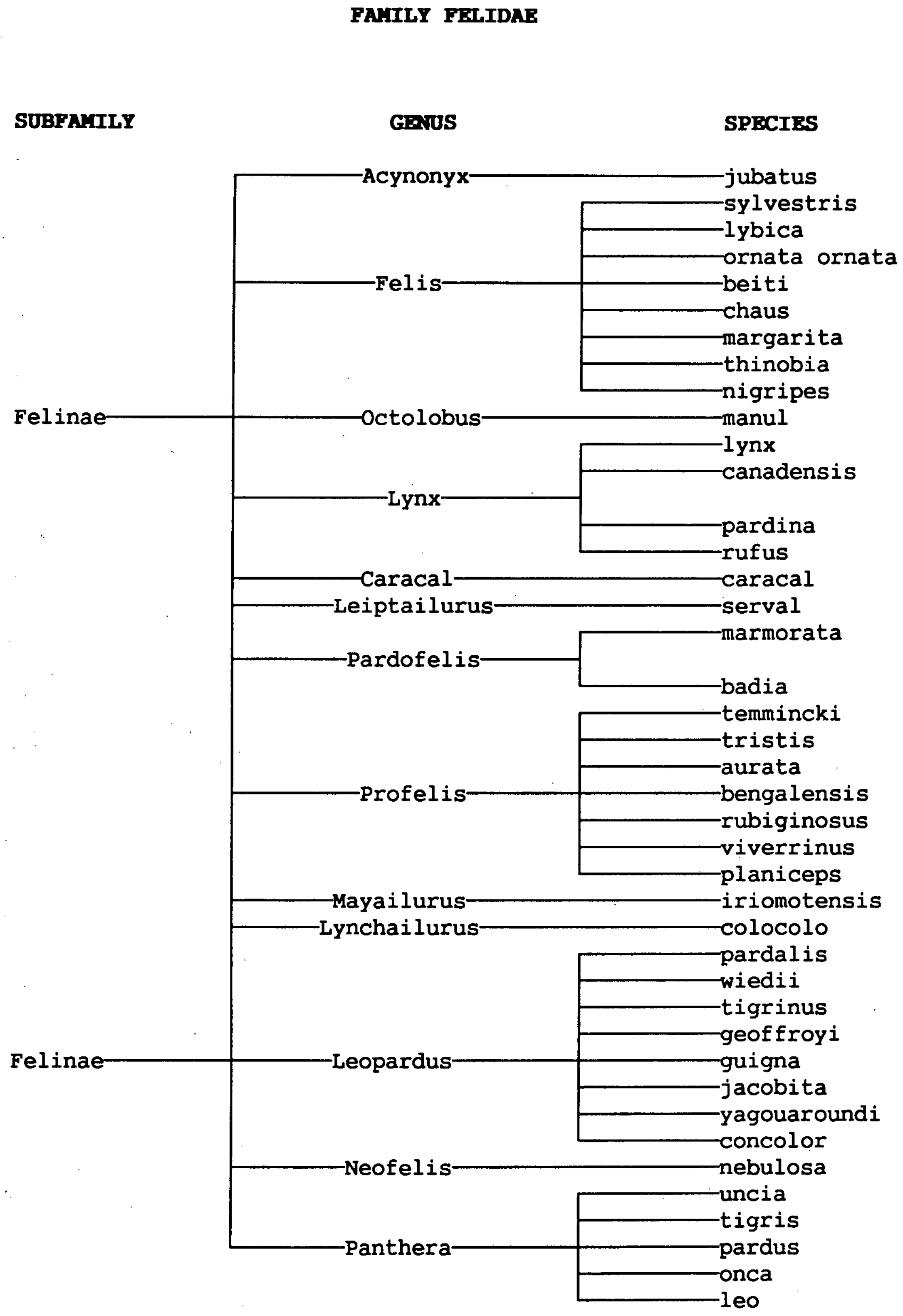


Fig. 1.4 Classification of the family Felidae according to Walker et al. (1964)

FAMILY FELIDAE

SUBFAMILY	GENUS	SPECIES	ORIGIN	COMMON NAME
FELINAE	Acinonyx	jubatus	Africa	Cheetah
	Panthera	onca	S.America	Jaguar
		pardus	Africa,Asia	Leopard
		leo	Africa	Lion
		tigris	Asia	Tiger
	Felis	caracal	Africa	Caracal(rooicat)
		libyca	Africa	African wild cat
		sylvestris	Europe	European wild cat
		catus	unknown	Domestic cat
		nigripes	Africa	Small spotted cat
		serval	Africa	Serval(tierboscat)
		planiceps	Malaya,Borneo	Flat-headed cat
		marmorata	S.Asia	Marbled cat
		temmincki	SE Asia	Golden cat
		aurata	W.Africa	Golden cat
		manul	Central Asia	Pallas’ cat
		concolor	America	Puma
		tigrina	America	Little spotted cat
		yagouaroundi	America	Jaguarundi
		bengalensis	S.&E.Asia	Leopard cat
		bieti	E.Asia	Chinese desert cat
		minuta	Phillipines	
		weidii	S.America	Margay
		pajeros	S.America	Pampas cat
		viverrina	S.Asia	Fishing cat
		geoffroyi	S.America	Geoffroy’s cat
		rubiginosa	India,Ceylon	Rusty spotted cat
		chaus	S.Asia	Jungle cat
		margarita	Sahara,Near east	Sand cat
		badia	Borneo	Borneo marbled cat
		pardalis	America	Ocelot
		wiedii	America	Margay
		guigna	S.America	Kodkod (huina)
		jacobita	S.America	Mountain cat
	Neofelis	nebulosa	S.Asia	Clouded leopard
	Uncia	uncia	Asia	Snow leopard
	Lynx	lynx	Europe	European lynx
		canadensis	Canada	N.Am.lynx
		rufus	N.America	Bobcat

Fig. 1.5 **Classification of the Family Felidae according to Nowak and Paradiso (1983)**



evolutionary relationships of the *Felidae*. He found five basic karyotypes. The two main ones were the old world cats which included the jaguar and jaguarundi; and the new world cats. The old world cats had 19 pairs of chromosomes whereas the new world cats had 18 pairs. In contrast with the morphological and behavioural studies (Walker *et al*., 1964; Ewer, 1973; Hemmer, 1978; Leyhausen, 1979; Nowak and Paradiso, 1983) he grouped the big cats such as the clouded leopard, lion, tiger with the smaller cats under genus *Felis*.

Kratochvil (1982) further analyzed the data based on karyological studies and baculum morphology to derive his own classification system. On the basis of karyological data he divided the cats into two main groups. The first group was comprised of the *Pantherinae* (big cats), *Neofilinae* (clouded leopard), *lynchinae* (lynxes) and *acynonychinae* (cheetah). The second group consisted of all the old world cats (*Felini*) and new world cats. This led Kratochvil (1982) to group the cats into 18 genera. He separated the caracal from the rest of the lynxes, a finding which supports the classification schemes by Ewer (1973), Hemmer (1978) and Nowak and Paradiso (1983). This finding is in contrast with the classification scheme by Leyhausen (1979) in which the caracal is classified under the *genus profelis*.

The study by Werdelin (1983) was based on multivariate statistical methods (eg. principal components analysis and correspondence analysis) to see if there are any patterns of skull shapes within the family *Felidae*. The results showed differentation between the larger cats and the smaller cats. All the lynxes fell into a common group with the exception of the caracal. The two oddities were the puma and the clouded leopard, which did not seem to fit anywhere (Kitchner, 1991).

In order to try and resolve the phylogeny of this taxonomically difficult family, several attempts have been made using molecular approaches. These include estimation of genetic distances between taxa using albumin immunological distances (Collier and O'Brien, 1985) and allozyme electrophoretic studies (Randi and Ragni, 1991).

O'Brien and Collier used immunological distances to construct a phylogeny of ten feline types (Collier and O'Brien, 1985). They concluded that the cats radiated about twelve million years ago into two main groups, namely the South American and the Old World felines (Collier and O'Brien, 1985). Approximately 4-6 million years ago, a *pantherine* lineage evolved from the Old World *felids*. The cheetah was found to be closely related to the other *pantherine* cats which is surprising as it has previously been assigned to a genus of its own (Walker *et al*., 1964). O'Brien *et al*., (1987) further compared the results from albumin immunological distance studies, karylogical studies and isozyme genetic distances. He found that in general, the results corresponded with Collier and O'Brein (1985) earlier phylogeny.

The identity of the ancestor of the domestic cat, *Felis catus* has also been a subject of controversy. The proposal that the African wild cat, *Felis lybica* was the ancestor, was based originally on reports of the domestication of the latter in Egypt approximately 5000 to 3300 years ago (Corbert, 1966; Zeunner, 1967; Rosevear, 1974; Smithers, 1983). Different workers have suggested that the true ancestors are respectively the leopard cat, *Felis bengaleuse's* and *Palla's* cat, *Felis manul* (Kratochsvil and Kratochvil, 1976; Hemmer, 1978) whereas Searle (1968) has suggested that it may be the jungle cat, *Felis chaus*. Compounding the problem is the dispute as to whether the African wild cat and the European wild cat *Felis sylvestris* are conspecific (Harrison, 1968; Randi and Ragni, 1991) or distinct, separate species (Pocock , 1951; Rosevear, 1974; Smithers, 1983; Ansell and Dowsett, 1988). The results obtained from allozyme electrophoretic distance estimation suggested that the African and European wild cats are conspecific, as no fixed allelic differences were found between domestic cat and European wild cat from Sardinia and African wild cat from Sicily, the central Apenine region and the North Eastern Alpine - Pre Alpine region (Randi and Ragni, 1991).

In the present study mt DNA was used in order to make a further contribution towards constructing a fine phylogeny of the family *Felidae*. The objectives of the study were fivefold:

1. Mitochondrial DNA restriction mapping was used to construct a phylogeny of eight members of the southern African *Felidae*, namely, the African wild cat, domestic cat, European wild cat, caracal, black-footed cat, lion, leopard, and the cheetah.
2. The genetic relationship between the African wild cat and the domestic cat was determined, since interbreeding between the latter have been reported and concern was expressed that this might effectively cause the disappearance of African wild cat (Grey, 1954 and Smithers, 1983).
3. To determine the relationship between African wild cat and the

European wild cat.

4. To determine the origin of the domestic cat.

5. To complement the restriction mapping data, the methodology for sequencing part of the cytochrome b region of mt DNA following PCR was developed.

1.2 MOLECULAR SYSTEMATICS

Traditionally, taxonomy of animals has largely been based on morphological and behavioural characters, which when used alone are not always sufficient for definitive classification of various animal groups (Thorpe, 1982; Ferris and Berg, 1985; Avise, 1989). This is partly due to the fact that such characters are extensively modified by selective environmental pressures (Ferris and Berg, 1985; Harley, 1988b; Avise, 1989).

In recent years, molecular studies, either on proteins or DNA have become standard means for assisting in resolving problems in systematics (Halliday and Arano, 1991). Compared with studies on morphology, the main advantages are the extensive data sets and the assumption that these macromolecules are less susceptible to modification by environmental factors (Hillis, 1987; Harley 1988b; Halliday and Arano, 1991). While it is difficult to estimate an evolutionary time scale for morphological trees, it can be more readily done for molecular trees (Nei, 1986). Molecular approaches are, however, not without limitations. Some of the major problems that complicate the interpretation of molecular data are:

(1) mutations follow a stochastic process and,

(2) the effect of backward and parallel mutations (Hillis, 1987)

which can result in homoplasies.

Since random variation is more significant when short DNA sequences are examined or when there are few restriction sites mapped, its contribution can be minimized by increasing the length of the DNA sequences and/or the number of mapped restriction sites. The first problem (stochasticity) mostly affects tree building methods based on distance estimation whereas the problem of homoplasies is more relevant to cladistic methods.

Considerable debate has arisen over the relative merits of morphological and molecular data as sources of information for phylogenetic reconstruction. Comparative studies have indicated that both approaches have distinct advantages and disadvantages (Hillis, 1987; Hillis and Moritz, 1990). To a large extent, this is due to the fact that systematists from either camp (morphologists or molecular biologists) are unable to address questions and problems that can be addressed by systematists from the other (Hillis, 1987). Morphological analyses are particularly suitable for the examination of museum material and in the case of extinct species, may be the only feasible approach. It is probable that analyses utilising both approaches will produce the most accurate phylogenetic and evolutionary interpretations (Hillis, 1987; Hillis and Moritz, 1990). The advent of molecular biological approaches to the field of systematics in the early 1960's resulted in the use of protein studies as the principal methods of phylogenetic analysis. These chiefly involved the determination of the degree of amino acid sequence variation among proteins by measuring protein electrophoretic mobility and immunological distance studies. More recently, studies on DNA have been used as a method for phylogenetic analysis. These molecular biological techniques are described below.

1.3 THE USE OF PROTEINS IN MOLECULAR SYSTEMATICS

1.3.1 Protein Electrophoresis

Protein electrophoresis which was first used by Tiselius in 1937 to distinguish multiple fractions of serum proteins (Avise, 1975) came into widespread use in the late 1960's (Lewotin and Hubby, 1966; Harris, 1969; Hubby and Lewontin, 1967).

The technique involves the migration of soluble proteins under the influence of an electric current through a matrix which can either be starch gel, polyacrylamide gel or agarose gel (Andrews, 1983). This rate of migration of proteins depends on several factors, such as the size of the molecules, shape of the molecules and their net charge (Kunker and Trautman, 1953; Andrews, 1983). For systematic analysis there are two general forms of protein data which can be gathered from electrophoresis, namely, identification of isozymes and allozymes. Isozymes are products of distinct genes which catalyze the same reaction but differ in their electrophoretic mobility. Allozymes on the other hand are protein products of a single genetic locus which differ in electrophoretic mobility (Avise 1975; Hillis and Moritz, 1990). Both forms of data have played a major role in molecular systematics.

The amount of genetic information and the high precision of the data obtained when using the technique makes it very attractive compared to morphological approaches (Avise, 1975).

The technique has, however, several drawbacks: Many nucleotide changes may occur without altering the amino acid sequence, and in many amino acid substitutions molecular charge is conserved (Avise, 1975). It is being estimated that approximately 30% of the possible nucleotide substitutions code for amino acids with different changes and this bias with chance identities in band mobilities, will cause underestimates of protein difference between populations (Avise, 1975; Halliday and Arano, 1991). In addition, when using this technique comparisons of enzyme electrophoretic patterns are limited to those of extant species and this is a restriction that is not faced by morphologists when fossil records are available (Avise, 1975).

1.3.2 Immunological Distances

Sarich and Wilson (1966, 1967) developed the method of studying genetic distances between species. They showed that the immunological distance is approximately proportional to the time of divergence between a pair of species (Nei, 1987).

The aim of the immunological distance determination is to quantitatively estimate the extent of immunological cross-reactivity between proteins from different taxa (Thorpe, 1974). Several immunological techniques are available and these include immunodiffusion and quantitative microcomplement fixation. Quantitative microcomplement fixation is by far the most widely used technique for systematic work, especially on animals (Thorpe, 1982).

There are several proteins that can be used in microcomplement fixation. These include transferrin, lysozyme, and serum albumin. Most researchers in molecular systematics have used serum albumin immunological distances, primarily because this protein has been shown to evolve faster than the other proteins (Maxson and Wilson, 1974).

The average rate of protein evolution has previously been estimated to be one amino acid substitution per 100 residues per 10^7 years, while that of albumin appears to be twice as fast (Maxson and Wilson, 1974). For these reasons, it seems to be a useful protein for detecting evolutionary differences among closely related species (Wallace and Wilson, 1972).

The technique is based on the specificity of the reaction of mammalian immunoglobulins to antigenic proteins (Thorpe, 1982). In principle, it involves the production of antisera (protein) by immunizing a mammal (usually rabbits) with purified serum albumin from one animal to be tested. This rabbit will thus develop an immune response to injected serum albumin. The antiserum produced, strongly cross reacts with the albumin from this organism (rabbit) (homologous antigen) but less strongly with that of another species (heterologous antigen) for a given concentration of antisera. The degree of antigenic differences between two albumins can then be measured and this is known as an index of dissimilarity (I.D.). This dissimilarity is assumed to be an indication of an evolutionary time that has elapsed since the divergence of the taxa concerned. The log ID index of dissimilarity, has been shown to be linearly related to the time since divergence between two organisms tested (compared). This is denoted by d where

$$d = 100 \times \log_{10} ID \text{ and,}$$

$$d = \text{immunolgical distance}$$

In many proteins d is thought to be linearly related to the proportion of different amino acids between two sequences compared. According to Maxson and Wilson (1974), one unit of d corresponds to roughly one amino acid difference in albumin.

Its degree of sensitivity, which approaches the detection of a single amino acid substitution between homologous albumins when pooled heterologous antisera are employed makes it a method of choice compared with other immunological techniques such as spot-plate (Champion *et al.*, 1974; Prager *et al.*, 1978; Wilson *et al.*, 1977). Furthermore, its great versatility and economy of materials make it a method of choice in evolutionary studies (Wilson *et al.*, 1977; Thorpe, 1982).

There are however several disadvantages associated with the technique. There is some evidence to suggest that only a small proportion of the total structure of the molecule may be significant in the binding of antibodies (Thorpe, 1982). Additionally, the approach unambiguously demonstrates differences between distantly related taxa, but not for closely related ones (Halliday, 1991). In the latter case, protein structures are often identical and therefore no differences in protein sequence cannot be observed by this method (Thorpe, 1982; Halliday, 1991).

Despite its inaccuracies, the technique has been applied extensively in evolutionary studies. These include the classification of humans and Apes (Nei, 1987); *Plethodontid* Salamanders (Maxson and Maxson, 1979); Mammalian pancreatic ribonucleases (Prager and Wilson, 1978) tree frogs in the *Hyla eximia* species groups (Maxson and Wilson, 1974), the relationships among *Eurasian Bufo*, albumin evolution and its phylogenetic implications in toads of the *genus Bufo* (Maxson, 1981).

1.4 USE OF DNA IN MOLECULAR SYSTEMATICS

1.4.1 Nuclear DNA Versus Mitochondrial DNA

Two types of DNA are used to generate data for taxonomic studies, namely nuclear DNA or organelle (mitochondrial and chloroplast) DNA. Nuclear DNA is best for comparison between taxa at the supra-generic level (orders and families) i.e. within 20-100 million years. Organelle DNA, principally mitochondrial DNA in animals, is more suitable for comparisons over the time scale of 0.5 to 20 million years and is therefore the method of choice for comparisons between sub-species, species and genera (see Fig. 1.6). Chloroplast DNA on the other hand exists as multicopies of a circular double-stranded molecule. On average, it is about 150 kb (Carroll, 1989) which is several magnitudes smaller than the nuclear DNA but larger than mitochondrial DNA. The overall size of chloroplast DNA is conserved with no histones bound to DNA but in contrast to the mt DNA, it possesses intervening sequences (introns) although the latter are

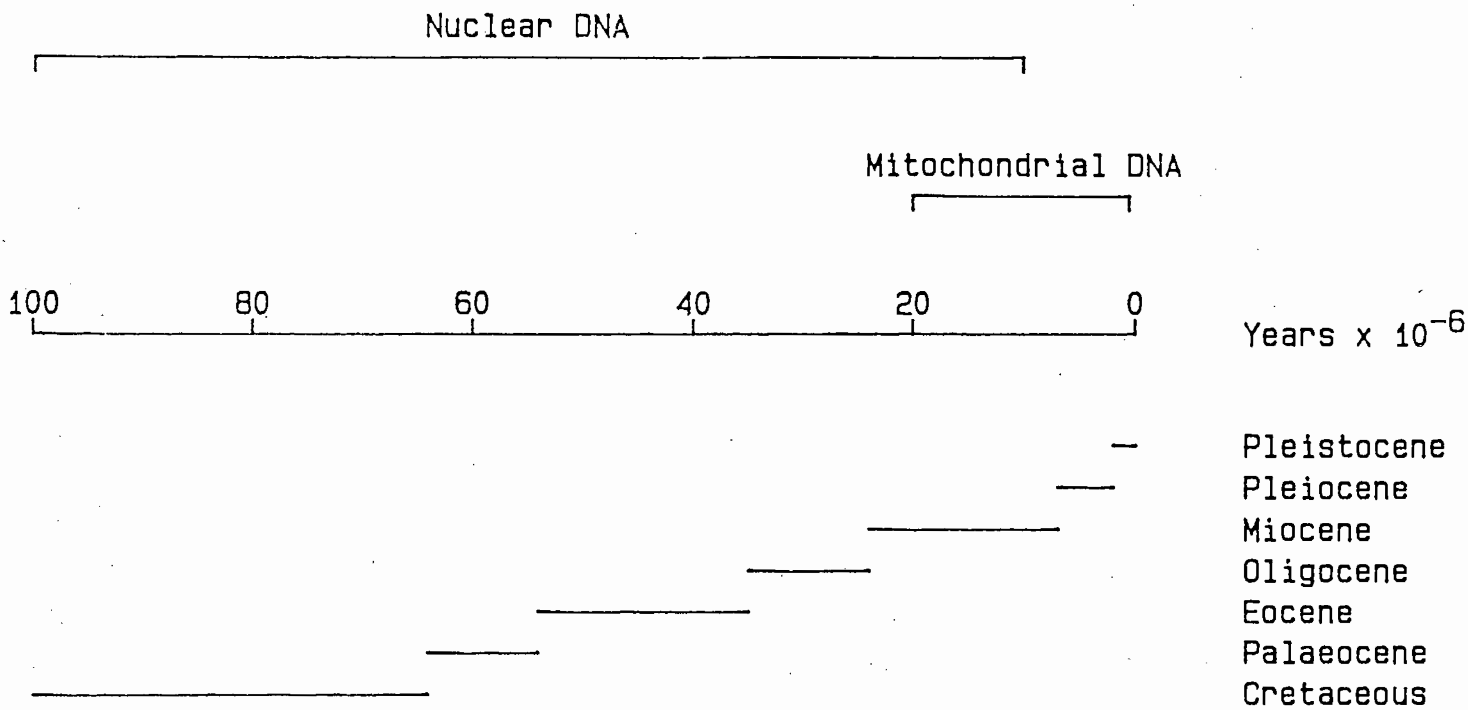


Fig. 1.6 The time scale over which nuclear DNA and mitochondrial DNA analysis can be used in relation to the geological past (taken from Harley, 1988)

documented to be rare. This molecule appears to evolve at a relatively slow rate, and is therefore generally useful in systematic studies at or above the species level (Stirton, 1987). Since the members of a single family are compared in this study, only the use of mitochondrial DNA will be reviewed below.

1.4.2 **Mitochondrial DNA**

Avise *et al.* (1987) has stated that "*if one were to specify the properties desired for an ideal molecular system for phylogenetic analysis, the wish list might include the following*":

The molecule should be:

i) distinctive, yet ubiquitously distributed, so that secure homologous comparisons could be made among a wide variety of organisms.

ii) It should be easy to isolate and assay.

iii) It should have a simple genetic structure lacking complicating features such as repetitive DNA, transposable elements, pseudogenes and introns.

iv) It should exhibit a straightforward mode of genetic transmission, without recombination or other genetic rearrangements.

v) It should provide suites of quantitative character state whose phylogenetic interrelationships could be inferred by reasonably parsimony criteria, and for purposes of microevolutionary analysis.

vi) It should evolve at a rapid pace such that new character states commonly arise within the lifespan of a species.

To a remarkable degree, the mt DNA of higher animals meets all of the above criteria.

Animal mt DNA possesses several unique features that makes it useful for studies in molecular evolution (Brown, 1983). It is a relatively small, covalently closed circular molecule of conserved size of approximately 16 000 - 18 000 base pairs in vertebrates (Fig. 1.7).

Its size makes it easy to purify and characterize and it can be prepared from small amounts of tissue with relative ease free of significant nuclear DNA contamination (Brown, 1980; Johnson *et al.*, 1983; Clary *et al.*, 1985). The arrangement of genes in mammalian mt DNA is totally conserved with no introns or intervening sequences except in the non-coding control, displacement loop (D-loop) region, in which a few examples of large insertions and duplications are documented (Clary *et al.*, 1985; Harrison, 1989). This region which is responsible for the initiation of replication and transcription is of variable lengths in different species (Cann *et al.*, 1978; Moritz *et al.*, 1987). Rates of sequence evolution vary along mt DNA molecule with the D-loop region being the most rapidly evolving region. Ribosomal RNA (rRNA) genes on the other hand are known to evolve a lot slower than other genes (Moritz *et al.*, 1987). The remainder of the mt DNA molecule consists of 22 tRNA genes, 2 rRNA genes, 13 genes which code for subunits of the electron transport chain proteins or proteins involved in ATP synthesis. It should be noted that the previously unidentified reading frames (denoted by URF on the diagram) have in fact been identified as representatives of NADH dehydrogenase. In most cases, there are no intervening sequences between genes (Lewin, 1983).

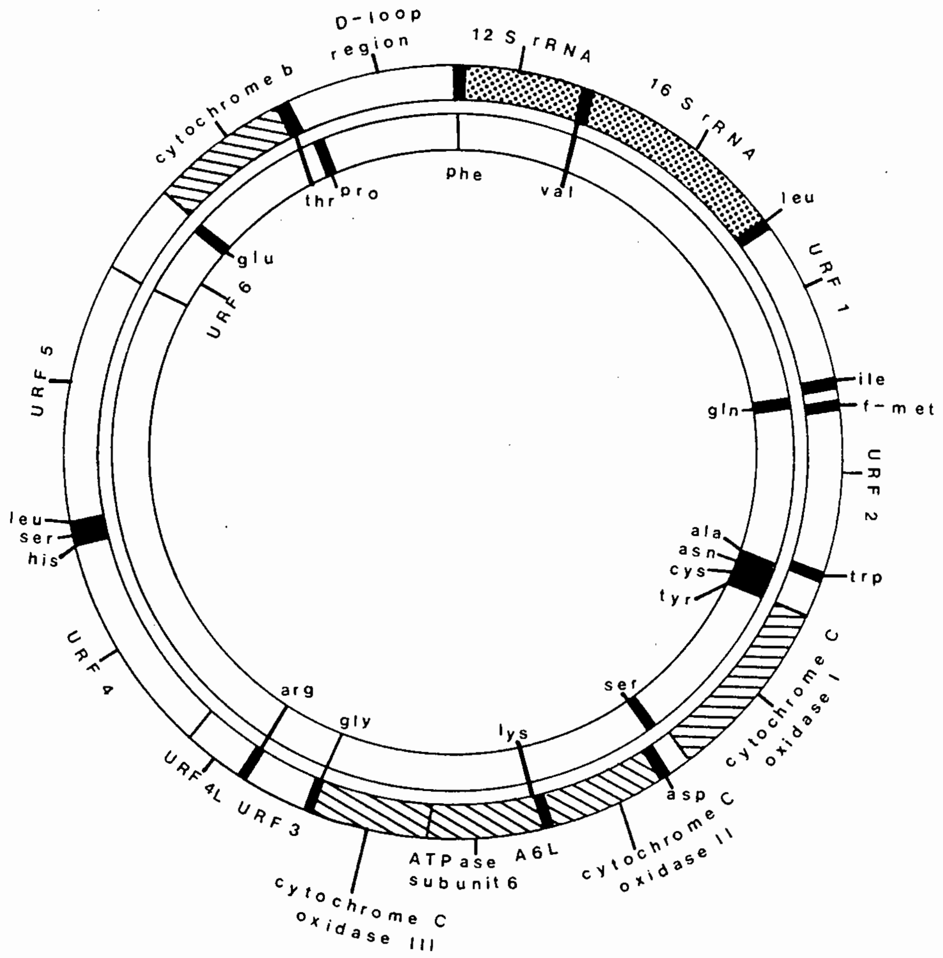


Fig. 1.7 **Schematic representation of mammalian mitochondrial DNA. Cyt10; cytochrome b; COI, COII and COIII; Cytochrome oxidase subunits I, II and III. URF : Unidentified Reading Frame (NADH dehydrogenase). OH : origin of replication of the H-strand. OL : origin of replication of the L strand (taken from Lewin, 1983)**

Animal mt DNA exhibits remarkable conservation of gene content, This was demonstrated in a recent report on the complete nucleotide sequences of human, bovine and mouse mt DNAs which revealed a highly conserved gene order with some unique characteristics (Lewin, 1983).

Mt DNA is maternally inherited (Clyde *et al.*, 1974; Brown, 1980; Johnson *et al.*, 1983; Clary *et al.*, 1985; Moritz *et al.*, 1987; Arctander, 1988) which therefore confers as a genealogical marker: a half fraction of the population pass their mt DNA on to the offspring and the effective population size for mt DNA is smaller than that for nuclear genes. Consequently, stochastic processes will be particularly important in determining frequencies of mt DNA genotypes.

Mitochondrial DNA does not undergo recombination in contrast to nuclear DNA and this has the advantage that it does not form hybrid in crosses between species or sub-species but instead reflects the pure maternal line of inheritance (Harley, 1988). An exception to this is the displacement loop (D-loop) in which recombination has been shown to occur (Wilson *et al.*, 1977).

Although its overall size, structure, and gene order are highly conserved in vertebrates, with the encoded proteins being very highly conserved, the rate of accumulation of single base changes is five to ten times faster than in nuclear DNA (Ferris and Berg, 1987). This makes mt DNA particularly useful for comparing more closely related groups of animals (sub-species) (Cann *et al.*, 1987).

Mt DNA evolutionary changes are mainly caused by nucleotide substitutions, particularly single base substitutions. Almost all these changes are neutral since most changes occur at the third codon position which does not bring about amino acid replacement and hence no phenotypic change in the organisms. Since only the third codon position is subject to rapid change, the rate of change in mt DNA is markedly reduced at values greater than 20% sequence divergence owing to the strong constraints against changes in the first two positions of the codon. These constraints are clearly illustrated when a dot-matrix homology search is made from known sequences of bovine (Harley, 1986) and xenopus (Roe, *et al.*, 1985) (see Fig. 1.8) when the line of homology and lack of rearrangements is very apparent, although these two lineages diverged over 400 million years ago (Romer, 1966).

It has also been suggested that other types of mutation, viz. deletions, insertions, and rearrangements can occur and some cladistic approaches use them as characters whereas certain analytical procedures such as DNA-DNA hybridization are sensitive to these changes, but the mathematical approaches required for calculation of sequence divergence and construction of phylogenies are simplified if single base changes alone are considered (Harley, 1988b).

There is evidence to suggest that in mt DNA, base transitions (purine/purine; pyrimidine/pyrimidine) occur more frequently (90%) than base transversions (purine/pyrimidine). This implies that back and parallel mutations may accumulate rapidly between two lineages generating homoplasies in cladistic analyses, and distance estimates will proportionally underestimate the level of sequence divergence (Ferris and Berg, 1985). As mutations arise, they are either lost in the population or become fixed through random drift. Any population, other than a highly inbred one arising from a small founder number,

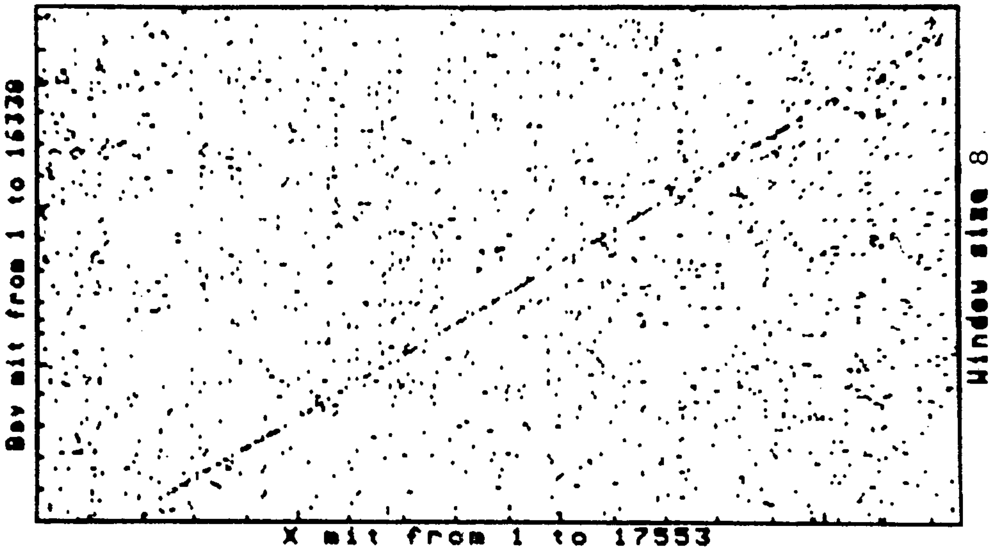


Fig. 1.8 Homology search of Bovine and Xenopus mitochondrial DNA sequences. The search window size is for complete homologies of a base pair segment reproduced from (Harley, 1988)

will show a certain level of polymorphism in mt DNA. For human mt DNA, this polymorphism is equivalent to an average sequence divergence of 0.32% (Cann *et al.*, 1987), with 133 different types of mt DNA having been identified. It is estimated that the ancestor of all surviving human mt DNA types existed approximately 140 000 (Brown, 1980) to 360 000 years ago (Cann *et al.*, 1987). The degree of polymorphism reflects population sizes, and in small populations mt DNA lineages can be lost rapidly. It has been calculated that for stable-sized populations of size, n, it is likely that all descendants will trace their mitochondrial ancestry to a single founder female within about 4n generations (Avise *et al.*, 1974). The polymorphism level of a population in making comparisons between different taxa should be subtracted from the between taxon sequence divergence estimate wherever possible. For some animal species this may not always be practical, but it is always necessary to be aware that estimates of mt DNA sequence divergence are measuring differences between mitochondrial lineages, rather than population lineages.

1.5 METHODS OF DNA COMPARISON

Differences in DNA between organisms can be studied using DNA-DNA hybridization, restriction fragment size comparisons, restriction site mapping and DNA sequencing.

1.5.1 DNA-DNA Hybridization

The technique operates on the assumption that the genetic relatedness of organisms is reflected in the similarity of their DNA base pair sequences (Sheldon, 1991). This similarity can be measured by hybridizing strands of DNA from different species followed by measuring the hydrogen bonding strength between the strands of these hybrids. The poorer the bonding strength, the more distantly related the organisms.

Experimentally, the single strands of DNA from two different species can be combined after denaturation by heating at 100°C allowing them to reassociate at proportionally decreasing temperatures. The melting temperature of such annealed duplexes can then be determined. This temperature is a measurement of the thermostability of DNA and is defined as the temperature (in °C), at which 50% dissociation of duplex DNA has occurred (Wilson *et al.*, 1977). The hybrid DNA duplexes from two different species contain base-pair mismatches due to their evolutionary divergence from a common ancestor (Thorpe, 1982; Sibley and Ahlquist, 1986). Since the melting temperature (Tm) of the duplex DNA is proportional to the number of hydrogen bonds between the two strands, such base-pair mismatches will lower the melting temperature of the hybrid DNA relative to that of perfectly matched double strands. The extent of nucleotide divergence can then be estimated from the thermostability of hybrid duplex DNA, which has been shown to decrease with an increasing proportion of mismatched nucleotides. The extent of reassociation in homoduplex DNA versus heteroduplex hybrid DNA reactions can be measured and the depression of the melting temperature in a heteroduplex hybrid relative to the homoduplex hybrid serves as an index of divergence between DNAs under comparison. One of the main advantages of the technique is its objectivity and the fact that it accounts for historically informative characters encoded in the DNA that are not necessarily expressed physically (Sheldon, 1991). The technique also has several drawbacks. It is very time consuming and gives a single estimate of the genetic distance based on the entire genome. It lacks the detailed resolving power of other methods such as restriction site mapping and sequencing. Also insertions, deletions and rearrangements (transpositions) of sequence within the strands compared will greatly affect the accuracy of the technique. Furthermore, the technique does not provide direct information on the sequence changes ie. nucleotides remain unidentified (Werman, Springer and Britten, 1990).

If the Tm is about 1 approximately 20% of nucleotides are substituted, the extent of annealing becomes unreliable and the technique cannot be used (Thorpe, 1982) and since DNA sequences differing by 20% or more do not usually anneal with high specificity (Britten and Kohne, 1968) this method cannot be used to compare distantly related species. Similarly, if species are too closely related, there may be no measurable reduction in thermal stability. Particularly where interspecies differences are small, the errors involved in making divergence time estimates by this technique may be considerable (Sharpe, 1982).

Despite the limitations, DNA-DNA hybridization has been employed in molecular systematics studies such as on the rodents and *artiodactyls* (Laird *et al.*, 1969) *Herous*, (Sheldon, 1987), and *Hominoids* (Van de Velde, 1972;). Sibley and Ahlquist (1986) have used this method extensively in a comprehensive attempt to resolve the phylogeny of the order Aves.

1.5.2 Restriction Fragment Size Comparison

In this technique, purified mitochondrial DNA is cleaved with restriction endonucleases and the restriction fragments separated by gel electrophoresis. Restriction fragment lengths are then estimated by comparing their electrophoretic mobility. Sequence divergence can be estimated from the pairwise proportion of shared fragments using a formula which is based on measuring the proportion of shared restriction fragments between DNA molecules in a pairwise manner (Nei, 1987). The fragments can be used as characters in cladistic analyses but this is generally considered undesirable owing to the non-interdependence of fragment and site losses (Hillis and Moritz, 1990).

The technique is simple and relatively easy and is useful for studying more closely related groups of animals ie. at sub-species level, but it becomes inaccurate once the proportion of shared restriction fragments become low as in more divergent species (Harley, 1988; Shields and Helmbychowski, 1991). Furthermore, fragment length comparisons are unreliable if the overall sequence divergence exceeds 10-15% (Moritz, 1987).

1.5.3 Restriction Site Mapping

This is considerably more time consuming than the above method but gives good resolution at the species and genus level. Restriction sites for species to be analyzed are usually determined using a variety of restriction endonucleases, at least 10 or more being necessary to obtain a useful number of sites. The technique allows more rigorous description of the mutations which produce various fragment patterns seen in the population and can therefore provide more accurate estimates of sequence divergence (Ferris and Berg, 1987). The problem is that it is still an indirect method which is an advantage in that it looks at the whole mitochondrial genome and therefore analyses over the total ±16 000 odd base pairs. Stochastic errors can be immunised by using more enzymes. Homoplasies are no worse, and may be better (3 in 1) than sequence data (less chance of research, at those parallel losses and common) due to their rapid change and limited character states (Moritz, 1987). Restriction site mapping was used in this study to generate data for classification of eight members of the southern African Felidae since it gives good resolution at the species and genus level.

1.5.4 Nucleotide Sequencing

This method has, perhaps the greatest potential and versatility of the DNA methods, since it allows direct comparison of homologous bases in the aligned region of the DNA provided the sequences under comparison are close enough to make correct alignment unequivocal. There is also no theoretical limit to the amount of DNA that can be sequenced, so large data sets can be obtained which minimize stochastic errors.

There are **two major problems** in using the technique:

1. Techniques have until recently been very complex and time consuming due to the requirements for cloning procedures prior to sequencing. However, the advyent of the polymerase chain reaction, a technique that is used to amplify DNA from as little starting material as a hair root to give microgram quantities of DNA in a few hours has really speeded up and simplified the

procedure (Higuchi *et al.*, 1988 and Vigilant *et al.*, 1989). The details of PCR will be discussed later.

2. An additional problem is that a region of DNA being sequenced is not necessarily a representative of the genome as a whole. In the present study mitochondrial DNA sequencing was used to generate a complementary data set for comparison with the restriction mapping data.

1.5.5 Polymerase Chain Reaction

The polymerase chain reaction is an invitro method for producing large amounts of specific DNA fragments of defined length and sequence from small amounts of complex templates (White *et al.*, 1989). The sensitivity, speed and versatility of PCR is having a profound impact on molecular biological approaches to problems in human genetics, forensic science, medical diagnostics and in evolutionary and developmental biology (Marx, 1988; Paäbo *et al.*, 1989; White *et al.*, 1989; Mullis and Faloona, 1989; Mullis, 1990).

In principle, PCR entails denaturation of double stranded DNA at very high temperature, typically 92-94°C, followed by the annealing of oligonucleotide primers to the single strands of DNA at 50-55°C (see Fig.1.9). These primers are added in vast excess relative to the target DNA. They hybridize to opposite strands of DNA . They are oriented with their 3′ ends facing each other so that the strand synthesis by DNA polymerase extends across the segment of DNA between them. The next step is the addition of a thermostable Thermus aquaticus DNA polymerase (Taq polymerase) which fascilitates the

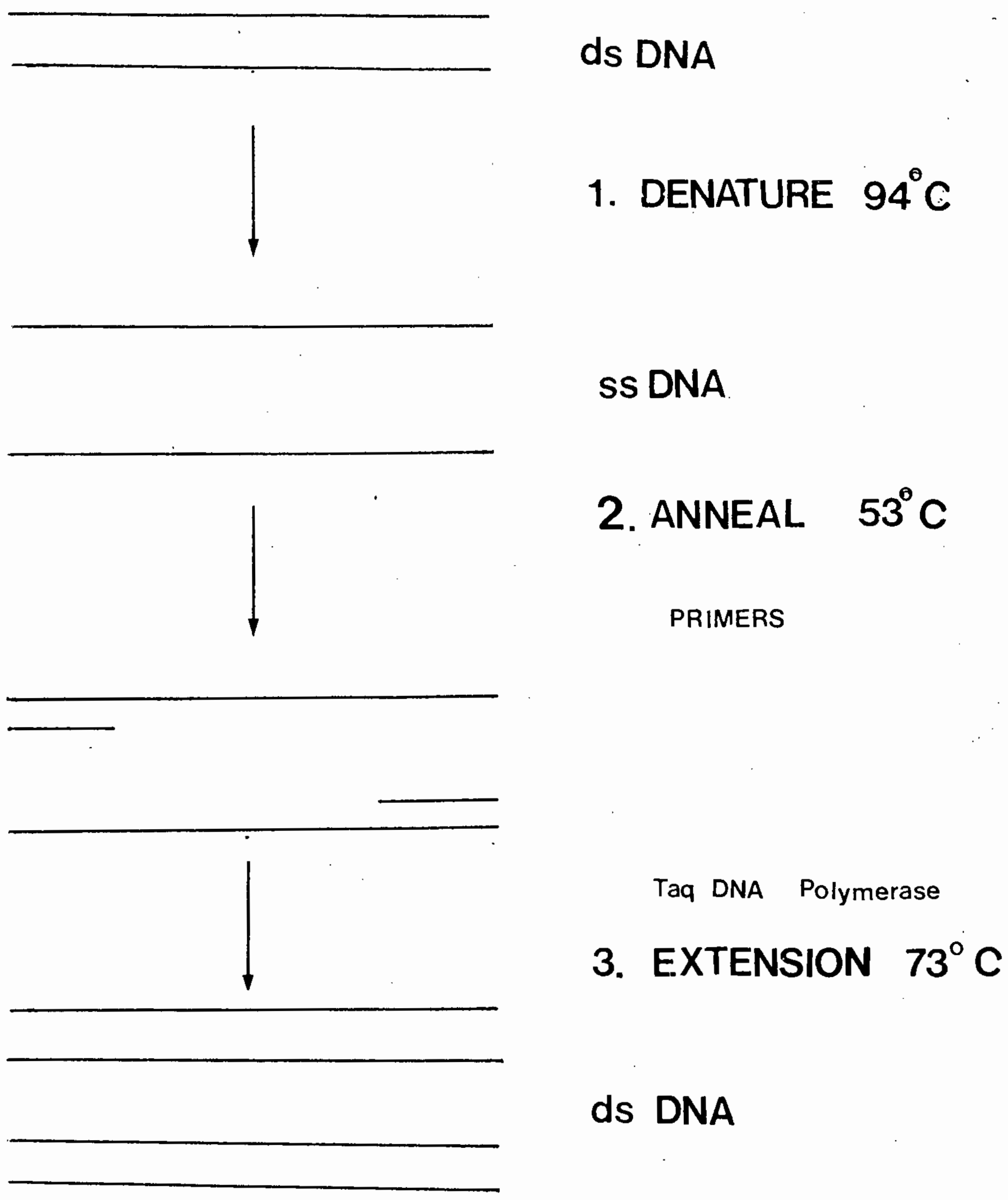


Fig. 1.9 Flow diagram of PCR

primer extension and the synthesis of new DNA strands. This step is usually performed at 72°C and the above three steps are known as a cycle or one round of DNA synthesis. One round of synthesis results in new strands of indeterminate length, which, like the original strands, can hybridize to the primers upon denaturation and annealing. These products accumulate exponentially with each cycle and the amount doubles with every subsequent cycle of denaturation, annealing and synthesis.

The speed of PCR is the main advantage over techniques such as gene cloning. Previously, researchers who wanted to produce sufficient quantities of a specific gene in order to perform sequence determination had to clone the gene, which could take weeks or longer whereas PCR takes a few hours (Marx, 1988).

The quality of DNA to be amplified is not very critical, for, PCR can work with unpurified as well as quite highly degraded DNA. Hence PCR becomes a method of choice in Molecular Biology over techniques like gene cloning and restriction mapping which require a highly purified DNA. As a consequence this has opened up new applications of the technique into analysis of ancient DNA samples (Paäbo *et al.*, 1989; Paäbo, 1989).

The quantity of DNA produced by PCR amplification is sufficiently large so that the high sensitivity imparted by the use of radioactive probes is no longer required and the amplified gene segments can be detected by means of non-radioactive probes or by ethidium bromide. Furthermore, a length of DNA as short as 50 - 100 base pairs of DNA can be amplified up (Marx, 1988; White *et al.*, 1989).

PCR can be used to enhance the sensitivity of prenatal diagnosis of many inherited disorders eg. sickle cell anaemia. Its application to human genetics has made it possible to analyze DNA sequences in single cells which, for example, can enable the determination of the sex of the human embryos fertilized invitro before implantation (White *et al.*, 1989).

1.6 PHYLOGENY RECONSTRUCTION

Phylogenetic trees can be constructed using both distance methods and cladistic methods. The two approaches are fundamentally different in principle. Cladistic approaches use shared derived characters (synapomorphies) to assign monophyletic groups. These approaches are mainly concerned with defining the topology (branching order) of the tree and generally are not used to provide branch lengths, from which pairwise sequence divergence tables can be constructed which in turn can be used to provide phylogenies with meaningful branch lengths and therefore estimation of the time. Distance methods on the other hand are based on the determination of the proportion of shared characters (eg. shared DNA fragments, restriction sites, or nucleotide) between different animals in a pairwise manner (Nei, 1987).

1.6.1 Cladistic Approaches

This approach uses qualitative characters (eg. spotted plumage) to compare members of a group. Such characters may be either primitive (*pleisomorphic*) or derived (*apomorphic*) (Sober, 1983).

The polarity of a character i.e. whether primitive or derived, can be determined by using an outgroup for comparison (Hillis, 1990). An outgroup is usually a taxon which is known to be slightly more distantly related to the group under study, than any of the taxa within the group are to one another. Characters which are ancestral to the whole group would be present in the outgroup as well as some members of the group under study and are described as *pleisomorphic* (primitive) whereas those that are unique to the group under study and therefore appear late in the development are classified as *apomorphic* (derived). Only *synapomorphies* (shared derived characters) are used to construct phylogenies in cladistic approaches. *Symplesiomorphies* (shared primitive characters) and *autapomorphies* (characters found in only one taxon) provide no information concerning phylogenetic affinities. For the DNA based methods, cladistic data is obtained by determining the number of phylogenetically informative characters (*synapomorphies*) deduced from either DNA fragment size comparison, or restriction site mapping or sequencing.

When assigning pylogenetically informative characters, characters which are common to all groups in question lack any phylogenetic information concerning the relationships between groups (eg. 1 in Fig. 1.10). Similarly, a character which is unique to a single taxon gives no additional information. The input data therefore consists of those phylogenetically informative characters, which are shared by at least two taxa and more than n-2 taxa, where n=total number of taxa under examination (eg. Fig. 1.10). In the example, 2 and 4 are phylogenetically informative because there are at least 2 shared characters whereas 1 and 3 lack phylogenetic information since in 1 there are no shared characters among all four taxa and 3 represents an

Fig. 1.10 **Determination of Phylogenetically Informative Characters**

	Position in sequence			
Taxa:	**1**	**2**	**3**	**4**
Taxon 1	G	G	T	C
Taxon 2	G	A	T	C
Taxon 3	G	G	A	T
Taxon 4	G	A	T	T

Taxa can be at any hierarchial level.

G, A, T, C = Nucleotides

Only position 2 and 4 are phylogenetically informative since at least two characters are shared by at least two taxa whereas 1 is uninformative since all of the characters are shared between the four taxa and 3 represent an autopomorphic change of a T to an A.

autapomorphic change of a T to an A.

Several cladistic methods are available, but the method of maximum parsimony is the most widely used method. The principle of the method is to infer the nucleotide sequences of ancestral species and choose a tree that requires a minimum number of mutational changes, i.e. the shortest tree (Felsenstein, 1983; Nei, 1983; Hillis and Moritz, 1990). Such a tree is known as the most parsimonious tree. (See in Fig. 1.11 for example). Trees constructed by this method are usually unrooted and this has a number of implications for example, although one can determine that an evolutionary change (mutation) has occurred, one cannot necessarily state the order or the direction of a change. An outgroup, can be used to establish an ancestral state and to give a root to the tree.

Several parsimony methods are available. These include Wagner parsimony, Camin-Sokal parsimomy, transversion and Dollo parsimony (Felsenstein, 1983b; Felsenstein, 1985; Hillis and Moritz, 1990). They are all based on minimizing some evolutionary significant quantity but differ in their evolutionary assumptions and methods (Felsenstein, 1983b). All of them use discrete character states ie. "0" and "1" for an ancestral state and a derived state, respectively.

Dollo parsimony assumes that the derived state is the more complex state and is allowed to rise only once, but can however revert to state 0 as many times as are necessary. Furthermore, the use of this method does not assume the existence of a hypothetical ancestor; it can also be applied to construct unrooted trees (Felsenstein, 1983). Although it might seem more appropriate for restriction site data than

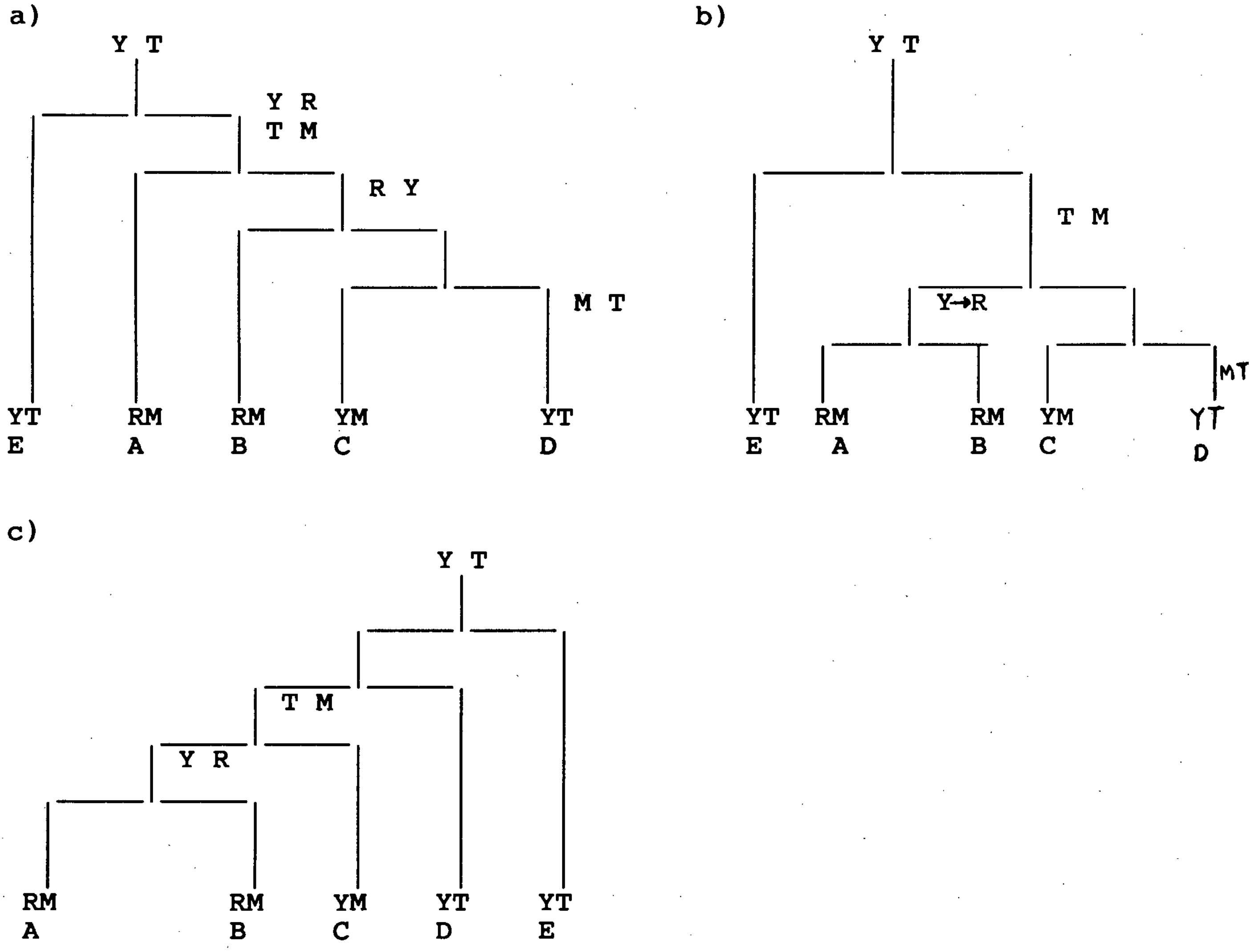
Fig. 1.11 Principle of the Cladistic Approach

A, B, C & D are taxa of say, birds, and E is an outgroup known to be slightly more distantly related to each of the taxa than any of the other are to each other.

A and B have red legs (R) while C, D and E have yellow legs (Y). A, B and C have massive bills (M), D and E have thin bills (T).

The outgroup, E is a reflection of what the ancestral state of the characters was (Yellow legs and thin beak) Y - Yellow; R - Red; T - Thin; M - Massive. The number of mutations required is 4, 3 and 2 for a, b and c, respectively therefore, C is the most parsimonious tree with respect to these two characters since it possesses the minimum number of mutations (substitutions).

Several possible different trees can be constructed from the number of mutations required by the topology (branching order) of the tree.



for sequence data, since parallel losses are theoretically more frequent than parallel gains, factors such as non-uniformity of sites with respect to frequency of change, and high transition bias in some DNA species, result in a significant frequency of parallel gains, which are disallowed in Dollo parsimony (Hillis and Moritz, 1990).

Cladistic methods tend to work better for analyses with greater than 4 taxa (4 being the minimum) and uses less of the data than distance methods since characters found in only one taxon (*autapomorphies*) are not used ie. cladistics uses only *synapomorphies*. On the other hand, the determination of sequence divergence in distance methods has been criticised for loosing the information given by the hypothesis of *synapomorphy*.

Algorithms for constructing trees and approximating the best topologies have been developed (Farris, 1970; Fitch, 1971; Felsenstein, 1983). Computer packages implementing these include Farris's Henning 86, Felsenstein's PHYLIP and Swofford's PAUP (Phylogenetic Analysis Using Parsimony) (Hillis and Moritz, 1990). These are all parsimony based packages. PAUP and Henning 86 algorithms are the most widely used and provide large increase in efficiency with which phylogenetic solutions can be obtained (Platnick, 1988).

Each of these algorithms implement both exact and approximate options. The exact algorithms use branch - and bound strategy to elimine suites of cladograms that cannot possibly be among the shortest arrangements. The efficiency with which these algorithms work is data dependent.

The approximate algorithms, on the other hand make an initial estimate of the best cladogram and then systematically rearrange their branches in search of shorter trees. Unlike the 'branch and bound', the use of these algorithms does not guarantee iteration to the shortest tree but in practise they do well and run quickly. For large data sets, they are consequently the only feasible approaches at the present time.

Weighting schemes which determine the phylogenetic information content of characters can be used to emphase or de-emphase different types of nucleotide character changes in phylogenetic analysis (Mindell and Honeycutt, 1990). A variety of weighting schemes can be used namely, equal weighting or a scheme where transversions are weighted more heavly than transitions (Halanych, 1991). This we used in cases where there is large degree of sequence divergence and transitions become less uninformative due to reversals. Alternatively stem-region nucleotide bases can be given heavier weighting than loop-region nucleotides (Mindell and Honeycutt, 1990; Halanych, 1991).

A further approach uses successive weighting in which weights are based on consistency or agreement, within the data set with no required prior decisions regarding weighting. Henning 86 algorithm possess such a facility for use in tree building (Mindell and Honeycutt, 1990). The PAUP package, implementing Wagner parsimony has been used extensively in evolutionary studies. Recent examples of its applications include the determination of the relative rates of DNA evolution in primates (Easteal, 1991) and evolution and application of rRNA in vertebrates (Mindell and Honeycutt, 1990).

1.6.1.1 Bootstrap Analysis

A recently developed statistical method called Bootstrapping can be used to test the robustness of the various branches of the phylogenetic trees constructed (Felsenstein, 1985; Benton, 1988).

Bootstrapping technique operates by resampling the data set by drawing points from it with replacement until a data set of the same size as the original are obtained (Felsenstein, 1985; Felsenstein, 1988; Moritz and Hillis, 1990). Each data set therefore consists of a new set of characters (restriction sites or nucleotide in this case), from the original data set by the process of random sampling. Within any new set, characters from the original set may be repeated or omitted. Each of these data sets can then be analyzed by tree-making algorithms to obtain an estimate of the phylogeny. A majority consensus rule tree is considered to be an overall bootstrap estimate of the phylogeny (Felsenstein, 1985).

1.6.2 Distance Approaches

For restriction site data the proportion of shared characters is calculated from formulae such as that of Nei and Li (1979):

$$S = \frac{Z}{x+y-z} \quad \text{or} \quad \frac{2Z}{x+y}$$

S = fraction of DNA restriction sites shared between two speci

x & y = the total number of restriction sites in taxon A and B respectively

Z = number of restriction sites shared by both taxon

The proportion of shared sites can then be used to estimate the percentage sequence divergence. The proportion of shared characters is represented in the form of a matrix (see Fig. 1.12 for example).

The sequence divergence is estimated by $\log_e - \underline{(s)}\ r$

Where r = Average number of nucleotides in the restriction cutting sequence, with the variance given by

$$\begin{array}{l} \underline{S(1-S)\quad -S^{2}\ (1-S^{0.5})^{2}} \\ ((\underline{x+y})\quad X\quad \underline{(rXs)^{2}}) \\ (\ Z\ ) \end{array}$$

A matrix of all pairwise values then serves as input data for the construction of a phylogenetic tree, which links taxa in a number of different ways, eg. UPGMA and neighbor joining. If these ratios of mutations in the DNA (and the taxa) under study have been calibrated in the acceptable way (Brown *et al.* 1979). against the geological or fossil record, then it may be acceptable to equate the branch lengths on the tree to an absolute time scale. Stochastic variation, which is a fraction of the number of changes recorded in the DNA, is the predominant source of error.

Several distance based methods of constructing phylogenies are available. These include UPGMA (unweighted Pair Group Method), Fitch-Margoliash's and Farris's distance Wagner methods and the Neighbor joining method of Saitou and Nei (1987).

UPGMA is one of the most commonly used methods. UPGMA works by assigning branch lengths (or mode levels) to the sum of squares of

Fig. 1.12 **Example to illustrate distance approach**

Example: For DNA sequences from 3 taxa:

Find the proportion of shared nucleotides in a pairwise mann

```
                                            Pairwise     Proportions
Taxon  1    A G C T T T G C A G  ---\
                                     |---7/10------\
Taxon  2    A C C T T A G C T G  ---/               |----5/10
                                   |-----4/10------/
Taxon  3    T G C A A A G G A G --/
```

These pairwise proportions may be tabulated in a grid form, sequence divergence values are easily calibrated as (1x) where x is the proportion value.

Taxa	1	2	3
1	1		
2	0.7	1	
3	0.5	0.4	1

It is evident that taxa 1 and 2 are more closely related to each other than either is to taxon 3.

differences between the observed and expected distances. Trees produced by this method can be rooted (Nei, 1985). Tateno *et al.* 1979; Nei *et al.* (1983) have shown that when distance estimates are subject to large stochastic errors, UPGMA is often superior to other distance methods in constructing the tree accurately. The method can be applied with relative ease to distance values estimated from restriction - site mapping and protein electrophoretic data. The only drawback in using this method is that when gene substitutions vary extensively from lineage to lineage, i.e. when there are differences in evolutionary rates the branch lengths cannot be unambiguously determined. This can result in the production of an incorrect topology.

Fitch and Margoliash's distance method, produces unrooted trees which will usually have similar topology to that of the UPGMA trees but with slightly different branch lengths. The topology of the trees constructed by Farris's Distance Wagner method also resemble the UPGMA and Fitch and Margoliash's trees with the main difference being that the branch lengths are considerably shortened near the roots (Felsenstein, 1983).

In this study, we used the PHYLIP phylogeny inference package, designed by Joseph Felsenstein (1983). Both 'FITCH' and 'KITCH' option were used (Fitch - Margoliash, 1967). The KITSCH algorithm assumes that the rate of evolution is constant and produced rooted trees with contemporaneous tips whereas the FITCH method is not based on this assumption and produced unrooted trees with varying terminal branch lengths.

CHAPTER 2 : GENERAL METHODS

2.1 MITOCHONDRIAL DNA ISOLATION

2.1.1 Mitochondrial DNA Extraction

Mitochondrial DNA (mt DNA) was extracted according to the method of Cummings et al. (1981) & Smith (1971) with minor modifications. Heart and liver were the tissues of choice for the isolation of mt DNA since the latter is especially abundant in these tissues. Approximately 30 g of tissue was cut into small cubes, resuspended in 4,5 ml x (mass of tissue) of extraction buffer (100 mM Tris, pH 7.4; 150 mM NaCl; 20 mM EDTA; 10% w/v sucrose) and centrifuged at 1000 g for 10 min at 4°C in a Beckman JA14 rotor. Subsequent steps were also performed at 4°C in order to minimize degradation of mt DNA by DNAses. The pellet recovered was resuspended in 3 x (mass of tissue) of extraction buffer and homogenized at full speed for 20 sec in a Phillips household blender in order to disrupt the cell membranes. To separate heavy nuclei from light organelle material, the homogenate was centrifuged at 1000 g for 10 min at 4°C, and the supernatant was filtered through cheese cloth to remove debris and residual fat particles. The filtrate was centrifuged at 10 000 g for 15 min in order to recover the mitochondrial pellet. The mitochondrial pellet was resuspended in 4 ml STE buffer (100 mM NaCl; 50 mM Tris, pH 7.4; 100 mM EDTA).

2.1.2 Mitochondrial DNA Purification by Ethidium Bromide Cesium Chloride Density Gradient Centrifugation

1% (w/v) SDS was added to the mitochondrial suspension in order to lyse the mitochondrial cell wall. The mixture was then made to 1.0M in cesium chloride and left for 10 minutes at room temperature to allow for contaminating proteins to precipitate out. The mixture was centrifuged at 10 000 g using a Sigma 2MK rotor (1200 rotor)/1000 rpm at 25°C. The mitochondrial DNA supernatant was decanted and cesium chloride added to a final concentration of 1 g/ml and the ethidium bromide added to a final concentration of approximately 80 µl/ml. The density of the mixture was further adjusted to 1,55 g/ml with solid cesium chloride. The solution was subsequently loaded into Beckman quickseal centrifuge tubes. The tubes were heat-sealed and the mixture centrifuged at 50 000 r.p.m. for 20 hours at 25°C using a VTi 65 rotor. The bands of DNA were visualised under ultra violet light at 375 nm wavelength. In most cases only the top nuclear band was visible. In such instances the fraction at approximately 4-10 cm below the nuclear band was collected which represented the area where supercoiled mt DNA would be expected to collect.

2.1.3 Removal of Ethidium Bromide

Ethidium bromide was removed from collected fractions by the addition of an equal volume of salt saturated isoamyl alcohol. This was mixed briefly and the two phases were allowed to separate. The upper, organic layer was discarded and the lower aqueous phase extracted until there were no visible traces of ethidium bromide. Sterile distilled water was added to the DNA preparation (2:1, v/v) followed by 100% ethanol (6:1 , v/v). This mixture was placed at -20°C for about 30min and DNA precipitated by centrifugation in a JA20 rotor at 10 000 r.p.m. for 30 min at 4°C. The mt DNA pellet was washed with 70% ethanol and centrifuged at 10 000 r.p.m. for 20 min at 4°C. Purified mt DNA was vacuum dried, resuspended in 400 µl of TE (1.0M Tris - HCl pH 8.0; EDTA pH), aliquoted and stored at -20°C.

2.2 RESTRICTION MAPPING

Purified mt DNA was subjected to cleavage by restriction endonucleases using either single or double digests. Restriction endonuclease digests were usually carried out according to the suppliers instructions (Boehringer-Mannheim, Amersham, New England Biolabs′ or Anglian Biotec). However, in certain instances partial digestion occurred and the incubation time was increased. In addition, care was exercised when working with certain restriction endonucleases, eg. Eco RI, since high restriction endonuclease concentrations, low ionic strength conditions and high glycerol concentrations relax their recognition sequence specificity leading to multiple cleavage (star activity).

Digests were performed by using 2 X KGB buffer (200 mM K^+ glutamate; 50 mM Tris - acetate (pH 7.6); 20 mM magnesium acetate; 100 µg/ml Bovine Serum Albumin; 1 mM 2 - ß Mercaptoethanol) since it proved adequate for both single and double digestions (Mc Clelland, 1988). The restriction endonuclease reaction mixture consisted of 2 µl (20 ng) mt DNA; 7 µl 200 mM KGB; 7µl double distilled water and 1 unit reaction of restriction endonuclease. The reaction was carried out in a microfuge eppendorf tube and was mixed briefly using a benchtop microcentrifuge followed by a two hour incubation period at 37°C.

2.3 END-LABELLING

Because of its high sensitivity and ease with which it can be performed, end-labelling was used in most of our experiments. Restriction fragments were labeled with Klenow fragment of DNA polymerase I, three unlabeled nucleotides (G , A, T, or C), followed by the fourth radiolabeled nucleotide triphosphate. Most restriction enzymes cut DNA to generate 5' or 3' overhangs, or blunt ends. For end-labeling purposes 5' overhangs are preferred, but in instances where 3' or blunt ends are produced, Klenow fragment of DNA polymerase I is used to generate 5' overhangs utilizing its 3' - 5' exonuclease activity (see fig. 2.1 for the mechanism of action of Klenow). Since dCTP was regularly used as the labeled nucleotide the activity of the Klenow was also necessary where the 5' overhang did not contain a complementary G (eg. EcoRI).

To the digested mt DNA, 0,5 units/ reaction of Klenow enzyme of DNA polymerase I was added and allowed to stand for 10 minutes at room temperature inorder to fascilitate the 3'-5' exonuclease activity to operate. 2 mM of each dTTP, dATP, dGTP, and 1µCi of α ^{32}P-dCTP was then added, the solution briefly mixed and incubated at room temperature for 15 minutes. Alpha ^{32}P-dCTP has a half life of 14 days, so after this period had elapsed the volume was doubled. The reaction was terminated by adding 4 µl of loading buffer (40% sucrose and 0.25% bromphenol blue).

2.4 AGAROSE GEL ELECTROPHORESIS

Restriction fragments were analyzed according to the method of Southern (1975). Agarose gel percentage used depended on the size of DNA restriction fragments to be analyzed. For good separation of large restriction fragments, typically 2kb, 2% agarose gels were used and for relatively small restriction fragments of approximately 200 bp, we used 1 -1.5 %. End-labeled mt DNA restriction fragments were

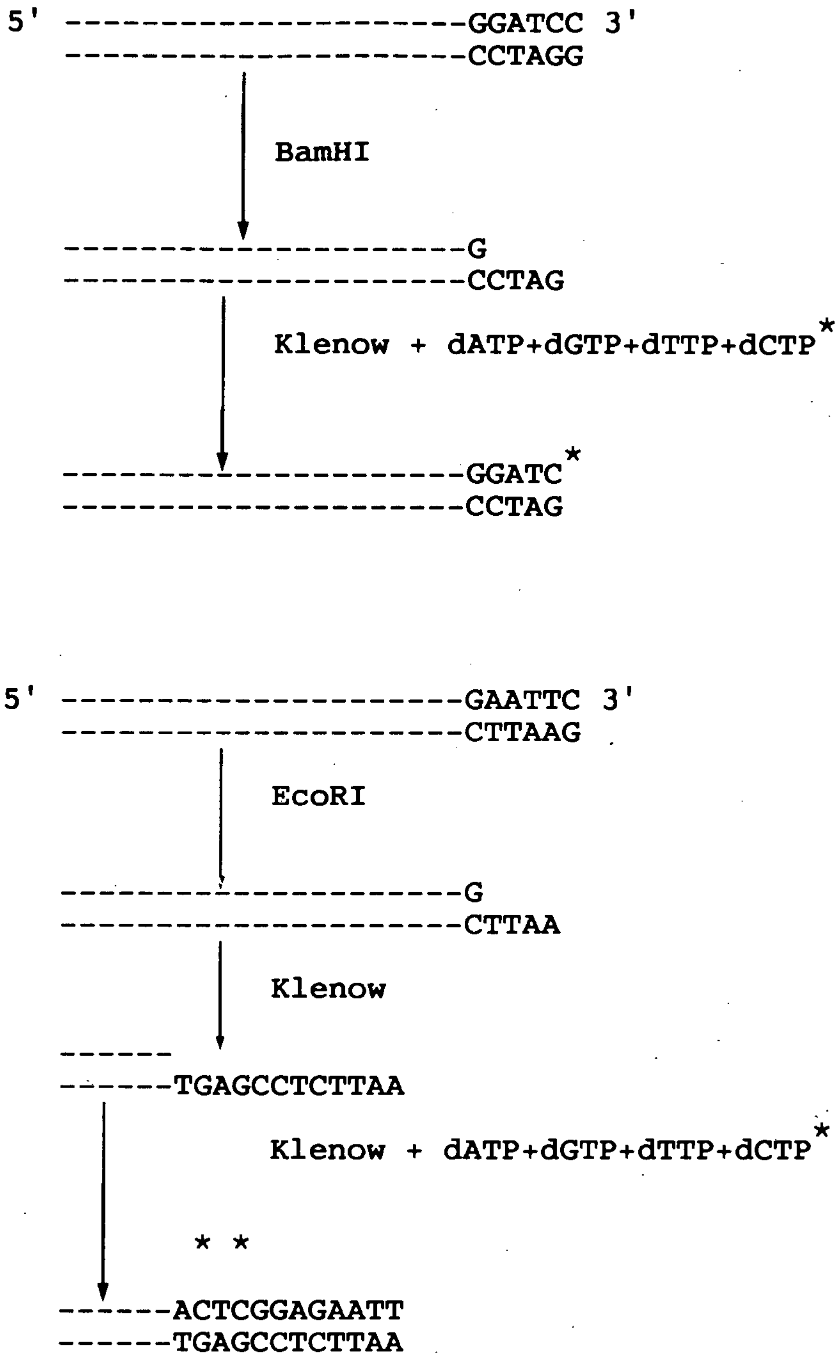


Fig. 2.1 **End-labelling with the Klenow fragment of DNA polymerase I**

loaded into the wells of the agarose slab gel and run in TAE buffer (Tris-acetate ; 0.002M EDTA) at a potential difference of 2 - 10 volts per cm applied across the gel. For good band resolution a low voltage, typically 35 V, was applied for a long period (overnight) while for quick checking gels a relatively high voltage was applied for a short period of time, typically 4-6 hours.

Agarose gels were dried under vacuum at 55°c and loaded onto cassettes. Autoradiography was performed using hyperfilm MP (Amersham).

2.5 CELL CULTURE METHODS

Due to the difficulty in getting post-mortem tissue for some of these species (eg. black-footed cat) and the fact that sometimes mt DNA was highly contaminated with nuclear DNA, total genomic (nuclear) DNA was extracted from either skeletal muscle or cell cultures. This genomic DNA was then used to carry out Southern blotting hybridization using mt DNA as a probe (Southern blotting hybridization to be discussed later).

2.5.1 Establishment of Fibroblast Cultures

Fibroblast cultures were established from fresh skin biopsies taken from the species in question. Particularly effective were small full thickness nicks taken from the pinna of tranquilised or (in the case of smaller species) restrained animals, a procedure which is quick and appears to entail minimal discomfoture to the animal. The tissue was cut into very small pieces, which were placed under glass cover-slips in plastic petri - dishes in Ham's F10 medium supplemented with 15% foetal bovine serum. After a few days, when cells were seen to be growing, the cover - slips were inverted and placed in different petri - dishes. When cells were confluent, they were trypsinised into 25 cm Falcon plastic flasks.

2.5.2 Maintenance of Cell Cultures

Cells were grown in Falcon plastic culture flasks in 5% CO at 37°C and were routinely maintained on Ham's F10 medium containing 15 % FCS with no antibiotics. The reason for the latter is that it helps avoid unrecognised contamination with mycoplasma introduced by poor manipulative techniques. The medium was changed twice a week. When the cells were confluent, they were trypsinised by adding 2 ml of trypsin, incubated at 37°C with intermittent shaking until they were detached from the flask. Approximately 10 ml of medium was added to the mixture and the resulting suspension was divided into two other flasks.

2.5.3 Cell Harvesting

When the cells were confluent, the medium was removed and the cells washed twice with 3 ml of saline. Two millimeters of trypsin was added to each flask and incubated at 37°C for 5min with intermittent shaking. The trypsin/cell mixture was transferred with a pippete to sterile 15 ml tubes. The latter step was repeated and the mixture topped up to the same level with saline. Cells were spun down at 1000g for 10 min and the supernatant was discarded. The harvested cells were then ready for total genomic DNA extraction.

2.6 TOTAL GENOMIC DNA ISOLATION

Genomic DNA was isolated from both cell cultures and skeletal muscle according to methods described below.

2.6.1 Total Genomic DNA Isolation from Cell Culture

The method of purification is as described in Maniatis et al. (1987). Freshly trypsinised skin fibroblast cells were resuspended in a solution of 0.8 ml TE buffer; 1% SDS; 100 µg proteinase K. This was followed by a two hours incubation at 37°C afterwhich 0,1M sodium perchlorate was added and extracted with an equal volume of phenol chloroform to remove proteins. Phenol chloroform extraction was repeated followed by two extractions with (1:1) chloroform octanol. To the aqueous phase, two volumes of cold absolute ethanol was added, and the mixture incubated at -20°C for 30 min. The suspension was centrifuged at 10 000g for 30 min at 4°C using a sigma 2MK centrifuge and the pellet washed with 70% alcohol, vacuum dried and resuspended in 100 µl of TE buffer.

2.6.2 Total Genomic DNA Isolation from Skeletal Muscle

Total genomic DNA was isolated and purified using the method described in Maniatis <u>et al</u>. (1987).

Approximately 5 g of tissue (skeletal muscle) was minced in a Philips blender, resuspended in 1,2 ml/100 mg digestion buffer (100 mM NaCl; 10 mM Tris-HCl, pH 8; 25 mM EDTA, pH 8; 0.5% SDS; 0,1 mg/ml proteinase K.) and incubated at 50°c for 12-15 hours in tightly capped pyrex test tubes. Protein was removed by extracting the suspension with an equal volume of phenol/chloroform/isoamyl alcohol followed by centrifuging at 1000 g using a Beckmann JA20 rotor. To the aqueous layer half as much 7,5 M ammonium acetate and twice the original volume of 100% alcohol were added to precipitate DNA which was recovered by centrifugation at 5000 g for two minutes and washing with 70% ethanol to remove any residual salts and phenol. The DNA was dried under vacuum, resuspended in 200 µl of TE buffer and stored at room temperature until it had completely dissolved. It was then stored at -20°C.

2.7 SOUTHERN BLOTTING HYBRIDIZATION

Total genomic DNA was cleaved with restriction endonucleases either single or in pairs. The restriction digestion reaction consisted of 20 µg of DNA, 3,0 µl of 1 X KGB buffer and 10 units of restriction endonuclease. The reaction mixture was incubated at 37°C for 6 hours. The restriction fragments were then separated by agarose gel electrophoresis at 35 V for 18 hours. In order to denature the DNA the gel containing the restriction fragments was incubated with shaking in an alkaline medium (1.5 M NaCl + 0.4 M NaOH) for 30 min at room temperature. This was followed by another 30 min incubation in a neutralization buffer (20 SSC ; Tris - HCl pH 8.0) at room temperature. The transfer was allowed to proceed at room temperature overnight (see fig. 2.2).

2.7.1 Preparation of the Radioactive Probe

The multiprime labeling system (Amersham) was used to radioactively label the DNA.

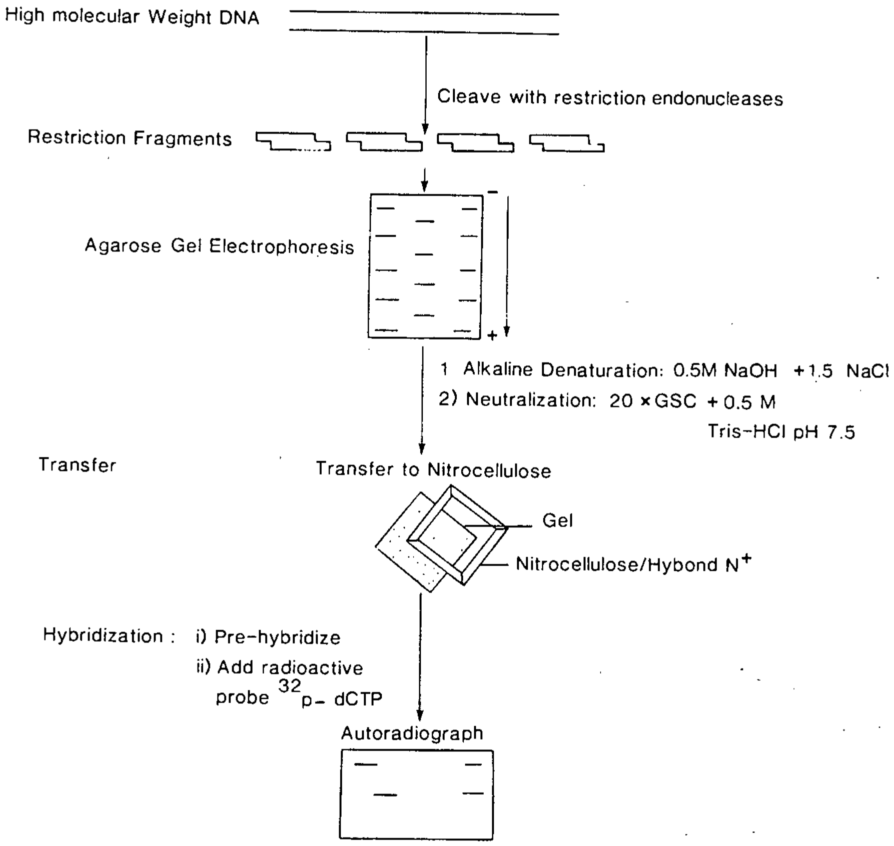


Fig 2: Flow Diagram Showing Southern Blotting Hybridization

Fig. 2.1 Flow diagram of Southern blotting hybridization

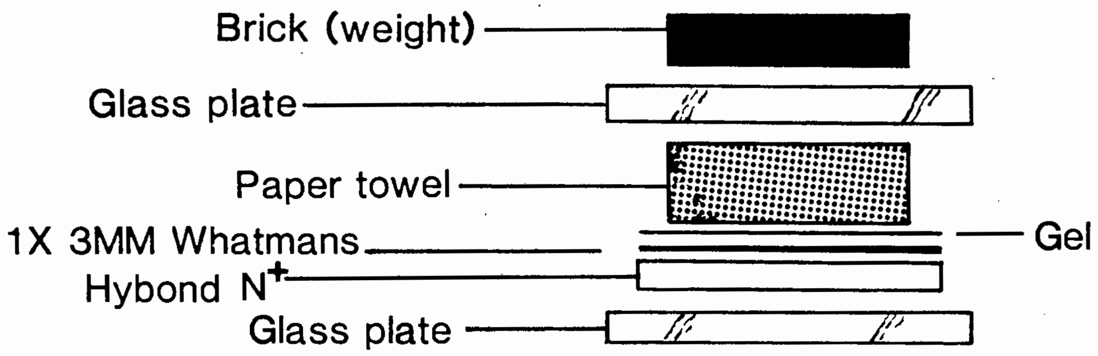


Fig X: Diagram of DNA Transfer Apparatus

Fig. 2.2 Southern blotting - transfer of DNA fragments to the membrane

Approximately 25 ng of mt DNA was linearised by digestion with five units of Hind III, followed by incubation at 37°C for two hours. Mt DNA was denatured by placing the tube in a boiling water bath for 5-10 min and then rapidly chilled in ice for 2 min to prevent renaturation. The next step was carried out on ice. Five microlitres of 25 ng linearised and denatured mt DNA was end-labeled by treatment with : 3 µl of a mixture of dATP, dTTP, dGTP; 2 µl of reaction mix; 50 µCi of α-^{32}P dCTP; 3 µl of sterile distilled water; 5 units of Klenow. The reaction was allowed to proceed for 2 hours at 37°C and terminated by the addition of 20 µl of 0.2 M EDTA pH 8.0.

2.7.2. Removal of Unincorporated Nucleotides from the Probe : Gel Filtration Chromatography using Sephadex G-50

One millilitre pasteur pipette was used as a column. This was packed with sephadex G-50, eluted with TE buffer and 150 µl fractions were collected. 1/100 volume aliquots were counted using Cerenkov radiation (Williams and Wilson, 1975) and the fractions containing the ^{32}P-labeled mt DNA were pooled. The specific activity of the radioactive-labeled mt DNA ranged from 1-2 x 10^7 dpm/µg.

2.7.3 Hybridization

Hybridization was performed as described by Johnson et. al.(1984). Prior to hybridization, a prehybridization step was performed. This entailed washing the membrane in 100 ml of a prehybridization solution (6 x SSC; 0.06% Na-PP; 0.25% blotto) at 65°C for at least 1 hour in a shaking water bath. The membrane, the prehybridization solution and the denatured radioactive probe were placed in a plastic bag. The plastic bag was heat sealed, incubated at 65°C and hybridization allowed to proceed overnight. After hybridization the membrane was washed with a solution of increasing stringency to remove non-specifically bound radioactive probe DNA. The first three washes were performed in 250 ml of 6 x SSC ; 0.1% SDS ;0.06% Na-PP ; at 65°C for 15 min. The fourth wash consisted of 3x SSC ; 0.1% SDS; 0.06% Na-PP, at 65°C for 15 min. This was followed by a 15 minutes wash at 65°C in 250 ml of 0.1 x SSC ; 0.1 % SDS ; 0.06% Na-PP which has a lower salt concentration, thus increasing the stringency of the wash. After the final wash, the membrane was exposed to an x-ray film and placed at -70°c for autoradiography.

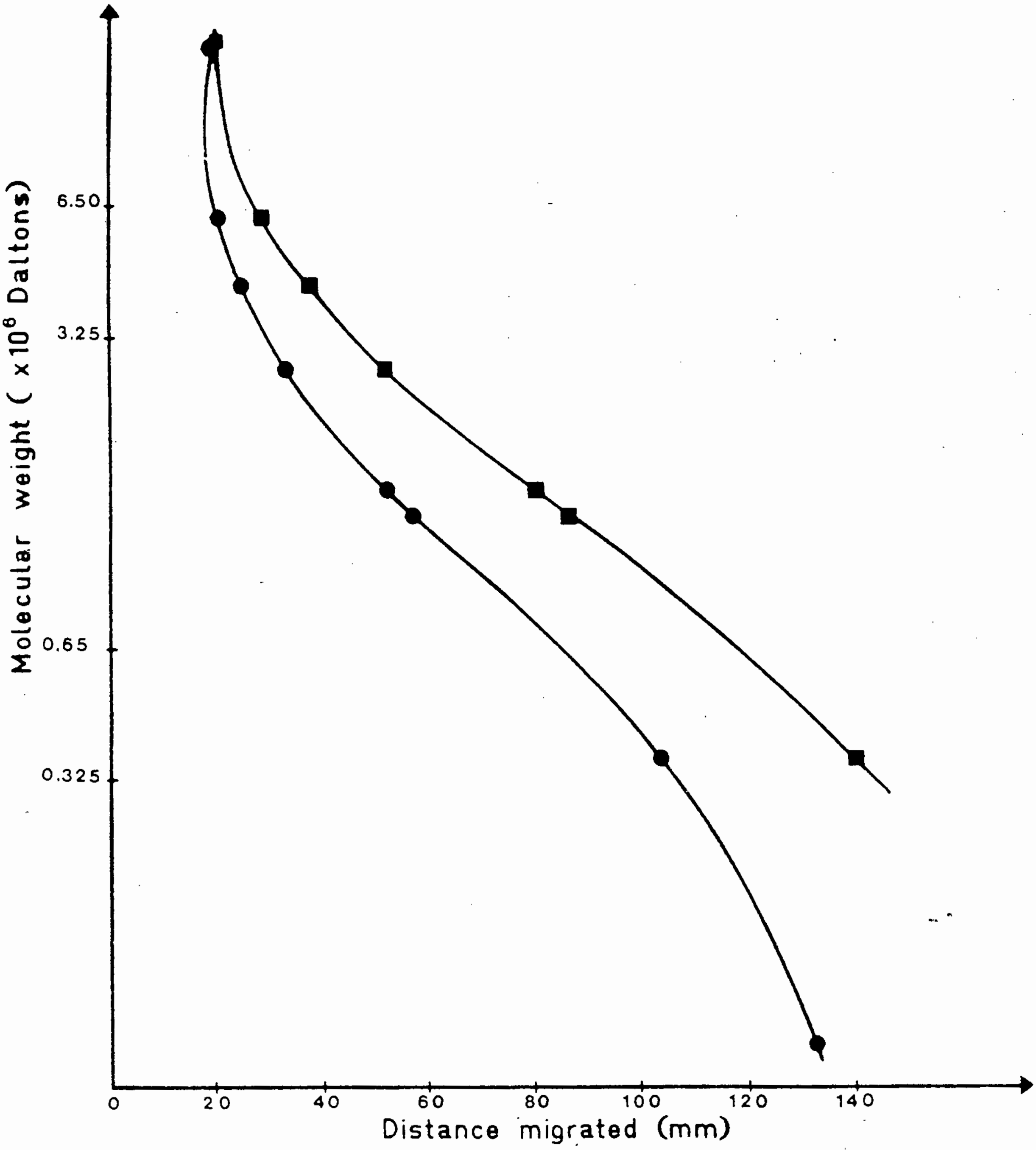
2.8 DETERMINATION OF MOLECULAR WEIGHT

Lambda DNA digested with Hind III was used as a marker for determination of restriction fragment sizes. A plot of lambda HindIII sizes (in base pairs) versus the distance of migration of restriction fragments was drawn on a semi-log graph paper and the restriction fragment sizes obtained by measuring the distance of migration of each restriction fragment from the calibration curve. (See Fig. 2.4 for the calibration curve). Fragment sizes of 9 000 base pairs and above were obtained by subtracting the accurately measured restriction fragments from 16 400 bp which is thought to be the average size of mt DNA in mammals (Lewin, 1983). The same strategy was applied to fragments smaller than 125 bp.

2.9 RESTRICTION MAPPING STRATEGY AND DATA MANAGEMENT

Restriction data obtained from single and double digestions of mt DNA

Figure 2.4 Molecular weight standard curve of bacteriophage lambda DNA digested with *Hind*III on semi-logarithmic paper, for a 1% (■) and 2% (●) agarose gel. 1 kb DNA is ~ 0.65 x 10^6 Da of double-stranded DNA (sodium salt) (Ausubel *et al.*, 1987)



were entered into the computer program, RESOLVE (Version 2.4; Harley, 1991 unpublished) which allows easy handling and manipulation of restriction mapping data. The data was initially stored on a temporary file which finds and stores solutions found to all double digests analyses. The program (then) allows for examination of all possible combinations of digestion for a specified triplet of restriction endonucleases (three-way analysis, see appendix II) and can then find a compatible solution to all such temporary solutions at a specified percentage error. Only those maps produced at 5% error or less were entered into the final map file.

2.10 ALIGNMENT AND ORIENTATION OF CLEAVAGE MAPS

The alignment of cleavage maps were achieved by using two SacII sites which cleaves mt DNA to produce a 1680 bp fragment which is invariant in virtually all vertebrates surveyed thus far (Carr, 1978) with the exception of the bats. Restriction endonuclease ClaI was used for the orientation of cleavage maps since it cleaves the 1680bp SacII fragment asymmetrically at position 990.

2.11 PHYLOGENETIC ANALYSIS OF CLEAVAGE MAPS

2.11.1 Cladistic Methods

In the present study, both PAUP and Henning 86 were utilized for phylogenetic analysis. To test the robustness of the topologies produced by the latter methods, the phylogenetically informative data (restriction mapping or sequencing) were run, with 1000 iterations, through the Bootstrapping program in J. Felsenstein's PHYLIP (Phylogeny Inference Package, version 3.1).

From the percentage sequence divergence, the phylogenetic relationships were established.

2.11.2 Distance Methods

PHYLIP phylogeny inference package was used to construct our phylogenies. Both FITCH and KITSCH were the main methods of analysis. Neighbor-joining method (NJ TREE), which was developed by Saitou and Nei (1987) was also used to construct a phylogeny.

CHAPTER 3 : MITOCHONDRIAL DNA RESTRICTION SITE ANALYSIS

3.1 INTRODUCTION

Despite being more time consuming than restriction fragment size comparisons, it is well documented that restriction mapping approach produces more precise and reliable data sets. Its high precision at the genus and species level makes it particularly useful as the basis of comparison in this study since the classification problem in the family *Felidae* is mostly at the genus level.

Restriction maps for eight Southern African *felidae* were analysed from which we attempted to construct trees using both distance and cladistic methods. Four members of the larger cats namely; lion, leopard, cheetah and caracal and four members of the smaller cats namely: African wild cat, domestic cat, European wild cat and black-footed cat were analyzed. These had not been studied previously using mt DNA and their larger size provides a rich source of biological material to allow for successful mt DNA extraction and purification.

3.2 MATERIALS AND METHODS

The source of biological material for all eight species used are summarized in table 3.1.

Detailed procedures of methodologies have been described above (see section 2.3).

Table 3.1 **Source of biological material of eight members of Southern African Felidae**

Name of Species	**N**	**Material**	**Source of Material**
A. jubatus	2	Post-mortem	1-Tygerberg Zoo 1- North Windhoek near Okahaja, Namibia
C. caracal	2	Post-mortem	Roadkill in Simonstown, Cape Town. South African Museum (Mr Rau)
F. catus	6	Post-mortem	St. Francis veterinary hospital, Cape Town, (Mr Briggs)
F. lybica	4	Post-mortem	1-Napier - western Cape Museum, Cape Town
F. sylvestris	2	Post-mortem	National wildlife Research Centre in Taif, Saudi Arabia (J. Flammand)
P. leo	1	Post-mortem	Etosha, Namibia
P. pardus	1	Post-mortem	Grahamstown, Eastern Cape
F. nigripes		Cell culture	Victoria West, Central karroo

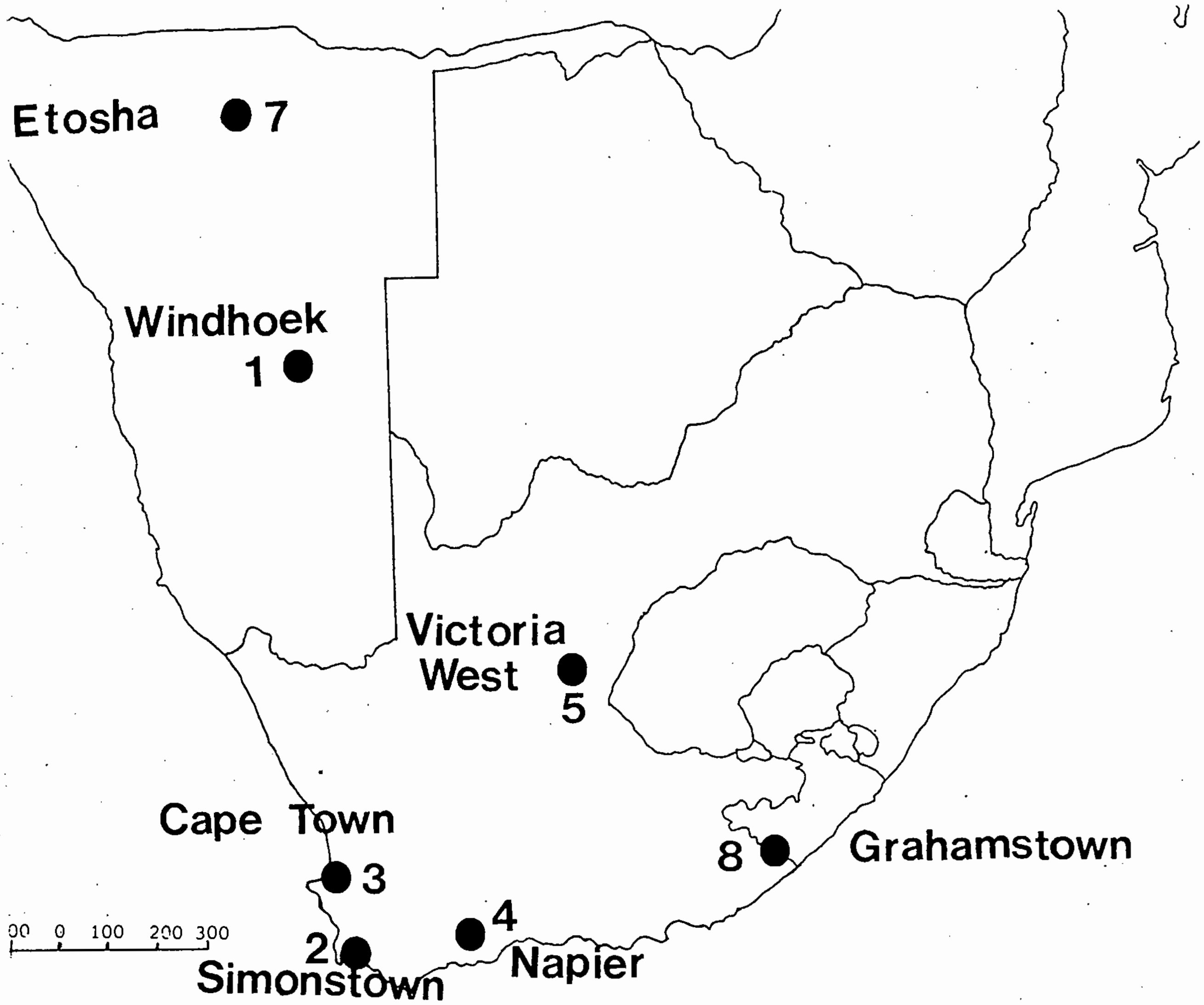


Fig. 3.1 **Map of southern Africa to illustrate the localities of eight southern African Felidae**
The number represent the names of species: 1 = Acinonyx jubatus; 2 = C. caracal; 3 = F. catus; 4 = F. lybica; 5 = F. nigripes; 6 = F. sylvestris; 7 = P. leo; 8 = P. pardus; F. sylvestris was found in Saudi Arabia.

3.2.1 Restriction Mapping

Purified mt DNA of each species was treated with a total of 16 restriction endonucleases (Asp 718, Bam HI, Bgl II, Cla I, EcoR I, EcoR V, Hind III, Nco I, Pst I, Pvu II, Sac I, Sac II, Sal I, Stu I and Xho I). Restriction fragments were end-labeled with ^{32}P-dCTP. In some cases, mt DNA was highly contaminated with nuclear DNA and therefore unsuitable for direct end-labelling with ^{32}P-dCTP. Approximately 5 µg of such contaminated preparations were run per lane on agarose gels, transferred to hybond N^+ membrane and probed with purified, radioactively labeled domestic cat mt DNA. In cases where there was no post-mortem material for mt DNA extraction (eg. black-footed cat) the same method of autoradiographic visualization was employed on total genomic DNA extracted from cell cultures.

Fragment sizes were determined and restriction maps constructed independently for all eight species. Restriction maps were aligned with Sac II sites which are invariant throughout the vertebrates studied to date with the only exception being that of the bats. Maps were oriented by using Cla I, which assymetrically cleaves the conserved Sac II fragment at site position 990.

3.2.2 Phylogenetic Analysis

3.2.2.1 Cladistic Methods

Table of phylogenetically informative sites were produced from restriction maps using the RESOLVE program (Harley 1991, unpublished), selecting sites which cluster within an error of 2% of the mitochondrial genome length (ie. about 320 bp) to be in homologous sites. The Hennig 86 computer program was then used to construct a cladogram by maximum parsimony procedure from the table of phylogenetically informative sites.

3.2.2.2 Distance Methods

The pairwise proportion of shared sites was calculated using the RESOLVE program, assuming, as above, that "homologous" sites aligning within 2% of the overall map length, were in fact homologous. The program also calculates a corresponding table of sequence divergence values using Nei and Li's formulae (1979). The PHYLIP (phylogeny inference) package designed by Joseph Felsenstein (1981) was used to produce distance dendograms for employing both FITCH and KITSCH options of the software package (Fitch and Margoliash, 1967). The KITSCH algorithm assumes that the rate of evolution is constant ie. DNA behaves like a molecular clock, while FITCH does not require this assumption. The Neighbour-joining method (NJ TREE) (Saitou and Nei, 1987) was also used as an illustrative distance based approach. It has been claimed that this can be as or more efficient than the maximum parsimony for producing the correct tree topology for restriction site data (Jin and Nei, 1991).

3.3. RESULTS

3.3.1 Mitochondrial DNA Extraction and Purification

Heart muscle was found to be the tissue of choice since it provided good mt DNA yields even if there is a delay in several hours after death before the tissue can be frozen or processed. A higher yield of mt DNA was extracted from fresh (unfrozen) tissue compared to that of frozen tissue (see Fig 3.2 for comparison). Figure 3.2 illustrates the comparison of the mt DNA yields obtained using fresh (unfrozen) and frozen tissues. Lanes 1 and 3 show higher mt DNA yields which were extracted from fresh leopard heart and digested with BamHI and EcoRI, respectively. Lane 2 represent relatively low mt DNA extracted from frozen cheetah heart and cleaved with BamHI.

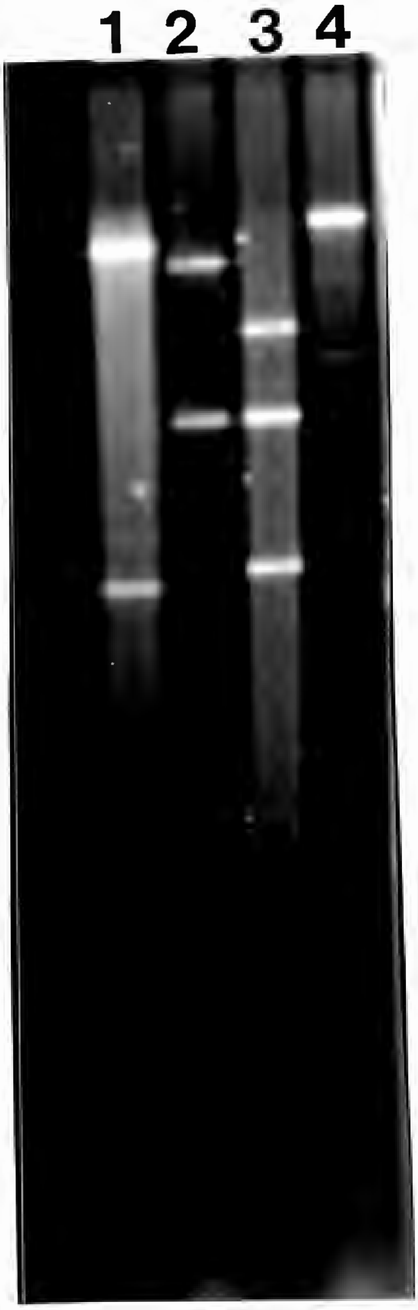
3.3.2 Autoradiographic Visualization

End-labelling of restriction fragments with ^{32}P-dCTP was used in most of our experiments since it is not as laborious and as time consuming as Southern blotting hybridization. DNA from black-footed cat was labeled using the latter procedure since it was not possible to obtain adequate quantities of purified mt DNA from cell cultures, which was the source of material in the case of other species. Figures 3.3 and 3.4 show typical single and double digests of end-labeled mt DNA restriction fragments and Southern blots of total genomic DNA from the caracal and the black-footed cat, respectively. Domestic cat mt DNA was used as a probe.

3.3.3 Restriction Maps

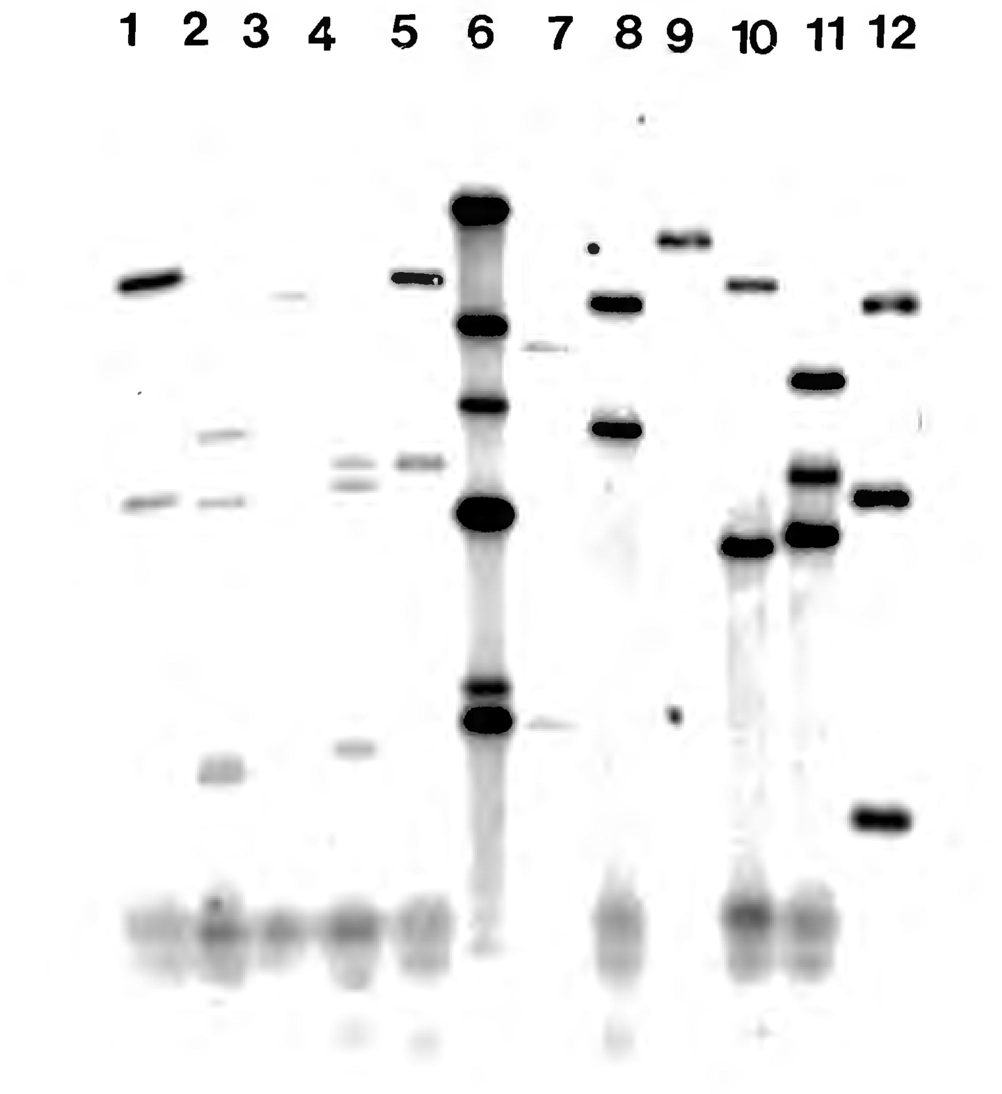
Table 3.2 gives the results of size determinations for eight *Felidae* species. The mitochondrial DNA genome did not differ significantly in size among the species compared and none of the values differs significantly from 16 400 base pairs, a value which is well documented to be an average size of mammalian mitochondrial genome (Lewin, 1983).

Fig. 3.2 **Autoradiography of cheetah and leopard mt DNA**



Lane 1: Leopard mt DNA (extracted from fresh tissue) digested with BamHI
Lane 2: Cheetah mt DNA (extracted from frozen tissue) digested with BamHI
Lane 3: Leopard mt DNA (extracted from fresh tissue) digested with EcoRI
Lane 4: Cheetah mt DNA (extracted from fresh tissue) digested with EcoRI

Fig. 3.3 **Autoradiography of single and double digested caracal mitochondrial DNA**



Lane 1 : EcoRI
Lane 2 : EcoRI and SacI
Lane 3 : SacI
Lane 4 : SacI and PstI
Lane 5 : PstI
Lane 6 : Lambda DNA digested with Hind III
Lane 7 : SacI and BamHI
Lane 8 : BamHI
Lane 9 : SalI
Lane 10: SalI and EcoRI
Lane 11: SalI and PstI
Lane 12: SalI and BamHI
Lane 13: SacII
Lane 14: SacII and EcoRI
Lane 15: SacII and PstI

1. Hind III; 2. Hind III + Pvu II; 3. Pvu II; 4. Hind III + Sal I; 5. Sal I; 6. Hind III; 7. EcoRI; 8. EcoRV; 9. EcoRV + PvuII; 10. EcoRV + Sal I; 11. DNA digested with Hind III; 12. Bam HI; 13. Bam HI + Pvu II; 14. Bam HI + Sal I

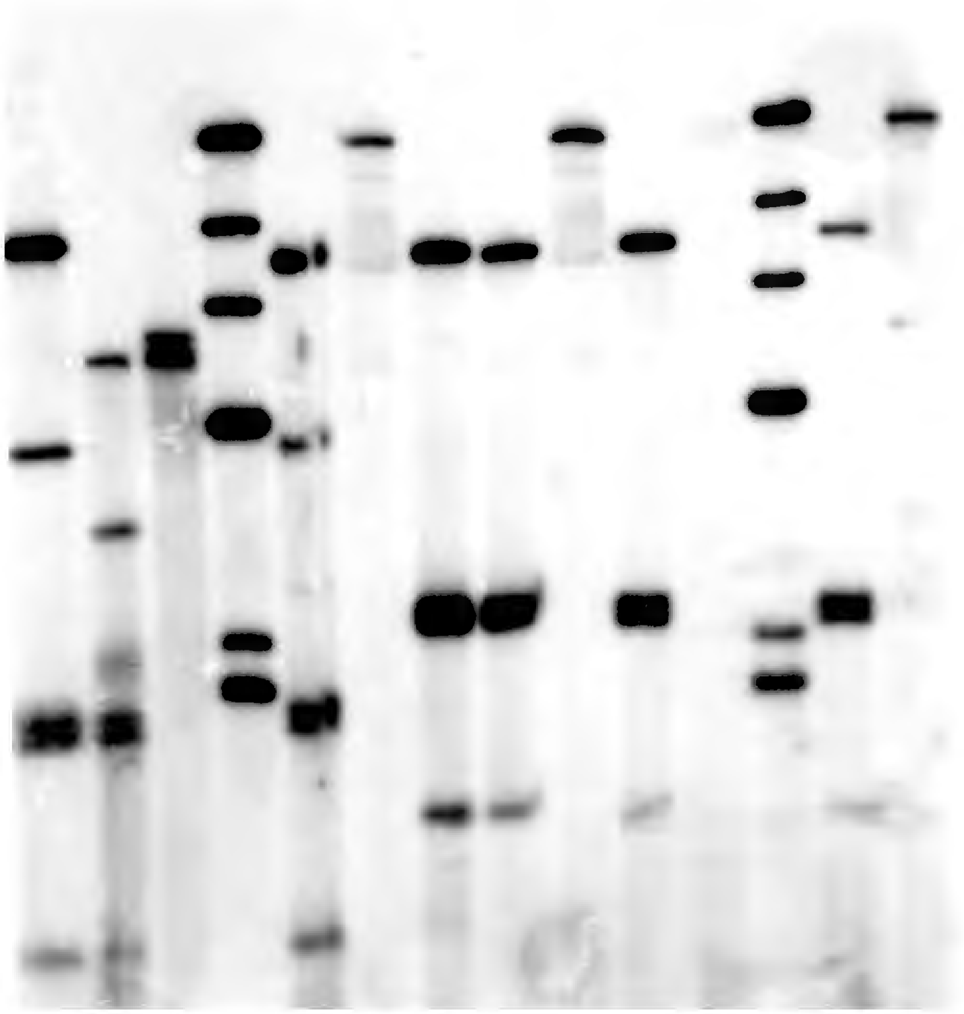


Fig. 3.4 Autoradiography of the hybridization of random primed domestic cat to black-footed cat total DNA extraction

Table 3.2 **Average genome sizes of 8 southern African Felids. Averages were determined from 5 values which were randomly selected from accurately sized sets of single or double digested DNA.**

Species	Genome sizes, (bp)
F. catus	16 560 ± 470
F. lybica	16 560 ± 470
F. sylvestris	16 500 ± 230
F. nigripes	16 500 ± 220
C. caracal	16 500 ± 270
P. leo	16 200 ± 480
P. pardus	16 520 ± 220
A. jubatus	16 349 ± 430
Average genome size	16 430 ± 350

Genome sizes are given ± 2 S.D.

Restriction maps were constructed independently for all the eight species, using a combination of various restriction endonucleases in pairs (Fig. 3.3). Restriction endonucleases yielding simple patterns (eg. those producing 2-4 fragments) were mapped first since such double digests produced a manageable number of fragments (see figs. 3.3 and 3.4). These data were first mapped manually and then mapped using the three way analysis option of the RESOLVE restriction mapping and management program (see Appendix II) (Harley, unpublished). Complex restriction patterns (6-10 restriction fragments), were mapped manually, using a number of single or double endonuclease cutters. Such data were then edited into the final map file and stored by the program.

The number of restriction sites mapped varied from a minimum of 36 sites for the cheetah to a maximum of 46 sites for the black-footed cat (see fig. 3.5). The maps in Fig. 3.5 show complete identity between the African wild cat and the domestic cat but there are four restriction site differences between these two species and the European wild cat. To check for accuracy of alignment of restriction sites, the site alignment option of RESOLVE was used to align individual restriction restriction enzymes for all eight species (Fig. 3.6). This option also gives an indication of the relative frequencies of autapomorphic sites as well as phylogenetically informative sites shared by several taxa. Very few shared restriction sites were observed amongst the larger cats (lion, leopard, cheetah). Xho I did not cleave the cheetah mt DNA.

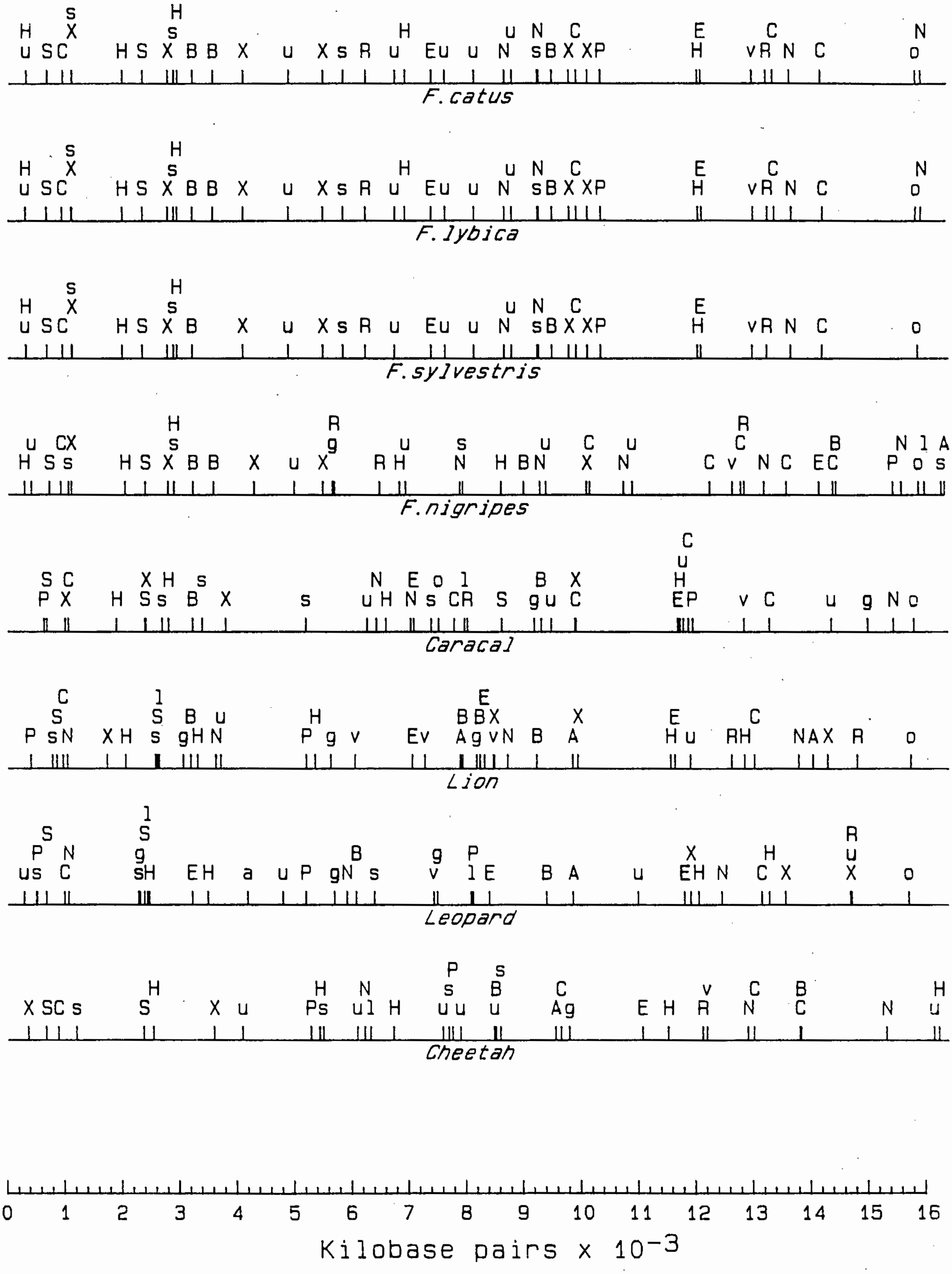


Fig 3.5 Restriction endonuclease cleavage maps of mt DNA of Lion, Leopard, Cheetah, Caracal, black-footed cat, African wild cat, European wild cat and domestic cat aligned on the invariant SacII site at position 676.

The complete names of restriction endonucleases are listed under abbreviations.

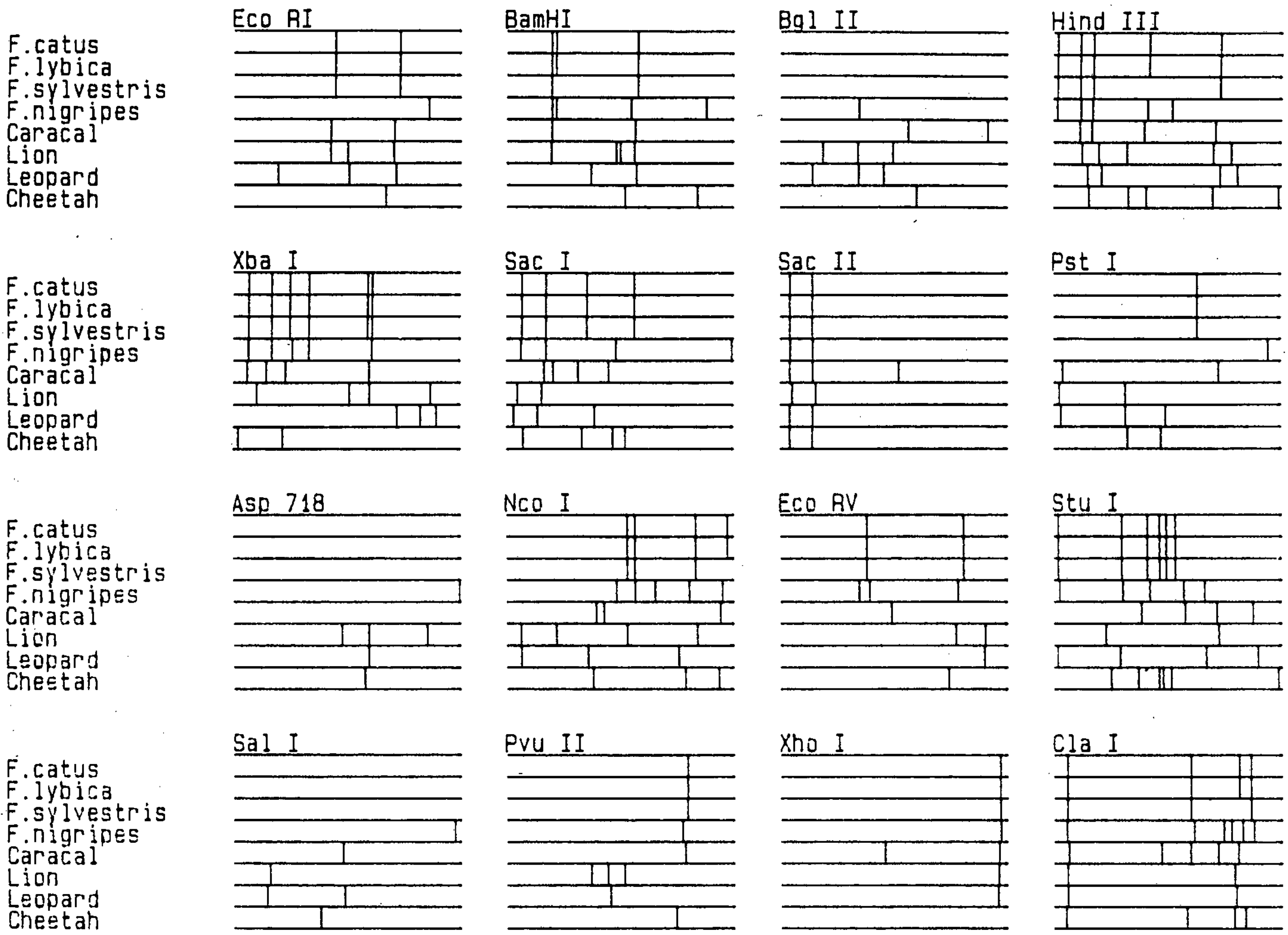


Fig. 3.6 **Restriction site alignment of individual restriction endonucleases.**

Horizontal lines represent mt DNA for each species in the linearized form, while vertical lines represent restriction sites.

3.3.4 Phylogenetic Analyses

3.3.4.1 Distance Methods

The proportion of shared restriction sites and hence sequence divergence was determined by using the formulas of Nei and Li (1979) (see introduction 1.4.1). These are built into the computer program RESOLVE and used to construct a distance matrix for all the species in question. Table 3.3 shows the proportion of shared restriction sites (above the diagonal) and the sequence divergence values (below the diagonal). Sequence divergence values ranged from 0 between the African wild cat and the domestic cat to 28.3 between the leopard and the black-footed cat.

Three algorithms, namely: FITCH, KITSCH and Neighbor-joining, the first two of which are available in the PHYLIP package were used to construct dendograms (distance trees) (see fig. 3.7). The topologies of the trees generated by FITCH and Neighbor-joining algorithms are essentially the same, the only difference being with the placement of caracal (fig. 3.7 B and C). KITSCH places the caracal as a sister taxon to the other members of the genus *Felis*, whereas neighbor-joining groups it with *Panthera* (larger cats). It should be noted that in the Neighbor-joining tree the longer branch length for the leopard relative to the lion does not imply that the former evolved faster than the latter, but is simply the result of stochastic variation. The same applies to the variations in the branch lengths between the caracal and the cheetah.

3.3.4.2 Cladistic Methods

Restriction endonuclease maps were analysed for phylogenetically

Table 3.3 **Distance matrix for eight Southern African Felidae**

	1	2	3	4	5	6	7	8
F.catus	**41**	1.000	.949	.598	.444	.286	.253	.23
F.lybica	.0	**41**	.949	.598	.444	.286	.253	.23
F.sylvestris	.8	.8	**37**	.554	.442	.300	.240	.21
F.nigripes	8.6	8.6	9.9	**46**	.395	.292	.190	.24
Caracal	13.6	13.6	13.7	15.6	**40**	.361	.231	.36
Lion	21.2	21.2	20.4	20.8	17.2	**43**	.469	.22
Leopard	23.3	23.3	24.2	28.3	24.9	12.7	**39**	.21
Cheetah	24.7	24.7	25.8	24.0	16.8	25.1	26.1	**36**

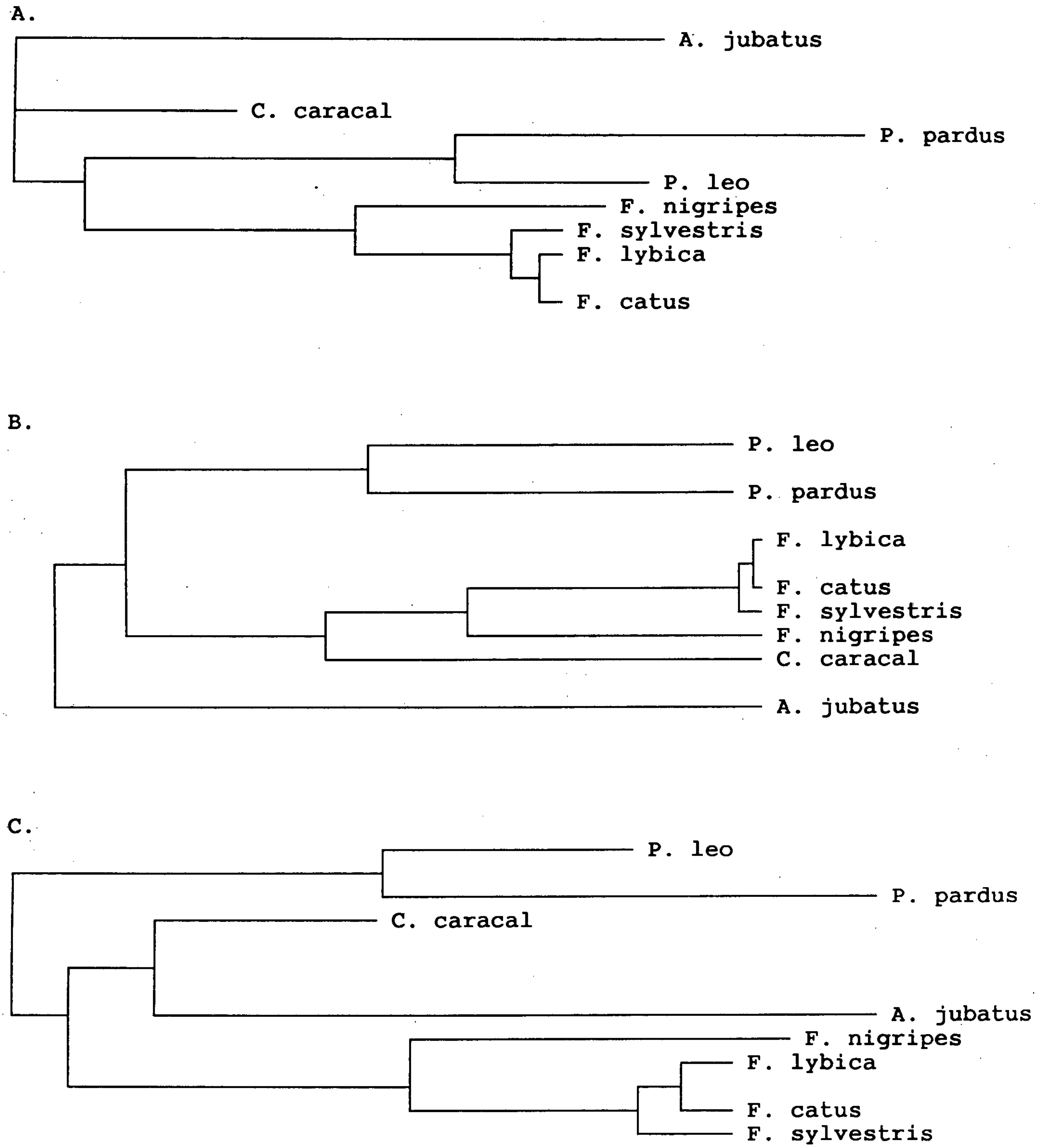


Fig 3.7 Distance dendograms: A. FITCH tree, B. KITSCH tree, C. Neighbor-joining (NJ) tree

informative sites using the procedure built into the RESOLVE computer program. These were set to an alignment of within 2% of the total map length as a criterion for identity of a site. Table 3.4 documents the phylogenetically informative site positions and their character states. The numbers, "1" indicate the presence of a site and "0" indicates its absence. A total number of 71 phylogenetically informative sites was found.

A single most parsimonious tree was then constructed using the HENNIG 86 computer package (Fig. 3.8.A). Cheetah was used as an outgroup since previous studies based on traditional morphological criteria have suggested its distant relation to the other Felids (Ewer, 1972; Hemmer, 1978; Leyhausen, 1979; Nowak and Paradiso, 1983). On the other hand studies on albumin immunological distance studies (Collier and O'Brien, 1985) seem to suggest that the cheetah did not diverge early on in the *Felid* history but may constitute a monophyletic group along with *Felis serval, jaguarundi*, golden cats and the puma. Hennig 86 produced a tree with 110 steps which is approximately 39 steps longer than the minimum length, which would be equal to the number of phylogenetically informative sites. Placing the caracal as a sister group to the larger cats (*Panthera*), using the xx option of Hennig 86, resulted in the generation of a cladogram which is three steps longer whereas placing the caracal with the smaller cats resulted in the generation of a cladogram which is only one step longer suggesting the close relationship of caracal to the smaller cats. The topology of the single most parsimonious tree obtained is consistent with the neighbor-joining tree, but only differs with the KITSCH tree with respect to the position of the caracal. To assess the robustness of

Table 3.4 **Table of phylogenetically informative sites and informative character states.**

'1' indicates the presence of a restriction site and '0' indicates the absence of a restriction site.

Informative site positions and character states

```
Site number                  1111111111222222222233333333334
                    1234567890123456789012345678901234567890
Enzyme code         uuuuuuuuuuXXXXXXXBBBBBssssssssEEEEECCCCC
                            11      1                11 1111
                     466788901123459033899 12255797781293334
Position            3917618498077158125324614838520437190372
                    1087350454890104128778312955568061016300
                    6080400409626008529882969207000755353320
                    ----------------------------------------
F.catus             1101111000110111111001010101010100110101
F.lybica            1101111000110111111001010101010100110101
F.sylvestris        1101111000110111110001010101010100110001
F.nigripes          1100000110110010111000010100000000000010
Caracal             0010000101101001010010000010101001010100
Lion                0000000001000001010110101000001011001000
Leopard             1100000010000000000001101000000011001000
Cheetah             0010100000001000000100010010100000001010

Site number         4444444444555555555566666666667
                    1234567890123456789012345678901
Enzyme code         HHHHHHHHHHNNNNNNNNRRRRvvPPPllgA
                            11    1111 111 1  1
                     22235661216893355623472 502859
Position            2059446950006206492727395235068
                    9001007155479759655355650353366
                    6220423601523781000509597132690
                    -------------------------------
F.catus             1101000101001101011010010010000
F.lybica            1101000101001101011010010010000
F.sylvestris        1101000001001101001010010010000
F.nigripes          1101000100000110100100000000010
Caracal             0101001000000000100000011000100
Lion                0100110010101001000101101101011
Leopard             0010100001110000000001101101111
Cheetah             0010011010010010100000000100000
```

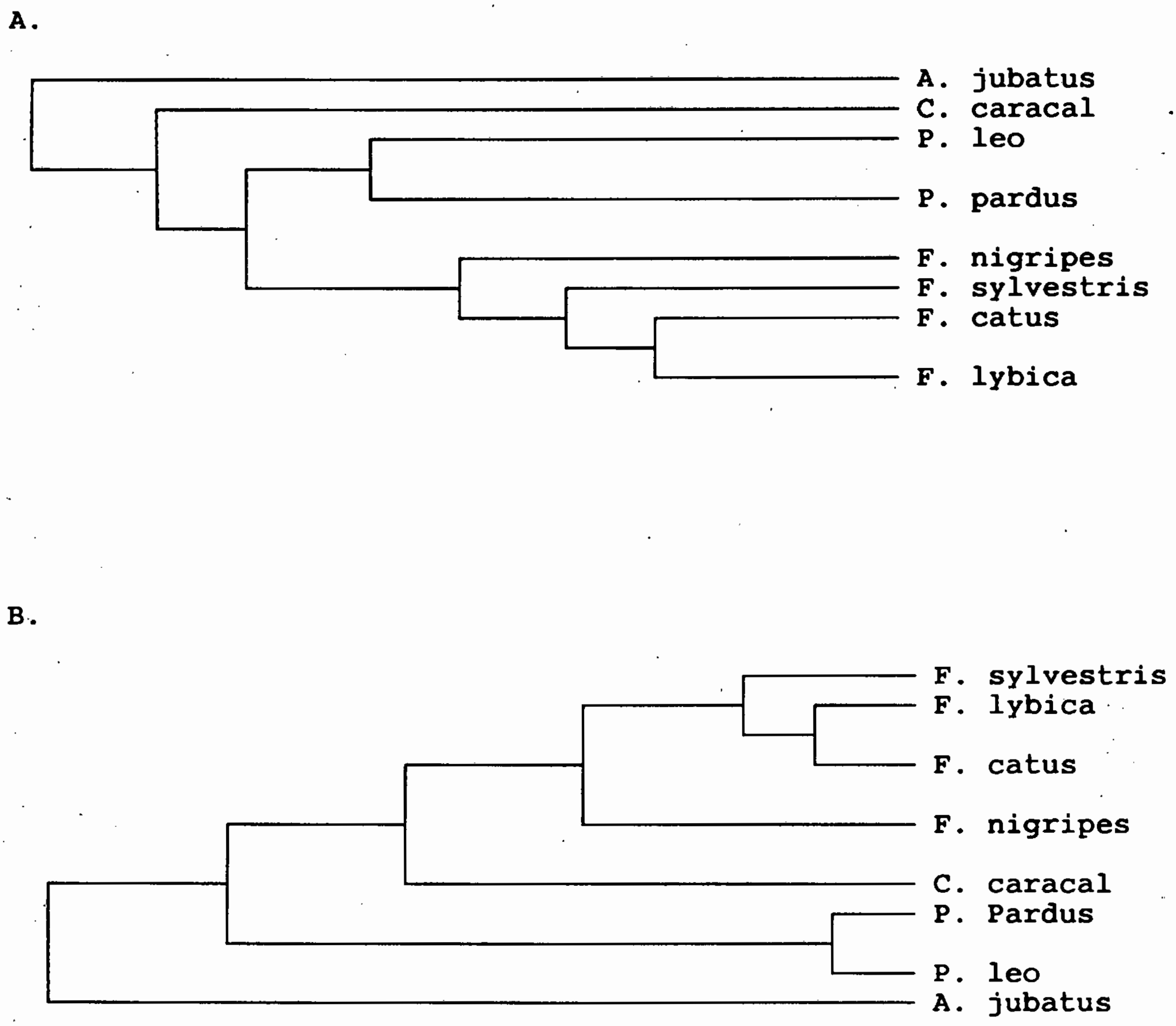


Fig. 3.8 Cladistic analyses using the cheetah as an outgroup: A. maximum parsimony tree, B. Bootstrap analysis using 1000 replicates

the maximum parsimony tree, bootstrap analysis using 1000 replicates was performed (fig. 3.8.B). Its majority consensus rule tree gave similar topology as the maximum parsimony tree except for the placement of caracal. In contrast with KITSCH and Neighbor-joining trees, such a tree resulted on a relatively low bootstrap value of 508 out of a total of 1000 replicates when the caracal is placed between the small cats (*Felis*) and the large cats (*Panthera*). This implies that the caracal is weakly grouped with either of the two genera and would strongly suggest a separate generic status of the caracal. Such a tree produced 1000 out of a total of 1000 replicates for the placement of the European wild cat with both the African wild cat and the domestic cat suggesting the monophyletic relationship between the latter three species. Similarly when the black-footed cat is grouped with the small cats a value of 935 out of a total of 1000 replicates is obtained implying a close relationship between the latter species and the black-footed cat.

3.4 DISCUSSION

3.4.1 Restriction Mapping

Mt DNA restriction maps for the African wild cat and domestic cat show complete identity (see fig.3.5). These differ with the European wild cat by four restriction sites. No polymorphic sites were found between individuals within any of the three populations.

Our analyses were based on a total of four wild cats showing the characteristic morphological features of the African wild cat, and coming from areas remote from human habitation. All four showed identical restriction maps, which would then appear to render very unlikely the possibility that our African wild cat mt DNA preparations all derive from domestic cat via interbreeding with feral cats. The identity between the domestic cat and the African wild cat supports a very recent origin for the former from the latter, probably within the last few thousand years.

Analysis of mummified cat-skulls, as well as cat forms and patterns on the walls of monuments and tombs of Ancient Egypt, have supported the idea that the African wild cat was first domesticated in Egypt (Smithers, 1983; Armitage and Clutton-Brock, 1981). Once domesticated it probably spread from Egypt to Europe via the military and commercial exploits of the Romans. From Europe, it would then have returned to various parts of Africa with the arrival of the European settlers in the seventeenth century (Smithers, 1983). The morphological differences between the domestic cat and the African wild cat presumably reflect man's propensity for selective breeding for desirable characters. No inbreeding depression is likely to follow interbreeding between the two species. However, since they are likely to show no assortative mating i.e. preferences against interbreeding, the chances of the wild cat phenotype being lost by dilution into the domestic cat is high. With increasing encroachment of human settlements (and therefore of domestic cats) into African wild cat habitat, an extensive conservation effort would be required to avoid loss of this phenotype. Justification for such an effort would need to be based on the desirability of retaining aesthetically desirable features of the African wild cat phenotype, since the only genetic justification, in face of evidence for such recent divergence between the African wild cat and the domestic cat, would be the retention of the few morphological, and possibly adaptive, characteristics which separate the African wild cat from the domestic cats. The sequence divergence value of 0.9% (see table 3.3) separating the African from the European wild cat mt DNA is several times greater than typical values for intra-species sequence divergence in mammals (Taylor *et al.*, 1986; Cann *et al.*, 1987; Brown, 1980; Ashley and Willis, 1987). This is be consistent with their current classification as different species.

Assuming a calibration of 1% per million years for the rate of change of mammalian mt DNA, (Brown and Wilson, 1979) and assuming no significant differences in mutation rate for feline mt DNA, then 0.9% sequence divergence points to the existence of a common ancestral mt DNA for the two wild cats about 450 000 years ago. This value is of the same order of magnitude as the divergence time of 2 million years inferred from albumin immunological data (Collier and O'Brien, 1985) but is not be consistent with an estimate from allozyme data of only 20 000 years (Randi and Ragni, 1991).

It should, however, be borne in mind that although restriction mapping eliminates problems of convergent fragment lengths (which are encountered when restriction fragment size comparison is employed), the sites themselves might be convergent. Another problem in using restriction mapping is the asymmetry in the probabilities of gaining and losing sites where the probability of convergent site losses is far greater than that of convergent site gains. A site loss may be caused by a point mutation within a cleavage site, whereas a site gain requires a specific base substitution at a specific base pair (Templeton, 1983; Li, 1986). These limitations should therefore be considered when using restriction enzyme sites.

3.4.2 Phylogenetic Analysis

Both cladistic and distance measurements yielded identical trees in accordance with morphological, ethological, physiological and molecular data (Walker *et al.*, 1964; Ewer, 1973; Hemmer, 1978; Leyhausen, 1979; Nowak and Paradiso, 1983; Collier and O'Brien, 1985). The smaller cats, namely, African wild cat, domestic cat, European wild cat and black-footed cat are grouped together and so are with the larger cats namely, lion, leopard and cheetah. The most significant difference is the placement of the caracal. FITCH and Neighbor-joining and the maximum parsimony trees group the caracal with the larger cats while KITSCH places it with the small cats. This is not surprizing since the latter has always proved difficult to classify with some taxonomists (Ewer, 1973; Hemmer, 1978; Kratochvil, 1982; Werdelin, 1983) putting it into a genus of its own, while others (Walker *et al.*, 1964; Nowak and Paradiso, 1983) place it under *genus Felis* (smaller cats). It should be noted that although Nowak and Paradiso grouped the caracal with the smaller cats, they placed it into a separate sub-genus Caracal. Leyhausen (1979) on the other hand, classified the caracal with the larger cats : this is consistent with our neighbor-joining dendogram and the maximum parsimony tree using the cheetah as an outgroup. These trees imply that the caracal and the cheetah are monophyletic. The close relationship between the latter two species was also inferred from recent studies on protein albumin immunological distances (Collier and O'Brien, 1985) and on isozyme genetic distances (O'Brien *et al.*, 1987). KITSCH's dendogram supports Nowak and Paradiso's view that the caracal is more closely related to the smaller cats than to the larger cats. Due to the inconsistency in these results, the precise placement of the caracal has not been resolved but its deep rooting would be more consistent with a separate generic status, rather than its inclusion within either the *Felis* or *Panthera genera*. Our conclusion is supported by the relatively low bootstrap value of 508/1000 when the latter species is placed between the small cats (*Felis*) and the larger cats (*Panthera*).

The close relationship amongst the larger cats inferred from the present study is supported by recent studies on both protein albumin immunological distances (Collier and O'Brien, 1985) and isozyme genetic distances (O'Brien *et al.*, 1987). Our data show a monophyletic relationship between the lion and the leopard (Fig. 3.7), supporting the previous findings implied by morphological and ethological results (Walker *et al.*, 1964; Ewer, 1973; Nowak and Paradiso, 1983). This is further supported by relatively high bootstrap value of 945/1000 replicates.

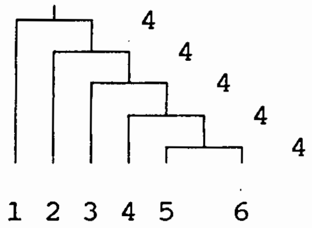
The genus *Felis* (smaller cats) inferred from our data corresponds with the five-species that Ewer (1973) grouped together in the genus *Felis*. This clustering is further supported by the results of Benveniste and Todaro (1974) and Benveniste *et al.* (1975), who detected endogenous DNA sequences related to RD114 and FELV retroviral genomes in the cellular DNA of six of the seven species of the small cat lineage (with the exception of the pallas cat), but not in any other *Felidae*. These gene families apparently originated from a cross-species infection of an ancestor of the domestic cat at a time subsequent to the earlier divergence of the *Felidae*, but prior to the radiation of the modern small cats in this lineage.

Independent confirmatory evidence of the domestic cat lineage can be inferred from the results of Wuster-Hill and Centrewall (1982), who found complete identity between the karayotypes of all the small cats (*Felis*). These karyotypes were also shown to be distinct from all other Felid species. The homogenous karyotypes of the small cat lineage then lend support to the hypothesis of the monophyletic relationship of the small lineage reported in the present study. The close relationship between these smaller cats, African wild cat, domestic cat, European wild cat and black-footed cat, was also suggested by Collier and O'Brien (1985). Taken together, these data support the monophyleticity of the small cat lineage.

It should, however, not be claimed that either of these phylogenies is the best since both distance and cladistic approaches are not without limitations. Cladistic approaches are prone to *homoplasies* due to convergence or back mutations. These *homoplasies* tend to bring divergence sequences together as sister groups in an unrooted tree. This phenomenon is well illustrated in the SEVONA computer simulation (Fig 3.9) in which *homoplasies* occur more frequently. The presence of *homoplasies* also to complicates molecular data when a high transition frequency is used (Fig 3.9), typically 90:10 transition transversion ratio as with mammalian mt DNA in which more than 90% of mutations are transitions (Brown *et al.*, 1982). The example in figure 3.9 has 16 phylogenetically informative sites (marked "X") of which no less than 9 are *homoplasies* (marked *). The lower casing indicates that a mutation has occurred. Distance methods, on the other hand, exhibit stochastic variation which can be minimized by increasing the size of the data set eg. 1000 sequenced base pairs show much less stochastic variation than 100. Distance measures are most useful for biogeographical arguments but are usually considered to be inherently weaker for determining the topology than cladistic methods.

Despite all these limitations, for molecular data, given a molecular clock with equal mutation rates and provided the stochastic effects are not great, distance measurements are almost as effective as cladistic approaches for constructing the true tree. Since both methods have advantages and disadvantages, the best approach when conducting a study would seem to be the use of cladistic methods for defining the topology and distance methods for determining the branch lengths and hence the time in which a pair of species diverged from the common ancestor.

Tree topology & Branch lengths



```
Ancestral sequence :
   CGACCATCCCGGAACCCTTAGTACTGAAAAAGGTGGACGGTGACTGAGCGTCTGATGGCCGTGCGAGTGCGTGC

Extant sequences :
 1 CaACCAcCgCGGAAtCCTcAGTACTGAAAAAGGTGGAtGGTGACTGAGtGTCTGATaGtCGTaCGAGgGCGTGC
 2 CGACCAcCCCaGAACCtTTgaTgCTGAAAAAGGTGGAtGaTGACTGAGtGTtTGATaGcCGcGCGAaTGCGTGC
 3 CGACCATCCCGGggCtCTTAGTACTGAAAAAGGTGGACGGTGACTGAaCGcCTaAcGGCtGcGCGAGTGCGTGC
 4 CaACCgTCCCGGAACCCTTAGcACTGAAAAtcGTGGACaGTGACTGAGCGcCTGATGGCaaTaCGAGTGCGTGt
 5 CaAtCAgCCCGGAACgCTTAGTACTGAAAAAcGTGagCGGTGACTGAaCGcCTGATaGCCGTGCGAGTaCGTGt
 6 CaAtCAgtCCGGAACgCTTAaTACTGAAAAAcGTGGgCGGTGACTGAGCGcCTGATaGCCGTGCGAGTaCGTGt
    X X  X        X    X          X    XX         XX X     X    XX     X    X
    *    *             *                *         **       *    **
```

Fig 3.9 **Simulation by the SEVONA program (Harley, unpublished) of an ancestral sequence (75 base pairs) evolving in a pectinate fashion to 6 progeny species. Transition frequency =90%. "X" represents phylogenetically informative sites. Lower casing indicates that a mutation has occur.**

CHAPTER 4 : PCR AND MITOCHONDRIAL DNA SEQUENCING

4.1 INTRODUCTION

In order to complement mt DNA restriction mapping, we have attempted to develop the methodology for sequencing part of the cytochrome b region of mt DNA with the aid of the polymerase chain reaction. Mt DNA sequencing can provide very large amounts of data and therefore has a potential to add to the precision of phylogenetic analysis. The cytochrome b region was selected since it has a rate of accumulation of base changes which is intermediate between that of the D-loop region and that of the more conserved ribosomal RNA region. As a start, a region of approximately 140 base pairs of cytochrome b from position 15521 to position 15682 was sequenced, from which a phylogeny of eight members of the southern African *Felidae* was constructed. It must be appreciated that this is a very short region for adequate phylogenetic analysis, and more will be required, but will serve as a beginning and suffice to demonstrate the principles of the approach.

The oligonucleotide primers flanking the region of target DNA were selected based on a combination of Paäbo (1989) and Kocher *et al.*'s previous amplification experiments (1989) and later by our own comparative studies to select short conserved regions appropriate for primer synthesis by comparing the cytochrome b sequence of man, Xenopus, bovine, and chicken. (See fig 4.1 for the comparison of sequences for the selection of oligonucleotide primers). The bold printed regions illustrate the positions for the primer selection. The size of the primers was 17 base pairs for primer 1 and 20 base pairs for primer 2. The primer sequences were:

i) CAGAACATTTGGTCTCT
ii) TGACCAACAGGAGGTTAAGT

```
HCB    TTCTCTCCTT  AATGACATTA  ACACTATTCT  CACCAGACCT  CCTAGGCGAC  CCAGACAATT
BCB    ....AG.TC.  ...ACT.C..  GT.......G  ....C.....  ...C..A...  .....T..C.
MCB    .CT.AATTC.  C..A..CC..  GT.T....T.  TC......A.  A.....A...  ........C.
GCB    .CAC.C.A..  CC.A...C..  G.C.......  .C..CA....  ......A...  .....T..C.
XCB    ..ACAG.AC.  T.CTCTCC..  G.CA....T.  .C...A....  TT....A...  ..........
                                                                         840
HCB    ATACCCTAGC  CAACCCCTTA  AACACCCCTC  CCCACATCAA  GCCCGAATGA  TATTTCCTAT
BCB    .C....C...  ...T..AC.C  .....A..C.  .T........  A.....G...  ..C...T...
MCB    .C.TA.C...  T..T..AC..  ........A.  ....T..T..  A.........  ..........
GCB    TC....C...  A.....AC..  GTA.....C.  .A..T.....  A..A......  .....T....
XCB    T.....C...  T..T..TC..  .T........  .A..T..T..  A..A......  ..C.......
                                                                         900
HCB    TCGCCTACAC  AATTCTCCGA  TCCGTCCCTA  ACAAACTAGG  AGGCGTCCTT  GCCCTATTAC
BCB    .T..A...G.  ...CT.A...  ..AA....C.  ..........  ...A..A..A  ......GCCT
MCB    .T..A...G.  C.....A..C  ..AA....C.  .T........  ...T.....A  ...T..A.CT
GCB    .......TG.  C..C..A..C  ...A....C.  .......T..  ...T..A..A  ......GC.G
XCB    ........G.  T..C..T...  ...A.---A.  ..........  C..A..GT.A  ......G.C.
                                                                         960
HCB    TATCCATCCT  CATCCTAGCA  ATAATCCCCA  TCCTCCATAT  ATCCAAACAA  CAAAGCATAA
BCB    .C..T.....  A..T..T..T  C........C  .A..A..C.C  C.........  .G........
MCB    ....T.....  A..TT....C  C....A..TT  ....T....C  C..A..G...  .G....C...
GCB    CC..AG....  ......CTTC  C........T  .......C.A  ...T......  .G..CA....
XCB    ..........  A...T....C  C.C..A..AT  .A.....C.C  ...A......  .G....C.T.
                                                                        1020
HCB    TATTTCGCCC  ACTAAGCCAA  TCACTTTATT  GACTCCTAGC  CGCAGACCTC  CTCATTCTAA
BCB    ....C..A..  ...C......  .GC..A.TC.  ..GC.....T  A........A  ..G.CA..C.
MCB    ....C.....  .A.C.CA...  ATTT.G..C.  ..A......T  A..CA....A  ..T..CT...
GCB    CC..C..A..  ...CTC....  A.C..A.TC.  ....T....T  A..CA....T  ..T..C....
XCB    ....C..A..  .T.T.CA...  ATCA.A.T..  ..GC.....T  T.....TACA  ..A..C....
                                                                        1080
HCB    CCTGAATCGG  AGGACAACCA  GTAAGCTACC  CTTTTACCAT  CATTGGACAA  GTAGCATCCG
BCB    .A.....T..  ..........  ..CGAAC...  .A.A..T..C  ...C......  C.......T.
MCB    .......T..  G..C......  ...GAAC...  .A....TT..  ......C...  C....C...A
GCB    ..........  .A.C......  ...GAAC...  .C..C.T...  ......C...  A........C
XCB    .......T..  ...T......  ...GAAG...  ...A......  A........G  T....C..A.
                                                                        1140
HCB    ATGAAAACCT TTTTCCAAGG
BCB    .A
MCB    .....G.T.. .C.CTT
GCB    C....G..TC CACC.TTCTT
XCB    .G.CT..AAC CC.C.TC.A. ACT
% Fully aligned sites matching reference sequence = 49.2
```

Fig. 4.1 **Sequences of part of cytochrome b region of mt DNA for human (HCB), Bovine (BCB), mouse (MCB), chicken (GCB) and Xenopus (XCB)**

The percentage G-C was estimated to be 50% using the formula:

$$Tm = 77.3°C + 0.41 (\%G + \%C) - 0.61$$

This is important for the estimation of the annealing temperature as it is dependent on the nucleotide sequence (G-C content) of the primer. For high G-C content, the annealing temperature should be correspondingly higher, due to the presence of 3 hydrogen bonds between G and C. As there are only 2 hydrogen bonds between A and T, annealing temperature should be lower for primers with a lower G-C content. The annealing temperature can be as high as 55°C and as low as 37°C (Maniatis, 1988). Practically, a lower annealing temperature than is appropriate is revealed by the presence of primer-dimers or other bands caused by priming at inappropriate sites.

4.2 MATERIALS AND METHODS

4.2.1 Polymerase Chain Reaction

The polymerase chain reaction was performed either directly from purified mt DNA eg. in the case of the leopard, lion, cheetah, African wild cat, domestic cat and the European wild cat or from total genomic DNA preparations eg. in the case of the black-footed cat and the caracal. PCR experiments were performed as described in Current Protocols in Molecular Biology (1988) and according to the manufacturer's instructions (Cetus, 1990).

PCR was performed in 100 µl reactions containing 200 uM each of dATP, dGTP, dCTP, dTTP, 10 x reaction buffer [100 mM Tris-Hcl; 500 mM KCl; 15 mM MgCl; 0,01% (w / v) gelatin] ; 30 pmol of each oligonucleotide primer, 2 ng/µl DNA and 2 units of Taq polymerase source. This was briefly mixed and 60 µl of mineral oil was layered on top of the reaction cocktail to prevent evaporation. Prior to the addition of Taq DNA polymerase, the reaction mixture was heated at 94°C for 3-4 min to degrade proteases that might be present in the template preparation and to denature the DNA template. Amplification controls in which water was substituted for the target DNA were performed throughout. Such controls are imperative to ensure that the sequences being amplified are from the specimens being examined and not due to laboratory contamination.

The magnesium chloride concentration required varied depending on the quality of the target DNA and required initial titration experiments to find the optimal conditions. Amplification products were analysed by electrophoresis on 2% agarose gels. The gel was run at 60V for 3 hours. The gel was stained using 1 µg/ml of ethidium bromide to visualize the amplified products under ultra violet light (235 nm).

4.2.2 Generation of Single Stranded DNA for Sequencing

Sequencing of single stranded PCR products has proven to be generally easier and more convenient than double stranded products (Maniatis, 1988) since the double strands of DNA can bind to each other thus making it difficult for subsequent binding and extensions to occur. The traditional method of generating single stranded DNA is to clone the PCR product into M13 plasmid DNA. However, due to the fact that cloning is very time consuming, alternative methods are available. In this study, the two methods used are outlined below:

i) Asymmetric PCR

This technique is similar to the conventional PCR method of amplification except that one primer is added in excess with a typical ratio being 100:1 or 50:1. This works under the premise that the limiting primer will be depleted after approximately 25 cycles and is therefore not available to initiate the second strand synthesis by Taq polymerase.

ii) The second method used was to use one primer instead of both. This method is similar to the conventional PCR amplified product. The technique involved performing PCR experiment under the conditions described previously (i), subsequent to an initial amplification using both primers.

4.2.3 Purification of PCR Amplified DNA

After PCR experiments there is unwanted material associated with the product of interest, for example unincorporated primers, dNTPs and buffer components. Three methods of purification were used to remove these:

(i) ammonium acetate

(ii) millipore filter units

(iii) DEAE cellulose membrane.

4.2.3.1 Precipitation by Ammonium Acetate

After removal of mineral oil from the PCR product, 100 µl of product was mixed with 100 µl, 10 M ammonium acetate; 200 µl of 2-propanol was added, mixed briefly, and the sample was placed at room temperature for 10 min. This was then spun down for 10 min using a microcentrifuge. The supernatant was removed and the pellet washed carefully with 50 µl of 70% ethanol. The pellet was dried under vacuum was dissolved in 10 µl of TE buffer.

4.2.3.2 **Purification by Membrane Filtrations**

The PCR product was placed in a millipore chamber containing a membrane that retains DNA fragments of 200 bp or longer. The nucleotides and other substances contained in the PCR product mixture which are smaller than 200 bp would be eluted through the membrane when the unit was centrifuged at a low speed spin in a benchtop microfuge for 4 min. The DNA retained on the membrane was resuspended, after flushing twice with ultra pure water, in 20 µl of double distilled water.

4.2.3.3 **Purification by DEAE Cellulose**

This involved running PCR products on 2% agarose gel for 2 hours at 70 volts after which the gel was stained with 0.5 µg/ml of ethidium bromide. After 15 min the gel was visualized under ultra violet light. A slit was cut in the gel immediately ahead of the PCR band and a piece of DEAE cellulose membrane was inserted. Prior to the insertion, the DEAE cellulose membrane was washed with 10 mM EDTA and 0.5M NaOH for 5 min each and finally 6 times with double distilled water. The gel was then run at the same voltage until the DNA fragment had run onto the DEAE cellulose membrane. The membrane was removed and rinsed using a low salt buffer (50 mM Tris-Cl; pH 8.0; 0.15 M NaCl; 10 mM EDTA pH 8.0) followed by incubation at 65°C for 30 min. The buffer was transferred to a fresh tube to which a high salt buffer (50 mM Tris Cl pH 8.0; 1 M NaCl; 10 mM EDTA pH 8.0) was added. The reaction was followed by incubation at 65°C for further 15 min. The second aliquot was then mixed with the first one and the DEAE cellulose membrane was discarded.

4.2.4 **Mitochondrial DNA Sequencing**

Sequencing was performed subsequent to the purification of PCR products. Experiments were performed as described by the suppliers (see appendix I).

The Sanger-dideoxy method was used because the reactions are quicker and harmful chemicals are not required. In principle, the method relies on premature termination of DNA synthesis by incorporation of nucleotide analogues into the newly synthesized DNA chains. The basic requirement is for the single stranded template to contain a primer-binding site immediately upstream of the DNA to be sequenced. The basic method of Sanger-dideoxy chain termination is illustrated on the flow diagram in Fig. 4.2.

4.2.4.1 **Annealing Reaction**

The primer was annealed to the template by mixing 5 µl of template DNA with 1 µl 0.54 pmol of an oligonucleotide primer in a 1:1 ratio. For double stranded sequencing the primers used for sequencing were the same as for PCR. For single stranded sequencing, the primer used was different from that used for PCR. To the above reaction mixture, 2 µl of sequencing buffer (200 mM Tris-HCl pH 7.5; 10 mM mgCl; 250 mM NaCl) was added and the reaction made up to a final volume of 10 µl with double distilled water. The reaction mixture was heated to 95°C for 4 min followed by `snap cooling' in which the eppendorf tube was rapidly placed on dry ice containing 100% alcohol for 10 min, to prevent reannealing of the two complementary strands.

4.2.4.2 **Labelling Reaction**

The labelling reaction was prepared by mixing the following : 1 µl of 0,1M DTT; 2 µl of 1 : 15 dilution of labelling mixture (dNTPs); 1 µl of [α - 35 S] dATP ; 2 µl of diluted sequenase (5 U/µl) . For double stranded DNA, 0.85 µl of 0.5% 10x Triton-X100 and 0.85 µl of 0.5% NonidetP-40 were included in the reaction cocktail. This was incubated at 12°C for 4 min afterwhich the primer - template reaction mixture was added, followed by incubation at 37°C for 3 min. 3.5 µl of this labelling reaction was transferred to each of the termination reactions which consisted of 2.5 µl of each dideoxynucleotide (ddATP, ddCTP, ddGTP, ddTTP). For double stranded DNA sequencing, the termination reactions also contained of 0.25 µl of 0.5% Triton-X100 and 0.25 µl of 0.5% Nonidet-P40. Four microlitres of formamide was added to terminate the reaction. Prior to loading , the reactions were denatured by heating at 90°C for 4 min.

4.2.4.3 **Polyacrylamide Gel Electrophoresis**

Polyacrylamide gels are used to obtain good separation of DNA fragments less than 1 kb in size (Hillis and Moritz, 1990) and are therefore used to analyze sequencing reactions. Depending on the length of DNA sequence to be analyzed, the percentage of polyacrylamide in the gels varies: for short sequences, a lower

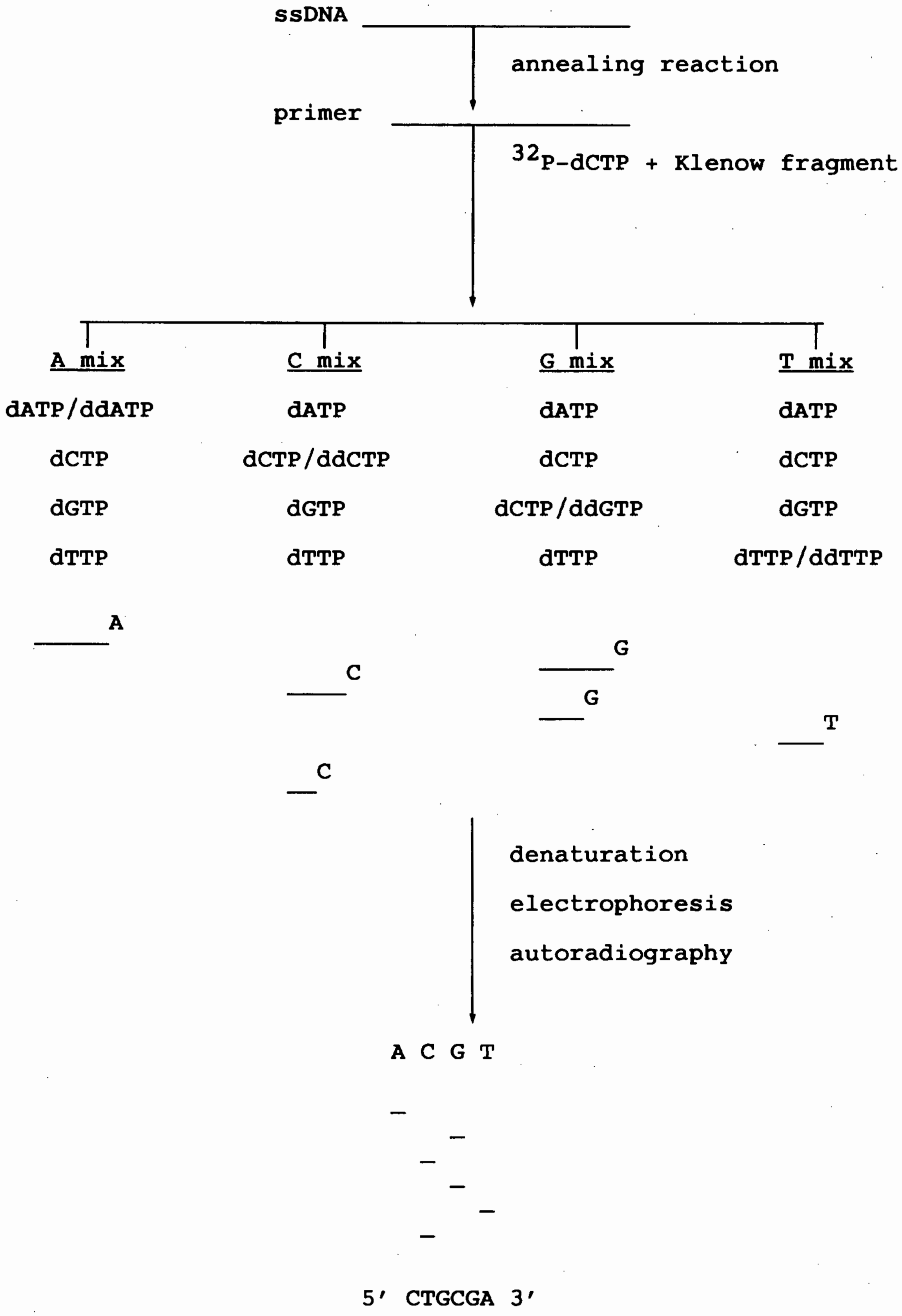


Fig. 4.2 Flow diagram of the sequencing reaction

percentage gel is used whereas for longer sequences a higher percentage gel is employed.

In this study, 6% polyacrylamide gels were used. This was prepared by dissolving 10X NNB buffer, double distilled water and urea in a microwave oven for 2 min. After this had cooled, 50% acrylamide 2.5% bis acrylamide solution was added, followed by filtration through 3 MM Whatman's filter paper. Urea was included in the solution as a denaturing agent to avoid reannealing of DNA strands.

350 µl (2.5%) of freshly prepared ammonium persulphate (polymerizing agent) and TEMED (catalyst) were added to fascilitate polymerization of acrylamide. The gel was poured between the glass plates with great care being taken to avoid the formation of air bubbles. The comb was inserted in between the glass plates into the top of the gel where the samples were to be loaded and the solution allowed to polymerize for an hour. The gel was allowed to stand at room temperature overnight. The comb was removed and the well at the top washed with 1X NNB buffer to remove unpolymerized acrylamide and urea. A sharkstooth comb was then placed into the slot, inorder to form wells into which the samples could be loaded.

4.2.4.4 Gel Preparation, Running and Drying

Two systems of sequencing gel electrophoresis, namely, the BRL Model SO and the Hoefer. The BRL model was used in these experiments. The glass plates were well washed and then thoroughly cleaned with 70% ethanol. Sigmacote was applied to one of the plates to prevent adhesion of the gel to the plate during separation.

Before loading the samples the gel was pre-run at 80 Watts with the power kept constant. The wells were washed with 1 X NNB buffer to remove urea which can prevent sharp band formation. Three microlitres of sequencing reaction were loaded and the gel run for two hours at 80 Watts. After electrophoresis the plates were gently separated with the gel sticking to one of them. The gel was fixed for an hour, in a solution containing 60 ml glycerol, 300 ml glacial acetic acid, and 300 ml methanol. The gel was dried for an hour at 80°C and autoradiographed. The exposure times varied from 3 days to a week.

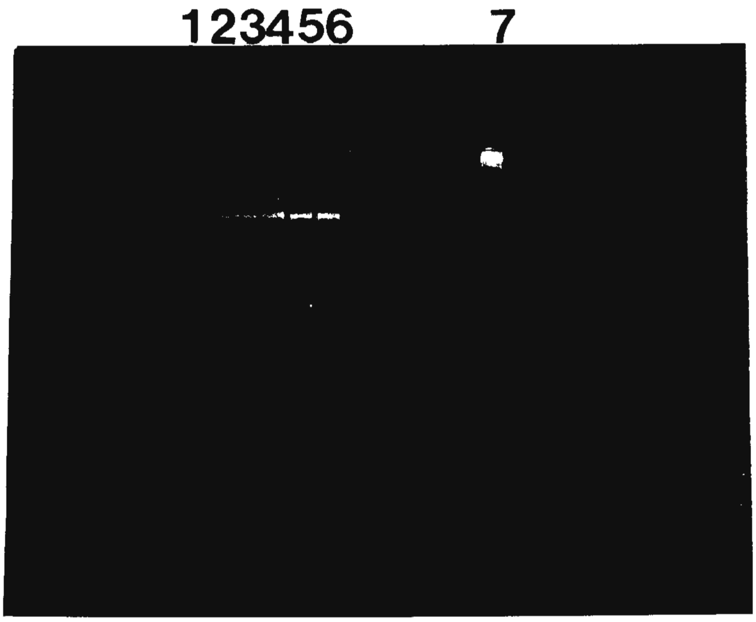
4.3 RESULTS

4.3.1 Polymerase Chain Reaction

The results of PCR were visualized on ethidium bromide stained agarose gels. Generally, approximately 1 ng of target DNA amplified produced about 500 ng of DNA for each 100 µl reaction in 35 cycles (Fig. 4.3). In all cases, a single DNA fragment of approximately 140 bp resulted with no apparent differences in size being evident amongst species (see Fig. 4.3).

The PCR product was purified using various methods. Of all the methods used, the millipore filter unit proved to be the method of choice since it is quick and easy to use and it was possible to recover good and clean DNA within a very short space of time (~10-20 min). Other purification methods such as ammonium acetate precipitation and the DEAE cellulose membrane were attempted. Direct ammonium acetate precipitation method was sometimes complicated if there are primer-dimers present on the PCR product which seems to be due to unoptimized PCR conditions. The DEAE cellulose membrane method

Fig. 4.3 **Ethidium Bromide stained Agarose Gel Electrophoresis of PCR Amplification for six Felid species**



Lane 1 : Leo
Lane 2 : Pardus
Lane 3 : Jubatus
Lane 4 : F. Caracal
Lane 5 : F. Nigripes
Lane 6 : F. Catus
Lane 7 : Lambda DNA digested with Hind III

is very laborious and time-consuming, although sometimes most effective in receiving pure DNA (results not shown).

4.3.2 **Mitochondrial DNA Sequencing**

The eight *Felid species* sequenced all showed variation between one another. Figure 4.4 shows a typical polyacrylamide gel electrophoresis of single-stranded caracal DNA. The African wild cat and the domestic cat sequences show one base change difference (Fig. 4.5), in contrast with the restriction maps which showed complete identity between these two species (Ref. to chapter 3). This variation could be a result of stochasticity. Such sequences were read manually, entered into the DNA computer program (Harley, 1992 unpublished) which allows the alignment of sequences with one another. Fig. 4.6 shows aligned mt DNA sequences of the eight species in question. The domestic cat sequence is shown in its entirety, whereas only nucleotides at variable positions are presented for other species. The caracal was found to possess an amino acid deletion at the position denoted with three dashes in the sequence (Fig. 4.6). This is unique since among all eight species examined it is the only one that possesses this deletion, and in addition, it is not found in any other vertebrates where this region has been sequenced. Had it been found in more than one species it would have represented a very strong synapomorphic character.

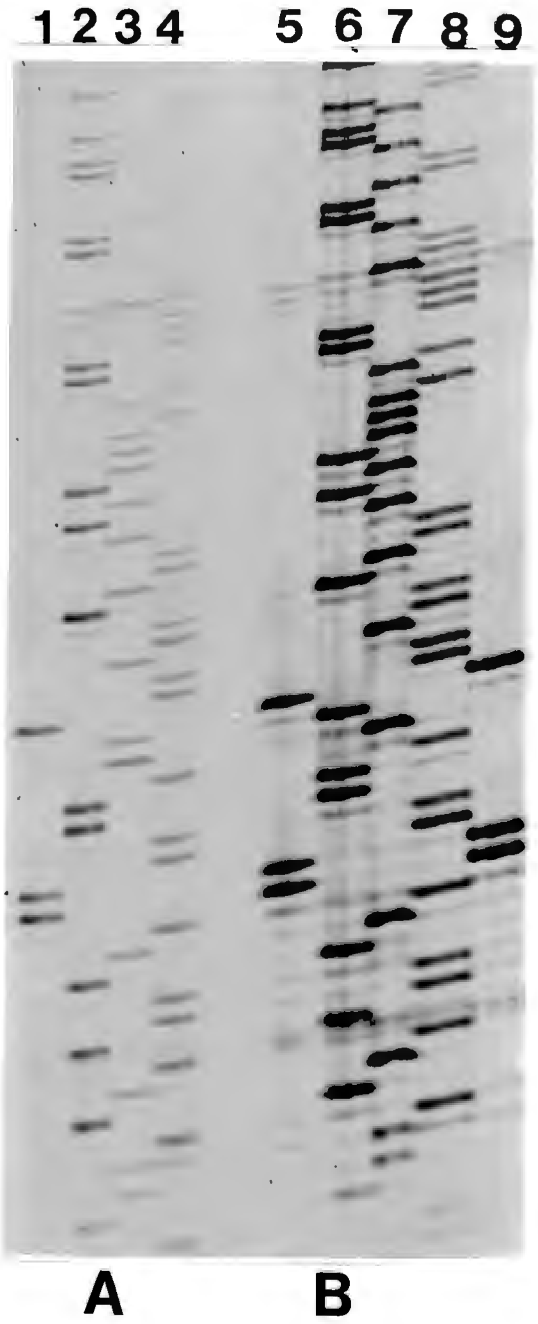
4.4 **PHYLOGENETIC ANALYSIS**

4.4.1 **Distance Methods**

The computer program RESOLVE enabled us to construct a distance matrix

Fig. 4.4 Polyacrylamide Gel Electrophoresis of single stranded caracal mt DNA sequence.

Lane 1 : Guanine
Lane 2 : Adenine
Lane 3 : Thymine
Lane 4 : Cytosine
Lane 5 : Guanine



1. Guanine; 2. Adenine; 3. Thymine; 4. Cytosine;
5. Guanine; 6. Adenine; 7. Thymine; 8. Cytosine;
9. Guanine

A = African Wild cat; B = Domestic cat

Fig. 4.5 Polyacrylamide gel electrophoresis of African wild cat and the domestic cat mt DNA sequences

```
  catus     ACATCCATTC ATCACCATCG GCCAACTAGC CTCCATCCTA TATTTCTTAA CCCTCCTAAT
 lybica     -••••••••• •••••••••• •••••••T•• •••••••••• •••••••C•• ••••••••••
sylvestr    ---••••••• •••••••••• •••••••••• •••••••••• •••••••C•• ••••••••••
nigripes    -••C•••••T •••••••••• •••••••••• T••••••••• ••C••••C•• •T•••••G••
caracal     ---------- ---------• •••••••••• ---••••••• ••C••••C•• ••••••••G•
  lion      --------•• •••••••••• •••••••••• •••••••••• ••C••••CC• •T••T•••••
leopard     ------••C• ••T••••••• •••••••••• •••••••••• ••C••••CC• T•••T•••G•
cheetah     ---------- ---------- ---------- ---------- ---------• T•••T•••G•

  catus     CCTAATACCC ATCTCAGGCA TTATTGAAAA CCGTCTACTC AAATGAAGAG TCTTTGTAGT
 lybica     •••••••••• •••••••••• •••••••••• •••••••••• •••••••••• ••••••••••
sylvestr    •••••••••• •••••••••• •••••••••• •••C•••••• •••••••••• ••••••••••
nigripes    •••••••••• •••••••••• •••••••••• •••••••••T ••G••••••• ••••••••••
caracal     •••••••••T ••T••••••• •C••C••••• •••C••C••A •••••••••• ••••••••••
  lion      •••••••••• •••••••••• •••••••••• •••C••C••• •••••••••• ••••C•••••
leopard     •••••••••• •••••••••• •C••C••••• •••C••C••A •••••••••• ••••••••••
cheetah     •••••••••• •••••••••• •••••••••• •••C••C••T •••••••••• ••••C•••••

  catus     ATA
 lybica     •••TAAAATA CTTTGG
sylvestr    •
nigripes    •••TAAAAT
caracal     •••TAAAATA
  lion      •••TAGAATA C
leopard     •••TAAAATA C
cheetah     •••TAAAATA C
```

Fig. 4.6 **Aligned mt DNA sequences of catus, lybica, sylvestris, nigripes, caracal, lion leopard and cheetah. Dashes refer to gaps. The catus sequences is shown in its entirety, whereas only nucleotides at variable positions are presented for other species.**

for eight *Felid species*. Percentage sequence divergence values are documented in Table 4.1. The sequence divergence varies from a minimum of 1.7 between the European wild cat and African wild cat or the domestic cat, to a maximum of 13.9 between the black-footed cat and the leopard. Phylogenies constructed using FITCH, KITSCH and Neighbor-joining approaches are presented in Fig.4.7. The FITCH and Neighbor-joining trees are topologically equivalent but the KITSCH tree differs in its placement of lion, which is placed in an unlikely position splitting the monophyly of the small cat (Fig.4.7).

4.4.2 **Cladistic Methods**

Eight Felidae sequences were analyzed for phylogenetically informative characters using RESOLVE computer program. Table 4.2 shows phylogenetically informative characters with the numbers 1, 2, 3, 4 and 5 indicating nucleotides (dATP) A, (dTTP) T, (dGTP) G, (dCTP) C and a gap, respectively. Phylogenetically informative characters are indicated by "X". The Henning 86 computer package was used to construct the phylogeny (Fig. 4.8), using the cheetah as an outgroup. The overall topology generated by Henning 86 resembles the dendograms produced by both KITSCH and Neighbor-joining approaches with the only significant difference being the monophyletic relationship between the lion and the cheetah. The FITCH dendogram, however differs from the cladogram by grouping the lion with the smaller cats, (Fig 4.6.A.). Bootstrap analysis using 1000 replicates was performed in order to assess the robustness of the tree. Its majority consensus rule tree

Table 4.1 **Distance Matrix for eight Felidae species**

		1	2	3	4	5	6	7	8
1.	**F. CATUS**	.0	1.7	1.7	7.5	10.1	7.0	11.3	9.7
2.	**F. LYBICA**	1.7	.0	1.7	7.5	10.1	7.0	11.3	9.7
3.	**F. SYLVESTRIS**	1.7	1.7	.0	7.6	8.1	5.3	9.6	8.3
4.	**NIGRIPES**	7.5	7.5	7.6	.0	11.1	8.8	13.9	12.5
5.	**C. CARACAL**	10.1	10.1	8.1	11.1	.0	10.1	5.1	11.1
6.	**P. LEO**	7.0	7.0	5.3	8.8	10.1	.0	7.9	5.6
7.	**P. PARDUS**	11.3	11.3	9.6	13.9	5.1	7.9	.0	5.6
8.	**A. JUBATUS**	9.7	9.7	8.3	11.5	11.5	5.6	5.6	.0

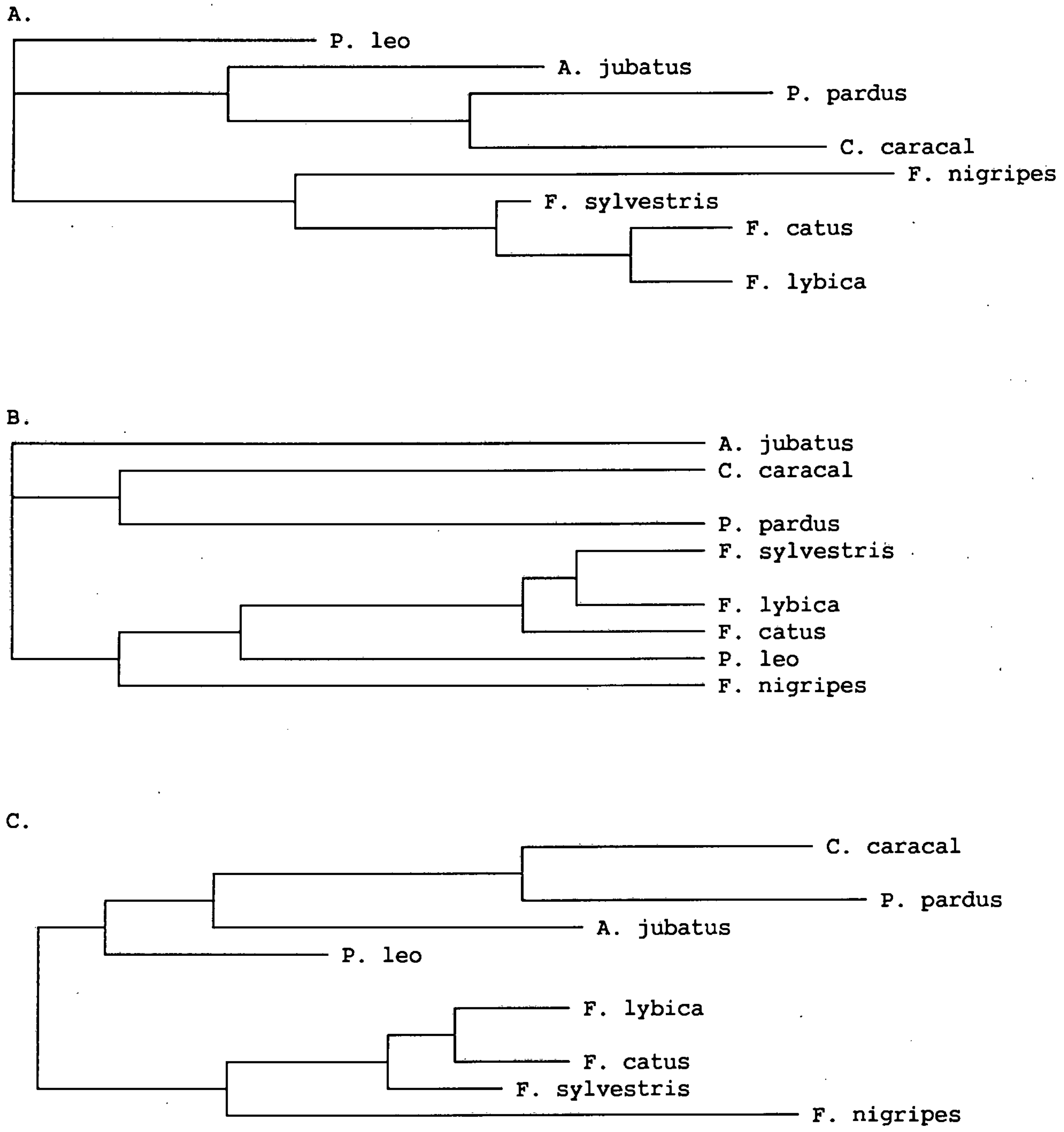


Fig. 4.7 Distance dendograms:

A) FITCH tree
B) KITSCH tree
C) Neighbor-Joining tree (NJ)

Table 4.2 **Table of phylogenetically informative sites for eight felid species with 1 = d(ATP)A; 2 = (dTTP)T; 3 = (dGTP)G; 4 = (dCTP); 5 = gap**

'X' indicates the phylogenetically informative positions.

```
ATCCATTCATCACCATCGGCCAACTAGCCTCCATCCTATATTTCTTAACCCTCCTAATCCTAATACCCATCTCAGGCA
TATTGAAAACCGTCTACTCAAATGAAGAGTCTTTGTAGTATA
ATCCATTCATCACCATCGGCCAACTTGCCTCCATCCTATATTTCTCAACCCTCCTAATCCTAATACCCATCTCAGGCA
TATTGAAAACCGTCTACTCAAATGAAGAGTCTTTGTAGTATATAAAATACTTTGG
--TCCATTCATCACCATCGGCCAACTAGCCTCCATCCTATATTTCTCAACCCTCCTAATCCTAATACCCATCTCAGGCA
TATTGAAAACCGCCTACTCAAATGAAGAGTCTTTGTAGTA
ACCCATTTATCACCATCGGCCAACTAGCTTCCATCCTATACTTCTCAACTCTCCTGATCCTAATACCCATCTCAGGCA
TATTGAAAACCGTCTACTTAAGTGAAGAGTCTTTGTAGTATATAAAAT
------------------GGCCAACTAGC---CATCCTATACTTCTCAACCCTCCTAGTCCTAATACCTATTTCAGGCA
ATCGAAAACCGCCTCCTAAAATGAAGAGTCTTTGTAGTATATAAAATA
------TTCATCACCATCGGCCAACTAGCCTCCATCCTATACTTCTCCACTCTTCTAATCCTAATACCCATCTCAGGCA
TATTGAAAACCGCCTCCTCAAATGAAGAGTCTTCGTAGTATATAGAATAC
-----ATCCATTACCATCGGCCAACTAGCCTCCATCCTATACTTCTCCATCCTTCTAGTCCTAATACCCATCTCAGGC
ATCGAAAACCGCCTCCTAAAATGAAGAGTCTTTGTAGTATATAAAATAC
------------------------------------------------ATCCTTCTAGTCCTAATACCCATCTCAGGC
TATTGAAAACCGCCTCCTTAAATGAAGAGTCTTCGTAGTATATAAAATAC
                                         X     X XX   X   X
X  X        X  X  X              X
```

Table of phylogeneticly (cladisticly) informative sites

1	2	2	2	4	4	4	4	5
2	1	1	1	1	1	4	4	5
3	4	4	4	4	4	4	2	2
4	4	4	4	2	4	2	4	4
5	4	4	4	4	4	2	2	2
6	1	1	1	1	3	1	3	3
7	2	2	2	2	4	2	4	2
8	2	2	2	2	4	2	4	2
9	2	2	4	2	4	4	4	4
10	1	1	1	1	4	4	4	4
11	4	4	4	2	1	4	1	2
12	2	2	2	2	2	4	2	4

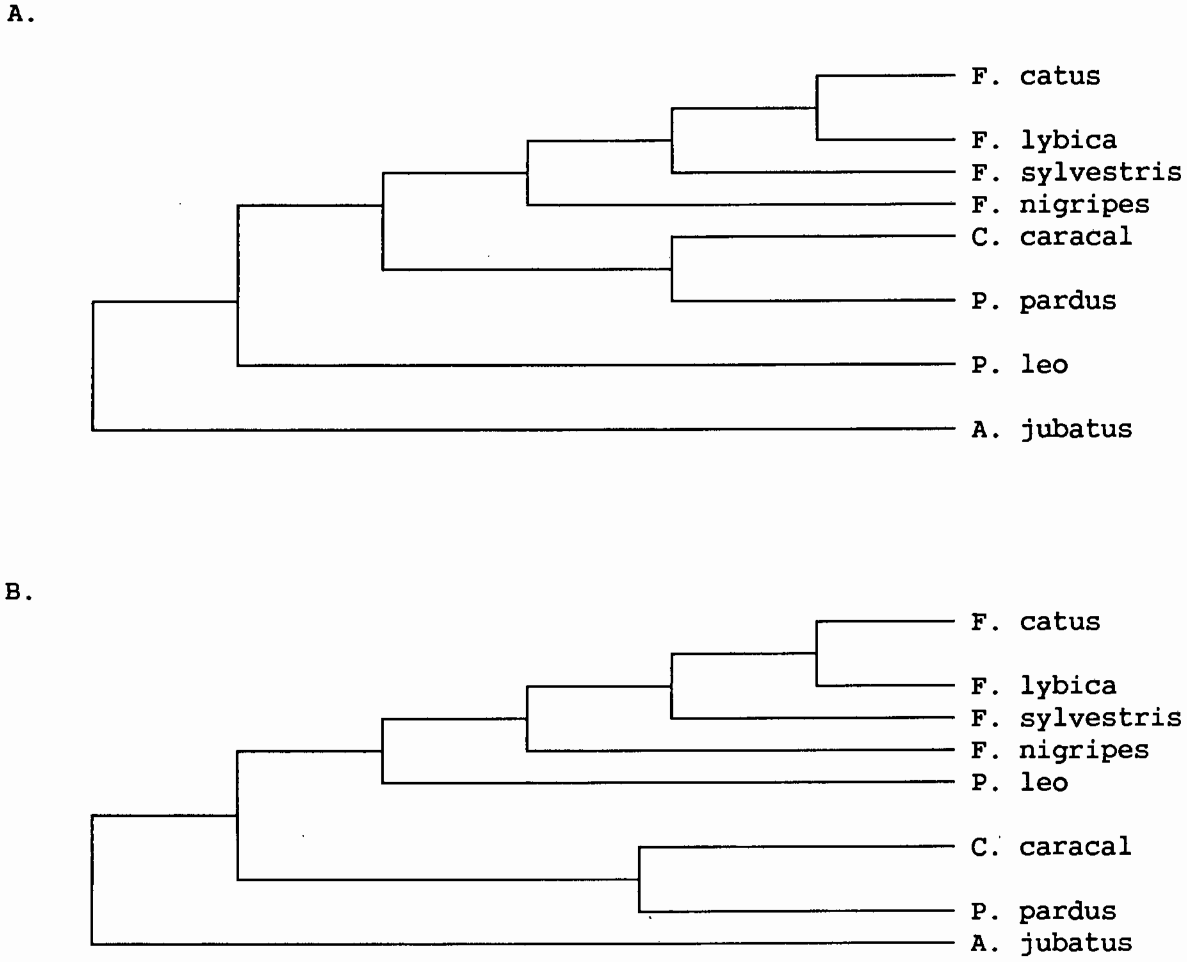


Fig. 4.8 **Cladogram of eight southern African Felidae using the cheetah as an outgroup: A. Maximum parsimony; B. Bootstrap Analysis using 1000 replicates**

resembles the trees generated by both KITSCH and Neighbor-joining methods but differs with the FITCH tree by grouping the lion with the larger cats. The bootstrap values are relatively lower than those obtained with the restriction mapping data but the topology is essentially the same. The most significant difference is the monophyletic relationship between the caracal and the leopard, with 567 out of a total of 1000 replicates. This observation could be attributable to the short length of DNA being sequenced.

4.5 DISCUSSION

To complement the mt DNA restriction mapping data, the methodology for sequencing part of the cytochrome b region of mt DNA with the aid of PCR was developed. Our results demonstrate the inadequacy of sequencing a short region (140 bp) of mt DNA for phylogenetic analysis between closely related species such as the members of the family Felidae. This is evidenced by the small number (12) of phylogenetically informative characters when a total number of eight species are examined as compared with the 71 characters given by the restriction mapping data. Furthermore, the topologies obtained from both cladistic and distance methods are implausible and this together with low bootstrap values (Fig 4.8B) show that this quantity of sequence data is inadequate. However our results are not completely uninformative, since both dendograms and cladograms seem to have resolved the small cat lineage (Fig 4.7 and 4.8). An exception to this is the FITCH dendogram in which the lion is grouped with the small cats i.e. the lion is placed between the black-footed cat and the African wild cat. The close relationship between the members of the smaller cats is consistent with the previous studies based on morphological, behavioural and ethological criteria (Walker *et al.*, 1964; Ewer, 1973; Hemmer, 1978).

The inconsistent results for the larger cats could be attributable to the short region of DNA sequenced. It is suggested that for the purpose of resolving the pantherine lineage (large cats), a longer region of mt DNA of approximately 300-400 bp would have to be sequenced. This will not only help generate a large number of phylogenetically informative sites for cladistic analysis, but would also minimize the stochastic variation when distance approaches are employed. Additionally, for closely related species such as the member of the family Felidae, a more rapidly evolving region of mt DNA such as the displacement loop (D-loop) might be more appropriate.

CHAPTER 5

5. CONCLUSION

The major findings of this study can be summarized as follows:

1. Using the mitochondrial DNA restriction mapping approach, phylogenetic trees of eight southern African *Felidae* have been constructed.

2. The restriction maps for the African wild cat and the domestic cat are identical, emphasizing their close relationship and the African origin of the domestic cat. The restriction map for the European wild cat differs from that of either the African wild cat as well as the domestic cat by four restriction sites. This would appear to be consistent with its current separate specific status.

   The confirmation of the separate specific status of the European wild cat with respect to the African wild cat and the domestic cat, provide a potentially stronger genetic basis for a conservation programme to preserve the European wild cat than for the African wild cat. The greater genetic distance between the mt DNA's of the European wild cat and the domestic cat implies that there is a greater chance for outbreeding depression, (ie. a decrease in hybrid vigour) if hybridization between the European wild cat and the domestic cats were to occur than in the case of hybridization between the African wild cat and the domestic cat.

3. Both distance and cladistic approaches strongly support the monophyletic relationship amongst the small cats, viz. African wild cat, domestic cat and the European wild cat. This is evidenced by a high bootstrap value of 1000 out of a total

of 1000 replicates between the European wild cat and the African wild cat or the domestic cat. Similarly a bootstrap value of 935/1000 between the black-footed cat and the other small cats would seem to favour the close relationship between the latter species.

4. The lion and the leopard are monophyletic in both distance and cladistic approaches.

5. The precise placement of the caracal has yet to be resolved but its deep rooting would be more consistent with a separate generic status rather than its inclusion within either *Felis* or *Panthera*. To elucidate the phylogenetic placement of the unresolved caracal would require additional work such as more mt DNA sequencing or more mapping studies.

6. We have attempted to develop the methodology for sequencing part of the cytochrome b region of mt DNA for eight *Felid* species following PCR amplification. This was successful and about 140 base pairs of sequences were obtained in all eight taxa. From these sequences phylogenies were constructed as a demonstration of the approach but were not conclusive due to the short length of DNA being compared. Only 12 phylogenetically informative sites were obtained from this length of sequence, compared with 71 from the restriction maps. Future work in this family would involve sequencing a longer region of cytochrome b using our personally-designed oligonucleotide primers as well as other described primer for this region.

6 : REFERENCES

ACTANDER P. (1988) Comparative studies of avian DNA by restriction fragment length polymorphism analysis : convenient procedures based on blood samples from live birds. J. of Ornithology *129*:205-216

ALLARD MW & HONEYCUTT RL. (1991) Ribosomal DNA variation within and between species of rodents, with emphasis on the genus onychomy. Mol. Biol. Evol. *1*:71-84

ANDERSON S, BANKIER AT, BASELL BG, DE BRUIJN MHL, COULSON AR, DROUIN J, EPRON JC, NIERLICH DP, ROE BA, SAUGER F, SCHREIER PH, SMITH AJH, STADEN R & YOUNG JG. (1981) Sequence and organization of the human mitochondrial genome. Nature *290*:457-474

AQUADRO CF & GREENBERG BD. (1983) Human mitochondrial DNA variation and evolution : Analysis of nucleotide sequences from seven individuals. Genetics *103*:287-312

ARMITAGE DL & CLUTTON-BROCK J. (1981) A radiological and histological investigation into the mummification of cats from ancient Egypt. J. Archaeol. Sci. *2*:185-196

ASHLEY M & WILLS C. (1987) Evolution *41*:854-863

AUSUBEL FM, BRENT R, KINGSTON RE, MOORE DD, SEIDMAN JG, SMITH JA & STRUHL K. (1988) Current protocols in molecular biology. John Wiley & Sons Inc. New York

AVISE JC, JOHN C. (1975) Systematic value of electrophoresis data. Syst. Zool.

AVISE JC, NIGIEL JE AND ARNOLD J. (1984) Demorgraphic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. *20*:99-105

AVISE JC, JOHN C. (1989) The role of molecular genetics in the recognition and conservation of endangered species. Tree *4*:279-281

BENTON MJ. (1988) The relationships of the major group of mammals: New approaches. Tree *3*:40-45

BENVENISTE RE, SHERR, CJ & TODARO GJ. (1975) Evolution of type C viral genes : Origin of Feline leukemia virus. Sci. *190*:886-888

BIER M. (1959) Electrophoresis Theory, Methods and Applications

BROWN TA. (1983) Gene cloning

BROWN WM, GEORGE, M & WILSON AC. (1979) Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA *76*:1967-1971

BROWN W. (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc. Natl. Acad. Sci. USA *77*:3605-3609

CANN RL, STONEKING M &WILSON AC. (1987) Mitochondrial DNA and human evolution. Nature *325*:31-36

CARR SM, BROTHERS AJ & WILSON AC. (1987) Evolutionary inferences from restriction maps of mitochondrial DNA from nine taxa of xenopus frogs. Evol. *41*:176-188

CHAMPION AB, PRAGER EM, WACHTER D & WILSON AC. (1974) Microcomplement fixation, in C.A. Wright (ed.) 397-416

CLARY DO & WOLSTENHOLME DR. (1985) The mitochondrial DNA molecule of Drosophila Yakube : Nucleotide Sequence, Gene Organization, and Genetic Code. J. Mol. Biol. *22*:252-271

COLLIER GE & O'BRIEN SJ. (1985) A molecular phylogeny of Felidae : immunological distance. Internat. J. of Organic Evolution *39*:473-487

CORBERT GB. (1966) The terrestrial mammals of Western Europe. G.T. Foulis and Co. Ltd. London 155-157

CROWE TM. (1988) Molecules vs morphology in phylogenetics : A non-controversy. Trans. Roy. Soc. S. Afr. *46*:317-334

CUMMINGS OW, KING TC, HOLDEN JA & LOUW RL. (1987) Purification and characterization of the potent endonuclease in extracts of bovine heart mitochondria. J. Biol. Chem. *262*:2005-2015

DARNELL, LODISH & BALTIMORE . (1986) Molecular cell biology

DE BRY RW & SLADE NA. (1985) Cladistic analysis of restriction endonuclease cleavage maps within a maximum - Likelihood Framehood. Syst. Zool. *34*:21-34

EASTEAL S. (1991) The relative rate of DNA evolution in primates. Mol. Biol. Evol. *8*:115-127

ESSOP MF, EMMANUEL M & HARLEY EH. (1988) Mitochondrial DNA analysis for the molecular taxonomy of South African mammals. Trans. Roy. Soc. S. Afr. *46*:291-293

EWER RF. (1973) The Carnivora 406-411

FELSENSTEIN I. (1985) Confidence limits on phylogenies : An approach using bootstrap. Evol. *39*:783-791

FERRIS SD & BERG WI. (1985) Parsimony in systematics : Biological and statistical issues. Ann. Rev. of Ecol. Syst. *14*:313-333

FERRIS SD, WILSON AC & BROWN WM. (1981) Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc. Natl. Acad. Sci. USA. *78*:2432-2436

FITCH WM & MARGOLIASH E. (1967) Construction of phylogenetic trees. Sci. *155*:279-284

FLOWER WH & LYDEKKER R. (1891) An introduction to the study of mammals. Living and Extinct. Adam and Charles Black, London

GEORGE JM & RYDER OA. (1986) Mitochondrial DNA evolution in the genus Equus. Mol. Biol. Evol. *3*:535-546

HAGELBERT E & SYKES B. (1989) Ancient bone DNA amplified. Nature *324*:485

HALANYCH KM. (1991) 5S Ribosomal RNA sequences Inappropriate for phylogenetic reconstruction. Mol. Biol. Evol. *8*:249-253

HALLIDAY T & ARANO B. (1991) Resolving the phylogeny of the European newts. Tree *6*:113-117

HARLEY EH. (1986) A general DNA analysis program for the Hewlett packard Model 86/87 microcomputers. Nucleic Acids Res. *14*:467-477.

HARLEY EH. (1988a) DNA approaches to molecular taxonomy. Trans. Roy. Soc. S. Afr. *46*:283-290

HARLEY EH. (1988b) The retrieval of the quagga. S. Afr. J. of Sci. *84*:158-159

HARRISON, & RICHARD, G. (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Tree *4*:7-11

HARTL DL. (1983) Principles of population genetics

HASEGAWA M, KISHIMO H, HAYASAKA K & HORAI S. (1990) Mitochondrial DNA evolution in primates : Transition rate has been extremely low in the lemur. J. Mol. Evol. *31*:113-121

HEMMER H. (1978) Were the leopard cat and the sand cat among the ancenstry of domestic cat races. Carnivore *1*:106-108

HENDERSON S, DE BRUIJN MHL, COULSON AR, EPERON IC, SANGER F & YOUNG IC. (1982) Complete sequence of mitochondrial DNA. J. Mol. Biol. *156*:638-717

HIGUCHI R, VAN BEROLDINGER CH, SENSABAUGH GF & ERLICH HA. (1988) DNA typing from single hairs. Nature *332*:543-546

HILLIS DM. (1987) Molecular versus morphological approaches to systematics. Ann. Rev. of Ecol. Syst. *18*:23-42

HILLIS DM, LARSON, ALLAN, DAVIS SK & ZIMMER EA. (1990) In molecular systematics by Hillis & Moritz

HONEYCUTT R & MINDELL DP. (1980) Ribosomal RNA in vertebrates : evolution and phylogenetic applications. Ann. Rev. of Ecol. Syst. *21*:541-566

HOYER BH, VAN DE VELDE NW, GOODMAN M & ROBERTS RB. (1972) Examination of hominid evolution by DNA sequence homology. J. of Human Evol. *1*:645-649

HUTCHINSON III, CA, NEWBOLD JE, POTTER SS, EDGELL MH. (1974) Maternal inheritance of mammalian mitochondrial DNA. Nature *221*: 536-538

JIN L & NEI M. (1991) Relative Efficiences of the Maximum-Parsimony and Distance-Matrix Methods of Phylogeny Construction for Restriction Data. Mol. Biol. Evol. 8:261-281

JOHNSON MJ, WALLACE DC, FERRIS SD, RATTAZZI MC & SFORZA LL. (1983) Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. J. Mol. Evol. *19*:255-271

KING JL & JUKES TH. (1969) Non-Darwiman evolution. Sci. *164*:788-798

KRATOCHVIL J & KRATOCHVIL Z. (1976) Zool. Listy. *25*:193-208

KRATOCHVIL J. (1982) The karyotype and system of the family Felidae (Carnivore Mammalia) Fol. Zool. *31*:289-304

LAYON SM. (1988) The stochastic mode of molecular evolution : What for systematic investigation? The AUK *105*:565-573

LAIRD CD, McCONAUGHY BL & McCARTHY BJ. (1969) Rate of fixation of nucleotide substitutions in evolution. Nature *224*:149-154

LANSMAN RA, SHADE RO, SHAPIRA JF & AVISE JC. (1981) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. J. Mol. Evol. *17*:214-226

LEYHAUSEN P. (1979) Cat behaviour. Garland STPM Press, New York

MARX JL. (1988) Multiplying genes by leaps and bounds. Sci. *240*:1408-1410

MAXSON LR & WILSON AC. (1974) Convergent morphological evolution detected by studying proteins of tree frogs in the Hyla Eximia species group. Sci. *185*:66-68

MAXSON LR & MAXSON RD. (1979) Comparative albumin and biochemical evolution in *Plethodontid* salamanders. Evol. *33*:1057-1062

MAXSON LR. (1981a) Relationships among Eurasian *Bufo*. Copeia, 579-583

MAXSON LR. (1981b) Albumin evolution and its phylogenetic implications in toads of the genus *Bufo*. Copeia, 579-583

MAXSON LR, SONG A & LOPATA R. (1981) Phylogenetic relationships among North American toads, genus *Bufo*. Biochem. Syst. Ecol. *9*:347-350

MAXSON LR. (1984) Molecular probes of phylogeny and biogeography in toads of the widespread genus *Bufo*. Mol. Biol. Evol. *1*:345-356

McCARTHY BJ & CHURCH RB. (1970) The specificity of molecular hybridization reactions.

McCLELLAND M, HANISH J, NELSON M & PATEL Y. (1988) KGB : A single buffer for all restriction endonucleases. Nucleic Acids Res. *16*: 364

MINDELL, DP & HONEYCUTT, R. (1990) Ribosomal RNA in vertebrates : Evolution and phylogenetic application. Ann. Rev. of Ecol. Syst. *21*:541-566

MORITZ C, DOWLING TE & BROWN WM. (1987) Evolution of animal mitochondrial DNA : Relevance for population biology and systematics. Ann. Rev. Ecol. Syst. *18*:269-292

MULLIS KB & FALOONA FA. (1985) Specific synthesis of DNA *invitro* via a polymerase catalyzed chain reaction. Sci. *230*:1350

MULLIS KB. (1990) The unusual origin of the polymerase chain reaction. Scientific American:36-43

NATHANS D & SMITH HO (1975) Restriction endonucleases in the analysis and restructuring of DNA molecules. Ann. Rev. Biochem. *44*:

NEI M. (1977) Standard error of immunological dating of evolutionary time. J. of Mol. Evol. *9*:203-211

NEI M & KOEHN RK. (1983) Evolution of genes and proteins

NOWAK RM & PARADISO JL. (1983) Walker's mammals of the world, 4th Edn. John Hopkins University Press, Baltimore

O'BRIEN SJ, WILOT DE, BUSH M, CARD TM, FITZGIBSON C, AGGONDEY I & LEAKEY RE (1987) East African cheetahs : Evidence for two population bottlenecks. Proc. Natl. Acad. Sci. *84*:508-511

O'BRIEN SJ, ROELKE ME, MARKER L, NEWMAN A, WINKLER CA, MELTZER D, COLLY L, EVERMANN JF, BUSH M & WILDT DE. (1985) Genetic basis for species vulnerability in the Cheetah Science. *227*:1428-1434

PAÄBO S, GIFFORD JA & WILSON AC. (1988a) Polymerase chain reaction : ancient DNA amplification. Nucleic. Acids Res. *16*:

PAÄBO S, HIGUCHI RG & WILSON AC. (1988b) Ancient DNA and polymerase chain reaction. J. of Biol. Chem. *264*:9709-9712

PAÄBO S. (1989) Ancient DNA : Extraction, characterization, molecular cloning and enzymatic amplification. Proc. Natl. Acc. Sc. *86*:1936-1943

PLATNICK NI. (1988) Programs for quicker relationships. Nature, *335*:310

POCOCK RI. (1951) Catalogue of genus Felis. Trustees of the British museum. Nat. Hist., London

PRAGER EM, WELLING GW & WILSON AC. (1978) Comparison of various immunological methods for distinguishing among mammalian pancreatic ribonucleases of known amino acid sequence. J. of Mol. Evol. *10*:293-307

RANDI E & RAGNI B. (1991) Genetic variability and biochemical systematics of domestic and wild cat populations (Felis sylestris : Felidae). J. Mamm. *72*:79-88

RIDDLE BR & HONEYCUTT RL. (1990) Historical biogeography in North American arid regions : An approach using mitochondrial DNA phylogeny in grasshopper mice (Genus Onychomys). Evol. *44*:1-15

ROBINSON R. (1976a) Homologous genetic variation in the Felidae. Gen. *46*:1-31

ROBINSON R. (1976b) Cytogenetics of the Felidae. In RL Eaton. The worlds cats, Carnivore Res. Inst., Washington DC. *3*:15-28

ROE BA, MA D, WILSON RK & WONG JFW (1985) The complete nucleotide sequence of the Xenopus Laevis mitochondrial genome. J. Biol. Chem. *260*:9759-9774

ROSEVEAR DR (1974) The carnivores of West Africa. Trustees of the British museum (Natl. Hist.), London:373-393

SARICH VM & WILSON AC. (1966) Quantitative immunochemistry and the evolution of primate albumins : Microcomplement fixation. Sci. 154-1563

SARICH VM & WILSON AC. (1967) Immunological time scale for hominid evolution. Sci. *158*:1200

SARICH VM. (1969) Principed origins and the ... of evolution in carnivore albumins. Syst. Zool. *18*:286-295

SATTA Y & YAKAHATA N. (1990) Evolution of Drosophila Mitochondrial DNA and the history of the Melanogaster subgroup. Proc. Natl. Acad. Sci *87*:9558-9562

SEALY PG & SOUTHERN EM. (1982) Gel Electrophoresis of nucleic acids

SEARLE AG. (1968) In comparative genetics of coat colour in mammals 138-140 (Logos Press Ltd., London)

SHELDON FH (1987) Phylogeny of Herons estimated from DNA-DNA hybridization data. The AUK *104*:97-108

SHIELDS GF & HELMBYCHOWSKI KM. (1991) Mitochondrial DNA of birds

SIBLEY CG & AHLQUIST JE. (1985) The relationships of some groups of African birds, based on comparisons of the genetic material, DNA. Proc. Intern. Symp. African Vertebr. Bonn.

SIBLEY CG & AHLQUIST JE. (1986) Reconstructing bird phylogeny by comparing DNAs. Sci. American :68-78

SMITHERS RHN. (1983) The mammals of the Southern African subregion. University of Pretoria, Pretoria.

SMITHERS RHN. (1986) In South African Red Data Book : Terrestrial mammals 77-79 (South African National Scientific Programmes Report 125, Pretoria, Council for Scientific and Industrial Research)

SOBER, E. (1983) Parsimony in systematics: philosophical issues. Ann. Rev. Ecol. Syst. *14*:335-357.

SOUTHERN EM. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*:503-517

TAMURA K, AOTSUKA T & KITAGAWA O. (1991) Mitochondrial DNA polymorphisms in the two subspecies of Drosophila Sulfurigaster : Relationship between geographic structure of population and nucleotide diversity. Mol. Biol. Evol. *8*:104-114

TAYLOR JW, SMOLICH BD & MAY G. (1987) Evolution *40*:716-739

THOMAS RH, SCHAUFFER W, WILSON AC & PAÄBO S. (1989) DNA phylogeny of the extinct marsupial wolf. Nature *340*:465-467

THOMAS WK, PAÄBO S, VILLABLANCA FX & WILSON AC. (1990) Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. J. Mol. Evol. *31*:101-112

THORPE JP. (1982) The molecular clock hypothesis : Biochemical evolution, genetic differentiation and systematics. Ann. Rev. of Ecol. Syst. *13*:139-168

VAN WAGNER CE & BAKER AJ. (1990) Association between mitochondrial DNA and morphological evolution in Canada geese. J. Mol. Evol. *31*:373-382

WALKER EP, WARNICK F, HAMLET SE, LANGE KI, DAVIS MA, NIBELE HE & WRIGHT PF. (1964) Mammals of the world. John Hopkins press, Baltimore. *2*:1268-1282

WALLACE DG & WILSON AC (1972) Comparison of frog albumins with those of other vertebrates. J. Mol. Evol. *2*:72-86

WALLACE DG, KING M & WILSON AC. (1973) Albumin diffirences among ranid frogs : Taxonomic and phylogenetic implications. Syst. Zool. *22*:1-13

WERDELIN L. (1983) Morphological patterns in the skulls of cats. Biol. J. Linn. Soc. *19*:375-391

WHITE TJ, ARNHEIM N & ERLILCH HA. (1989) The polymerase chain reaction. TIG. *5*:

WILLIAMS BL & WILSON K. (1975) Radioisotope techniques in : A biologists guide to principles and techniques of practical biochemistry. First ed. William Clowes and Sons, London:170-198

WILSON AC, CARLSON SS & WHITE JJ. (1977) Biochemical Evolution. Ann. Rev. Biochem. *46*:573- 639

WILSON AC, CANN RL, CARR SM, GEORGE M, GYLLENSTEIN VB, HELMBYCHOWSKI KM, HIGUCHI RG, PAWMBI SR, PRAGER EM, SAGER RD & STONEKING S. (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. of the Linnean Soc. *26*:375-400

WILSON AC, OCHMAN H & PRAGER EM. (1987) Molecular time scale for Evolution. TIG. *3*:241-247

WUSTER-HILL DH & CENTREWALL WR. (1982) The interrelationships of chromosome banding patterns in canids, mustelids, hyena and Felids. Cytogenet. Cell. Gen. *34*:178-192

ZEUNER FE. (1967) History of domesticated animals. Hutchinson and Co. Ltd., London:387-400

APPENDIX I

BUFFERS

MITOCHONDRIAL DNA EXTRACTION, PURIFICATION AND END-LABELLING

EXTRACTION BUFFER:	100 mM Tris pH 7.5 150 mM NaCl 20 mM EDTA 10% W/V sucrose
STE BUFFER, PH8:	10 mM Tris (pH8) 1 mM EDTA pH8 100 mM NaCl
TE BUFFER:	10 mM Tris pH8 1 mM EDTA pH8
TAE BUFFER:	(1 l) - 242 g Tris Base 57.1 ml Glacial Acetic Acid 100 ml 0.5 M EDTA (pH8) 5 ml 10% Na-PP
(KBG) K^+ GLUTAMATE BUFFER:	2 x KGB 200 mM C^+ Glutamate 50 mM Tris Acetate pH 7.6 20 mM Magnesium Acetate 2 mg/ml^{-1} or 100 µl/ml BSA 1 mM 2x Mercaptoethanol of 50 Mercapto EtOH

BUFFERS FOR SOUTHERN HYBRIDIZATION AND nDNA ISOLATION

DIGESTION BUFFER:	100 mM NaCl 10 mM Tris HCl pH8 25 mM EDTA 0.5% SDS 0.1 mg/ml proteinase K	
20X SSC:	For 1 l: 175.3 g NaCl 88.2 g NaCitrate pH to 7 with 10 M NaOH	
NEUTRALIZATION BUFFER:	20 X SSC 0.5 M Tris HCl pH 7.5	
DENATURATION BUFFER:	1.5 M NaCl + 0.5 M NaOH	
PRE-HYBRIDIZATION BUFFER:	6x SSC 0.1% SDS 0.06% Na-PP 0.25% Blotto	
POST HYBRIDIZATION STRINGENCY WASHES:	1st Wash -	6x SSC 0.1% SDS 0.06% Na-PP
	2nd Wash -	3x SSC 0.1% SDS 0.06% Na-PP
	3rd Wash -	0.1x SSC 0.1% SDS 0.06% Na-PP

ENZYMES AND KITS: Enzymes were supplied by Amersham International pIc, Amersham Place, Little Chalfort Buckinghamshire, England, HP7 9NA; Bethesda Research Laboratories (BRL) Gaitherburg, Maryland 20877; Boehringer Mannheim GmbH, Biochemica, P O Box 310 120, 06800, Mannheim 31, West Germany.

Multiple Kit was supplied by Boehringer Mannheim.

Sequenase version 2.0 was supplied by Amersham

Gene Amp DNA Amplification Reagent Kit was supplied by Perskin Elmer Cetus, 761 Main Avenue, CT 06859

REAGENTS AND MATERIAL:

Hyperfilm MP - Amersham
^{32}P-dCTP - Amersham

Klenow Enzyme or Large Fragment of DNA polymerase I - Amersham
Tag DNA polymerase - Cetus, New England Biolabs and Promega

SOLUTIONS FOR DNA ISOLATION FROM TISSUE CULTURE

PROTEINASE K: 5 mg (0.005 g) to 5 ml H_2O
use 100 µl per vial of cells

SODIUM PERCHLORATE: 1 M solution:
14 g in 100 ml
double distilled H_2O

LYSIS BUFFER:
10 mM Tris-HCl
5 mM $MgCl_2$
1% v/v Triton x-100

SEQUENCING BUFFERS/SOLUTIONS

10X NNB :		1L
	Tris base	162 g
	Boric Acid	27.5 g
	EDTA	9.3 g
	Distilled H_2O	800 ml

50% Acrylamide; 25% Bis Acrylamide (20:1)

STOCK: 500 ml:-
250 g Acrylamide
12.5 g bis-acrylamide
200 ml distilled H_2O

1X NNB + Acrylamide - WORKING SOLUTION

6% Gels:

50% Acryl, 2.5% bis	12 ml
10X NNB	10 ml
Distilled H_2O	42 ml
Urea 8 M	48 g

TEMED + Ammonium Persulphate.

DEAE CELLULOSE EXTRACTION METHOD, BUFFERS AND SOLUTIONS:

LOW SALT BUFFER:
50 mM Tris-CHl (ph 8.0)
0.15 M NaCl
10 mM EDTA (pH 8.0)

HIGH SALT ELUTION BUFFER:
50 mM Tris-HCl pH 8.0
1 M NaCl
10 mM EDTA pH 8.0

APPENDIX II

COMPUTER SIMULATIONS

SEVONA

Operates by generating a string analogous to a DNA sequence (of special length) as the ancestral sequence. It allows the sequence (a choice between a random sequence and one with a G-C bias is offered) to evolve in a manner simulating the stochastic accumulation of point mutations in DNA. The sequence generates progeny sequences at specified intervals to produce ultimately a set of sequences analogous to those found in set of contemporary related taxa, but for which the true phylogeny is known. The evolution of the progeny sequences may follow either pectinate or dichotomous branching patterns.

Base changes can be set to be either random (where there is an equal chance for any base to change to any of the three bases), or non-random, in which case a value for frequency of transitions can be set. This may be a high transition bias (20%), as expected with mtDNA, or if the mutations were random the fequency of transitions would be 33.3%. This has a profound effect on the frequency of homoplasies in a set of taxa.

The programme allows specified and variable - but stochastic amounts of sequence divergence (1% or 2%) between the nodes of the tree. The progeny sequences generate extant sequences which can then be analyzed by any of the standard cladistic or distance - based programmes, to see whether they can reproduce the correct phylogeny. In this way, the progress of change of individual characters through the generations can be followed.

Fig.3.9 shows a typical simulation, where lower case letters represent the occurrence of any mutations. Homoplasies are indicated and phylogenetically informative sites are marked with X.

Simulations using an ancestral sequence of, say, 75 bp, stand much less chance, for a given amount of mutation between the nodes, for inferring the correct trees than when using a sequence of 300 bp. This is due to stochastic variation. Therefore, the larger the set of data, the more reliable the outcome of phylogenetic analysis. Similarly, if there are too few phylogenetically informative sites for cladistic analysis, the resulting phylogenies appear to be less well-resolved.

Resolve Version 2.4, Harley - Unpublished

Resolve, a computer program which allows easy handling and manipulation of restriction mapping data was originally written for the Hewlett-Packard model 86 but has subsequently been re-written in True Basic, in an IBM compatible format.

The program of offers three main functions:

i) Construction of restriction maps of DNA from single and double digestion data.

ii) The management of sets of mapped data from different DNA's after catalogueing, editing and manipulating, where necessary.

iii) The comparison of maps from related taxa and the construction of sequence divergence tables and phylogenetically informative sites.

Detailed Example of a Three-way Analysis : Reproduced from Dr MF Essop Ph.D Thesis, 1992, UCT, SA

A set of dummy restriction data will be used to illustrate a three-way analysis for a hypothetical Test sequence of 16400 base pairs. The Test sequence restriction data for BamHI, PstI and EcoRI, in various combinations, is shown in Table 2.1. Restriction data for each double digestion was individually entered into the program, using option 1 of the main menu layout (Fig. 2.2). This led to the file status being displayed as is shown in Fig. 2.3, and appropriate option was selected (in this case option 6). The data of one of the double digestion, i.e. EcoRI/PstI was then entered at a specified error of 4% (Fig. 2.4).

The algorithms subsequently proceeded to match-up each of the single digestion fragments with various sets of the double digestion fragments (Fig. 2.5). Partial solutions which satisfied data were then stored in the temporary file of the program (Fig. 2.3). The EP1 designation displayed in the temporary file in Fig. 2.3, would indicate that one partial solution was found for the EcoRI/PstI double digestion. The restriction data for BamHI/PstI and BamHI/EcoRI double digestions was analyzed in a similar manner, and their partial solutions also stored in the temporary file (Fig. 2.3). The partial solutions found for BamHI/EcoRI, BamHI/PstI and PstI/EcoRI are illustrated in Fig. 2.6. Such plots can be readily produced using options 2 and 10 of the main menu layout (Fig. 2.2). Option 7 of the main menu layout was then used to collectively analyze the 3 data sets stored in the temporary file, and map the PstI, BamHI and EcoRI restriction sites relative to one another. Such data was then stored in the final file (Fig. 2.3). The final map positions for PstI, BamHI and EcoRI are shown in Fig. 2.7).

Table 2.1 **Restriction fragment sizes (in base pairs) for a three-way analysis of the Test sequence using BamHI, PstI and EcoRI**

BamHI	BamHI + PstI	PstI	PstI + EcoRI	EcoRI	BamHI + EcoRI
12 000	8 500	8 500	7 900	16 400	6 500
4 400	4 400	7 900	5 500		5 500
	2 500		3 000		4 400
	1 000				

Fig. 2.2 **Display of the Main Menu Layout of the RESOLVE Program**

```
                         M A I N   M E N U


              1.   Display management file status
              2.   Display temporary solutions
              3.   Display final maps
              4.   Edit DNA, or restriction enzyme files
              5.   Edit temporary solution files
              6.   Edit final map files
              7.   Map new enzymes to temporary files
              8.   Three enzyme consensus analysis (temp. to final maps)
              9.   Analyses of final maps
             10.   Toggle display/print modes
             11.   Exit program (always end session with this)

                Enter appropriate number (1-11)
```

Fig. 2.3 **A Typical Display of the RESOLVE Program's File status. The Complete Names of Restriction Endonucleases are listed under Abbreviations**

```
     DNA         R.E. pairs in temporary file   R.E.s in final file

1.   Buffalo     vB3 Bg2 vu2 ................   EN
2.   Kudu        Bu2 ........................   E
3.   Eland       gu2 gB2 Bu2 ................   NvgHRB
4.   Nyala       ............................
5.   Bushbuck    ............................
6.   Test        EP1 BE1 BP1 ................   EBP
```

Fig. 2.4 Display of Restriction Data (PstI/EcoRI) being Entered into the RESOLVE Program

```
Enter number of fragment given by restriction enzyme EcoRI -

Enter number of fragment given by restriction enzyme PstI  -

Enter % error allowed (between 2 and 5% is reasonable to start with)

Now enter the fragment sizes for EcoRI (in any order)
?
16400

Now enter the fragment sizes for PstI
?
8500
?
7900

Now enter the fragment sizes for restriction with both enzymes
? (Expecting 3 fragment entries)
7900
?
5500
?
3000


                EcoRI          PstI       EcoRI + PstI

                16400          8500          7900
                               7900          5500
                                             3000
                -----         -----         -----
Totals          16400         16400         16400
```

Fig. 2.5 Display of Fragment Analyses and Comparison by the RESOLVE Program

```
A no 1   =  16400 pairs with fragments 1 & 2 & 3 = 16400
            Total fits = 1

Partial solution 1:

A fragment 1 consists of AB fragments 1 & 2 & 3

One partial solution found for A



B no 1   =  8500 pairs with fragments 2 & 3 = 8500
B no 2   =  7900 pairs with fragment 1      = 7900
            Total fits = 2

Partial solution 1:

B fragment 1 consists of AB fragment 2 & 3
B fragment 2 consists of AB fragment 1

One partial solution found for B


 A partial solution compared with B partial solution 1

               Full solution found
This consists of A partial solution 1 with B partial solution 1


Search complete, with 1 solution(s) found

                            OPTIONS

               1.  Store solution(s) in temporary file

               2.  Try again with a different error value

               3.  Exit


Enter appropriate number
?
```

Fig. 2.6 Partial solutions for BamHI/EcoRI, PstI/BamHI and PstI/EcoRIfor the Test Sequence. The complete names of restriction endonucleases are listed under Abbreviations

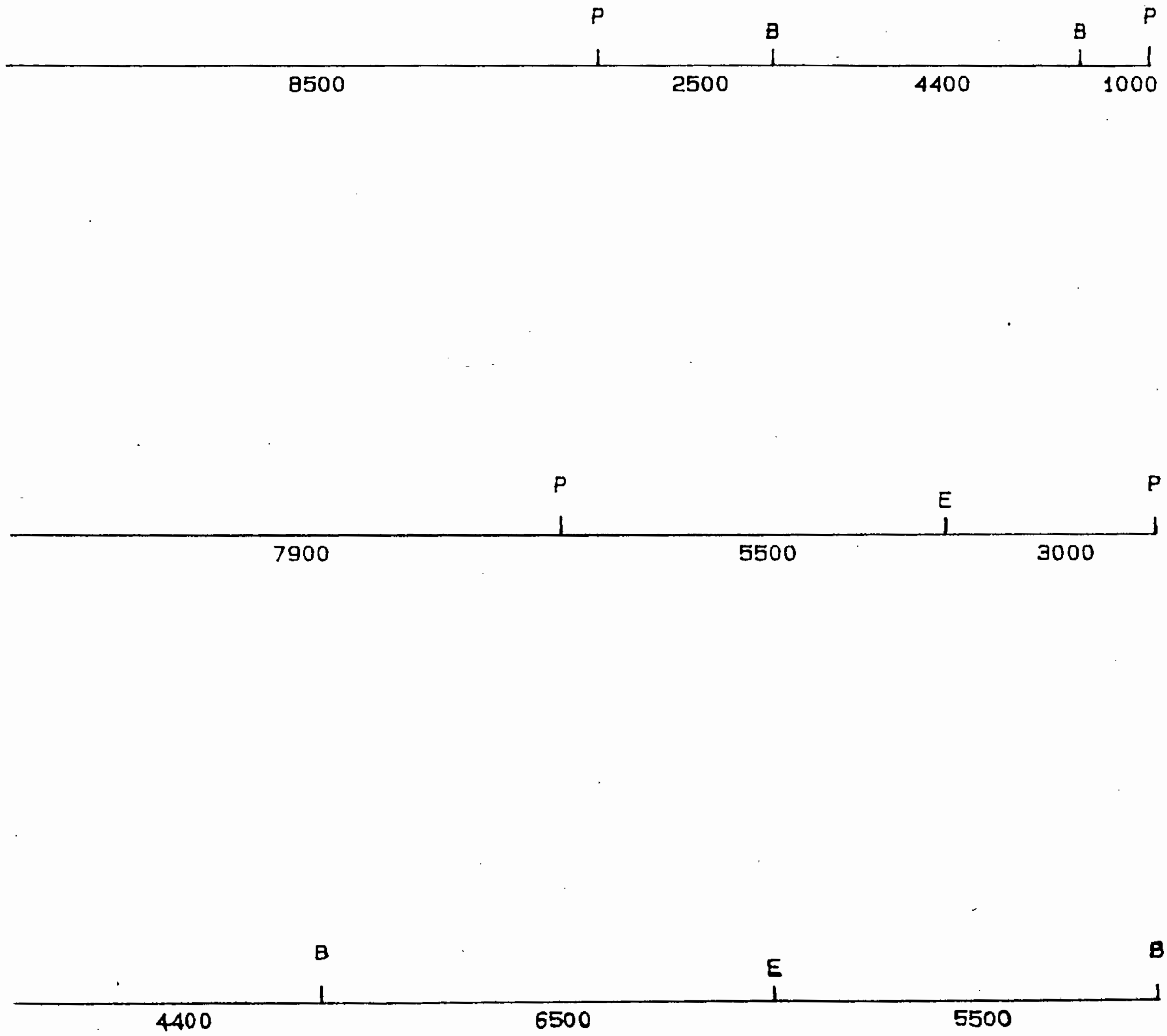


Fig. 2.7 **Final Restriction map of the Test Sequence for BamHI, PstI and EcoRI (error = 4%)**