SYSTEMATIC RELATIONSHIPS IN SOUTHERN AFRICAN FRANCOLINS AS DETERMINED FROM MITOCHONDRIAL DNA

by

Mariola Barbara Jakutowicz

Submitted in fulfilment of the requirements for the degree of MASTER OF SCIENCE (MED) in the department of Chemical Pathology in the Faculty of Medicine at the University of Cape Town.

June, 1991

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

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

Signed by candidate

23 April 1991

(Date)

Lugete, o Veneres Cupidinesque,
et quantum est hominum venustiorum.
passer mortuus est...
qui nunc it per iter tenebricosum
illuc, unde negant redire quemquam.
at vobis male sit, malae tenebrae
Orci, quae omnia bella devoratis;
tam bellum mihi passerem abstulistis.
vae factum male! vae miselle passer!

(Catullus III)

κακοῦ κόρακος κακὰ ᾠά

(Greek Proverb)

ACKNOWLEDGEMENTS

I wish to thank Professor E.H. Harley for his guidance and valuable criticism in the supervision of this project, and for the marvellous computer programmes of his making, which have been vital to the progress of this study.

Thanks also to Assoc. Professor T.M. Crowe of the Fitzpatrick Institute of African Ornithology for initiating the project and for the use of the Francolin macrocharacter data used to complement the results of this mitochondrial DNA analysis.

Sincere thanks to my colleagues in the laboratory, especially Colleen O'Ryan, for their education in the way of techniques and trouble-shooting, and for their support, both academic and emotional.

Thanks are extended to Dave Woolley for assistance with photography, and Daryl-Anne Freeman for her help in the typing of this dissertation.

Special thanks go to my friends and to my family, especially my parents, for their help and support.

The financial support of the CSIR is gratefully acknowledged.

ABSTRACT

The Francolins constitute the largest genus in the Galliform family Phasianidae. There is little accord concerning the taxonomic classification of its members. In the past, information on this group has been provided by morphological and palaeontological evidence.

An investigation into the molecular history of this group is presented, using mitochondrial DNA (mtDNA) as an evolutionary tool. A comparison of mtDNA restriction fragment lengths has been used to help define the phylogenetic relationships between 13 southern African Francolin species and a selected outgroup, the Japanese Quail.

Both cladistic and distance-based analytical methods have been used to construct phylogenies from the molecular fragment data. The trees relating the Francolins are in general agreement with the traditional classification based on morphological, behavioural and morphometric studies, but differ in the branching order of two species, F.levaillantii and F.hartlaubi. A recent proposal for the partitioning of the genus into two monophyletic assemblages of quail-like "partridges" and pheasant-like "francolins" is supported by mtDNA fragment data, with the exception of the two aberrant taxa. On the basis of the initial fragment size comparison, F.hartlaubi and F.levaillantii constitute part of an unresolved quadrichotomy at the base of the tree.

A restriction endonuclease site mapping approach has been utilized to provide a deeper resolution for the molecular phylogeny. Detailed mtDNA restriction endonuclease maps of F.levaillantii, F.hartlaubi, two species representing the "partridge" and "francolin" monophyletic groups respectively, and also of the Madagascar Partridge, have been constructed.

Phylogenetic analysis of this data has helped to resolve the problematic placement of the two aberrant taxa by showing an early separation of F.levaillantii from the "partridge" lineage, and of F.hartlaubi from the "francolin" lineage. The Madagascar Partridge was anticipated to be a likely sister-taxon to the whole group, but instead appears to have close relationships within the "partridge" lineage.

LIST OF ABBREVIATIONS

1. Nucleic Acid Terms

dATP (A)	Deoxyadenosine triphosphate
dCTP (C)	Deoxycytidine triphosphate
dGTP (G)	Deoxyguanosine triphosphate
dTTP (T)	Deoxythymidine triphosphate
^{32}P-dCTP	Cytidine triphosphate radioactively labelled with Phosphorus-32 (a position)
kb	kilobase pair
bp	base pair
DNA	Deoxyribonucleic acid
mtDNA	Mitochondrial deoxyribonucleic acid
RNA	Ribonucleic acid
tRNA	Transfer ribonucleic acid
rRNA	Ribosomal ribonucleic acid
scnDNA	Single copy nuclear DNA
DNAse	Deoxyribonuclease
RFLP	Restriction fragment length polymorphism
D-loop	Displacement loop
PCR	Polymerase chain reaction

2. Units

k, m, μ, n	kilo-, milli-, micro-, nano- (prefixes)
m, l, g	metre, litre, gram
Ci	Curie
°C	Degrees celsius
mol	moles
M	Molar
N	Normal
g	Centrifugal force
V	Volt

3. Chemicals

EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium Bromide
EtOH	Ethanol
SDS	Sodium Dodecyl Sulphate
SSC	Sodium Chloride - Sodium Citrate buffer
Tris	Tris (hydroxymethyl) amino methane
TE	Tris EDTA buffer
STE	Saline Tris EDTA buffer
HCl	Hydrochloric acid
NaCl	Sodium Chloride
H_2O	water
NaOH	Sodium Hydroxide
KGB	Potassium Glutamate buffer
BSA	Bovine Serum Albumin
CsCl	Cesium Chloride
TAE	Tris Acetate EDTA buffer
NaPP	Sodium Pyrophosphate

4. Species Names

Cot	Coturnix coturnix japonica
SW	Francolinus swainsonii (Swainson's Francolin)
LU	F.leucoscepus (Yellow-Necked Francolin)
RN	F.afer (Red-Necked Francolin)
AD	F.adspersus (Red-Billed Francolin)
CA	F.capensis (Cape Francolin)
NT	F.natalensis (Natal Francolin)
HT	F.hartlaubi (Hartlaub's Francolin)
OR	F.levaillantoides (Orange River Partridge)
SH	F.shelleyi (Shelley's Partridge)
GW	F.africanus (Greywing Partridge)
RW	F.levaillantii (Redwing Partridge)
CR	F.sephaena (Crested Partridge)
CQ	F.coqui (Coqui Partridge)
MP	Margaroperdix madagascariensis (Madagascar Partridge)

5. Miscellaneous

min	minutes
rpm	revolutions per minute
dpm	disintegrations per minute
Tm	melting temperature at which duplex DNA becomes single stranded
RT	room temperature
UV	ultraviolet
w/v	weight in volume
v/v	volume in volume
RE	restriction enzymes
Ma	million years
y.b.p.	years before present

TABLE OF CONTENTS Page

SECTION C: METHODS FOR INFERENCE OF PHYLOGENY

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GENERAL INTRODUCTION

SECTION A: PHYLOGENETIC HISTORY OF THE FRANCOLINS

The phylogenetic relationships among the Galliformes (Class Aves) are a matter of much debate, and there is even less accord concerning relationships among the members of the family Phasianidae. Its diverse members include the Guineafowl, Jungle Fowl, Peafowl, Turkeys, Grouse, Pheasants, Partridge, Francolins and New and Old World Quail.

One of the main reasons for this is that, despite their highly divergent external morphology and behavioural characteristics, these taxa all exhibit a remarkably uniform skeletal anatomy (Crowe et al., 1990, submitted). Consequently, the Phasianidae comprise a plethora of small and frequently monotypic and bitypic genera of uncertain phylogenetic relationships. The single exception to this is Francolinus.

The Francolins constitute the largest genus in the order Galliformes (having 41 recognized species) and form one of the largest genera in the Class Aves (Bock and Farrand, 1980). These sedentary and rather unremarkable gamebirds occupy a very diverse range of habitats and are extensively hunted for food and sport. However, relatively little is

known about Francolinus, as few species have been studied in detail. Indeed, the phylogenetic interrelationships of the Francolins have long been a matter of debate (Verheyen, 1956; Cracraft, 1981; Sibley and Ahlquist, 1985; Johnsgard, 1986; Crowe, 1988).

In her study of speciation in Francolins, Hall (1963) acknowledged 41 species (36 African and 5 in Asia) and fused these morphologically, ecologically and behaviourally diverse birds into a single monophyletic group. From this, she assigned all but 4 species to 8 sub-monophyletic groups, 7 of which are represented in Africa (Fig.1). She further proposed an Asian origin for the genus, during the Oligocene (ca. 25-35 x 10^6 y.b.p.) and hypothesized that extant species of African Francolins evolved as recently as 10^4-10^5 y.b.p. DNA-DNA Hybridization studies by Sibley and Ahlquist (1985) substantiated Hall's Oligocene origin for Francolinus, but did not support her suggestion of a late Pleistocene origin for extant Francolins. Instead, Sibley and Ahlquist estimated an approximate divergence time of 9 x 10^6 y.b.p. for F.capensis from F.natalensis.

Milstein and Wolff (1987) agreed with Hall's phylogenetic designation of the Francolins, but argued for the partition of the genus into two major clades comprising "partridges"

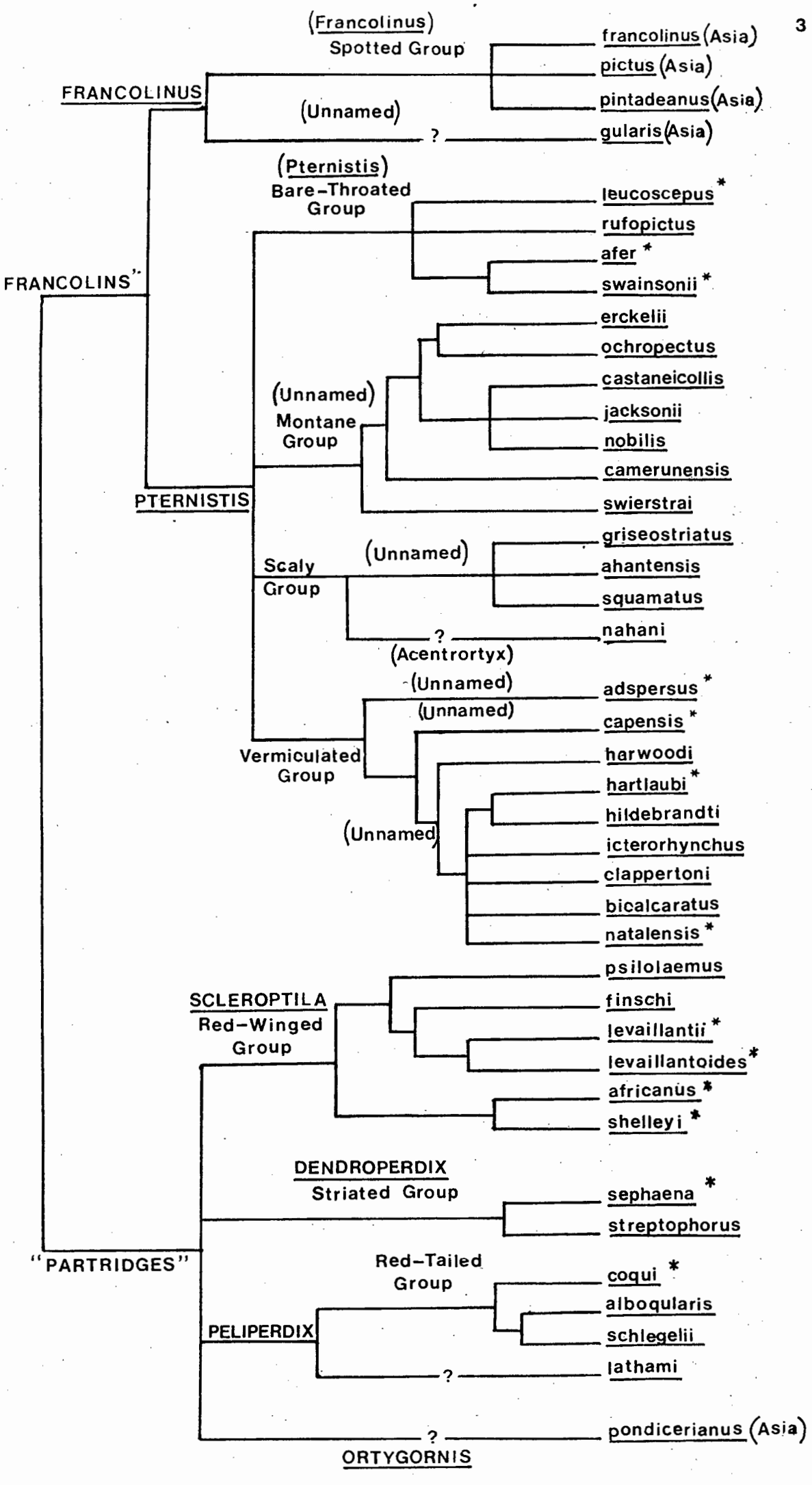


Fig 1 Phylogenetic relationships of Francolinus spp according to Hall (1963). Group common names follow Hall (1963), the "PARTRIDGE" - "FRANCOLIN" dichotomy follows Milstein and Wolff (1987). Scientific names follow Wolters (1975). MtDNA was studied from taxa marked with asterisks.

and "francolins". Their proposal was based on several observed differences between the two groups in their general behaviour, calls of adults, chick incubation periods, spur development on the legs, natal down (dorsal plumage of the downy young) and sexual hybridization.

Thus, the "partridges" (Red-Winged, Red-Tailed and Striated Groups in Fig.1 (see also Fig. 2(a)) are small, ground-roosting birds with striped, quail-like dorsal plumage, and give clear, tonal whistles.
By contrast, the "francolins" are larger in size and roost in trees, tending to utter harsh, atonal calls. They have dark dorsal plumage, vermiculated with white or buff. They comprise the Spotted, Bare-Throated, Montane, Scaly and Vermiculated Groups in Fig.1 (see also Fig. 2(b)).

Crowe and Crowe (1985), using morphological data, were unable to support Hall's monophyly of the Francolins. They did however accept the monophyly of her Spotted, Red-Winged, Red-Tailed, Bare-Throated and Montane Groups, but rejected the monophyly of her Scaly, Vermiculated and Striated Groups, further proposing a system of subgenera similar to that of Wolters (1975) - who divided the genus into 6 genera and 9 subgenera (some of which are unnamed; see Fig. 1).
Moreover, Crowe and Crowe did not acknowledge Milstein and Wolff's "partridge" - "francolin" dichotomy, instead

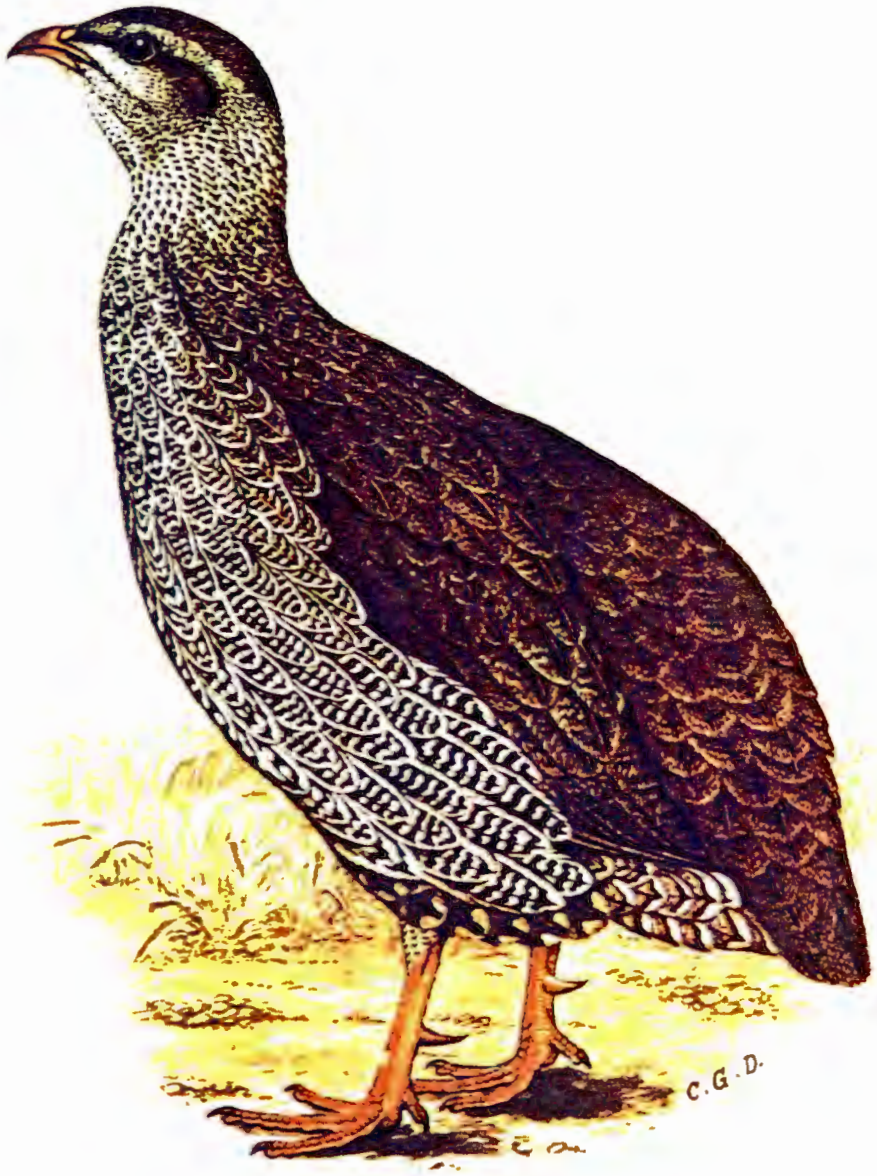


Fig. 2(a) A typical representative of the "francolins" (Milstein and Wolff, 1987), Francolinus natalensis, commonly known as the Natal Francolin. Also a member of Hall's (1963) Vermiculated Group. (Taken from *A Guide to the Terrestrial Gamebirds of the Transvaal*, Transvaal Provincial Administration, Nature Conservation Division.

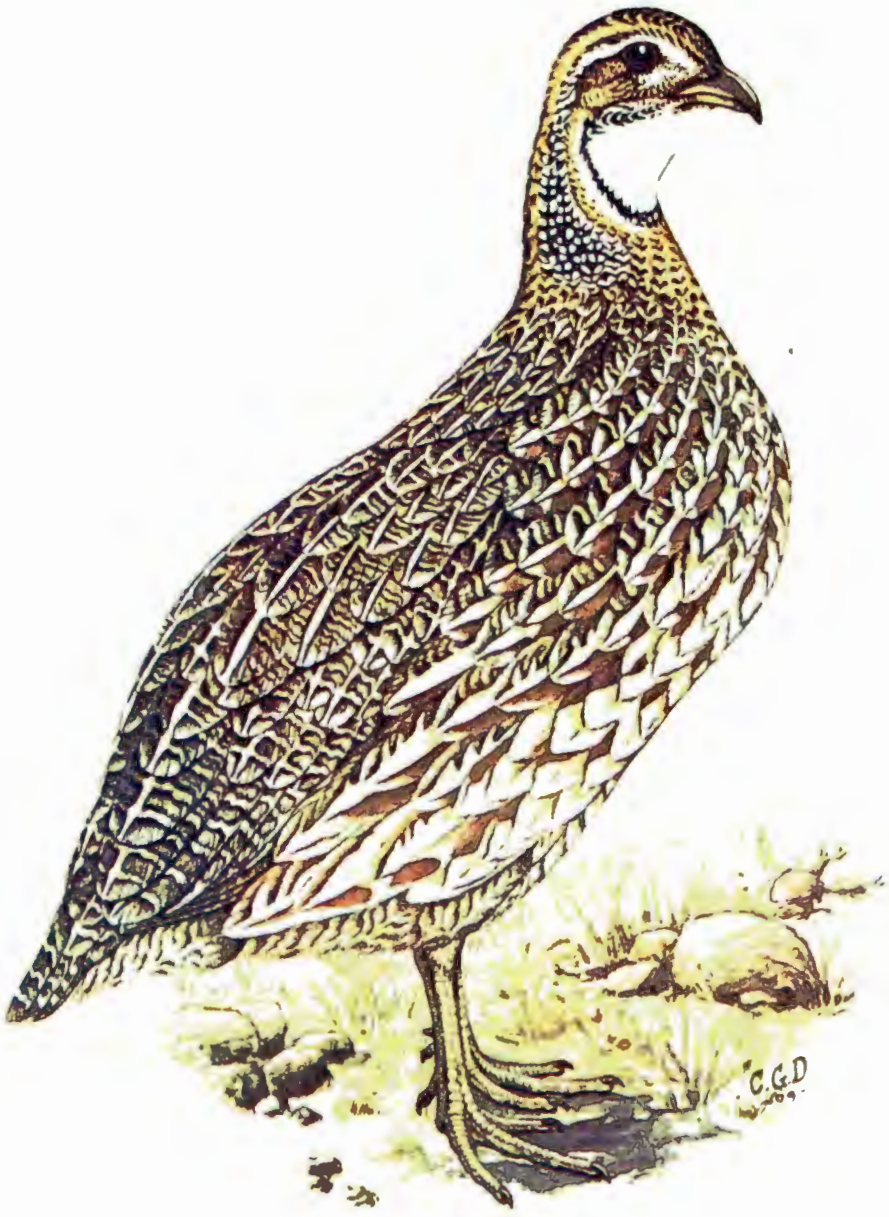


Fig. 2(b) A typical representative of the "partridges" (Milstein and Wolff, 1987), Francolinus levaillantoides, commonly known as the Orange River Partridge. Also a member of Hall's (1963) Red-Winged Group. (Taken from *A Guide to the Terrestrial Gamebirds of the Transvaal*, Transvaal Provincial Administration, Nature Conservation Division.

suggesting that the "partridges" are a paraphyletic assemblage.

On the basis of morphological and ontogenetic studies, Crowe and Crowe (1985) hypothesized for the Francolins a small, quail-like migratory Phasianid as ancestor. In fact, they speculated that the ancestral Francolin most probably resembled the Common quail (Coturnix coturnix).

Thus far, our knowledge of Francolin taxonomy is derived primarily from morphological, behavioural and ecological evidence. Molecular evidence could provide important additional insight into, or better still, resolve the phylogeny of this group.

Mitochondrial DNA was deemed an appropriate molecular variable to investigate, on account of its rapid rate of evolution and simplicity of preparation and analysis (see Section B below and Chapter 2).

SECTION B: MOLECULAR APPROACHES IN SYSTEMATICS

The last decade has brought with it spectacular progress in the field of systematic and evolutionary biology. This has become possible with the advent of sophisticated biochemical techniques, particularly those at the DNA level. Phylogenetic trees, until the 1960's, were based primarily on the analyses of behavioural and morphological characteristics. Although these approaches are still utilized (with increasing sophistication), the study of biological macromolecules has provided additional insight into the evolutionary changes of genes and populations.
A diversity of molecular techniques is currently available, offering varying degrees of resolution along a taxonomic hierarchy. Molecular applications to systematics utilize either protein or nucleic acid data.

1.1 Protein-based Approaches

1.1.1 Immunological and Amino Acid Sequence Approaches

The earliest protein methods, using immunological techniques, were pioneered by Nuttal (1904) and the approach achieved major recognition in the 1960's with studies of the relationships and times of divergence among hominids (Goodman, 1961; Sarich and Wilson, 1966).

Immunological methods entail the production of antibodies against an antigen (X_A, the homologous antigen) from a particular species (A). These can then be used to test relative cross-reactivities of that antigen to other antigens ($X_{B,C,D}$, the heterologous antigens) from a series of related species (B,C,D). The degree of immunological cross-reaction obtained with the heterologous antigens, relative to that obtained with the homologous antigen can be used as a general measure of the genetic relationship between species. A variety of immunological techniques exists. These differ primarily in the way antisera are produced and the means by which the immunological cross-reactions are measured. Whether this measure is an estimate of sequence difference or is a measure of the binding avidity of the dominant antigenic site, must be evaluated for each technique.

The most powerful quantitative immunological assay, which provides an estimate of the relationship between immunological reactivity and sequence differences, is that of Microcomplement Fixation (MC'F). This measures reactions between a soluble antigen and antibodies in dilute solution under conditions in which only high-affinity antibodies react. Antisera of broad specificity are primed to a single purified antigen, and are used to estimate the number of unmodified antigenic sites and thereby estimate the sequence differences between homologous proteins from different

species. Homologous reactions are standardized and those with heterologous antigens are measured relative to the homologous reaction. Complement measures the amount of antibody that is bound to antigen, and it assays this reaction by lysing sensitized sheep red blood cells. The fixation of complement requires tightly bound antibody and antigen. In this way, antigens from different species are compared and the results converted to units of immunological distance.

MC'F has been applied to the measurement of divergence (rate of change) in amino acid sequence in serum albumins (for example, see Benjamin et al., 1984; Collier and O'Brien, 1985; Maxson and Maxson, 1986).
Using the albumin immunological distance approach, Collier and O'Brien (1985) estimated the phylogenetic distances between 34 of the 37 extant species of Felidae. There appears to be a proportionate relationship between immunological distance and evolutionary time, which forms the basis of an Albumin Molecular Clock hypothesis. Such a clock has been reported for carnivores and primates (Sarich, 1969) and has been calibrated in a number of taxa (Thorpe, 1982; Wilson et al., 1977).

Direct comparisons of amino acid sequences have also been extensively used in establishing phylogeny (Goodman et al., 1987). However, this technique, like all other protein

methods, provides only an indirect measure of mutations at the gene level, i.e. the proteins isolated are removed from the actual information source (DNA). The proteins would also not be indicative of base changes that are neutral - those which have no significant effect on the phenotype. Neutral changes include those at the third codon position (wobble position), which often do not alter the amino acid coded for by the nucleotide triplet, as well as certain changes that occur in non-coding regions, for example, in transcription control regions, introns and spacer regions.

1.1.2 Protein Electrophoretic Approaches

Protein electrophoresis, which began modestly with starch gel electrophoresis in the 1950's (Smithies, 1955) is a widely used technique in molecular systematics today. The classic studies of Harris, and of Hubby and Lewontin in the 1960's contributed greatly to our understanding of molecular evolutionary processes (Harris, 1966; Hubby and Lewontin, 1966). Protein data can be obtained from either the electrophoresis of isozymes (functionally similar forms of enzymes), or allozyme electrophoresis (a subset of isozymes which are variants of polypeptides representing different allelic alternatives of the same gene locus). Although protein electrophoresis detects only a fraction of amino acid changes in proteins (as already noted, all base changes do not necessarily result in amino acid changes), the

technique has established that most natural populations have a high degree of genetic variation at the protein level (Lewontin, 1974).

Protein methods have often been utilized in intraspecific questions regarding population structure (Mihok et al., 1983; Bonell and Selander, 1974). They have supplied much information on inbreeding, outcrossing and dispersal of populations (using allozymes - Ryman and Utter, 1987), paternity studies (Tilley and Hansman, 1976) and species boundaries (Wake and Larson, 1987). Interspecific applications (i.e. phylogenetic systematics) involve the use of allozymes, and to a lesser extent, isozyme data.

Johnson and co-workers (1984) provided a comprehensive phylogeny for birds using allozymes, and Gutierrez et al. (1983) have also used allozyme analysis to evaluate levels and patterns of genetic differentiation among 10 species of galliform birds in the Phasianidae and Tetraonidae. Studies on passerines by Barrowclough and Corbin (1978) revealed that these birds possess considerably lower levels of genetic (allozymic) differentiation than other vertebrate taxa, at comparable taxonomic levels. However, the reasons for this remain unclear (Avise et al., 1980).

Isozyme electrophoresis is still widely used for studies of mating systems, population structure and heterozygosity estimates, which require analysis of many individuals at

many loci and therefore are ideally suited to isozyme techniques.

Protein electrophoresis has its limitations. For example, there is often a problem in finding a sufficient number of measureable changes. Standard laboratory routines measuring dozens of enzymes will only detect mutations that change the overall electric charge of these enzymes. These represent about half of the mutations in a miniscule portion of the coding sequences, which themselves constitute less than 10% of the genome. A further complication lies in the suggestion that birds exhibit an exceptionally high degree of enzyme conservation (Avise and Aquadro, 1982).

The nuclear alleles which encode allozymes segregate and recombine with sexual reproduction, thus continually recreating genotypes. Back mutations can give rise to homoplasies (convergences). Therefore, it may be difficult to link together these nuclear genotypes in an altogether convincing phylogenetic tree. Methods utilizing mitochondrial DNA such as Restriction Fragment Length Polymorphism (RFLP) studies provide an alternative approach (see Chapter 3). Analysis of mitochondrial DNA results in trees in which mutational changes separating adjacent genotypes may easily be counted.

1.2 Nucleic Acid-based Approaches

Perhaps the most significant and exciting molecular advancement in phylogenetic analysis in the last ten years has been that of the manipulation and analysis of the nucleic acid, DNA. Techniques include DNA-DNA hybridization, the use of restriction endonucleases (RE's) to locate nucleotide mutations and rearrangements, and the direct sequencing of DNA.

DNA-DNA hybridization is a quantitative assessment of the relatedness of biological species using nuclear DNA. Initially developed in the late 1960's/early 1970's (Wetmur and Davidson, 1968; Kohne, 1970), the technique was widely applied by Sibley and Ahlquist in their work on birds and primates (Sibley and Ahlquist, 1983, 1987b). Maximum efficiency was achieved through automation in the form of a thermal elution device, the DNAnalyzer.

The procedure entails the isolation of "tracer" DNA from a species to be tested, by the shearing of nuclear DNA into approximately 500bp fragments, and then denaturation into single-stranded DNA. Multiple repeat sequences do not evolve at the same rate as single copy nuclear DNA and will hybridize rapidly because of their high copy number. For this reason they are removed by a preliminary hybridization step. Denatured tracer DNA is radioactively labelled, added to an excess of single-stranded "driver" DNA (from the second species in the comparison) and the mixture is allowed

to anneal. The duplex thus formed is subjected to a thermal gradient (60°C - 90°C) which progressively separates the heteroduplex hybrid DNA fragments.

Theoretically, if tracer and driver DNA's are from the same species, 100% hybridization should occur, and it would require a high temperature to melt the duplex. In genetically diverged species, however, hybridization will not be as strong, due to base pair mismatch and a lower temperature would be required to release tracer fragments. The decrease in reassociation temperature of a DNA mixture of two animals is measured from plots of the melting curves of homo- and heteroduplexes.

The technique employs several parameters which are derived from differences between these melting curves. These are T_m, T_{mode}, NPH (Normalized Percentage of Hybridization) and $T_{50}H$, which can then be used in phylogenetic analyses (T_{50} is the reference temperature at which duplex melting is 50% complete).

Much controversy surrounds the use of DNA-DNA hybridization in molecular phylogenetics (Lewin, 1988a, 1988b; Hillis and Moritz, 1990). The first criticism levelled at the technique, is its failure to identify individual characters such as nucleotides. Individual shared derived characters are not isolated and described, and therefore cladistic approaches cannot be used. Proponents of the method believe

that the number of nucleotides under comparison is compensation enough for this shortcoming. Accuracy of this approach has been widely questioned: the large number of fragments arising in comparisons between species effectively suppresses distance fluctuation due to sampling. Furthermore, several sources of error, such as incomplete removal of multiple repeat sequences, could give rise to erroneous melting curves, and thus ambiguous distance information. The choice of statistical analysis of raw data provides another point of contention: which of T_M, $T_{50}H$ and T_{mode} are the appropriate measures to use?

Insertions, deletions and transpositions of sequences within the compared strands will also greatly affect the accuracy of the technique, i.e. they would affect the kinetics of interaction between different DNA molecules. In practice, initial duplex formation, even with DNA of identical species, is never complete. This fraction of unhybridized DNA increases as genetic distance increases. Does it contain valuable phylogenetic information or not, and should it be used in statistical measurements or not? The debate continues.

Despite these criticisms, DNA-DNA hybridization has been successfully used in several studies: (Sheldon, 1987; Springer and Krajewski, 1989; Caccone and Powell, 1987) Sibley and Ahlquist have generated a very impressive and

comprehensive phylogeny of the class Aves (Sibley and Ahlquist, 1983, 1985). Britten et al. (1978) in their study of the sea urchin, Strongylocentrotus purpuratus, used hybridization techniques to assess interindividual and interpopulational sequence divergence and variation.
Accuracy of the technique can be improved by closer attention to technical details, and automation has greatly enhanced precision. DNA-DNA hybridization does have an advantage in accounting for historically informative characteristics in the DNA sequence that are not always expressed phenotypically (i.e. as proteins), e.g. pseudogenes and regulatory sequences.

1.3 Mitochondrial DNA

This small, covalently closed circular molecule has become the one of the most widely studied components of the eukaryotic genome, and plays an increasingly important role in evolutionary biology and population genetics. Several unique features, as described in the following subsections, make mitochondrial DNA a particularly attractive choice for molecular evolutionary studies.
One major practical advantage is that mitochondria yield relatively large amounts of DNA for analysis. A number of factors contribute to its easy isolation and purification; its occurrence in an organelle other than the nucleus, its high copy number, and the buoyant density it exhibits when

mixed with ethidium bromide in an isopycnic density gradient (see Chapter 2, Section 2.1.4.1). A typical yield is approximately 1μg per gram of liver or heart tissue.

By contrast, the purification of homologous nuclear DNA sequences has typically required the laborious construction and screening of genomic libraries for each individual species under investigation. This process has recently been greatly facilitated by the development of the Polymerase Chain Reaction (P.C.R.) for isolating specific DNA fragments.

1.3.1 Genetic Composition of Mitochondrial DNA

Gene order in mitochondrial DNA (mtDNA) is highly conserved among the vertebrates, with a very stable arrangement of 37 genes in its compact genome. Mitochondrial DNA size in animals ranges within and between species from approximately 14 kilobasepairs (kb) to 30kb (Moritz et al., 1987), although in plants and fungi, it is much larger and more complex in organisation. Galliforms appear to possess a mitochondrial genome of between 16.3 and 16.4kb (Shields and Helm-Bychowski, 1988).

The vertebrate mitochondrial genome comprises two rRNA genes, 22 tRNA genes and 13 genes coding for polypeptides (Anderson et al., 1981; see also Fig. 3). Animal mtDNA also lacks many of the complicating features of nuclear DNA, such as introns and intergenic sequences. A "control" region is

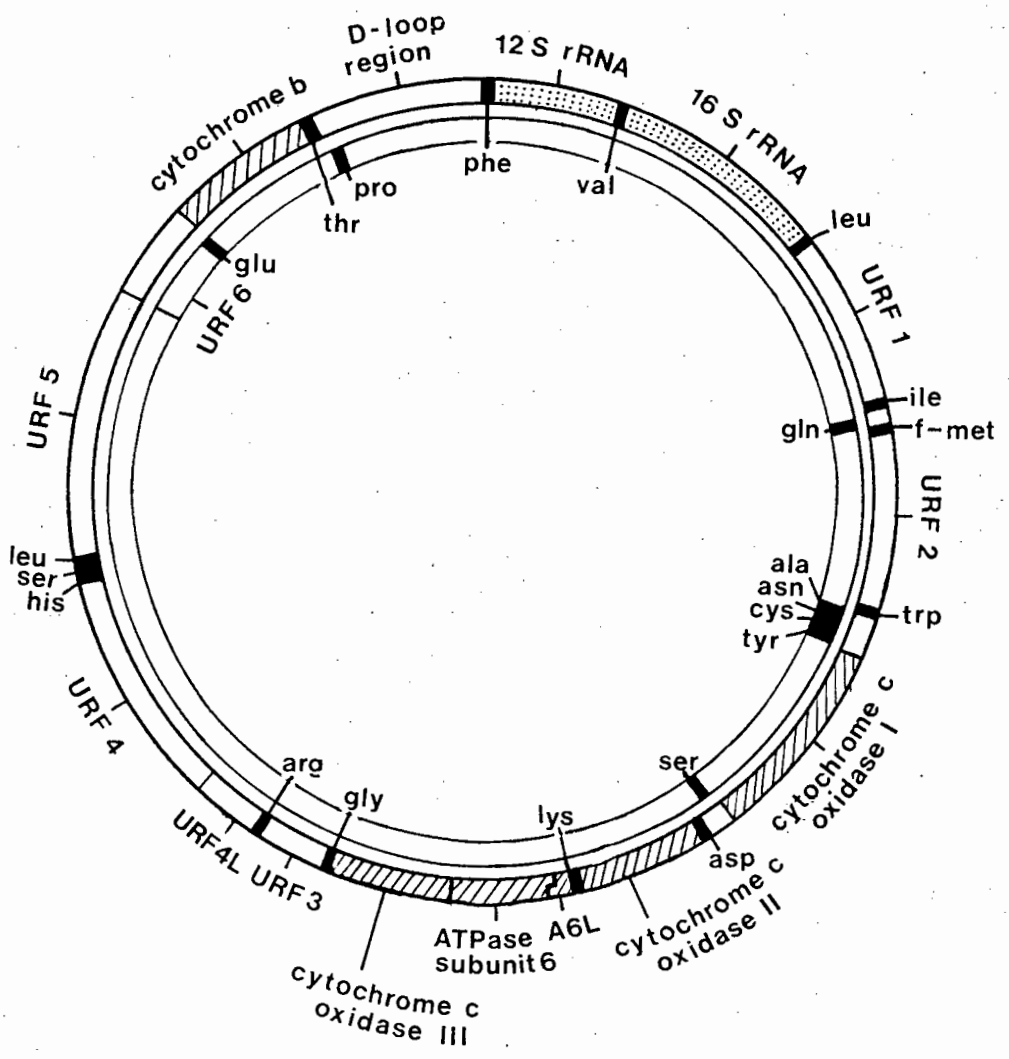


Fig. 3 Human Mitochondrial Genome (Taken from Grivell, 1983). The DNA is shown as concentric double circles (one circle for each strand of the double helix). Diagonally striped segments show the extent of genes for known proteins. Dotted segments show the genes for ribosomal RNA. The genes for transfer RNAs are shaded black, each of which is labelled with the abbreviation for its specific amino acid. Clear segments indicate unassigned reading frames (U.R.F.) which are presumably genes for proteins that have not been identified.

present, which lacks structural genes, but does contain certain sequences for the initiation of replication and transcription. Echinoids (Matsumoto et al., 1974) and vertebrates (Kasamatsu et al., 1971) have within their mitochondrial control region a displacement loop (D-loop) which appears to function in replication. Glaus (1980) positioned the D-loop at about 700-900 base pairs (bp) from the unique Xba I site in chicken mtDNA and postulated it to have a double-stranded size of approximately 700bp. Greenberg et al. (1983) found two "hot spots" for divergence within the D-loop and propose that the extent of divergence within this region is at least ten times higher than for the mtDNA as a whole.

1.3.2 Maternal Inheritance and Heteroplasmy

MtDNA is haploid and maternally inherited (Lansman et al., 1983a; Gyllensten et al., 1985). Recombination is therefore absent, and this gives evolutionary biologists access to clearly defined maternal genealogies. There appears to be no contribution of mtDNA from sperm cells to the next generation. Gyllensten et al. (1985) crossbred two species of mouse and found that multiple generations of back-crossing of females to males of the paternal parental type gave no evidence of paternal contribution.

Mitochondria occasionally exhibit the phenomenon of heteroplasmy - where two or more mitochondrial genotypes coexist within an individual. Heteroplasmic states, though rare in mammals, are believed to arise from mutations (e.g. length differences or point mutations) within a cell line, rather than from paternal contribution of mtDNA from the sperm cell (Lansman et al., 1983a; Gyllensten et al., 1985). Heteroplasmy appears to be a transitory state, due to rapid segregation of mitochondrial genotypes in germcell lineages (Rand and Harrison, 1956; Ashley et al., 1989). The extent of heteroplasmy is largely undetermined in birds - the first instance of avian heteroplasmy was reported by Avise and Zink (1988) in their study on rails.

1.3.3 Rate of Sequence Divergence

Because of the crucial dependence of animals on mitochondrial functions, one might expect the rate of mitochondrial evolution to be very slow, with a highly conserved sequence. In primates, however, this is not the case; the rate of mitochondrial sequence divergence is rapid, exceeding that of single copy nuclear DNA (scn DNA) by 5-10 times (Brown et al., 1979). In their study, Brown et al. calculated the mean rate of divergence, averaged over the whole mtDNA molecule, to be about 2% per million years for primates. Helm-Bychowski (pHD thesis, 1984) arrived at a similar rate for gallinaceous birds. These estimates have

been calibrated both with fossil material and independent dating based on proteins. MtDNA has also been reported to evolve faster than nuclear DNA in rodents (Miyata et al., 1982) and Xenopus (Carr et al., 1987; Dawid, 1972), but the rates of mitochondrial and nuclear evolution appear to be similar in echinoids (Vawter and Brown, 1986) and fruit flies (Powell et al., 1986).

Evolutionary changes in mtDNA are primarily due to base substitutions, with transitions greatly outnumbering transversion events and length mutations.
In comparison to organelle genomes of fungi, protista and plants, mtDNA's of animals show relatively little length variation, with minor length differences (150bp) accumulating predominantly in the small non-coding sequences and in the control region, perhaps a result of duplication or deletion of sequences, consequent upon slippage during replication.

The control region changes rapidly both within and between species (Fauron and Wolstenholme, 1980) whereas the tRNA genes evolve slowly (Hixson and Brown, 1986). As already pointed out, the control region of human mtDNA has shown "hot spots" for base substitutions (Greenberg et al., 1983)

Taking into account the remarkable conservation of size, structure and gene order of mtDNA, its rapid rate of

evolution appears paradoxical. However, it has been found that almost 90% of the base substitutions occurring are neutral, the third codon positions being much more susceptible to change than the first two positions.

A plot of mtDNA nucleotide sequence divergence against time becomes curvilinear - the initial rapid rate of mtDNA sequence divergence (2% per Ma), due to such third-position codon changes, slows down considerably after about 8-10 million years, and eventually reaches a plateau. After reaching about 20% sequence divergence, differences in mtDNA accumulate much more slowly (Fig.4; Brown et al., 1979).

The initial rapid rate of mtDNA evolution can be attributed to any of the following: firstly, there is an apparent inefficiency of the mitochondrion (compared with the nucleus) in repair of DNA damage and replication error - this elevates the effective mutation rate (Wilson et al., 1985). Secondly, mtDNA does not appear to produce proteins involved directly in its own replication, transcription or translation mechanisms (Lansman and Clayton, 1975). As opposed to the translation apparatus in the cell sap, one would expect translational accuracy to be tolerable in the mitochondrion.

Lastly, mtDNA has a higher turnover rate than scn DNA in tissue; this increases the number of replication rounds in which errors could arise (Rabinowitz and Swift, 1970).

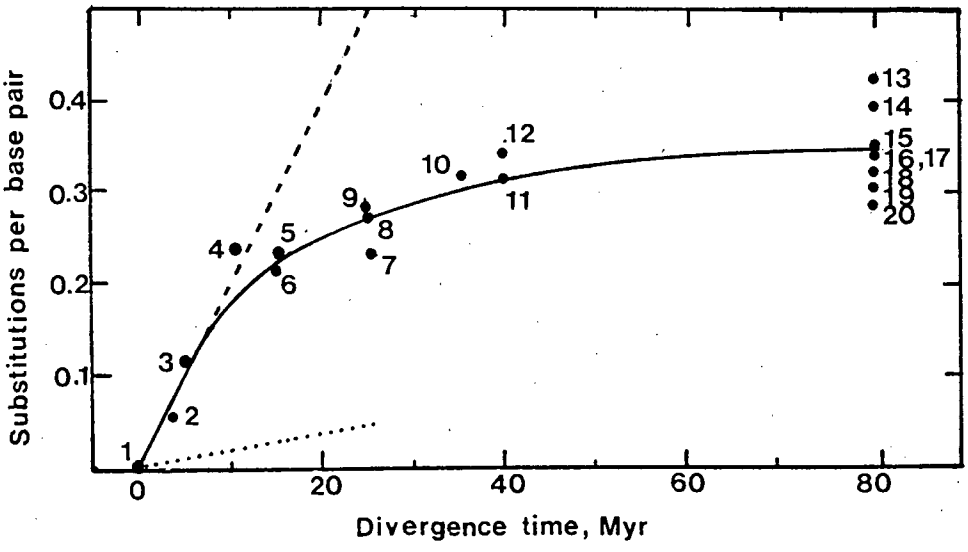


Fig. 4 Dependence of sequence divergence in mtDNA upon time of divergence (Taken from Brown et al., 1979) The y-axis shows the estimated number of base substitutions that have accumulated per base pair (p) for each species compared. This number is calculated from restriction map comparisons by use of equations 1 and 3 in Brown et al.(1979). The rate of substitution for mtDNA is obtained from the initial slope of the curve, indicated by the broken line. The rate for scn DNA is obtained from the slope of the dotted line. Each point on the graph corresponds to a comparison of 2 species and of individuals within a species. 1, mean difference among humans; 2, goat and sheep; 3, human and sheep; 4, baboon and rhesus; 5, guenon and baboon; 6, guenon and rhesus; 7, human and guenon; 8, human and rhesus; 9, human and baboon; 10, rat and mouse; 11, hamster and mouse; 12, hamster and rat; 13-20, rodent-primate species pairs. Both fossil and protein data were used to estimate the times of divergence.

1.3.4 Rate of Sequence Divergence in Birds

A controversy has arisen over the rate of molecular evolution in avian mtDNA; it was suggested that birds may have experienced a slower rate than other vertebrate groups (Britten, 1986). This allegation results from immunological data studies (Prager et al., 1974), DNA-DNA hybridization results (Sibley and Ahlquist, 1984) and mtDNA observations (Kessler and Avise, 1985).

1.3.5 MtDNA as a Tool

By virtue of its maternal inheritance and relatively rapid rate of evolution, mtDNA has become a widely used marker for studies of female-mediated gene flow, and the history of species, including hybrid taxa.
Cann et al. (1987), in their study on humans, have shown that mtDNA may provide valuable insight into historical patterns of migration and colonization, including founder events. They have concluded that all modern humans have an African origin, and suggest that each non-African population is the result of multiple colonization events.

Due to its considerable variation among individuals both within and between populations, mtDNA has proved to be an effective marker of population structure and patterns of

intraspecific geographic variation (Avise et al., 1987). In their review, Avise et al. show how mtDNA bridges the gap between population genetics and systematics: an analysis of the distribution of mtDNA genotypes has led to the geographic structuring of certain populations - In particular the horseshoe crab (Limulus polyphemus), the pocket gopher (Geomys pinetis) and the freshwater bowfin fish (Amia calva).

MtDNA can usefully be applied to certain problems in nature conservation. Wildlife managers wanting to establish the extent of genetic divergence within or between populations could use mtDNA for this purpose.

1.4 Restriction Endonuclease Analysis

Restriction endonuclease (RE) analysis has proved an efficient and rapid means of characterizing DNA and thus, in estimating phylogeny. MtDNA may be used in three types of DNA analyses: these assays for sequence variation provide discrete characters in the form of DNA fragments (see Chapter 3), restriction sites (see Chapter 4) or nucleotide sequences (see 1.5 below).

Brown and Vinograd (1974) were the first to utilize RE's in the analysis of sequence relatedness of different animal mtDNA's.

1.5 Nucleotide Sequencing

No doubt the most powerful DNA approach is that of nucleotide sequencing. It provides the most detailed and acccurate data for calculations of sequence divergence. However, it is also very time-consuming, and until recently, involved lengthy cloning procedures. The advent of PCR has rendered sequencing more practical. This technique amplifies specific DNA sequences without the need for cloning (Scharf et al., 1986; Mullis and Faloona, 1987), and these can then be sequenced by a variety of methods (Wrischnik et al., 1987). Sequencing provides an excellent way of investigating the mtDNA of endangered or extinct animals (Plante et al., 1987; Higuchi et al., 1987; Paabo, 1989). The technique could also be used in studies of intraspecific variation and population genetics, but as these require examination of large numbers of individuals, it may best be directed at resolving particularly difficult questions.

The most significant drawback to sequencing is that, when applied to DNA, nuclear or mitochondrial, it addresses only a tiny fraction of the genome, and the fragment sequenced may not be representative of the genome as a whole. A combination of the various DNA analyses available, therefore, would best be able to provide high resolution information, with high efficiency.

1.6 Application of Molecular Techniques to the Taxonomic Hierarchy

Analysis of mitochondrial DNA is best-suited for comparisons over the time scale of from 1 to 20 million years ago, and therefore provides a very good method of choice for comparisons at the subspecies, species or genus level. Closely-related species (diverged within the past 5 million years) can also be studied by examining relatively fast-evolving isozyme loci. Other techniques are generally not sensitive enough to detect sufficient changes over such a short time scale.

Nuclear DNA, on the other hand, is more appropriate for comparisons at the supra-generic level (i.e. families or orders), e.g. DNA-DNA hybridizaton is useful for covering the period of from 20-100 million years ago.

1.7 DNA Fingerprinting

The most effective approach for inferring relatedness between individuals is DNA-fingerprinting. This technique was developed to establish relationships between first degree relatives (Lynch, 1988) - for example, it is used in paternity testing and forensic analysis in humans. The technique can be applied to other mammals, birds, and most likely, to other eukaryotes (Vassart et al., 1987). It may

also be useful in determining the degree of inbreeding in populations (e.g., in foxes - Gilbert et al., 1990, and in house-sparrows - Wetton et al., 1987).

DNA fingerprinting relies on the existence of families of hypervariable minisatellite regions in nuclear DNA, each of which contains tandem arrays of short (10-60bp) repeat sequences. Slippage in replication gives rise to variation in the number of tandem repeats per array. Variable sequences may be demonstrated with the appropriate radioactive probe to give a pattern of multiple bands upon gel electrophoresis. In humans, with the exception of identical twins, each individual has a unique banding profile or DNA fingerprint (Jeffreys et al., 1985).

The application of DNA fingerprinting to population studies has a twofold advantage over isozyme analyses: the average number of alleles per locus is very much greater than in the case of enzymatic loci, and the markers for many loci are visible on a single fingerprinting gel. Thus the technique, although technically demanding and expensive, is quick and has the ability to provide much more information than isozyme analyses.

1.8 Conclusion

The diversity of molecular techniques available to evolutionary biologists is considerable - each of the methods discussed has been successfully applied to questions

of phylogeny, with the appropriate techniques varying from study to study. It should not be claimed that any of these is necessarily superior to the other for systematic purposes. Each molecular approach contributes differently towards establishing a tier in the taxonomic hierarchy.

SECTION C: METHODS FOR INFERENCE OF PHYLOGENY

1.1 Molecular Clock Controversy

One of the most controversial and significant concepts to emerge from the comparative study of molecules, is that of the "Molecular Clock". This relatively well-established tenet of molecular evolution holds that random mutations accumulate at a constant rate, subject only to stochastic variation. First introduced by Zückerkandl and Pauling in 1962, the model's strong correlation between estimates of sequence divergence and of divergence time, raised the possibility that molecular comparisons could provide divergence times for the extinct ancestors of extant species and monophyletic groupings.

Molecular and morphological evolution do not progress at a corresponding rate: a base substitution in DNA may give rise to a large phenotypic change, a small change, or none at all (neutral changes), depending on the location of the point mutation. In addition, whereas most morphological change is subject to selection, most of the measured DNA sequence changes are neutral. Unlike molecular evolution, therefore, the evolution of morphology would appear to progress rather irregularly.

Nonetheless, can the ticking of the clock be described as regular? It appears that molecular evolutionary rates are

far from being established as constant, and there is considerable rate variability among taxa (Wilson et al., 1977, 1985; Nei, 1987). Indeed, as previously noted, it should not be assumed that rates are equal (Gillespie, 1986a,b; Goodman, 1985; Britten, 1986). However, the greatest disparity in rates is between widely divergent taxa (at the class or phylum level) and it may well be true that rates within, say, the family level may be much more constant.

Although there appear to be no deviations from the 2% divergence rate for primates and gallinaceous birds, known for the mtDNA molecule as a whole, one particular region exhibits departure from the molecular metronome. This example of "accelerated co-evolution" (Wilson et al., 1985) occurs along a lineage leading to higher primates: the subunit II of cytochrome oxidase, which is mitochondrially encoded, has indicated a five-fold acceleration in rate of evolution at the amino acid sequence level (Brown and Simpson, 1982; Cann et al., 1984). Parallel to this is an acceleration in the evolutionary rate of nuclearly-encoded protein cytochrome C, which interacts directly with the oxidase subunit in the electron transport chain. The reason for such acceleration in rate remains a mystery. However, the rate at which the molecular clock ticks is not only irregular in different sectors of DNA, as seen above, but also along the mtDNA molecule itself, and from molecular

site to molecular site (Aquadro and Greenberg, 1983). Several subtle factors contribute to the latter type of perturbation of the clock - the probabilities that mutations at any nucleotide site will become fixed are very varied, and depend on what type the mutation is, whether it occurs in a coding region, and if so, which of the three bases in the codon are more resistant to change than others.

Also, processes within the mitochondrial genome itself play a role in making the clock tick at an irregular rate; unequal crossing-over events, gene conversion, slippage and transposition events have several effects on the steady accumulation of mutations - such processes can sometimes spread new sequences within a genome throughout a local population, or sometimes eliminate them (Lewin, 1988).

Even though a broad correlation between molecular divergence and time is accepted by most molecular biologists, these departures from clock-like behaviour should be borne in mind when intending to use mtDNA as an evolutionary dating-device. In general, molecular clocks are calibrated by dividing the average estimate of the age of the last common ancestor, by the average measure of molecular divergence.

Assumptions of a molecular clock have influenced the methods of phylogenetic analysis used to process molecular data. Nonetheless, such analyses can often proceed without this

strong an assumption; many recent molecular systematists have chosen analytical methods that are applicable to both the constant rate of a molecular clock, as well as to data with varying divergence rates (Avise et al., 1979; Hillis, 1985, Baverstock et al., 1979; Goodman et al., 1979).

1.2 The Neutralist-Selectionist Controversy

The constancy of rates is a basic assumption of Kimura's Neutral Theory of Molecular Evolution (Kimura, 1968, 1983) which hypothesizes that most evolutionary change and genetic polymorphism arise from neutral mutations. Another long-standing controversy is one between the Neutral Theory supporters and the Selectionists. The latter argue for natural selection as the basis of evolutionary change and polymorphism, rather than neutral nucleotide substitutions (Gillespie, 1984, 1986a, 1986b).

The Neutral Theory does not rule out natural selection against deleterious mutants, and it argues that most differences in the rate of evolution between different molecules and different parts of the genome are accounted for by conservation of biologically significant sequences. Although many protein, chromosome and DNA variants are undoubtedly acted upon by natural selection, selectionists have not, to date, been able to produce a general, testable, theory of molecular evolution based on selection, and the controversy thrives.

1.3 Methods for Inference of Phylogenies

In recent years, methods of tree construction for the inference of phylogenies have rapidly proliferated. As viewed by Felsenstein (1988), phylogeny inference can best be looked at from a statistical point of view - it can be seen as ".....making an estimate of an unknown quantity, in the presence of uncertainty, and using a probabilistic model of the evolutionary process". Each method of phylogenetic inference has its own constraints and limitations, and may be as acceptable as the next.

The two fundamental approaches for preparing molecular data for tree construction are the cladistic method and the method based on distance measures. The former presents either restriction fragments, restriction sites, or sequence data in the form of characters, coded simply by their presence or absence, or by the frequency with which each character appears. The distance approach sums similar molecular data into an overall distance index, using methods such as that of Nei and Li (1979). Again, a variety of analytical programmes for the inference of phylogenies from either approach is currently available.

1.3.1 Cladistic Approach

Once an input data set is obtained, any of three general cladistic methods may be followed: Maximum Parsimony (Felsenstein, 1983), Maximum Likelihood (Felsenstein, 1981) and Compatibility Analysis (Templeton, 1983a, 1983b). By far the most widely used of these is Maximum Parsimony. Cladistic methods, unlike distance-based approaches, give only the branching order of the tree and tend to avoid measurements of branch lengths. The number of character-state changes is sometimes indicated on trees, however. Cladistic methods can therefore not be calibrated using a molecular clock to estimate the time at which any pair of taxa diverged from a common ancestor.

1.3.1.1 Maximum Parsimony

The Parsimony criterion is one which constructs trees such that the branch lengths connecting taxa in a study are minimal, i.e. they represent the minimum number of mutations (transformations from one character state to another) required to explain the observed data. Algorithms have been developed to build such trees as well as to enhance the probability of approximating the best topologies (Farris, 1970; Fitch, 1971; Felsenstein, 1983). This is because a number of alternatives are possible, which increases exponentially as the number of taxa in the

comparison increases. First used in molecular studies by Eck and Dayhoff (1966), Maximum Parsimony has become popular in analyses of restriction enzyme data (Avise et al., 1979a, 1979b, 1983; Ferris et al., 1981, 1983b; Aquadro and Greenberg, 1983; Lansman et al., 1983b).

A tree based on this principle minimizes not only the total number of evolutionary steps, but also the number of additional steps (homoplasies such as convergences, parallelism or reversals) needed to explain the data.

The mathematical basis for Maximum Parsimony can be seen in the following equation (Swofford and Olsen, 1990):

From the set of all possible trees, find all trees τ' such that

$$L(\tau) = \sum_{k=1}^{B} \sum_{j=1}^{N} W_j = \text{diff}\,(x_{k'j} ; x_{k''j}) \text{ is minimal}$$

where B = number of branches

N = number of characters

k', k'' are the two nodes incident to each branch

$x_{k'j}$, $x_{k''j}$ represent either elements of the input data matrix, or optimal character state assignments made to internal modes.

diff (y,z) is a function specifying the cost of transformation from state y to state z along any branch.

The coefficient W_j assigns a weight to each character. W_j is typically set to 1, but this need not be the case.

The definition of "optimal character state assignments" may include restrictions on the nature of permissible character-state changes.

Parsimony analysis consists of a group of related methods, the simplest of which are those of Fitch Parsimony and Wagner Parsimony (Fitch, 1971; Kluge and Farris, 1969). These methods find the minimal tree length under certain constraints on permissible character state changes, and they are distinguishable from the algorithm that finds optimal trees (to be discussed in a subsequent section). In these Parsimony methods one would assume that every possible tree can be evaluated, and each of these trees optimized according to the chosen criterion (Wagner or Fitch Parsimony, for example), and then ranked according to that criterion. Fitch Parsimony imposes no restrictions and Wagner Parsimony, only minimal constraints upon permissible character state changes.

The Wagner method (Kluge and Farris, 1969; Farris, 1970) can be applied to binary characters (e.g., restriction fragment

or site data) and ordered or unordered multistate characters (e.g., nucleotide sequence data), and it assumes that any transformation from one character state to another, also implies a transformation through intervening states. This parsimony approach is employed in both Hennig 86 and PAUP programmes. Fitch Parsimony such as that utilized in PHYLIP dictates that any state may transform directly to any other state. Both methods assume that the change of character states is equally probable in either direction, i.e. may transform from one state to another, and back again. One result of this assumption of reversibility, is that the tree may be rooted at any point, with no change in the tree length. Phylogenetic trees so constructed are not based on assumptions of evolutionary rate.

Hennig 86 and PAUP are both relatively efficient in attaining solutions, although distance-based programmes have the advantage of a very short computing time. Computing efficiency for Hennig 86 and PAUP is dependent on the kind of data available, and the number of taxa in the comparison. Clear data, with little disagreement among the characters (i.e. having a consistency index near 1.0), facilitate rapid results. The consistency index is the ratio between the minimum number of steps required to produce a fully homologous tree, and the actual number. For example, in a study by Platnick (1988), both PAUP and Hennig 86 obtained exact answers for a 21-taxon data set (consistency index =

0.82) in less than 15 seconds. Unrooted trees are presented, but both methods have options which allow for rooting at any specific position, and rerooting does not alter the length of the tree. Each programme offers many combinations of options for improvisation of an initial tree selected from the best solutions, e.g., the systematic rearrangement of branches. Both Hennig 86 and PAUP allow weighting of characters (See Section 1.3.1.6 below) and can treat multistate characters as additive (where a change from one state to another requires an intermediate state), or non-additive (such that change is directly from one state to another).

1.3.1.2 Rooting of Trees

Maximum Parsimony networks are often unrooted, i.e. the taxa are connected without defining the ancestral position. Without a root, it can be said that mutations have occurred along a lineage, but that the order and direction of those evolutionary changes are not established.

1.3.1.3 Outgroup

The most reasonable method of rooting a tree from any systematic analysis, is to include an outgroup in the data set. Based on studies apart from those in the data set being used, the outgroup is generally a taxon found to be

related to the other taxa as a whole, but nonetheless distinct from them. Addition of the outgroup at a particular location on the tree implies a root with respect to the remaining taxa, thus giving a direction to the evolutionary change.

1.3.1.4 Dollo Parsimony

As already noted, Parsimony criteria like those of Wagner and Fitch operate under the assumption of reversibility, i.e. that a transformation of state "0" to state "1" is equivalent to a change from state "1" to state "0". In the case of restriction site data, this is probably an inappropriate assumption to make, since the probability of losing an RE site is much greater than that of gaining one; if a particular 6-bp sequence is only one substitution away from becoming an RE recognition sequence, there is only a 1 in 18 chance that any one base will substitute for the correct base to convert the sequence to an RE site (since there are 6 bases, with 3 possible alternative states each). This is based on the assumption that all base changes are random and each has an equal chance of mutation. However, if there is a constraint of a 90-100% transition bias (since the transition rate greatly exceeds that of transversion) then the ratio decreases to 1 in 6 (since each of the 6 bases can only change into a purine or a pyrimidine). Moreover, a constraint based on the increased susceptibility

to mutation of the third base in a codon (wobble position) would hone the probability of RE site gain down to 1 in 2 possible mutations!

On the other hand, if the 6-bp sequence is already an RE recognition site, then a base-substitution at any of the 6 positions would cause the site to be lost.

Such asymmetry in the probabilities of gaining and losing sites may make the Dollo Parsimony method (Le Quesne, 1974; Farris, 1977; Debry and Slade, 1985) appropriate for restriction enzyme data.

The method is one of Parsimony - requiring the minimum number of steps - but operates under the constraint that every derived character state be uniquely derived, i.e. homoplasies (such as parallel or convergent site gains) are not allowed. In the context of restriction site data, each site may be gained only once with any number of site losses occurring to explain the given data. Therefore, a derived character state cannot be lost and then regained.

Construction of the tree in Fig.5(a) (Swofford and Olsen, 1990) requires 2 site gains to explain the data given. This is unacceptable under the Dollo criterion, which would reconstruct the tree with a single gain, followed by 2 site losses (Fig.5(b)). As with Fitch and Wagner Parsimony, Dollo trees can be unrooted, but the inclusion of outgroup taxa is convenient for RE site characters; if a site is

present in the outgroup as well as in some of the remaining taxa, then the most recent common ancestor is assumed to have had the site, and Dollo analysis searches for loss of that site over the whole tree. Sites not found in the outgroup, but only in some of the other taxa are assumed to have been absent ancestrally and a single gain would be hypothesized at an optimal location with respect to the ingroup (taxa excluding the outgroup).

Swofford and Olsen (1990), however, advise that a strict application of Dollo Parsimony for site data is not entirely realistic - if, albeit unlikely, a particular RE site does arise independently in two different lineages (Fig.5(c)), then Dollo can drastically overestimate the number of evolutionary changes, due to its intolerance for the regaining of RE Sites (Fig.5(d)). One way to circumvent this problem is to adopt a more relaxed Dollo criterion, e.g., a situation with one gain, two losses might be preferable to two independent gains, but two independent gains might be more desirable than one gain, ten losses - such a system can be found in a "generalised parsimony" method.

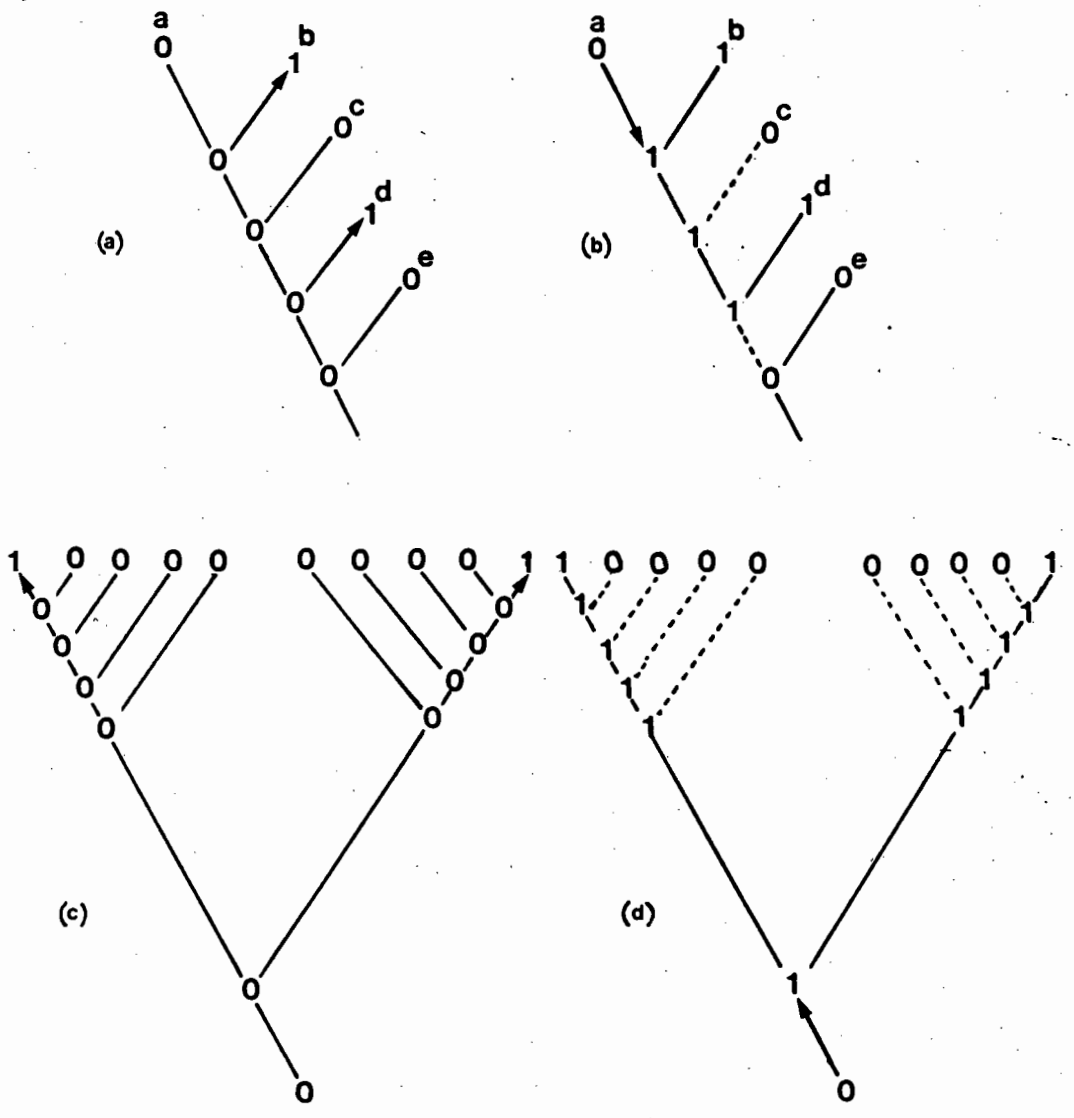


Fig. 5 Character state reconstruction demonstrating the Dollo Parsimony criterion (Taken from Swofford and Olsen, 1990). Branches on which site gains occur, indicated by arrows. Site losses are indicated on branches with broken lines.
(a) Most parsimonious reconstruction if multiple originations of state 1 are allowed. Two gains are indicated.
(b) Most parsimonious reconstruction under Dollo Parsimony, in which only single origination of state 1 is permitted. A single gain, followed by two losses is indicated.
(c) Demonstration of problems affecting Dollo Parsimony if multiple originations of the derived state occur. The "true" tree indicates two independent site gains (state 1).
(d) Reconstruction of (c) under Dollo Parsimony presents this tree which requires 11 steps (one site gain and 10 site losses, or reversal to the ancestral state 0).

1.3.1.5 Maximum Likelihood

The computation of Maximum Likelihood methods to find alternative trees is rather complex and time-consuming, and frequently involves taking products of many quantities of logarithms. This is perhaps the reason why these techniques are not very widely used. The approach was first studied by Cavalli-Sforza and Edwards (1967), using gene frequency as data, and the first application to molecular sequences was by Jerzy Neyman (1971).

The Maximum Likelihood approach works out the net likelihood that the given evolutionary model will yield the observed data, and the phylogenies with the highest likelihood are chosen. Likelihood programmes for sequence data are all currently based on two models - that of Jukes and Cantor (1969) and the Kimura 2-parameter model (Kimura, 1980).
The Jukes and Cantor Model assumes that all four nucleotides are equally frequent and all substitutions are equally likely, and the Kimura Model allows independent rates for transitions and transversions. RE site data have been subject to Maximum Likelihood analysis in several studies (Debry and Slade, 1985; Kaplan and Langley, 1979; Nei and Tajima, 1985; Li, 1986; Smouse and Li, 1987).

1.3.1.6 Weighting of Characters

In the study of the evolutionary history of taxa, not all the characters obtained are equally informative: some are unreliable and misinformative, and it is these which complicate phylogenetic analyses. These characters may violate the assumptions of a method or cause inconsistencies within the method,and therefore give rise to systematic errors. However, their influence can be minimized by assigning them less weight, while giving a higher weight to characters that are informative and reliable. In this way, homoplasies (convergence, parallelism and reversal) can be greatly decreased. One form of such character weighting is to assign misinformative characters a weight of zero and all reliable ones a maximum weight. This is the basis upon which the method of Character Compatibility functions.
One advantage of restriction fragment and site data is that all characters are unweighted.

1.3.1.7 Character Compatibility

This method of analysis (Felsenstein, 1981b) chooses the largest set of mutually "compatible" characters that can all evolve on the same evolutionary tree without homoplasy - in other words, a character is compatible with a phylogeny if its evolution can be explained without assuming

that any state arises more than once (e.g. Le Quesne, 1982; Estabrook, 1983). The method is also not widely used, possibly because of the complete exclusion of misinformative characters from the set of mutually compatible characters, i.e. these are regarded as no longer holding any useful information at all, and this is perhaps an unrealistic assumption to make. Penny and Hendy (1985, 1986) have described an approach that uses Compatibility as a weighting criterion (rather than estimate phylogenies directly) and Sharkey (1989) has developed a similar approach for binary characters - these are possibly more promising methods than that of Compatibility itself.

1.3.2 Distance Measure Approach

Distance methods offer an alternative way of representing evolutionary interrelationships: they all involve the determination of the proportion of shared fragments, sites or nucleotides, between different animal DNA's, in a pairwise manner.

Pairwise proportions can be used to estimate the amount of sequence divergence using a mathematical approach such as those of Nei and Li (1979). A matrix of all pairwise values then serves as input data for the construction of a phylogenetic tree, which links taxa by their similarity, or lack of it.

Phylogenies thus inferred, predict the distance for each pair of species as the sum of branch lengths in the path from one species to another in a tree. Distance methods may, or may not assume a molecular clock, which hypothesizes that mutations have accumulated along lineages at a relatively steady rate.

Since mutations occur in a stochastic (or random) fashion, one requirement for distance-based approaches is to record as many events as possible to reduce the stochastic error to as low a value as possible.

The first distance matrix methods were introduced by Fitch and Margoliash (1967) and independently by Cavalli-Sforza and Edwards (1967). Both of these were based on a measure of "least-squares".

Observed distance measures were compared to expected distances computed from the tree, and the discrepancy between them measured as a "lack of fit". This was formulated as:

$$\sum_{i,j} W_{ij} (D_{ij} - d_{ij})^2$$

where D_{ij} is the observed distance between taxa i and j;
d_{ij} is the expected distance, and
W_{ij} is the weight value assigned.

Distance data generally fall into two categories based on certain mathematical properties: additive or ultrametric distances.

Ultrametric distances produce trees that can be rooted, so that all the taxa are equidistant from the root - this means the assumption of a molecular clock. Ultrametric trees are the product of a collection of related techniques known as Cluster analysis, the most commonly used method being that of UPGMA (Unweighted Pair Group Method of Arithmetic Averages, Sneath and Sokal, 1973). This method implicitly assumes a constant rate of evolutionary divergence and is therefore considered to be inappropriate for reconstructing phylogenies (Farris, 1971). Notwithstanding this however, the method is attractive in its neat and simple computation of averaging the distance indices across taxa, and it is still used by molecular systematists (e.g. Hillis, 1985).

Additive distances can be fitted to an unrooted tree and there is a variety of techniques for the construction of additive trees. An additive distance matrix can even be transformed into an ultrametric matrix, and then a phylogeny inferred by using Cluster analysis - a well-known example of such is the Neighbour-Joining method of Saitou and Nei (1987). Additive trees can be rooted by the inclusion of an outgroup taxon. Alternatively, Midpoint-Rooting (Hillis, 1985) is utilized. This technique requires a rate-dependent

assumption (that average rates of change between the most diverse taxa are uniform). If the two most rapidly diverging lineages in the study have evolved at the same rate, then the appropriate root is at the midpoint of the distance connecting these taxa. Again, if this assumption is not true, then the rooted tree may be a poor representation of the phylogeny.

1.3.2.1 Neighbour-Joining

This method for inferring additive phylogenies was devised by Saitou and Nei (1987) and is related conceptually to Cluster analysis. Unlike the latter method, Neighbour-Joining does not assume the molecular clock. By contrast with Cluster analysis, Neighbour-Joining focuses upon nodes of the tree, rather than keeping track of the taxa, or clusters of taxa. The distance matrix used is modified, such that the separation between each pair of nodes is adjusted on the basis of their average divergence from all other nodes (Swofford and Olsen, 1990). The tree is constructed by joining the pair of nodes having the smallest distance (after matrix modification) between them. Their common ancestral nodes with their respective taxa are removed from the tree, in a kind of branch pruning process. As a result, the newly added common ancestral node now becomes a terminal node on the tree, now reduced in size. The process continues in this manner, with a single new node

replacing two terminal nodes at each step, until only two nodes remain, separated by a single branch.

1.3.2.2 Fitch-Margoliash Method

Another example of an additive method is that of Fitch and Margoliash, and is based upon the aforementioned Least-Squares method (Fitch and Margoliash, 1967). The technique concerns itself with defining the overall disagreement between a tree and the given data, in other words, it strives to find a lowest value for E, the error of fitting the distance estimates to a tree. This can be formulated mathematically as follows:

$$E = \sum_{i=1}^{T-1} \sum_{j=1}^{T} W_{ij} \left| d_{ij} - p_{ij} \right|^{\alpha}$$

Where T is the number of taxa in the study;

W_{ij} is the weight applied to separation of taxa i and j;

d_{ij} is the pairwise distance estimated, and

p_{ij} is the length of the path connecting i and j in the given tree;

The vertical bars represent the absolute value, and α = 1 or 2.

Both α and W_{ij} are given a particular value: setting α to 1 means that the weighted absolute deviations of the tree path lengths from the distance estimates, are minimized.

If $\alpha = 2$, then the equation becomes that of the Least-Squares criterion, but is weighted to accommodate particularly uncertain estimates with large error. If it is known which estimates are apt to be erroneous, then a very low weight is assigned to them by the Least-Squares method.

If it is not known which estimates are particularly susceptible to error, then using the minimum absolute deviations (where $\alpha = 1$) reduces the overall perturbation caused by the presence of such erroneous data values.

The Fitch-Margoliash method strives to find the phylogeny with the lowest **E** value and so must first optimize the branch lengths to find the smallest **E** consistent with a given tree topology, and then must find the topology with the smallest **E** of all trees.

Other distance methods exist which are not defined in terms of a measure of lack of fit - these include Farris' Distance Wagner Approach (Farris, 1972), Li's method (Li, 1981), Tateno et al's Modified Farris method (Tateno et al., 1982), and the aforementioned Neighbour-Joining method (Saitou and Nei, 1987). These each yield a tree which fits the distance data exactly.

1.3.3 Searching for Optimal Trees

Methods like Maximum Parsimony, the Fitch-Margoliash Distance approach and Maximum Likelihood all have particular optimality criteria which evaluate a particular tree, but do not actually find the optimal tree(s). The latter can be done either by evaluating every possible tree (Exhaustive Search) or by an exact algorithm that identifies optimal trees without requiring exhaustive searching, e.g. the Branch and Bound criterion.

1.3.3.1 Exhaustive Search

These procedures are satisfactory for small numbers of taxa only - because as the number of taxa increases, there is an exponential increase in the number of possible trees generated, e.g., for 7 taxa, there are 945 possible trees, but for 10 taxa, the number jumps to 2×10^6! (Felsenstein, 1978b). This limits exhaustive search procedures to a relatively small number of taxa. The exact number of taxa depends on the computer time available to the systematist.

1.3.3.2 Branch and Bound

First applied to evolutionary trees by Hendy and Penny (1982), this method provides an exact algorithm for

identifying all optimal trees for any criterion whose value is known to be non-decreasing as additional taxa are connected to a tree. It is commonly used for Maximum Parsimony criteria such as PAUP and although faster than the exhaustive search procedure, still requires extended computing time.

1.3.4 Investigatory Approaches

For a very large data set, where methods of finding optimal trees such as those mentioned above (Sections 1.3.3.1 and 1.3.3.2), are not suitable, other approaches can be applied: an initial tree is selected, and then improved upon be rearranging it in various chosen ways, such that it fits the chosen optimality criterion (e.g. Maximum Parsimony) more closely. Such approaches commonly begin with the stepwise addition of taxa to a growing tree, and then observation of the effects of such. Following this, a process of alternating positions of branches can improve the initial estimate by carrying out pre-defined rearrangements of the tree. Hennig 86 offers such options in the form of a Tree Editor, Dos Equis, which allows trees to be interactively modified, displayed and diagnosed (J.S.Farris, version 1.5, 1988). It also manipulates character data such that, if to begin with, 8 equally probable trees were available, after manipulation, there are two full-fit trees left of the original 8.

1.3.5 Consensus Trees

The conflict between proponents of morphological studies and those of molecular studies in systematics has led to the development of consensus and combination techniques for tree construction from character (cladistic) data. Of the consensus methods available, the most commonly used are Adams consensus (Adams, 1972), Strict consensus (Nelson, 1979) and Majority consensus (Margush and McMorris, 1981). The first approach collapses conflicting clades (from molecular vs. morphological data) to the first node of agreement between the competing phylogenetic hypotheses. Adams consensus trees are useful in the identification of taxa that cause conflict in characters, but are disadvantageous in that they can result in cladograms that are not indicative of any of the original trees. The Strict consensus method, true to its name, is extremely conservative in its merging of conflicting data, and represents all groups of species defined by every data set. Majority consensus identifies monophyletic groups recognized only in the majority of the conflicting phylogenies.

Alternatively, the combination of morphological and molecular data sets (Miyamoto, 1985), as opposed to consensus techniques, yields greater information content, although some have argued that the approach biases resulting

trees in favour of molecular data, particularly in the case of DNA sequences, where many more character states exist.

1.3.6 Resampling Methods

Even if a phylogenetic method has produced the desired trees, it is possible that some of the inferred phylogenies are incorrect, due to chance events (e.g., convergences or homoplasies). For this reason, statistical resampling methods have been developed to estimate the reliability of the results of a phylogenetic analysis.

Two such methods are those of the Bootstrap (Efron and Gong, 1983) and the Jackknife (Mueller and Ayala, 1982), which estimate the form of sampling distribution by repeatedly resampling data from the original data set. The Bootstrap and Jackknife differ in their method of resampling; Bootstrapping involves random resampling from one's own data, by drawing points from it, with replacement, until a data set the same size as the original is obtained. In a Bootstrap replication, some points are sampled several times, some only once, and some not at all. For each replication, the statistic of interest is computed - this is a binary variable that represents the presence or absence of, for instance, a particular monophyletic group of interest, on the tree(s) resulting from each replication. In this way, characters are weighted according to the number

of times they appear in each replicate sample. If the monophyletic group of interest arises in 95% or more of the trees resulting from bootstrapping replicates, then it appears that the occurrence of that group in the tree at that particular position, is significantly supported.

The Jackknife resamples the original data set by eliminating a certain number of data points at a time, and then recomputing the estimate from the remaining set of data points or characters. Usually, only one data point is dropped at a time, such that each of the remaining data points are dropped in turn. If the number of data points is large however, random sample groups of characters can be dropped, one at a time.
Either resampling method can be used on cladistic or on distance-based phylogenetic methods.

1.3.7 Computer Simulations

Each of the phylogenetic approaches discussed above, whether cladistic or distance-based, has its advantages and disadvantages. One very effective way of assessing the relative effectiveness with which different tree-building programmes produce the correct tree, is to perform computer simulations on fictitious data.

A programme designed to simulate evolutionary scenarios in this manner is SEvoNA (Simulating the Evolution of a Nucleic Acid sequence; E.H. Harley, 1991, version 1.2; Department of Chemical Pathology, University of Cape Town Medical School, RSA).

SEvoNA operates by generating a string analogous to a DNA sequence (of specified length) as the ancestral sequence. The programme allows this sequence (a choice between a random sequence and one with a GC 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 a 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 (90%), as expected with mtDNA, or if the mutations were random the frequency 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 several standard cladistic or distance-based programmes, to see whether they can reproduce the correct phylogeny. In this way, the progress or change of individual characters through the generations can be followed.
Fig.6 shows a typical simulation, where lower-case letters depict the occurrence of any mutations. Homoplasies are indicated and phylogenetically informative sites are marked with an X.

Simulations using an ancestral sequence of, say, 75bp, stand much less chance, for a given amount of mutation between the nodes, of inferring the correct trees than when using a sequence of 300bp. 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.

Fig. 4 Example of computer simulation of an evolutionary scenario using SEvoNA (E.H.Harley, 1991, version 1.2). Lower case letters in the progeny sequences indicate mutations. Examples of homoplasy at nucleotide positions during evolution are indicated by arrows. Phylogenetically informative sites are marked with an X. Mutations were random, with a transition frequency of 33.3%. Evolution of 6 progeny sequences followed a pectinate branching pattern.

(a) Table of phylogenetically informative sites and the phylogeny inferred by Hennig 86 analysis.
(b) Table of pairwise percentage sequence divergence values.
(i) Phylogeny inferred by FITCH analysis of pairwise sequence divergence values for the sequence data.
(ii) Tree constructed by Neighbour-Joining analysis of pairwise sequence divergence values.
(iii) Tree constructed by KITSCH analysis of pairwise sequence divergence values.

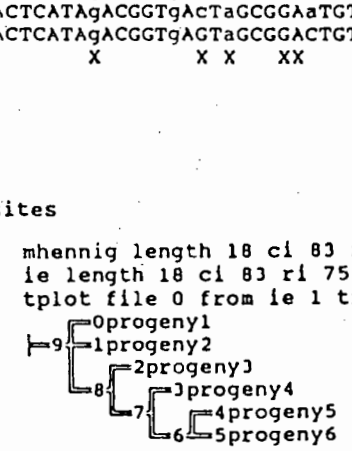
```
:voNA.       Seqence length = 75
:ctinate branching.   Progeny species = 6 .  Transition frequency =  33.3  %
>minal distance (sequence divergence) between nodes =  4 %

icestral sequence :
  CATCATATCAGATTGAATGCTAGTAATTTCCAATCCACTCATATACGGTCAGTCGCGGACTGTCCGTATACCAGA

xtant sequences :                          ↓          ↓  ↓
1 CATCATAgCAGATctAcTGCaAGTAATcTtCAATCCACTCATgTAtaGaCcaTaGaGGgCTtaCCGcATgCCAGA
2 CATgAacTCAGATTGtATGCTAaTAATTTtCAgTCgtCcCATATcaGGcgAGTCccTcgCTGTaCaTATACCAGA
3 CATCATtTCAGATTGAATGCTAGaAATTTCgAgTCaACTCATAgACGGTgAcTCaCGcACTGTaCGTATACCAGA
4 CATCcTtTCAGATTGAAgGCTAGcAAcTTCCAgTCgACTCATAgAACGGTgAGcaGCaGACTGTaCGTATACgAGA
5 CATCATtTCgGATTGAtTGCTAGcAATTaCtAgTCgACTCATAgACGGTgAcTaGCGGAaTGTaCtTATACgAGg
6 CATCATtcCgGATTGAtTGCTAGcAATTgCtAgTCgACTCATAgACGGTgAGTaGCGGACTGTaCGTATtCgAGA
          X      X      X     XX            X       X X   XX             X
```

(a)

Table of phylogenetically informative sites

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

```
mhennig length 18 ci 83 ri 75 trees 1
ie length 18 ci 83 ri 75 trees 1
tplot file 0 from ie 1 tree
```



(b)

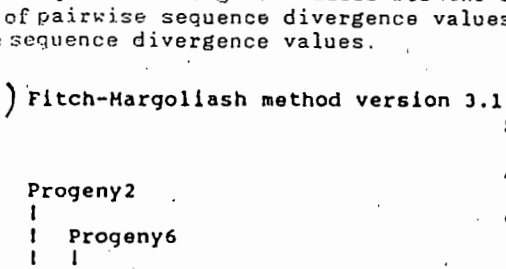
Pairwize sequence divergence table

1	2	3	4	5	6
.0	45.3	40.0	42.7	44.0	40.0
45.3	.0	26.7	32.0	36.0	34.7
40.0	26.7	.0	17.3	17.3	17.3
42.7	32.0	17.3	.0	17.3	14.7
44.0	36.0	17.3	17.3	.0	9.3
40.0	34.7	17.3	14.7	9.3	.0

ack mutations to ancestral sequence : nil

(i) Fitch-Margoliash method version 3.1

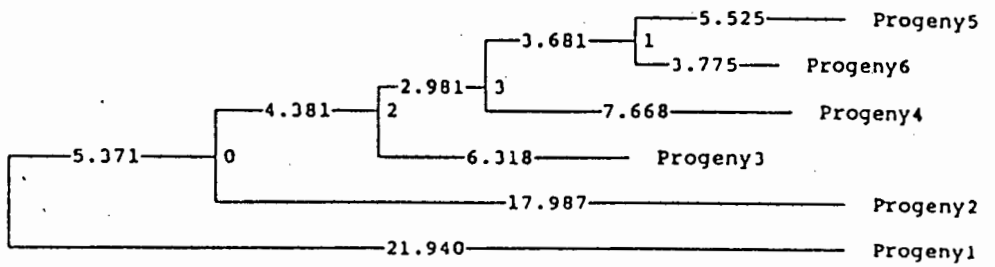
```
Sum of squares =      0.04486

Average percent standard deviation =
                                    4.00258
examined   28 trees
```



remember: this is an unrooted tree

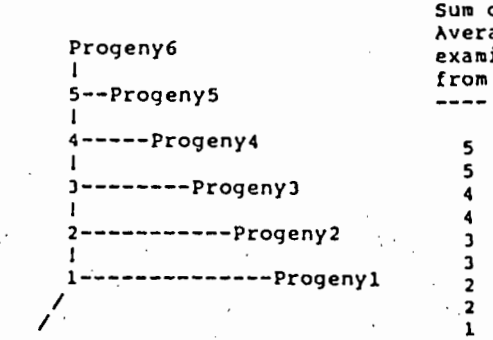
Between	And	Length
Progeny1	1	27.45134
1	Progeny2	17.84866
1	2	4.16502
2	3	3.43019
3	4	3.56516
4	Progeny6	3.86733
4	Progeny5	5.43267
3	Progeny4	7.70133
2	Progeny3	5.82593

(ii)



(iii) Fitch-Margoliash method with contemporary tips, version 3.1

```
Sum of squares =       0.165
Average percent standard deviation =    7.66947
examined  185 trees
```



from	to	length	time
5	Progeny6	4.65000	21.09369
5	Progeny5	4.65000	21.09369
4	5	3.24507	16.44369
4	Progeny4	7.89507	21.09369
3	4	0.75493	13.19862
3	Progeny3	8.65000	21.09369
2	3	7.08555	12.44369
2	Progeny2	15.73555	21.09369
1	2	5.35814	5.35814

1.4 Conclusion

This review of analytical programmes for inferring phylogenies and assessing their reliability shows that systematics is in an incomplete, but interesting state. Each approach has its own merits, limitations and assumptions, and the assumptions are frequently not valid. The challenge now lies in improving data analysis to deal with the influx of molecular information.

SECTION D: MOLECULES VERSUS MORPHOLOGY

A major controversy in systematics which arises from recent advances in molecular technology, is one of Molecules versus Morphology. Some systematists have claimed molecular data to be relatively weak in systematic analysis (Kluge, 1983) while others have claimed that morphological character data are misleading (Sibley and Ahlquist, 1987a).

Indeed, such "data chauvinism" is rife among proponents of molecular data themselves (Hillis and Moritz, 1990). When DNA-DNA hybridization and restriction enzyme analytical techniques were developed (see Sections 1.2 and 1.4 of Section B above), there were assertions of superiority as regarded their application to systematics - they were considered preferable to isozyme electrophoresis or Microcomplement Fixation (see Section 1.1.1 of Section B above).

However, some authors are of the opinion that this is an over-emphasized conflict (Hillis, 1987; Crowe, 1988; Hillis and Moritz, 1990): morphological and molecular data approaches each possess distinct advantages and disadvantages (Shoshani, 1986). Morphological techniques are applicable to a large range of museum and fossil material, and molecular data approaches provide a potentially very extensive data set - the maximum number of independent characters of an organism is limited only by the number of nucleotide pairs in its DNA.

Comparative studies show that morphological and molecular divergence are independent, responding to different evolutionary pressures (Wilson et al., 1974; 1977). Another conflict arising from this controversy concerns whether it is justifiable in phylogenetic analysis to assume a constant rate of mutational change through time, i.e. whether there is a molecular clock which keeps constant time (see Section 1.1, Section C above).

Certainly there is more validity for a regular molecular clock, at least within particular taxonomic groups, than for a morphological clock; morphological evolution does not exhibit clock-like behaviour, i.e. does not progress at a regular rate.

The conflict could be addressed by combining and comparing data from each approach and analysing them cladistically (since this analysis makes no assumption of a constant rate of evolution). This would maximise information content to give a comprehensive view of evolution.

SECTION E: OBJECTIVES OF THIS STUDY

The following chapters of this dissertation present a molecular approach towards helping define a Francolin phylogeny. Comparisons of both fragment and restriction endonuclease site data from the mtDNA of several Francolin species are presented. From these molecular data, phylogenetic relationships were reconstructed using different cladistic and distance-based approaches, and a comparison of individual cladistic and distance-based methods was made.

Cladistic analysis of mitochondrial DNA restriction fragment lengths, and restriction enzyme sites, required the selection of appropriate outgroup taxa for each approach, i.e. species that are related to, but distinct from the Francolins.

The birds chosen were Coturnix coturnix japonica and Margaroperdix madagascariensis, respectively. Selection of the former, commonly known as the Japanese Quail, was grounded on Crowe and Crowe's (1985) proposal of a quail-like ancestor for Francolinus.

Margaroperdix madagascariensis is generally referred to as the Madagascar Partridge and is endemic to the island. Milne-Edwards and Grandidier (1885), in a skeletal comparison of Margaroperdix with the partridge Perdix and the quail Coturnix, discovered that Coturnix showed a closer

affinity with the Madagascar partridge than did Perdix. Frost (1975) suggested that Coturnix also most closely resembles Margaroperdix in natal down pattern, a characteristic often used by researchers to indicate phylogenetic similarities and differences, since they tend to be very conserved. This link with Coturnix made the Madagascar Partridge a logical choice of outgroup for the site comparison.

Francolin mitochondrial DNA data from both the fragment and the restriction enzyme site comparisons were combined with Francolin morphological, behavioural and ecological data (termed macrocharacters; Crowe et al., 1990, submitted) and analysed cladistically. A general comparison of the classical Francolin phylogeny, based on morphological characters (Fig.1), was made with the phylogenies generated from analysis of Francolin mitochondrial DNA data and the combined data set; the soundness of Hall's (1963) assigned monophyletic groups was assessed and Milstein and Wolff's (1987) "partridge" - "francolin" postulate was tested.

It is hoped that the results of this study will illustrate the potential of molecular data in the field of systematics.

CHAPTER 2

THE ISOLATION OF FRANCOLIN MITOCHONDRIAL DNA AND ITS APPLICATION IN BIOCHEMICAL TECHNIQUES

2.1 Isolation and Purification of Mitochondrial DNA

2.1.1 Materials

Liver, heart and spleen tissue were taken from the Japanese Quail (Coturnix c. japonica), each of the 13 species of southern African Francolin, and from Margaroperdix madagascariensis. Table 1 shows areas of collection of each species and the number of individuals used in extraction of mitochondrial DNA.

2.1.2 Preparation of Tissue

Tissue samples preserved in 99% ethanol were found in trial extractions to give poor yields of mtDNA, if at all. This was thought to be due to damage of mitochondrial membranes.

In initial experiments, it was found that relatively equal concentrations of mtDNA were obtained from both frozen and fresh liver tissue, i.e. approximately 7-10ng/μl for each.

Table 1: Francolin species used in mtDNA fragment comparison and restriction enzyme site comparison. Asterisks denote those species used in the mapping approach.

"Francolins"	Species	Abbreviation	Common Name	Collection Locality	No. individuals
Vermiculated Group	Francolinus adspersus	AD	Red-billed Francolin	Omaruru District, Namibia	1
	F. natalensis	NT	Natal Francolin	Sabie District, Transvaal	1
	F. capensis*	CA	Cape Francolin	Cape Town District, Cape Province	9
	F. hartlaubi*	HT	Hartlaub's Francolin	Omaruru District, Namibia	
Bare-Throated Group	F. afer	RN	Red-Necked Francolin	Uitenhage District, Cape Province	1
	F. swainsonii	SW	Swainson's Francolin	Nylstroom District, Transvaal	1
	F. leucoscepus	LU	Yellow-Necked	Athi River District, Kenya	1
"Partridges"					
Red-Winged Group	F. africanus*	GW	Greywing Partridge	Ceres District, Cape Province*	5
				Molteno District, Cape Province	1
	F. shelleyi	SH	Shelley's Partridge	Nylstroom District, Transvaal	1
	F. levaillantii*	RW	Redwing Partridge	Sabie District, Transvaal*	7
				Grant's Castle Nature Reserve, Transvaal	1
	F. levaillantoides	OR	Orange River Partridge	Balfour District, Transvaal	1
Red-Tailed Group	F. coqui	CQ	Coqui Partridge	Nylstroom District, Transvaal	1
Striated Group	F. sephaena	CR	Crested Partridge	Nylstroom District, Transvaal	1

Therefore, all tissue samples were removed immediately after death of the individual, snap-frozen in liquid nitrogen and thereafter stored at -70°C for the sake of convenience.
When preparing for extraction of mtDNA, care was taken to avoid repeated freezing and thawing of tissue, as mitochondrial (and nuclear) membranes have been found to be prematurely lysed as a result of this (Shields and Helm-Bychowski, 1988).
The liver, containing very little fibrous connective tissue, was found to yield the most mtDNA per gram (approximately 0.5-1μg mtDNA/g tissue). Where soft tissue was not available, samples of pectoral or thigh muscle were utilized, although mtDNA yields were considerably lower and tended to include substantial amounts of contaminating nuclear DNA.
Detailed descriptions of buffers and solutions are presented in the Appendix. Unless otherwise stated, reagents were used at the concentrations given in the Appendix.

2.1.3 Isolation of Mitochondria

Mitochondrial DNA was extracted from between 1g and 10g of thawed tissue, according to procedures taken essentially from Brown (1980) and Lansman et al.(1981), with some modifications. Samples were finely minced with scissors and suspended in cold Extraction Buffer (see Appendix) at a ratio of 4.5ml to 1g tissue.

All subsequent steps were carried out on ice, or at 4°C. The suspension was homogenized in a Waring blender at full speed for 15 seconds to disrupt cell membranes. The homogenate was then centrifuged for 15 minutes at 1000g in a JA 14 rotor, which pelleted the larger organelles, cell nuclei and debris, leaving intact mitochondria in the supernatant. There was a significant reduction in nuclear DNA contamination if the supernatant was decanted carefully, and the centrifugation (1000g, 15 mins) repeated.
The supernatant, containing intact mitochondria, was then filtered through a layer of cheesecloth to remove surface debris (e.g. lipid particles). Mitochondria were crudely pelleted after a high-speed centrifugation at 10 000g for 15 minutes in a JA 20 rotor. The pellet was resuspended in 35ml cold Extraction Buffer and spun finally at 20 000g (JA 20, 15 minutes). The pelleted mitochondria were used immediately to prepare mtDNA.

2.1.4 Isolation and Purification of MtDNA

2.1.4.1 MtDNA Isolation Procedure

This procedure was carried out at room temperature (R.T). The mitochondrial pellet (see section 2.1.3 above), which contained whole mitochondria and some nuclear DNA, was resuspended in 3.5ml Saline Tris EDTA Buffer (STE Buffer; see Appendix). Sodium Dodecyl Sulphate (SDS; see Appendix)

was added to the mitochondria in suspension to give a final concentration of 1%, and incubated for 15 minutes to lyse the mitochondrial membranes Addition of CsCl salt to 1M and incubation for 15 minutes precipitated mitochondrial membranes and proteins. The suspension was then spun in a SIGMA 2MK benchtop centrifuge at 10 000g for 15 mins. One gram CsCl per ml solution was added to the resulting supernatant - less the amount of CsCl previously added. ethidium bromide (EtBr; see Appendix) was added to the solution of mtDNA, such that the final density of the mtDNA solution was approximately 1.55 - 1.56g/ml.

A CsCl density gradient was generated after centrifuging the DNA solution in heat-sealed plastic tubes, at 50 000g (Vti 65 rotor) for 18 hours. When viewed under UV light, mtDNA and nuclear DNA appeared as two discrete bands in the gradient. The supercoiled circular mtDNA is denser and bands below the linear nuclear (and damaged mt) DNA in the gradient (Smith et al., 1971).

The mtDNA band was extracted (using a 375nm UV light source) with either a peristaltic pump and some fine plastic tubing, or a sterile syringe. Where a mtDNA band was not visible, an estimated region of 2-7mm below the nuclear DNA band was removed for purification of mtDNA.

2.1.4.2 Purification of MtDNA

The mtDNA-CsCl solution was treated repeatedly with salt-saturated isoamyl alcohol, into which EtBr partitions. The resulting top phase, containing EtBr, was removed and discarded each time. The remaining aqueous solution was diluted with sterile distilled water at a ratio of 2 volumes H_2O to 1 volume DNA solution. This was to decrease the CsCl concentration and therefore to avoid precipitation of salt crystals with mtDNA. MtDNA was then precipitated (at -20°C, 1 hour) with 6 volumes of 100% Ethanol (EtOH) to 1 volume of the original mtDNA solution. The purified mtDNA was pelleted in a Sigma 2MK benchtop centrifuge at 10 000g for 20 mins, 4°C. The pellet was washed with 70% EtOH and recentrifuged. After vacuum drying, the mtDNA pellet was resuspended in 200µl sterile Tris EDTA (TE Buffer; see Appendix) and stored in aliquots at -20°C.

2.2 Restriction Enzyme Digestion

Only RE's recognising 6-bp sequences were utilized in restriction enzyme digests; 4-bp - recognising RE's resulted in very complicated mtDNA gel profiles which would have made fragment comparisons and mapping rather more difficult, particularly in identification of individual gains or losses of cleavage sites.

Approximately 1/100th of the total yield of mtDNA per individual was used in each digest (e.g. 2μl from 200μl final volume mtDNA). The concentration of mtDNA per digest was estimated visually to be about 5-30ng, the minimum concentration made visible by end-labelling (Section 2.4.1).

MtDNA was cleaved by RE's at 37°C for 2 hours. All RE's were used at a concentration of approximately 1 unit/μg DNA (see Appendix). Conditions for RE digestion followed the specifications of the supplier where necessary (Amersham International, Boehringer Mannheim, New England Biolabs). However, most incubations were carried out in 1x or 2x KGB Buffer (McLelland et al., 1988; see Appendix), which provided optimal digestion conditions for most of the RE's used. The final volume per digestion was 15μl. Enzymes such as EcoR1 and Pst1 sometimes exhibited star activity. To reduce this effect, digestion times were minimal for these enzymes.

2.3 Agarose Gel Electrophoresis

Restriction enzyme-digested mtDNA fragments were separated according to size by horizontal electrophoresis through large agarose gels (200mm x 150mm) made up in 1x Tris Acetate EDTA Buffer (see Appendix for preparation of agarose gels and TAE Buffer). The concentration of agarose varied between 1% and 2%, allowing relatively accurate

estimation of fragment size over a range of approximately 150bp-10000bp. MtDNA fragments were usually electrophoresed at 35V, overnight.

2.4 Visualization of RE fragments

DNA fragments could be detected either by direct staining with EtBr, or with the very sensitive end-labelling (Section 2.4.1 below) and transfer hybridization techniques (Section 2.4.3 below). MtDNA yields were generally not high enough for efficient use of EtBr in staining gels. Therefore end-labelling was the method of choice, and Southern DNA Hybridization (Section 2.4.3) was reserved for old, frozen, small or otherwise stubborn nuclear DNA-contaminated samples.

2.4.1 End-Labelling of MtDNA

Cleaved mtDNA fragments (preferably with 5′ overhangs or blunt ends) were subjected to the 5′-3′ exonuclease activity of large fragment E.coli DNA Polymerase 1 (Klenow Polymerase; see Appendix), at a working concentration of 1 unit enzyme/μl per reaction. This eliminates nucleotides sequentially from the 5′ end of the mtDNA fragments. This pre-incubation took place at R.T. for 10-20 mins. Thereafter dATP, dGTP, dTTP (to a final concentration of 2mM; see Appendix) and ^{32}P-labelled dCTP

(at a concentration of 1μCi/μl per end-labelling reaction; see Appendix) were added and the reaction mix was incubated for a further 10-20 min at R.T. This resulted in the radioactive labelling of individual mtDNA fragments, utilizing the 3'-5' polymerase function of the Klenow fragment.

Because each fragment has the same number of ends, radioactive intensity is independent of fragment size. Gel Loading Buffer (see Appendix) was added to the labelled, digested mtDNA fragments. This terminated the end-labelling process and provided a visible dye-front for tracking the progress of mtDNA fragments during electrophoresis.

The end-labelling technique allowed detection of nanogram quantities of DNA - therefore 10-50ng Francolin mtDNA were routinely used in these analyses.

2.4.2 Gel-drying and Autoradiography

After electrophoresis, those agarose gels not intended for Southern DNA Hybridization were vacuum-dried at 55°C to a piece of Whatman 3MM Chromatography paper for an hour. This rendered the gels much easier to handle for autoradiography and sharpened the fragment patterns.

Dried gels were loaded into an autoradiograph cassette fitted with intensifying screens, and used to expose Amersham Hyperfilm-MP X-ray film, at -70°C. This temperature enhanced the gel image. Exposure times depended on efficiency of the labelling reaction, amount of mtDNA labelled, and age of the gel. Generally, an overnight exposure was sufficient, although sometimes it was left for a few days.

After exposure, autoradiographs were developed, fixed, and allowed to dry.

2.4.3 Southern DNA Hybridization

Where samples were contaminated by nuclear DNA, the desired DNA fragments were visualized by Southern DNA Hybridization.

After gel electrophoresis, the DNA fragments were denatured (see Section 2.4.3.2) and transferred overnight to a Hybond-N Nylon membrane (Amersham International), using standard Southern Blotting procedure (Southern, 1975). The membrane was probed with mtDNA radio-labelled (see Sections 2.4.3.1 and 2.4.3.3 below) with ^{32}P-dCTP.

2.4.3.1 Nick Translation of Probe MtDNA

MtDNA digestion fragments were radio-labelled with ^{32}-P dCTP essentially as described by Rigby *et al.*, (1977)

but using a commercial nick-translation kit (Amersham International).

Approximately 500ng of Francolin mtDNA was nicked by DNAse I, and then treated with Klenow Polymerase, which adds nucleotides to the 3′ hydroxyl-end created by the nicking of one DNA strand. Simultaneously, the 5′-3′ exonuclease activity of Klenow enzyme eliminated nucleotides from the 5′ end. Following the specifications of the nick-translation kit, 50μCi of ^{32}P-labelled dCTP in a nick-translation buffer cocktail was added to make up the probe mtDNA. Probe mtDNA was then separated from unincorporated ^{32}P dCTP by elution through an erect column of sephadex G50, equilibrated in TE buffer (pH 8) and packed in a sterile nipped Pasteur pipette. The column was eluted once with sterile STE buffer and 100μl of probe DNA applied. The nick-translated mtDNA was fractionated, using TE buffer as eluent.

Fractions were scanned for radioactive counts, and the labelled probe and unincorporated nucleotides were seen as two separate peaks, respectively, after scintillation counting. Fractions from the first peak (having highest d.p.m.) were pooled and specific activity for the probe DNA calculated as follows:

Specific Activity =

(Sum of dpm in pooled fractions) x 100/0.4 = dpm/μg of DNA

Conditions of the nick-translation were such that there was about 30% incorporation of ^{32}P dCTP into the mtDNA. The probe mtDNA, after denaturing, recognises complementary DNA sequences in the total sample and binds to them. This allowed visualization of the radioactive bands upon autoradiography.

2.4.3.2 Southern Transfer

After electrophoresis, agarose gels were shaken gently for 30 mins at R.T. in 300ml basic Denaturing Solution (see Appendix) - this separated double-stranded DNA fragments. The gel was then neutralized by soaking in Neutralization Solution (see Appendix), in a similar manner.

Transfer of single-stranded mtDNA to Hybond-N membrane proceeded as follows: Three pieces of Whatman 3MM chromatography paper cut larger than the gel were soaked in 20xSSC (see Appendix) and placed onto plastic cling wrap on the flat bench top. The gel was placed upside down on the 3MM paper and a piece of Hybond-N membrane cut to the same size as the gel was placed on top of the gel. Care was taken to prevent the formation of air bubbles between the gel and the membrane. A single piece of 3MM paper cut to the same size as the gel was placed on top of the Hybond-N membrane. A stack of paper towelling was put on top of the

3MM paper and weighed down with a 1kg weight. Transfer was allowed to proceed overnight.

The resulting blot of the gel was air-dried and the denatured mtDNA irreversibly bound to the membrane by UV-irradiation (275nm) for 2-5 mins. Blots were used immediately for hybridization.

2.4.3.3 DNA-DNA Filter Hybridization

The membrane and 15ml preheated Prehybridization Solution (see Appendix) were sealed in a plastic sleeve and incubated for an hour at 65°C. Prehybridization solution contained 0.25% Protea milk powder to reduce nonspecific binding of the probe to the membrane.

The nick-translated probe mtDNA (see Section 2.4.3.1) was denatured to single-stranded DNA by boiling for 5 mins, followed by a quick transfer to ice to prevent re-annealing of the DNA. Denatured probe DNA was added to the prehybridized filter membrane in its Prehybridization Solution (to a concentration of 33ng/ml), and by gentle shaking at 65°C, allowed to hybridize to complementary single-stranded sequences over a period of 24-60 hours. Following hybridization the filter was immersed in each of three Stringency Washes with varying salt concentrations (see Appendix), again at 65°C. The first wash was repeated three times (15 mins each) and each of the following two

washes were performed once (15 mins each). These removed as much of the characteristic non-specific binding (appearing after autoradiography as a smudged dark background), as possible. Membranes were wrapped in plastic cling wrap and then subjected to autoradiography (see section 2.4.2). The Hybridization Solution with probe mtDNA could be removed, stored frozen and re-used in subsequent blots within two weeks (2 weeks being the ½ life of ^{32}p).
Filter membranes could be re-used after stripping off the old bound probe, as follows: Membranes were submerged in a solution of 0.1% SDS, which was brought to the boil, left for 30 to 60 mins, and cooled to R.T. This was rinsed in distilled water, and then checked for residual bound probe with a Geiger Counter.

2.5 Calibration of Molecular Weight

The molecular weights of mtDNA digested fragments were determined relative to a standard molecular weight marker. In all cases, this was an end-labelled Hind III digest of phage Lambda DNA. This exhibits a characteristic profile of fragments, having standard molecular weights (see Appendix II). The mobilities of the Lambda marker fragments during electrophoresis was plotted against their molecular weights on semi-log graph paper to give a standard curve (Fig. 7). As depicted in Fig. 7, the extent of curvature is a function of agarose concentration. The sizes of Francolin and

outgroup mtDNA fragments were estimated by reference to the standard curve. One percent agarose gels were used to size fragments in the 6.7kb to 16.4kb size range, 1.2% to 1.5% gels for fragments in the 2.0kb to 6.7kb range, and 1.8% to 2% gels to size fragments in the 0.15kb to 2.0kb range, the 125bp marker fragment of the Hind III lambda being well visualized on these gels.

Large RE digested fragments, (9.0kb to 16.4kb), often display ambiguous mobilities on agarose gels, and hence, anomalous molecular weights. This was taken into consideration when determining the size of bands, and dealt with in the following ways:

Sizes of questionable bands larger than 9kb were estimated from the sum of fragments resulting from double digests in mapping. Expected small fragments that were not visible (even on high percentage agarose gels), were sized by inference from size differences in products of double digests.

Despite the inevitability of some error in molecular weight determination from these approaches, a coherent estimation of relative molecular weights was obtained as required in mapping RE cleavage sites in mtDNA.

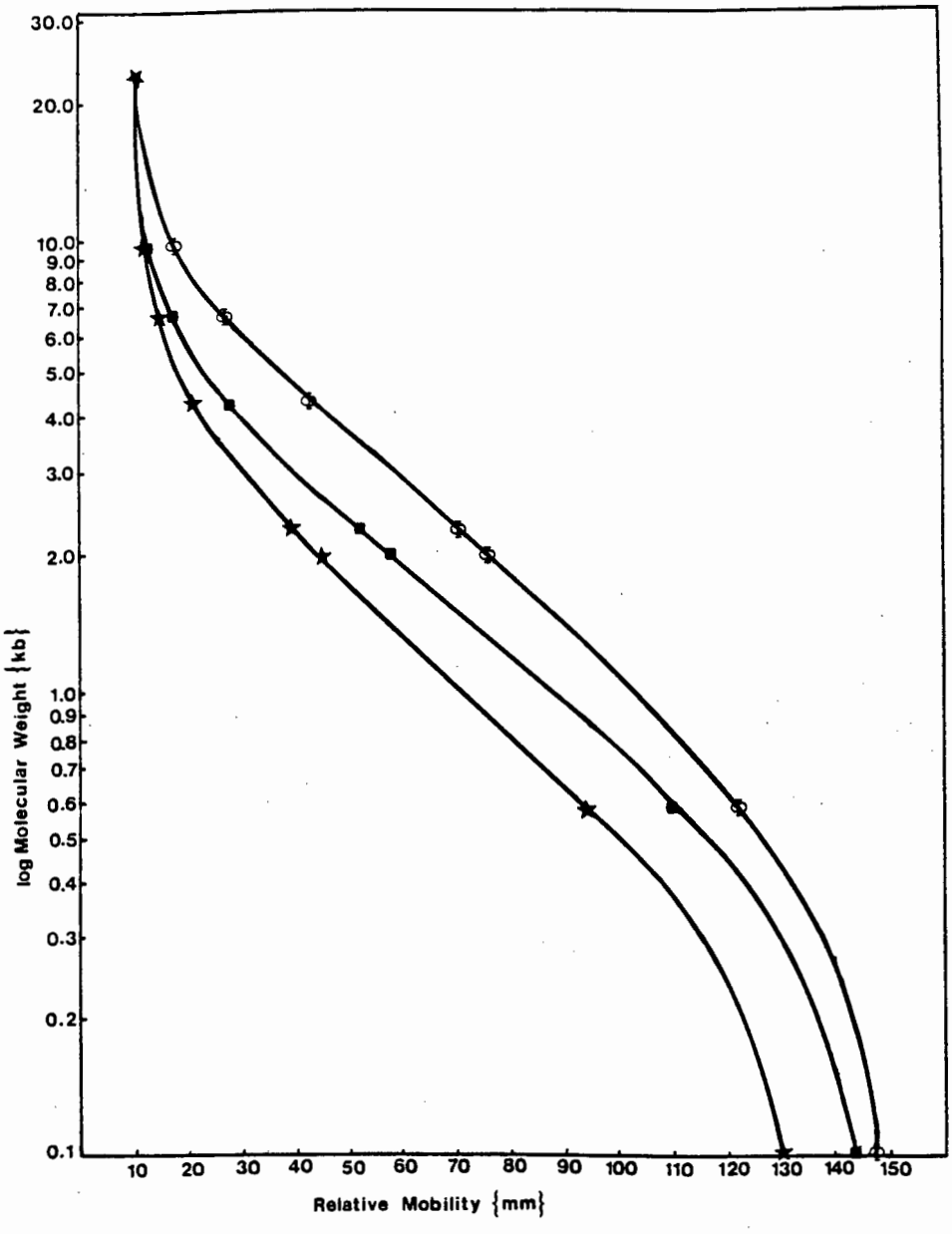


Fig. 7 Semi-logarithmic plot of the relative mobilities (mm) and molecular weights (kb) of restriction fragments resulting from a HindIII digestion of phage Lambda DNA; symbols used: ★ 2.0% agarose gel
■ 1.5% agarose gel
Φ 1.2% agarose gel

CHAPTER 3

COMPARISON OF FRANCOLIN MITOCHONDRIAL DNA FRAGMENTS

INTRODUCTION

In the study of systematics, the quickest molecular method (and by far the simplest) utilizing restriction endonuclease analysis is RFLP - or Restriction Fragment Length Polymorphism. In this approach, restriction enzymes cleave purified DNA at specific recognition sites, resulting in a number of fragments of varying lengths. These may be separated electrophoretically, forming a pattern characteristic of the animal under study. Fragment lengths are then estimated by comparing their electrophoretic mobility (a function of length) with the mobility of size standards of known molecular weight (see Sections 2.3 and 2.5 in Chapter 2). RE's recognizing 6-bp DNA sequences typically produce 1-10 fragments of mtDNA.

A comparison of mtDNA fragments indicates the homology and differences between fragments from the taxa in a study. Related mtDNA's will share some fragments (i.e. RE sites) and not others. Differences between the fragment patterns of animals under comparison are usually due to base substitutions that cause an RE site either to be gained or lost. Although uncommon in comparisons of closely related

species, differences may sometimes be the result of length mutations (e.g. in primates - Aquadro and Greenberg, 1983; Cann and Wilson, 1983, and in lizards - Densmore et al., 1985).

DNA fragment comparisons should preferably only be used for very closely related animals, e.g. species and subspecies, since the technique becomes very inaccurate once the proportion of shared fragments becomes low. Although RFLP's do provide insight into the nature and extent of differences between two DNA sequences, they offer less information on evolution of the sequence itself. Notwithstanding these reservations, the approach has been successful in defining relationships in a variety of genera (Potter et al., 1975; Avise et al., 1979; Ferris et al., 1983). In many cases, relationships defined on this basis are concordant with relationships derived from analysis of morphological and protein data (e.g. see Avise et al.,1979; Ferris et al., 1983b and Marshall and Sage, 1981).

METHODS AND MATERIALS

Unless specified otherwise, all biochemical methods and materials utilized in the mtDNA fragment comparison are identical to those described in Chapter 2.

3.1 Sources of Birds

Specimens of each of 13 species of Francolin were collected at various localities in Southern Africa(Table 1). Coturnix coturnix japonica (4 individuals) were obtained from the Fisheries Industrial Research Institute (FIRI), Cape Town.

Specimens of F.levaillantii and F.africanus were collected from populations of each in two geographically separated localities to provide an indication of intraspecific mtDNA polymorphisms (Table 1). Individuals of levaillantii were collected in the Sabie and Drakensberg districts and africanus specimens, in the Ceres and Molteno districts.

3.2 Isolation and Purification of MtDNA

The isolation and purification of Francolin mtDNA were performed as described in Section 2.1.4 of Chapter 2.

3.3 Restriction Enzyme Digestion

MtDNA's from each of the 13 Francolins, the Japanese Quail, and the additional F.africanus and F.levaillantii specimens from separate geographical locations (see Table 1), were digested with the following RE's: Asp 718, Bam HI, Bcl I, Eco RI, Eco RV, Hind III, Nco I, Pst I, Pvu II, Sac I, Sac II, Sca I, Stu I and Xba I. Conditions for restriction enzyme digests were as described in Section 2.2 of Chapter 2.

3.4 Agarose Gel Electrophoresis

Restriction-digested mtDNA fragments were separated according to size as described in Section 2.3 of Chapter 2.

3.5 Visualization of MtDNA Fragments

MtDNA fragments in the comparison were end-labelled as described in Section 2.4.1 of Chapter 2.
Fragment comparisons involving F.leucoscepus employed the technique of Southern DNA Hybridization (detailed in Section 2.4.3 of Chapter 2), since end-labelling of the mtDNA of this species always resulted in the bands being indistinguishable from the heavy background. This was primarily due to contamination by nuclear DNA.

Although any Galliform mtDNA would have sufficed as a probe, the mtDNA used for this purpose was that of F.capensis, since it was preferable to utilise a Francolin species and it was freely available in large quantities (Section 2.4.3.1 of Chapter 2).

3.6 Gel Drying and Autoradiography

Gel drying of gels not intended for Southern Blotting, and autoradiography were performed as described in Section 2.4.2 of Chapter 2.

3.7 Calibration of Molecular Weight

The molecular weights of mtDNA fragments were determined relative to the fragments produced by a HindIII digest of phage Lambda DNA (described in Section 2.5 of Chapter 2; see Appendix II).

RESULTS

3.1 Introduction

Fourteen sets of RE digestion patterns, representing the mtDNA of 13 Francolins and the Japanese Quail, were assessed for shared fragments (See Fig. 8). Southern Blotting of F.leucoscepus required much larger amounts of enzyme per digest than digests that were end-labelled, and there appeared to be a loss of binding efficiency for fragments smaller than 300bp (see Fig. 9).
Although end-labelling seems to be a more favourable technique in this case, Southern Blotting can be very useful in low cost blood sampling and preservation of DNA from live birds. Samples can be taken during ornithological field work such as bird-ringing without the need for freezing facilities (Arctander, 1988), particularly where the species in question are endangered.

A total of 211 distinct restriction fragments was thus produced. Satellite bands (see Results, Section 3.3 of this chapter) were excluded from this data set. Bands appearing to have a molecular weight larger than 9kb were also excluded, since they could well display ambiguous mobilities on the agarose gels, and therefore inaccurate molecular weights.

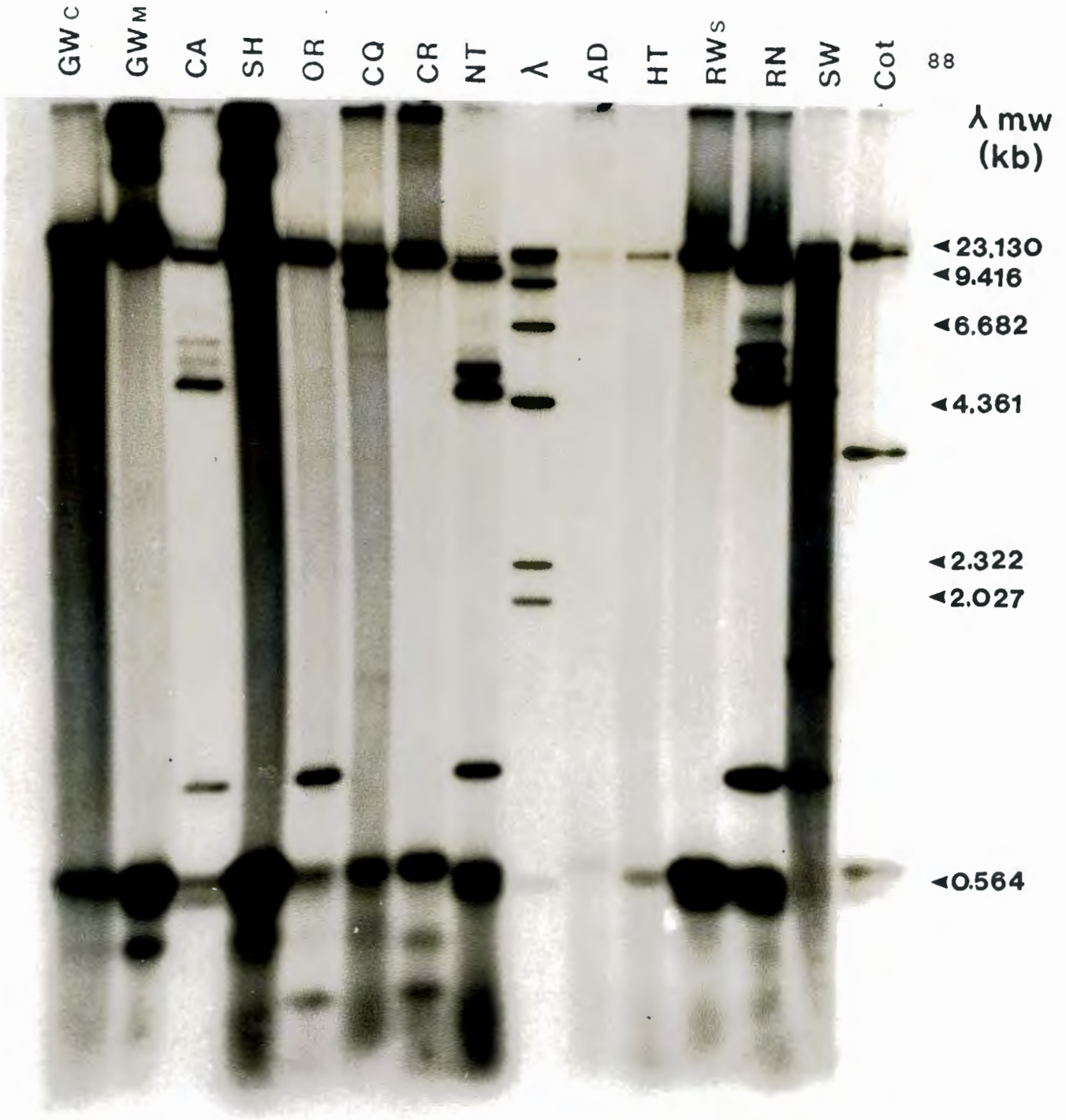


Fig. 8 Autoradiograph of the electrophoretic profiles of Francolin and Japanese Quail mtDNA, digested with XbaI, and end-labelled. The fragments were separated in a 1.2% agarose gel at approximately 2V/cm. End-labelled Lambda DNA (λ) digested with HindIII was used as a size marker. Cot = Japanese Quail; RWs = Redwing Partridge (Sabie); GWc = Greywing Partridge (Ceres); GWm = Greywing Partridge (Molteno); OR = Orange River Partridge; SH = Shelley's Partridge; CQ = Coqui Partridge; CR = Crested Partridge HT = Hartlaub's Francolin; NT = Natal Francolin; CA = Cape Francolin; AD = Red-Billed Francolin; RN = Red-Necked Francolin; SW = Swainson's Francolin

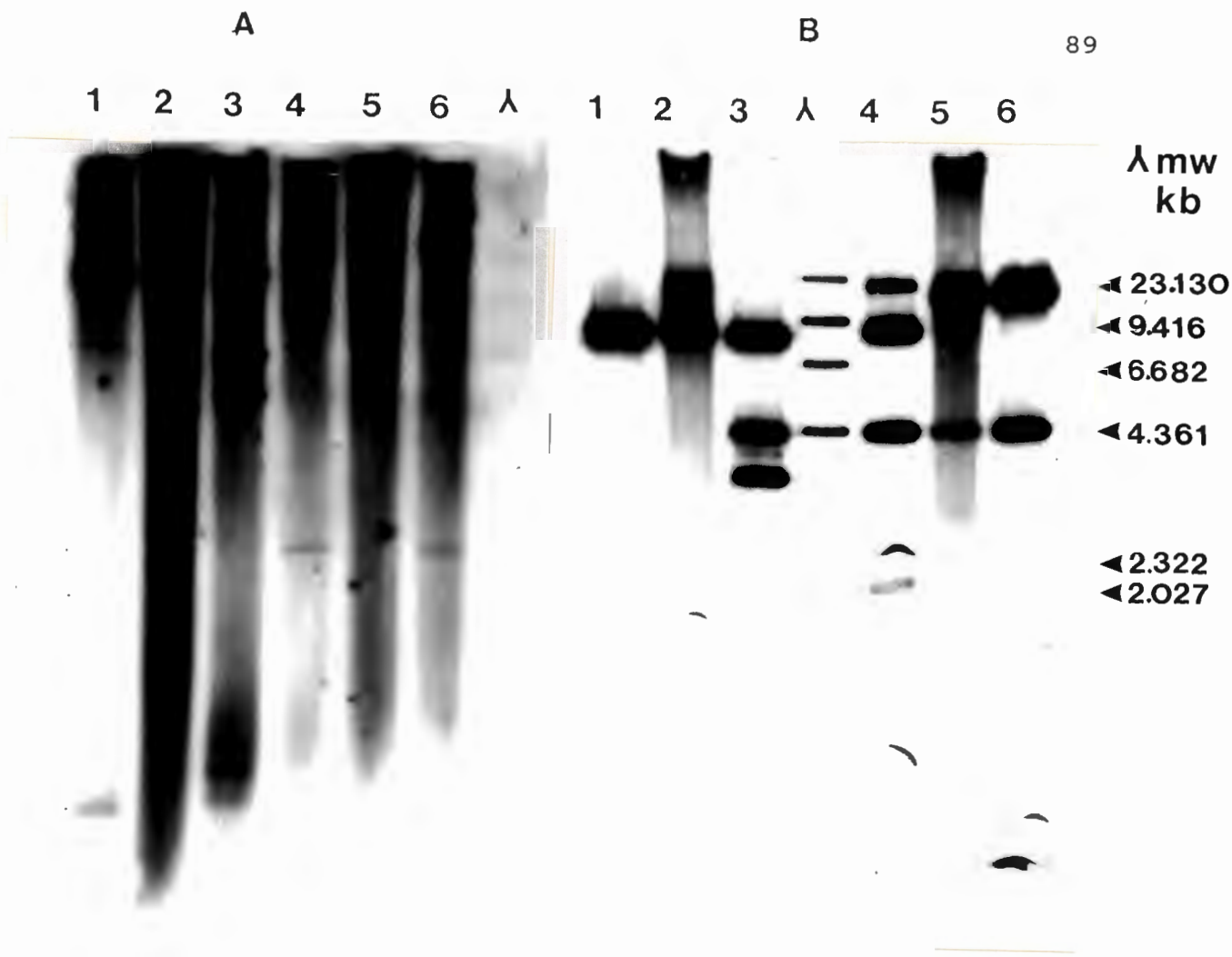


Fig. 9 A. 1.5% Agarose gel of end-labelled Francolin mtDNA subjected to electrophoresis at approximately 2V/cm, and autoradiographed.

B Southern Blot of similar 1.5% agarose gel, probed with F.capensis mtDNA, and autoradiographed. Lanes 1,2,3 in A and lanes 1,2,3 in B are NcoI digests of NT= Natal Francolin; Lu = Yellow-Necked Francolin and RN = Red-Necked Francolin. Lanes 4,5,6 in A and lanes 5,6,7 in B are SacII digests of NT LU and RN. Lambda DNA (λ), digested with HindIII and end-labelled, was used as the size marker.

3.2 Assumptions Made

Fragments of the same electrophoretic mobility were assumed to be homologous between individuals, i.e. to share flanking cleavage sites. It was also assumed that all the differences in the fragment patterns stemmed from the gain or loss of cleavage sites, as a result of base substitution, i.e. there appeared to be no insertions, deletions or duplications.

An example of possible site gain may be seen in Francolin mtDNA profiles for Hind III (Fig. 10). Most of the species exhibit a noticeable doublet of bands with approximate sizes of 2300bp and 2150bp. Howevever, the 2150bp band is absent in the Crested Francolin (CR, F. sephaena) profile. It is possible that its ancestral 2150bp fragment has gained a single Hind III recognition site, thus giving rise to two smaller fragments of about 1800bp and 350bp. Also evident are three bands of approximately 750bp, 700bp and 600bp. An alternative possibility is that the ancestral 2150bp fragment gained two internal HindIII sites which yield the aforementioned three smaller bands upon digestion.

Where differences in fragment length are due to site loss, it can be postulated that the restriction site at the junction of two fragments may have been lost as a result of a single nucleotide substitution within the enzyme's

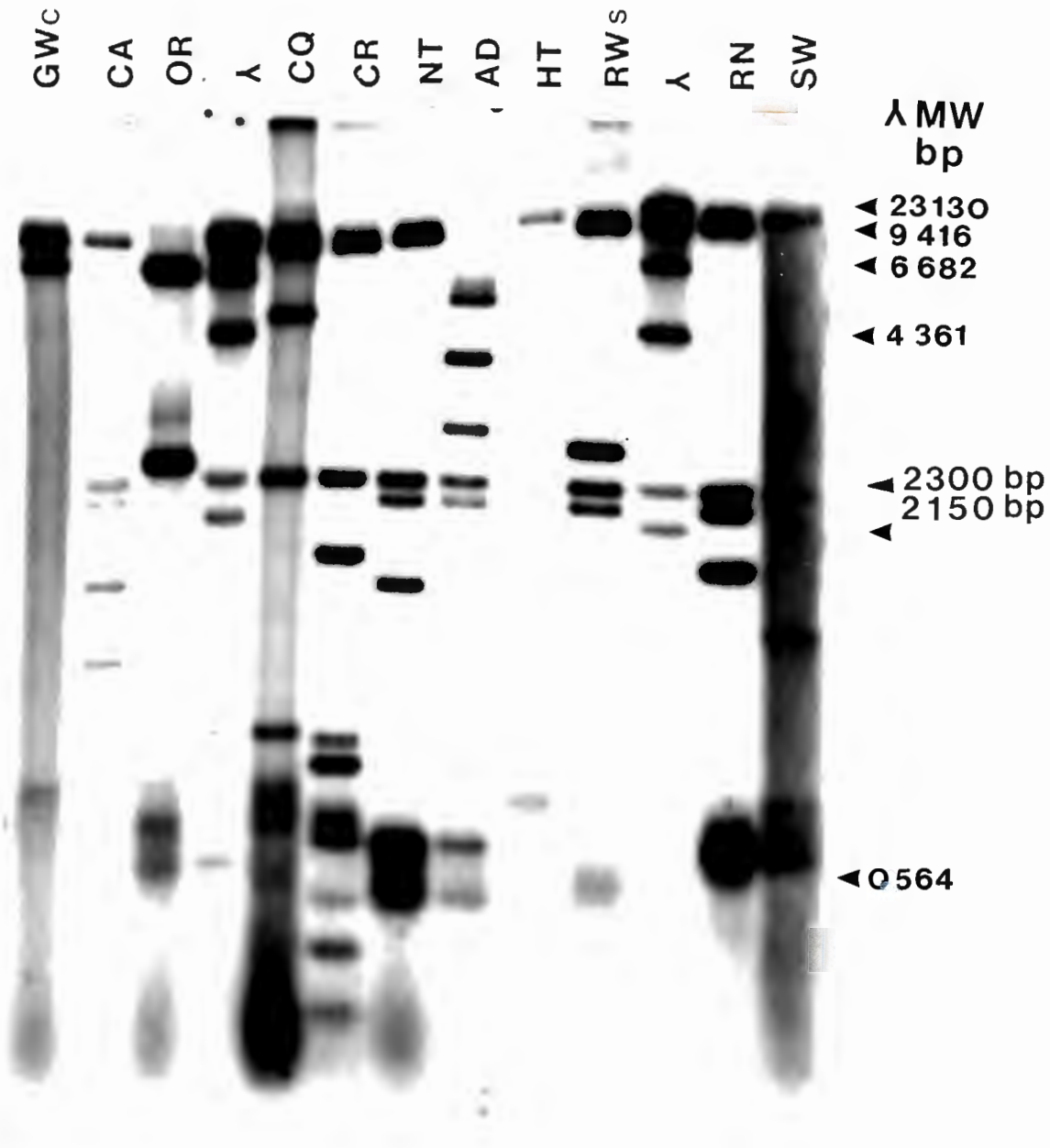


Fig. 10 Autoradiograph of electrophoretic profiles of Francolin mtDNA, digested with HindIII, and end -labelled. The fragments were separated in a 1.2% gel at approximately 2V/cm. Arrowheads indicate the doublet of mtDNA fragments (2300bp and 2150bp) common to most of the Francolin species studied. The size marker was an end-labelled HindIII digest of Lambda (λ) DNA. RWs = Redwing Partridge (Sabie); GWc = Greywing Partridge (Ceres); OR = Orange River Partridge;CQ = Coqui Partridge; CR = Crested Partridge; HT = Hartlaub's Francolin; NT = Natal Francolin; CA = Cape Francolin; AD = Red-Billed Francolin; RN = Red-Necked Francolin; SW = Swainson's Francolin.

recognition site. Only phylogenetic analysis can indicate whether site loss is more likely to have occurred than site gain, or vice versa.

3.3 Satellite DNA

Satellite DNA is highly repetitive nuclear DNA consisting of millions of copies of a single sequence - these repeated base sequences are of various lengths (Freifelder, 1983). It has been observed in many organisms and may make up 1-30% of the total DNA. Satellite DNA was evident in almost all of the gels in the fragment comparison, occurring usually in the region of 500-600bp. Its general appearance was that of a dark, fuzzy band, and caution was exercised when scoring fragments for the comparison, so as to exclude satellite bands from the data.

3.4 Outgroup

The Japanese Quail, Coturnix coturnix japonica, used as an outgroup species to define ancestral characters, appeared in the fragment comparison to contribute generally too few common bands with respect to any of the Francolins. However, it was included as an outgroup due to lack of any suitable alternatives.

3.5 Phylogenetic Analysis

3.5.1 Cladistic Approach

One way of estimating the degree of interrelationship between taxa is to present each distinct DNA fragment as a character. Each mtDNA character was coded by its presence or absence in a Francolin species, and the data set of characters was subsequently used to construct phylogenies, using appropriate cladistic programmes.

Those characters which were common to all the taxa in the comparison were considered to lack phylogenetic information concerning the relationships between them, i.e. it could be implied that the ancestor common to all the taxa possessed that character but that would represent the extent of the character's value. Similarly, a character unique to a particular Francolin species, termed an autapomorphy, bears no additional information in cladistic analyses. Therefore, the input data set from n taxa consisted of those characters which were shared by at least two taxa, and by no more than n-2 taxa. These characters are then collectively termed phylogenetically informative (Ferris et al., 1981). A table having 99 phylogenetically informative characters (See Table 2) was thus constructed, with the digit "1" denoting presence of a particular fragment, and "0" signifying absence of a fragment.

Table 2: Phylogenetically Informative mtDNA Fragment Characters

		Restriction Enzyme and Character No.						
		EcoRI	EcoRV	SacII	NcoI	BclI	PstI	StuI
			1111	111	1112	222222	22233	333333334444444444
		1234567	890123	456	7890	123456	78901	234567890123456789
	Coturnix	0000001	000000	001	0000	100000	01000	100000000010000100
"P A R T I D G E S"	sephaena	0100000	000000	001	0000	110100	10100	001100010010001000
	coqui	0100000	000000	001	0001	100000	00100	000000010010010011
	africanus	0000010	000000	001	0001	100110	10000	010000000110010100
	shelleyi	1000010	000000	001	0000	100110	10000	011000010010010000
	lev-oides	1000010	000000	001	0001	000010	10010	001010001010011100
	lev-antii	1100000	000000	001	0000	000000	00010	100110000010000000
"F R A N C O L I N S"	adspersus	1011100	100011	001	1110	011000	01000	100010001001000001
	natalensis	1011100	100011	110	0000	111000	01000	100011000101000001
	hartlaubi	1000001	000000	001	0000	010000	00000	100000000000100000
	capensis	1011100	100000	110	0000	111000	00000	100001000001100010
	afer	1001100	111000	001	1110	100001	01001	000000100010010010
	leucoscepus	1001100	100100	001	0010	100001	01001	100000100001000100
	swainsonii	1000000	011100	001	1010	000000	01001	000000000001000000

		XbaI	BamHI	HindIII	ScaI	SacI	PvuII	Asp718
		555	555555	5666666	6666777777	7777888	88888889999	999999
		012	345678	9012345	6789012345	6789012	34567890123	456789
	Coturnix	000	100000	1011010	0000000000	0001100	00100000000	000000
"P A R T I D G E S"	sephaena	000	000110	0001000	0010101000	0000000	10110100010	100000
	coqui	000	000110	0001000	0000001000	0000100	00000100000	000000
	africanus	000	001000	0000001	0011100100	0000000	10110100010	110110
	shelleyi	000	000000	0000101	0100100000	0100001	00000101000	110000
	lev-oides	100	110000	0000101	0101100100	0100001	00000101000	110110
	lev-antii	000	010001	0011010	0000010010	0000000	01000000000	100000
"F R A N C O L I N S"	adspersus	000	000000	0011000	1000000011	0010100	00001110100	100000
	natalensis	111	010001	0111000	1000000011	1010100	01001110000	001001
	hartlaubi	000	010001	1011000	0000000000	1000000	00000110001	000000
	capensis	111	000000	1111000	1000000011	0001100	00001110100	001001
	afer	111	000000	0011000	0000010010	0010010	10001110000	000000
	leucoscepus	111	001000	0011000	0000010010	0010010	00000110000	000000
	swainsonii	110	000000	0101000	0000010010	0000000	00000110001	000000

3.5.1.1 Cladistic Analysis of MtDNA Fragment Data

With Japanese Quail as outgroup, phylogenetically informative fragment characters (excluding autapomorphies) were analyzed cladistically using the programme Hennig 86 (J.S. Farris - based on Wagner Parsimony; version 1.5). In this analysis, the shortest possible phylogenetic tree(s) were found using the Implicit Enumeration (I.E.) option. This guarantees finding the shorter tree, or trees.

Hennig 86 analysis of the phylogenetically informative fragment characters resulted in 9 most parsimonious trees when rooted on Coturnix (see Fig.11(a)). A strict consensus tree was generated using Hennig 86's Nelsen option and is presented in Fig.11(b).

3.5.1.2 Comparison with Morphological Data

A study of morphological, behavioural and ecological characteristics of the same Francolin species was conducted by Crowe et al., 1990 (submitted). This yielded 24 phylogenetically informative qualitative characters (termed macrocharacters). The data set on which these analyses were based is available from T.M. Crowe (Fitzpatrick Institute of African Ornithology, University of Cape Town, RSA).

Morphometric investigations were performed by comparing the skeletons of 25 extant Francolin species (as well as other galliforms) to give 73 osteological characters. Similarly, morphometric studies were conducted on putative Francolin fossils from the mid-Miocene (ca. 15 x 10^6 y.b.p) and early Pliocene (ca. 5 x 10^6 y.b.p) deposits in Southern Africa. Results for these analyses are presented in more detail in Crowe et al., (1990, submitted).

The macrocharacter set for the same 13 Francolins was analyzed cladistically using the I.E. option of Hennig 86. One cladogram, rooted on a quail-like hypothetical ancestral outgroup (See Crowe and Crowe, 1985), was generated and is shown in Fig. 11(c).

The results of Hennig 86 analysis on fragment data were compared with trees from macrocharacter data for the same 13 Francolin species (see Discussion below).

The phylogenetically informative mtDNA fragment characters and Crowe's macrocharacters were combined and analyzed by Hennig 86, with the outgroup possessing a C.c.japonica mtDNA genotype and a macrocharacter phenotype of the hypothetical quail-like ancestor (Crowe and Crowe, 1985).

Four trees were produced, the strict consensus for which is shown in Fig. 11(d).

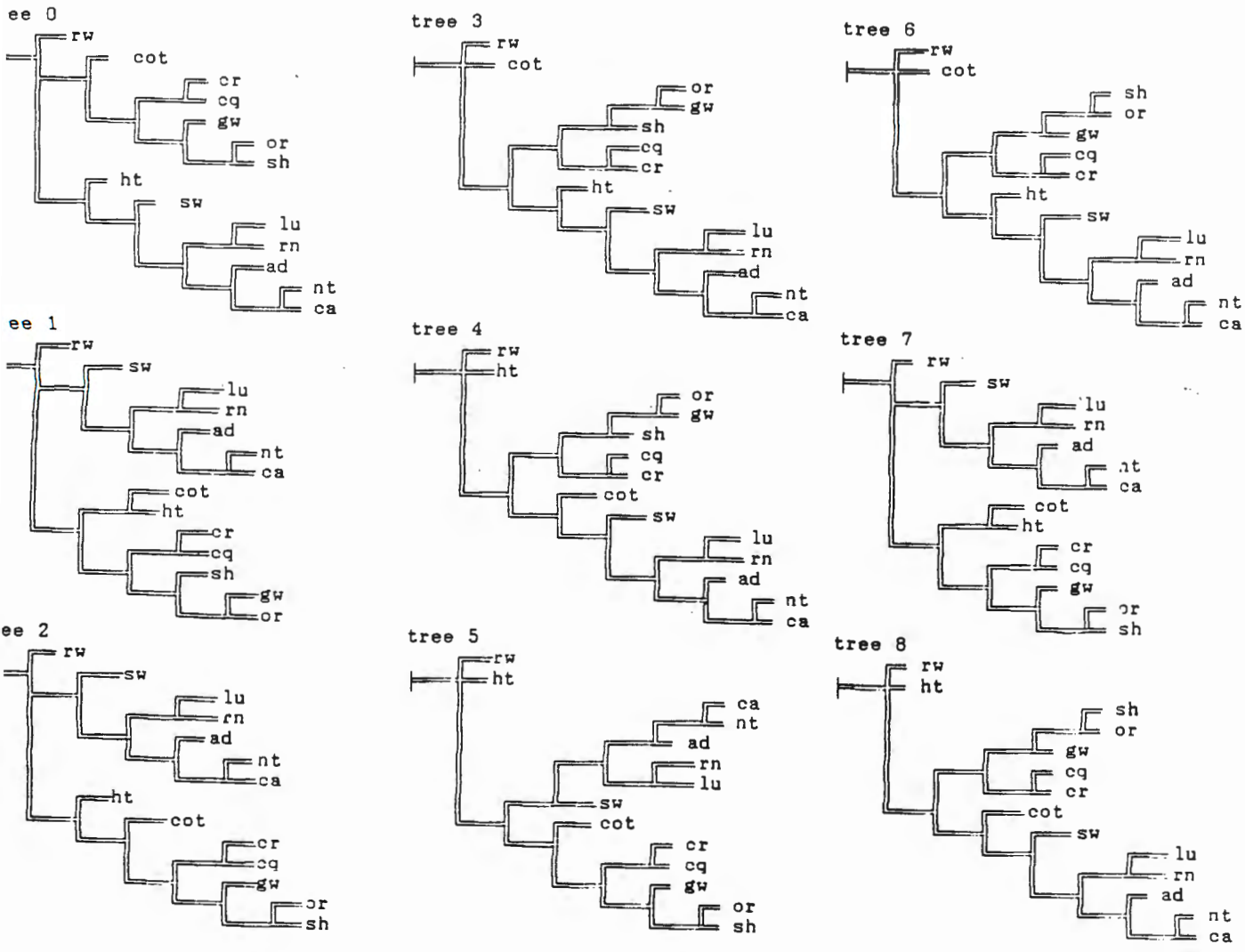


Fig. 11 (a) Nine cladograms inferred by Hennig 86 analysis (I.E. option) of phylogenetically informative Francolin mtDNA fragment characters.
Cot = Japanese Quail; RW = Redwing Partridge; GW = Greywing Partridge; OR = Orange River Partridge; SH = Shelley's Partridge; CQ = Coqui Partridge; CR = Crested Partridge; HT = Hartlaub's Francolin; NT = Natal Francolin; CA = Cape Francolin; AD = Red-Billed Francolin; RN = Red-Necked Francolin; SW = Swainson's Francolin; LU = Yellow-Necked Francolin. C.I. = 0.51.

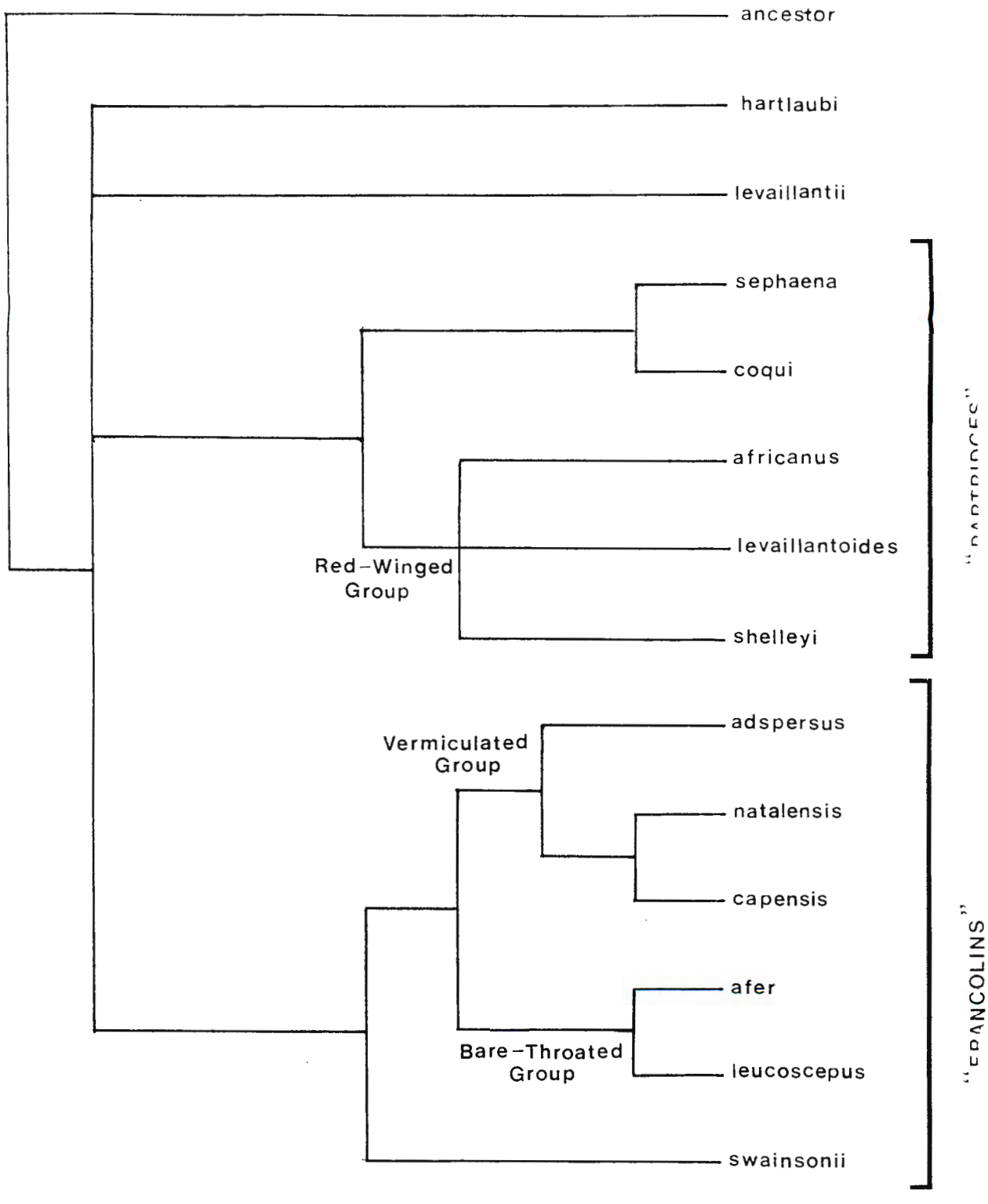


Fig. 11 (b) Strict consensus of 9 trees (Fig.11(a)) inferred by Hennig 86 analysis (I.E. option) of Francolin mtDNA fragment characters. C.I. = 0.51.

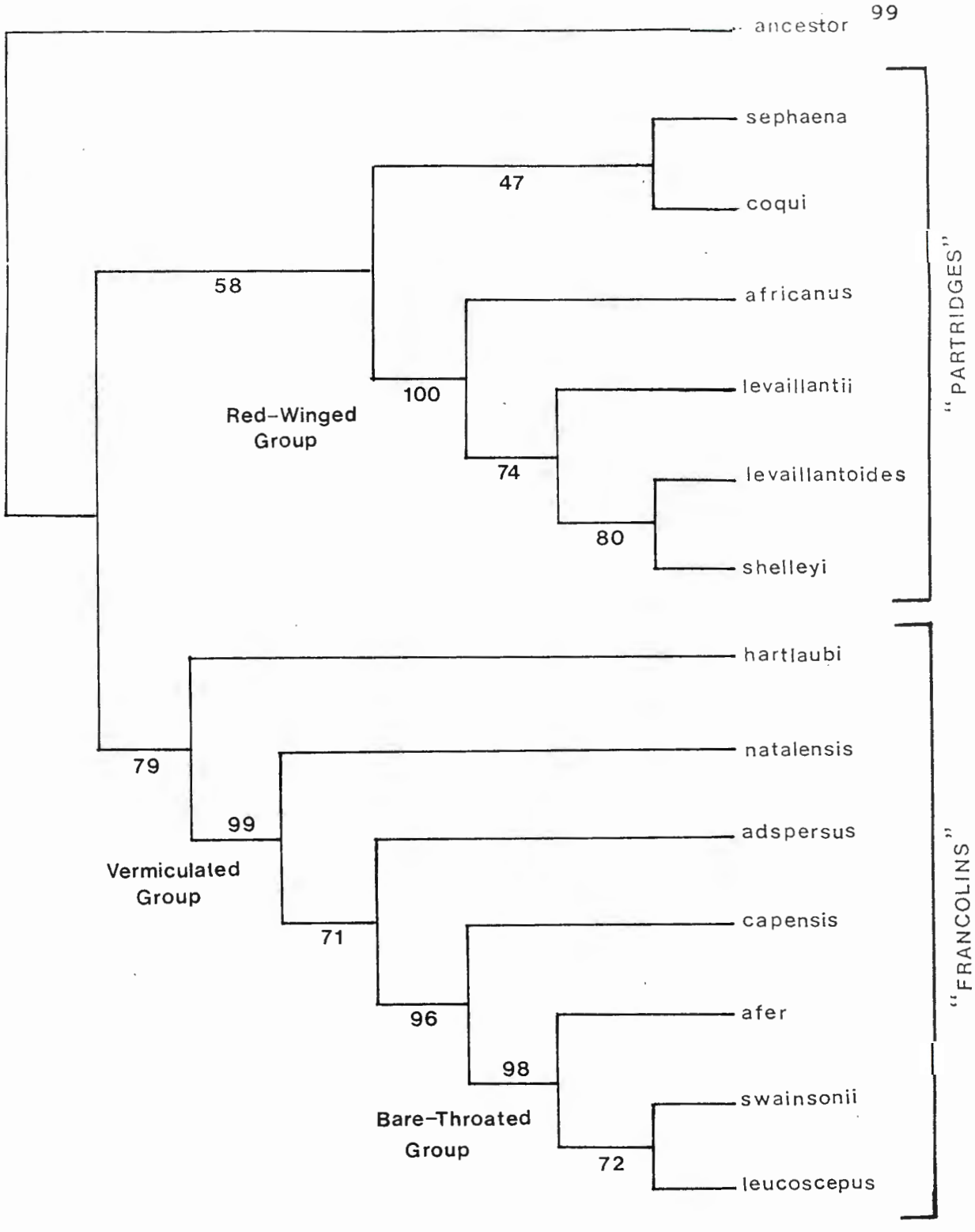


Fig. 11(c) Cladogram inferred by Hennig 86 analysis (I.E. option) of Francolin macrocharacters (Taken from Crowe et al., 1990, submitted). Bootstrapping frequencies are indicated beneath branches. C.I. = 0.77.

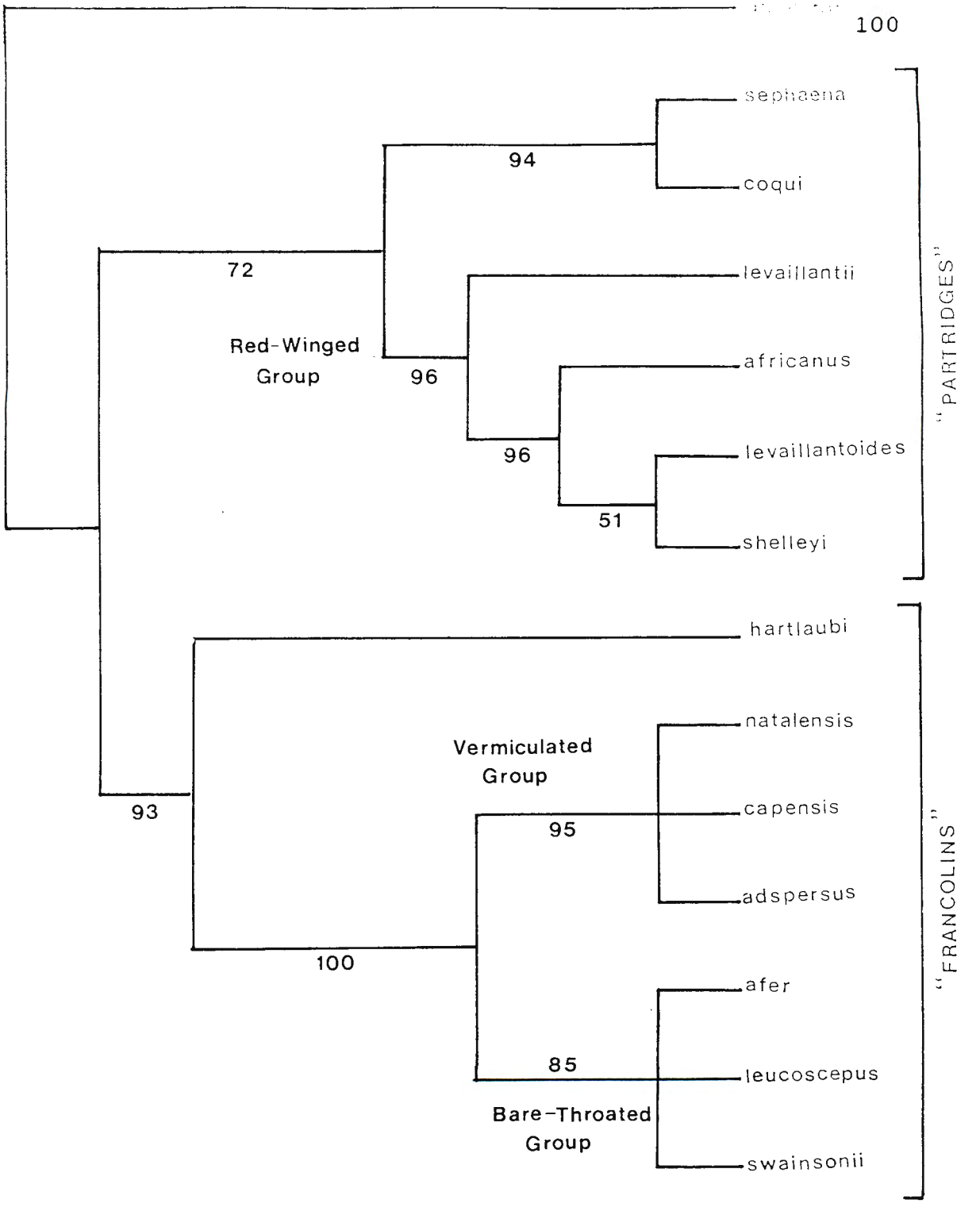


Fig. 11(d) Strict consensus of 4 trees inferred by Hennig 86 analysis of combined Francolin mtDNA fragment characters and macrocharacters (Crowe et al., 1990, submitted). Bootstrapping frequencies are indicated beneath branches. C.I. = 0.56.

3.5.1.3 Bootstrap Analysis

In order to test the soundness of the branch topology, the phylogenetically informative fragment data set was run, with 100 iterations, through the Bootstrapping programme (BOOT) in J. Felsenstein's PHYLIP (Phylogeny Inference Package, version 3.1). This programme resamples characters randomly, with replacement. Fig. 12 shows the tree and bootstrapping frequencies generated by BOOT upon analysis of fragment data.

Bootstrapping analysis was applied to the Hennig 86-produced phylogenies of the macrocharacter and combined data sets. In these analyses, all multistate characters were recoded in an additive binary fashion. Bootstrapping frequencies are indicated above the branches in the trees in Figs. 11(c) and (d).

3.5.2 Distance Measure Approach

An alternative to using the abovementioned approach is one based on the estimation of percentage sequence divergence (δ). In this approach, autapomorphies were included, but information from single cuts was excluded from analysis - Single-cutting restriction endonucleases, after cleavage at any single site, would each produce a linear fragment of uniform size (±16.4 kb) - without any

indication of the site location. Inclusion of these single linear bands in the analysis could result in a disproportionate number of such bands being scored as homologies.

3.5.2.1 Estimation of Percentage Sequence Divergence

To estimate the percentage sequence divergence between the base sequences of pairs of Francolin mtDNA, the following approach was used:
The method, based on proportion of shared fragments between Francolin mtDNA pairs, was taken from equation 20 of Nei and Li's (1979) paper. As mentioned above, this assumes that fragments of coincident mobility are homologous, and it takes no account of back mutations. The method is a revision and extension of that originally proposed by Upholt (1977).

Proportion of shared fragments (F) was calculated from:

$$F = 2n_{xy} (n_x + n_x)$$

where n_x and n_y represent the total number of fragments in species x and species y, respectively,
and n_{xy} denotes the number of fragments common in size to both species x and y.
Only 6 bp-recognizing restriction enzymes were used, which simplifies the ensuing calculations.

Bootstrapped mixed parsimony algorithm version 3.1
14 species, 99 characters
100 replicates, random number seed = 9
Wagner parsimony method

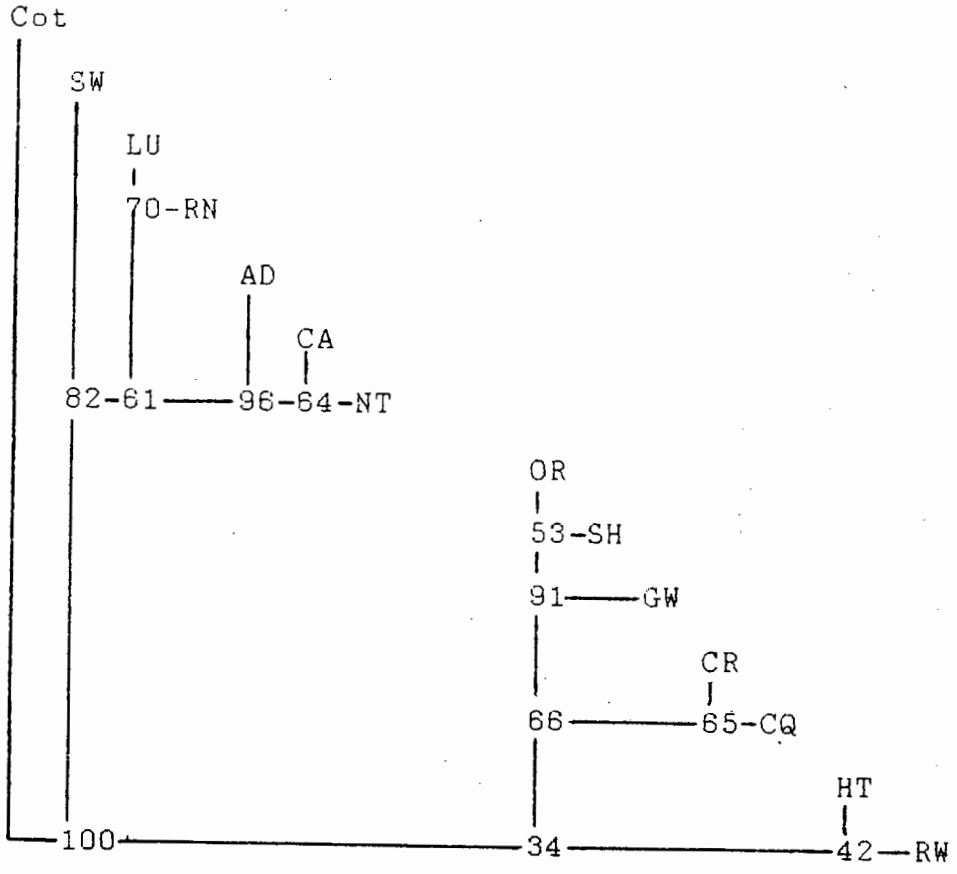


Fig. 12 Tree resulting from Bootstrap analysis of Francolin mtDNA fragment characters (BOOT, in PHYLIP). 100 replicates were made. Bootstrapping frequencies are indicated at the nodes.
Cot = Japanese Quail; RW = Redwing Partridge; GW = Greywing Partridge; OR = Orange River Partridge; SH = Shelley's Partridge; CQ = Coqui Partridge; CR = Crested Partridge HT = Hartlaub's Francolin; NT = Natal Francolin; CA = Cape Francolin; AD = Red-Billed Francolin; RN = Red-Necked Francolin; SW = Swainson's Francolin; LU = Yellow-Necked Francolin.

Table 3 shows all calculated F values for the pairwise comparisons, in the upper right-hand corner.

Nei and Li (1979) provide a graphic method for estimation of the percentage sequence divergence (δ) (See also Upholt, 1977). This required the plotting of a standard curve (see Fig.13) in which F was calculated from given values of δ, using the formula:

$$F \approx P^4 / (3 - 2P)$$

where $P = e^{-r\lambda t}$;

r = 6, the number of base pairs in the restriction enzyme recognition sequence;

λ is the mutation rate;

t = time and $\delta = 2\lambda t$

However, it should be borne in mind that since the relationship between F and δ is curvilinear, small errors at low F values give rise to large errors in the estimation of δ values.

Alternatively percentage sequence divergence values were computed via the Iterative method of Nei and Tajima (1983; see Fig. 14).

Thus fragment patterns of individuals were compared by the construction of a grid (See Table 3), which shows values of δ in the lower left-hand corner.

Table 3: Matrix of percent nucleotide divergence estimates (δ) (lower half matrix) and proportion of shared mtDNA restriction fragments (upper half matrix) (Nei and Li, 1979) for Francolinus spp and Coturnix c. Japonica.

		"PARTRIDGES"						"FRANCOLINS"							
		afr	lvi	levo	she	coq	sep	nat	ads	har	cap	afer	swa	leu	cot
"PARTRIDGES"	africanus		.154	.521	.508	.225	.448	.123	.105	.097	.113	.211	.123	.206	.212
	levaillantii	11.7		.265	.207	.212	.258	.342	.310	.316	.212	.282	.233	.317	.262
	levaillantoides	3.8	8.1		.606	.189	.257	.167	.177	.123	.135	.203	.176	.197	.203
	shelleyi	4.0	9.7	2.9		.250	.400	.135	.145	.109	.156	.232	.172	.197	.169
	coqui	9.2	9.6	10.3	8.5		.382	.171	.182	.127	.194	.234	.152	.174	.239
	sephaena	4.7	8.3	8.3	5.4	5.7		.128	.184	.169	.176	.219	.161	.185	.222
	natalensis	13.3	6.4	11.2	12.6	11.0	13.0		.621	.247	.707	.414	.316	.481	.208
"FRANCOLINS"	adspersus	14.4	7.1	10.8	12.1	10.6	11.3	2.7		.235	.597	.439	.338	.459	.250
	hartlaubi	14.9	6.9	13.3	14.1	13.0	11.1	8.5	8.9		.286	.206	.140	.200	.276
	capensis	13.8	9.6	12.6	11.6	10.1	10.8	2.0	3.0	7.6		.442	.333	.493	.269
	afer	9.6	7.7	9.8	9.0	8.9	9.3	5.2	4.9	9.7	4.8		.479	.649	.222
	swainsonii	13.3	8.9	10.8	11.0	11.8	11.4	6.9	6.5	12.4	6.6	4.3		.508	.164
	leucoscepus	9.7	6.9	10.0	10.0	10.9	10.5	4.3	4.6	9.9	4.1	2.5	4.0		.281
	Coturnix c. japonica	9.6	8.2	9.8	11.1	8.8	9.2	9.7	8.5	7.8	8.0	9.2	11.3	7.7	

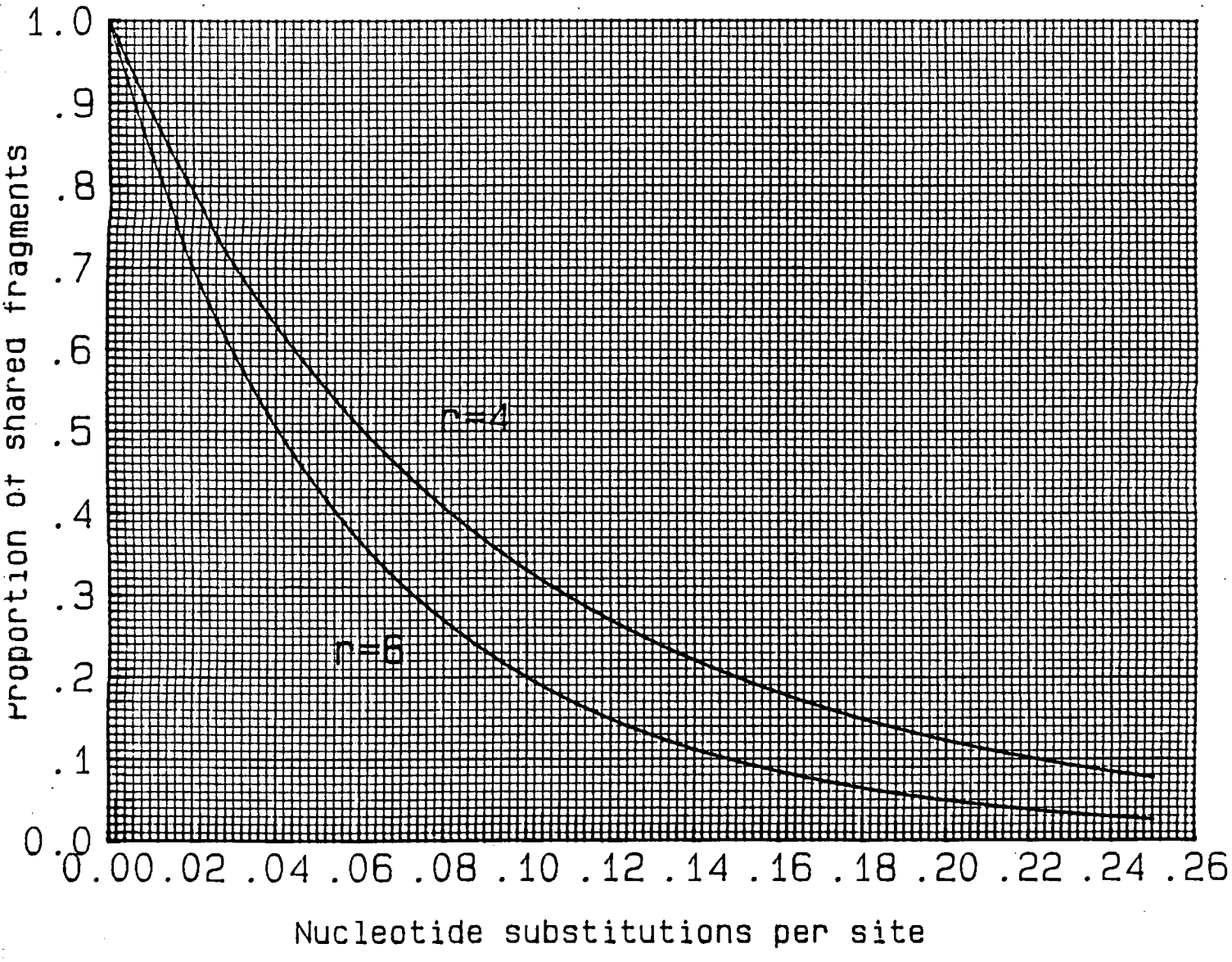


Fig. 13 Relationship between the proportion of shared DNA fragments and the number of nucleotide substitutions per site. r = the number of bases in the restriction endonuclease recognition sequence (Nei and Li, 1979). The curve used was for r = 6.

```
SUB  SDFROMG
     LET r=6
     LET G=F^.25
     DO
          LET X=G
          LET G=(F*(3-2*X))^.25
          ! PRINT G,X
     LOOP UNTIL IP(G*1E6)=IP(X*1E6)
     IF G=0 THEN LET d,SD(K,J)=0 ELSE LET
     d,SD(K,J)=ROUND(-(2/r)*LOG(G),3)
END SUB
```

Fig.14 True Basic subroutine used to compute percentage sequence divergence values.
r = number of bases in RE recognition sequence, 6 in this case.
F = proportion of shared fragments
d = percentage sequence divergence
SD = array into which sequence divergence values are put between taxa K and J.

3.5.2.2 Distance-based Analysis of MtDNA Fragment Data

The distance matrix served as input data for analysis by the distance measure approaches FITCH (which uses the Fitch-Margoliash method) and KITSCH (Fitch-Margoliash with contemporary tips) in PHYLIP. Each of these yielded a putative phylogenetic tree of best fit (Figs. 15(a) and (b)).

3.6 Intraspecific Polymorphism

The additional specimens of F.africanus and F.levaillantii, collected some hundreds of kilometres from the original collection localities, were tested for any significant intraspecific polymorphism in their mtDNA's (i.e. fragments present in some individuals and absent in others of the same species). Fragment comparisons between the individuals and those from the original collection localities, showed that they were virtually identical in their mtDNA's, with very few polymorphic fragments; digestion of levaillantii (from Sabie and Giant's Castle Nature Reserve (Drakensberg) districts, see Table 1) with Stu I showed a polymorphism in the form of an extra Stu I site which yielded a 600bp fragment in the Drakensberg individual. This fragment was not present in the Sabie specimen.

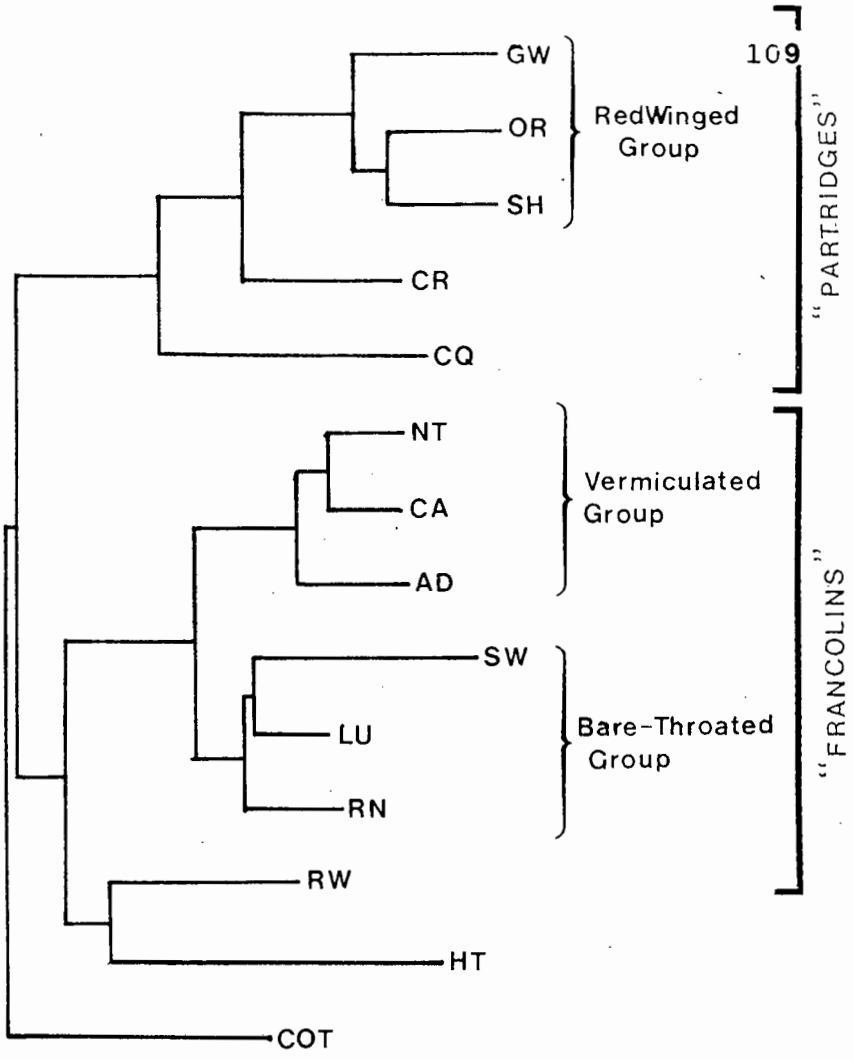
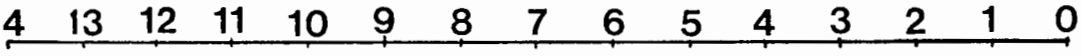


Fig. 15(a) Tree of best fit inferred by FITCH (in PHYLIP) analysis of Francolin mtDNA fragment data.
Cot = Japanese Quail
RW = Redwing Partridge
GW = Greywing Partridge
OR = Orange River Partridge
SH = Shelley's Partridge
CQ = Coqui Partridge
CR = Crested Partridge



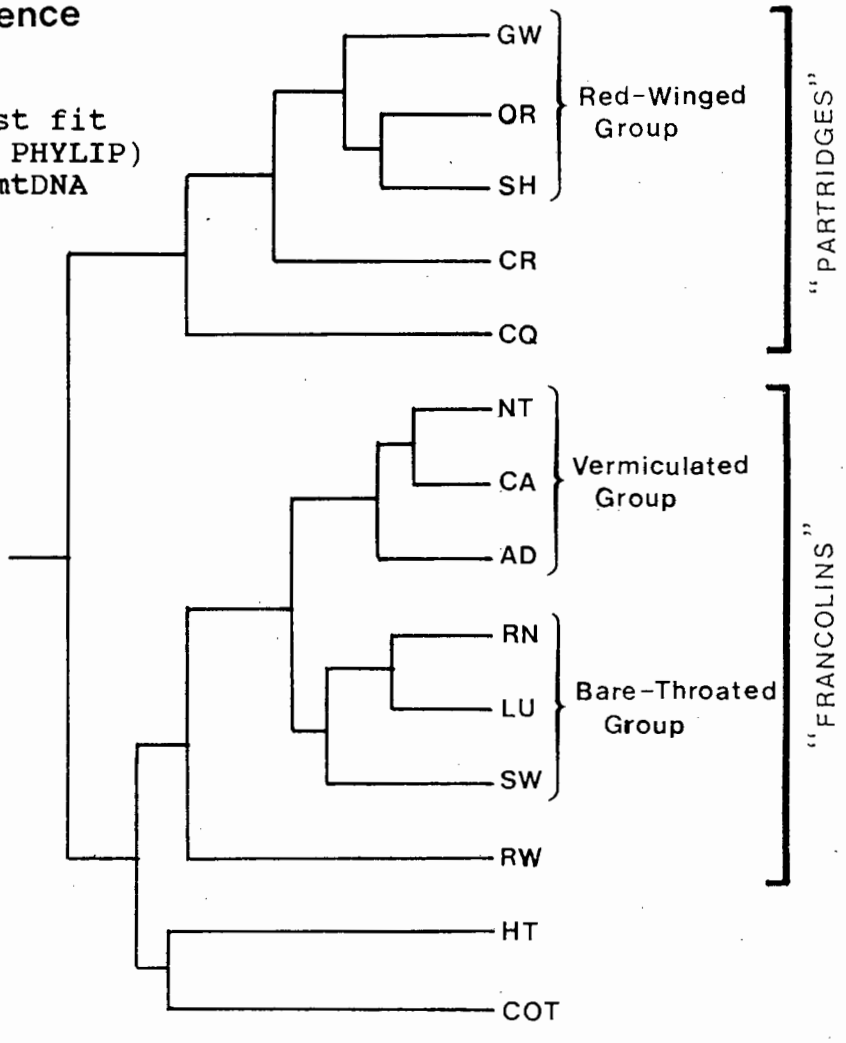


Fig. 15(b) Tree of best fit inferred by KITSCH (in PHYLIP) analysis of Francolin mtDNA fragment data.
HT = Hartlaub's Francolin
NT = Natal Francolin
CA = Cape Francolin
AD = Red-Billed Francolin
RN = Red-Necked Francolin
SW = Swainson's Francolin
LU = Yellow-Necked Francolin

Digestion of africanus (from Ceres and Molteno Districts, see Table 1) with Stu I showed the presence of an extra restriction site in the Molteno specimen, which was not evident in the Ceres individual. This polymorphic site resulted in the generation of an extra 800bp and a 300bp fragment upon digestion, which together made up a 1100bp fragment in the Ceres individual.

In addition, digestion of these two africanus specimens with Sac I showed the presence of an extra site in the Molteno individual. This resulted in two fragments of 10 000bp and 6 400bp, where the Ceres specimen exhibited only one Sac I site and therefore a single mtDNA fragment of 16 400bp.

Percentage sequence divergence values between the two Greywing Partridges, and the two Redwing Partridge representatives, were only equal to 0.3 and 0.1 respectively. This was an order of magnitude lower than those for most of the interspecific comparisons, suggesting that there is little geographic variation in the RFLP's of Greywing and Redwing Partridge.

Therefore, the additional specimens of africanus and levaillantii from Molteno and Drakensberg, respectively,were not included in Tables 2 and 3 or subsequent distance and cladistic analyses.

Table 1 lists the number of individuals from which mtDNA was extracted for each species, both male and female. Although

this was a limited survey, it was felt that the analysis of only single individuals in certain cases, for the fragment comparison, could be justified for the purpose of this systematics study, given such low levels of intraspecific polymorphism. Indeed, where the mtDNA's of more than one individual per Francolin species were extracted for mapping purposes (see Chapter 4 and Table 1), no polymorphisms were found.

DISCUSSION

The data resulting from a Francolin mtDNA fragment approach generated some interesting and rather controversial results. The two methods of phylogenetic analysis resulted in a major difference in topology between the mtDNA-based trees and the traditional morphological tree, in the position of F.hartlaubi and F.levaillantii.

3.1 Cladistic Analysis

Hennig 86 analysis resulted in an unresolved quadrichotomy towards the roots in the consensus tree (Fig.11(a)).

Although F.levaillantii is a "partridge" in phenotype, it appeared in the fragment comparison to most closely resemble a "francolin" in mtDNA genotype. This destroyed the monophyly of Milstein and Wolff's (1987) proposed "partridge" clade (see Fig.5). It also opposed Hall's designation of levaillantii to the Red-Winged Group (Fig.5).

Of the 11 fragment characters in F.levaillantii not shared with Coturnix (Table 2), 3 were shared with both "partridges" and "francolins", and 4 each with only "partridges" or "francolins". Furthermore, F.levaillantii tended to lack those fragments (e.g., characters 6, 25 and

95 in Table 2) which were synapomorphies for the other three Red-Wing species analysed here (africanus, levaillantoides and shelleyi).

The genotypic similarity of F.levaillantii to the "francolins" was very surprising, as Hall has named the Red-Winged Group after the common name of this typical "partridge".

If the fragment data is taken at face value, the following arguments would be needed to support the "francolin" genotype of F.levaillantii:

1) Perhaps as recently as 3 x 10^6 y.b.p. there was gene flow between "partridges" and "francolins" such that levaillantii males hybridized successfully with female "francolins", and "francolin" mtDNA introgressed into levaillantii through subsequent back-crosssing of fertile hybrid females with levaillantii males;

2) F.levaillantii represents a relict species having an ancestral quail-like phenotype and mtDNA haplotype; or

3) F.levaillantii is a "francolin" which has convergently taken on "partridge" macrocharacters.

Detailed mapping or nucleotide sequencing might narrow down these possibilities and provided a deeper resolution of the phylogeny.

The other aberrant Francolin, F.hartlaubi, appeared not to be as clearly associated with the Vermiculated monophyletic group as originally postulated by Hall (1963, Fig.5). F.hartlaubi appeared, in fact, to be paraphyletic to all the other Francolins. This would throw into doubt Milstein and Wolff's proposal of a true monophyly for the "francolin" clade.

At this point, it was surmised that F.hartlaubi might be the product of an early divergence from the Francolinus lineage, and could be a sister-species of the remaining Francolins. It was thought that a RE site comparison would establish whether hartlaubi is indeed a sister-species of the Francolins.

The cladograms inferred by Hennig 86 analysis of mtDNA fragment data thus appeared to support the monophyly of Hall's Bare-Throated Group, but not that of her Red-Winged and Vermiculated Groups - the latter groups being disrupted by the aberrant positioning of F.levaillantii and F.hartlaubi, respectively.
Milstein and Wolff's "partridge" - "francolin" hypothesis was generally supported by fragment data, except for the topology of the above two species.

The tree derived from macrocharacter data (shown in Fig.11(c)) displayed discrete "francolin" and "partridge"

clades. The consensus mtDNA phylogeny accorded well with the topology of this tree, with the exception of levaillantii and hartlaubi. The molecular cladogram did associate F.natalensis and F.capensis more closely than the macrocharacter phylogeny, and the same could be said for F.afer and F.leucoscepus which showed a paraphyletic arrangement in the macrocharacter cladogram (Fig.11(c).

Crowe et al.'s cladistic analysis of macrocharacter data supported the monophyly of Hall's Bare-Throated and Red-Winged Groups, and favoured the "partridge" - "francolin" dichotomy. The members of her Vermiculated Group, including F.hartlaubi, formed a paraphyletic assembly.

The cladogram of combined mtDNA and macrocharacters (Fig. 11(d)) presented a compromise phylogeny which, again, was in accordance with Hall's Bare-Throated and Red-Winged Groups, but not with the Vermiculated Group - although three of the four representatives of this group (F.adspersus, F.capensis and F.natalensis) clustered together monophyletically. Here a monophyletic "partridge" clade shows a macrocharacter data bias in incorporating levaillantii, and hartlaubi roosted comfortably within a "francolin" clade. The combined consensus phylogeny reflected a partiality toward molecular data in its more definite clustering of natalensis, capensis and adspersis, and of afer, leucoscepus and swainsonii into separate monophyletic groups. The topology of the tree

derived from the combined data set supported Milstein and Wolff's hypothesis.

Bootstrapping of fragment data is indicated in Fig.12. The frequencies for the occurrence of monophyletic groups are indicated at branch nodes.
The monophyletic grouping of the "francolins" was relatively robust, as shown by a figure of 82% for this group in the bootstrap. However, the corresponding figure of 34% for the "partridge" clade was relatively weak on account of the tendency for some taxa, presumably usually hartlaubi or levaillantii, to come out in the "francolin" clade in some individual bootstrap replicates.

The monophyly of Hall's Bare-Throated Group (swainsonii, leucoscepus and afer) was not supported, whereas that of the Vermiculated Group was supported strongly at 96%, as was the Red-Winged Group of levaillantoides, shelleyi and africanus, (but excluding levaillantii) with a figure of 91%.
The monophyly of sephaena and coqui was weakly supported, possibly because each hails from a separate monophyletic group in Hall's assignation of Francolins (Striated and Red-Tailed Groups, respectively). The monophyletic grouping of hartlaubi with levaillantii was weak.

Bootstrapping frequencies for the macrocharacter data and the combined tree are indicated at the branches in Fig.11(c) and Fig.11(d), respectively.

The tree representing a combined data set showed bootstrapping frequencies which endorsed the monophyly of a "francolin" clade more strongly than the "partridge" monophyly.

3.2 Distance-Based Analysis

The lowest percentage sequence divergence (δ) values between pairs of Francolin species ranged from 2.0 (between F.capensis and F.natalensis) to 6.9 (between F.levaillantii and F.hartlaubi). The mean δ's between Coturnix and the "partridges" and "francolins" were 9.5 and 8.9 respectively. These relatively high δ values fall into the range reported for inter-generic comparisons (Shields and Helm-Bychowski, 1988).

Percentage sequence divergence values for F.hartlaubi were also high (ranging from 11.1 to 14.9) relative to the "partridges" (excluding the anomalous F.levaillantii), as well as to the "francolins" (ranging from 7.6 to 12.4).

The placement of F.hartlaubi as sister-taxon to the Francolins was perhaps not as unexpected as that of F.levaillantii as a "francolin"; Hartlaub's Francolin is the

most distinctive species, genetically and morphometrically, from the other Francolins (Crowe et al., 1990, submitted), and has a highly specific habitat of isolated rocky hills surrounded by sub-desert conditions. It differs from the other Francolins in its small size and very complex antiphonal advertisement call (Komen, 1987). Mating behaviour is notably female-dominated during copulation, and like Coturnix spp., males have extremely large, ovoid testes 2 to 3 times the size of those of any other Francolin.

Distance-based phylogenies for the fragment data were generally in concordance with cladistic topology, where both FITCH and KITSCH analyses exhibited monophyly of Hall's Bare-Throated Group, but resisted the monophyly of the Red-Winged and Vermiculated Groups because of the aberrant positioning of levaillantii and hartlaubi.

FITCH analysis of fragment data resulted in an unrooted tree of best fit, since the programme does not assume a constant rate of evolution (Fig.15(a)). The branch lengths of the tree represent the amount of evolutionary change, and its topology was generally similar to the cladogram produced by Hennig 86, but again, discounted Milstein and Wolff's "partridge" - "francolin" dichotomy hypothesis by the tenuous positioning of hartlaubi and levaillantii. In fact, FITCH grouped these taxa together in a monophyletic group, distinct from the remaining species. Other minor

differences included a paraphyletic branching of coqui and sephaena where Hennig 86 clustered them, and a preference for swainsonii to group with leucoscepus where cladistic analysis favoured the clustering of afer with leucoscepus.

The KITSCH approach, assuming that the molecular clock runs true to time (but subject to stochastic variation), gave a rooted phylogeny having very similar branch lengths to those of the FITCH tree. A KITSCH tree presents branch lengths as being indicative of mutational change, but differs from the FITCH approach in making the branch tips contemporaneous, by averaging their distances (Fig.15(b)). The KITSCH tree generated from fragment data was very similar to that of FITCH, but showed rearrangement of the three taxa levaillantii, hartlaubi and Coturnix near the root of the tree.

3.3 General Discussion on Fragment Approach

Clearly, a fragment comparison was able to indicate the differences between the mitochondrial genomes of the individual Francolins studied. Analysis of data from this approach presented a very basic estimation of Francolin phylogeny.

It must be noted here that in the present comparison, the assumption of homology between co-migrating fragments in this comparison may sometimes be erroneous; fragments

differing only slightly in length will tend to co-migrate, and so small differences in length will not be detected. Often a tiny fragment making up the difference has migrated off the gel and is therefore not visible. Alternatively, bands of almost identical lengths may be produced in different parts of the mt genome, particularly in the case of RE's, like StuI, that cut in many positions and give rise to very numerous fragments.

This phenomenon of convergence is probably negligible in sequences from the more closely-related species. However, the likelihood of convergence increases as the sequences exhibit greater divergence (Shields and Helm-Bychowski, 1988). In this case, convergence would contribute to the homoplasic "noise" in the deeper branches of the phylogenetic tree, and therefore to uncertainty in the topology (See Chapter 4, Discussion Section 4.2).

A further caveat attached to the use of fragment data in phylogenetic analysis is that there is no assumption of independence of characters, e.g. if a new site evolves between two pre-existing sites, one longer fragment disappears and two shorter ones appear. Thus, even though two species may share two of the three sites, they have no fragments in common, and a potentially significant source of error arises.

Similarly, several assumptions (that are violated) accompany the Nei-Li (1979) distance measure model and should be borne in mind when utilizing their equations to estimate δ values. These include:

i) that common restriction sites originate from one original ancestor - Lansman et al.(1983b) noted the existence of hypervariable RE sites which switch back and forth between being "on" and "off" with respect to enzyme recognition. Taking into account the known high transition frequency, such sites usually alternate between two nucleotides, i.e. between two purines or two pyrimidines;

ii) that nucleotides are randomly arranged throughout the genome - this assumption is also violated (Adams and Rothman, 1982). However, Nei and Li (1979) claim that the distribution of nucleotides would have to be extremely non-random to affect their model;

iii) that all restriction fragments are detected - as previously noted, this may not be the case for very small fragments, however Nei and Tajima (1981) are of the opinion that "the effect of elimination of small fragments is generally unimportant".

Despite these violations, however, Quinn and White (1987) postulate that the consequences thereof are small when closely related mtDNA's, like those of the Francolins, are being compared.

The observed deviation of F.levaillantii and F.hartlaubi from the traditional morphological phylogeny, resulting in an unresolved quadrichotomy at the base of the molecular trees (e.g. Fig.11(b)), called for a more detailed resolution of the Francolin phylogeny. It was hoped that some clarification as to the phylogenetic identity of the two deviant Francolins would be provided by a Restriction Endonuclease site mapping approach.

CHAPTER 4

COMPARISON OF FRANCOLIN MITOCHONDRIAL DNA RESTRICTION ENZYME SITES

INTRODUCTION

A restriction endonuclease site mapping approach, in positioning RE cleavage sites on the Francolin mitochondrial genomes, would allow a more rigorous approach to the phylogenetic analysis of the Francolins.

Although this approach is considerably more time-consuming than mtDNA fragment analysis, it provides a more precise set of molecular data. This results in finer resolution of phylogenetic trees at the species and genus level.
In addition, mapping often allows the identification and localization of sequence rearrangements and duplications (Palmer et al., 1985; Moritz and Brown, 1987).

In the time available, it was not possible to map all the species examined in the fragment study, so the species chosen for mapping included the following: F.capensis (Cape Francolin) and F.africanus (Greywing Partridge) as typical representatives of the "francolin" and "partridge" groups, respectively; those species having less certain or more problematical placement in the fragment study, namely

F.levaillantii (Redwing Partridge) and F.hartlaubi (Hartlaub's Francolin), as well as a species which was obtained late in the study, Margaroperdix madagascariensis (Madagascar Partridge). It was hoped that Margaroperdix would provide a suitable outgroup, since it is not a mainland African species.

The construction of maps generally entails the use of various techniques: double digestion with two or more RE's together, followed by end-labelling, partial digestion or end-shortening with endonuclease Bal 31 to visualize the mtDNA fragments (See Avise et al., 1979b; Brown et al., 1979; Harley, 1988). The method of visualization adopted in this approach was that of end-labelling.

Sequencing studies have shown that the Nei-Li (1979) distance measure model makes a reasonable prediction of the average sequence divergence between closely related species, where at least 40 restriction sites per sample are examined, using several different enzymes (Brown et al., 1982; Aquadro and Greenberg, 1983), therefore an average of 50 RE sites per bird was mapped.

METHODS AND MATERIALS

Unless otherwise specified, all methods and materials utilized in the generation of Francolin mtDNA cleavage maps are identical to those described in Chapter 2.

4.1 Sources of Birds

Specimens of each of 4 species of Francolin were collected at various districts in Southern Africa (Table 1). These were as follows: F.africanus (Greywing Partridge), F.levaillantii (Redwing Partridge), F.hartlaubi (Hartlaub's Francolin), F.capensis (Cape Francolin), and M.madagascariensis (Madagascar Partridge). (Note: F.africanus and F.levaillantii samples were representatives of the original collection localities - Sabie and Ceres respectively; Table 1). The number of individuals used per species in mapping is listed in Table 1.
Margaroperdix madagascariensis was obtained from a private breeder in Pretoria, Transvaal.

4.2 Isolation and Purification of Mitochondrial DNA

Tissue samples were prepared and mitochondrial DNA was isolated and purified as described in Chapter 2, Section 2.1.4.

4.3 Restriction Enzyme Digestion

MtDNA's from each of the 4 Francolin species and the Madagascar Partridge were digested with the same 14 RE's (Chapter 3, Methods and Materials, Section 3.3) that were used in the fragment approach, with the addition of Bgl II, Hpa I, Nhe I, Sal I and Xho I.

Conditions for RE digests were identical to those described in Chapter 2, Section 2.2. 1x KGB Buffer (see Appendix) was compatible with requirements for double-digestions in mapping (see Section 4.7.1 below).

4.4 Agarose Gel Electrophoresis

Restriction enzyme-digested mtDNA fragments resulting from double-digests were separated according to size as described in Chapter 2, Section 2.3.

4.5 Visualization of RE Fragments

MtDNA fragments in the RE site mapping approach were end-labelled as described in Chapter 2, Section 2.4.1. Gel-drying and Autoradiography were as described in Chapter 2.

4.6 Calibration of Molecular Weight

The molecular weights of mtDNA double-digestion products were determined relative to a standard molecular weight marker. This was a Hind III digest of phage Lambda, as described in Chapter 2, Section 2.5. (see also Appendix II, and Fig. 7).

4.7 Construction of Cleavage Maps

MtDNA Restriction Endonuclease maps were constructed by determining the order of recognition sites for each RE, and their location relative to cleavage sites for other RE's. The method employed was that of double-digestion.

4.7.1 Double Digestion Experiments

The general strategy for RE mapping was initiated with a range of single digests on each of the Francolin and the outgroup mtDNA's. Nineteen RE's were used. The cleavage patterns for 14 of these enzymes were provided from the fragment comparisons in Chapter 3. The other 5 enzymes used were Bgl II, Hpa I, Nhe I, Sal I and Xho I.

Double-digests began with three pairwise combinations of three enzymes, using those which cut the least number of times (preferably one or two cuts per enzyme). For example,

Cape Francolin mtDNA displayed single cleavage sites for Bam HI, Pst I, Hpa I and Bgl II (See Fig. 16).
Initial pairwise combinations chosen for double-digestion in Cape Francolin were therefore as follows:

Bam HI Pst I

Bam HI Hpa I

Pst I Hpa I

Mapping of data was done either by hand, or by computer, using the programme RESOLVE (E.H. Harley, version 2.0), which catalogues, edits, manipulates and manages sets of unmapped data from double-digests.
Pairwise combinations giving digestion products with a total of 2-8 fragments were generally easily mapped by RESOLVE, which provides possible alternative solutions.
However, more complex double-digest problems giving 8 or more digestion products were above the programme's limit for accurate resolution, and so were solved manually.

Thus, the "core" RE's (initially chosen for their simple cleavage patterns), once mapped, were used in double-digests with the remaining RE's, in this way building up a map. For the Cape Francolin, therefore, each of the single sites for Bam H I, Pst I and Hpa I was mapped relative to the single Bgl II site (see Fig. 16 and Table 5). This provided

a relatively well-spaced set of recognition sites on the genome for the remaining RE sites to be mapped.

Finally, the relative positions of closely-spaced sites, generated by RE's with many recognition sites, were tested by selected double-digests. This was more easily done once the map was filled with well-spaced sites for the purpose of reference. The accuracy of the provisional map was tested in this manner.

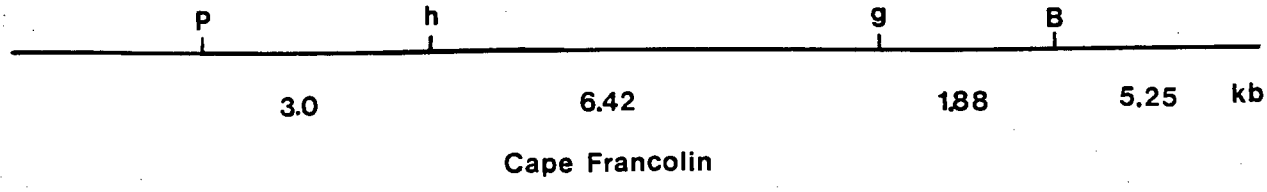


Fig. 16 Preliminary restriction map of Cape Francolin mtDNA, showing restriction sites for restriction enzymes B = BamHI; g = BglII; h = HpaI; P = PstI.

RESTRICTION ENZYMES USED	SIZES OF FRAGMENTS GENERATED
BamHI	16.40
PstI	16.40
HpaI	16.40
BglII	16.40
BamHI and PstI	11.3; 5.25
PstI and HpaI	13.40; 3.00
BamHI and HpaI	8.30; 8.10
BglII and PstI	8.88; 7.52
BglII and HpaI	6.42; 9.95
BglII and BamHI	1.88; 14.92

Table 4: Fragment sizes obtained in mapping of BamHI, PstI, HpaI and BglII sites in Cape Francolin mtDNA.

RESULTS

4.1 Generation of Restriction Maps

The cleavage maps for the mtDNA of four Francolin species and Margaroperdix madagascariensis (Fig.17(a)) provided a total of 253 RE sites (with an average of 50 sites per bird) after digestion with each of 19 RE's. Mapping was done either manually (as described in Methods and Materials, Section 4.7 above), or by computer, using the programme RESOLVE (E.H. Harley, Version 2.1, 1990). Of the RE sites mapped, 42 were phylogenetically informative. Fig. 18 shows a typical mapping gel.

4.2 Satellite DNA

End-labelling of the bird mtDNA's that were mapped, required reasonably well-purified DNA, due to rather severe interference from satellite DNA. Satellite DNA occurred frequently, presenting as dark, fuzzy bands in the region of 400bp to 600bp, thus making visualization of digestion products in that region rather difficult. As in fragment analysis, extreme caution had to be exercised in scoring digestion fragments in this vicinity (see Fig. 18).

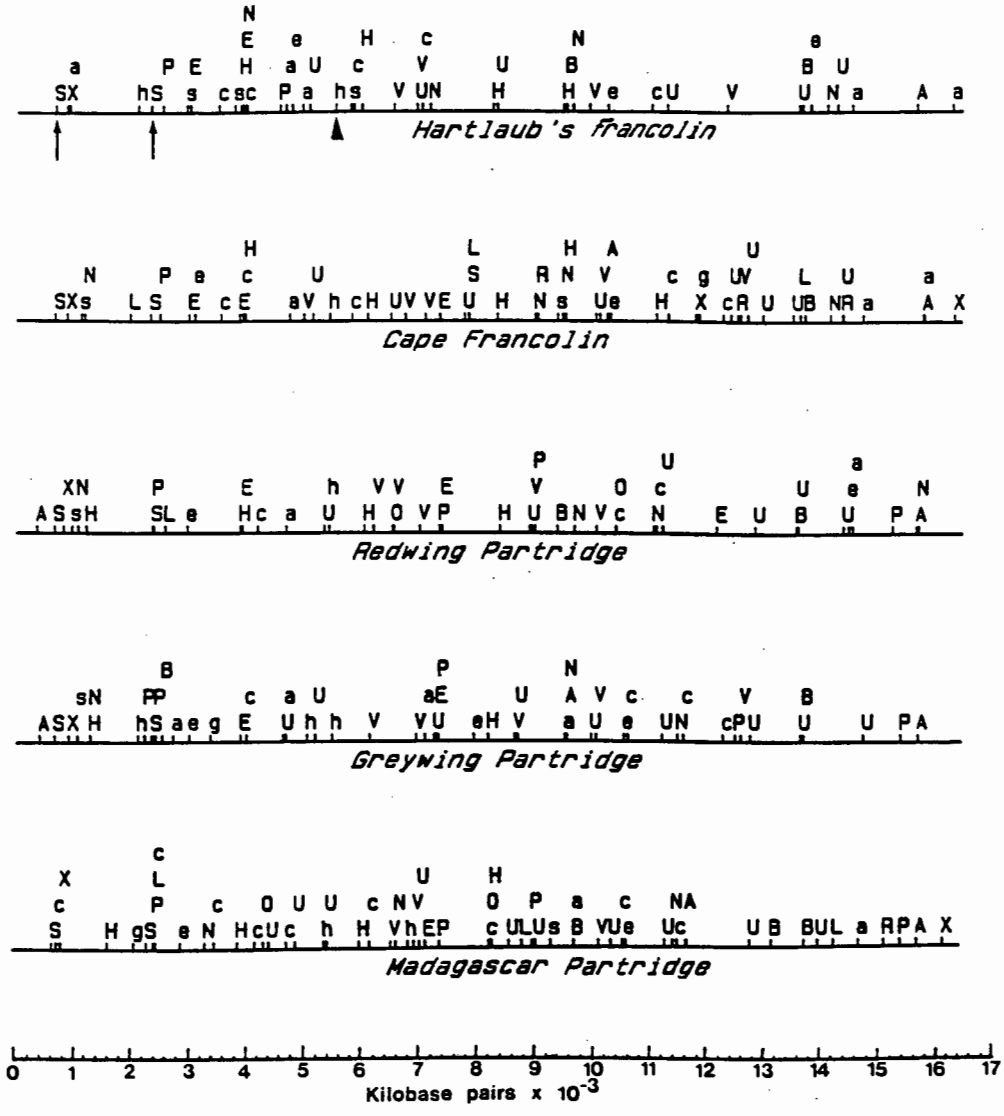


Fig. 17(a) Restriction maps of the mtDNA's of F.hartlaubi (Hartlaub's Francolin), F.capensis (Cape Francolin), F.levaillantii (Redwing Partridge), F.africanus (Greywing Partridge) and Margaroperdix madagascariensis (Madagascar Partridge). These have been aligned on the two invariant SacII sites at positions 676bp and 2356bp, indicated by long arrows, and orientated on the invariant HpaI site at 5540bp - indicated by an arrowhead on the map of Hartlaub's Francolin. Abbreviations for restriction enzyme cleavage sites are as follows: A = Asp718; B = BamHI; c = BclI; g = BglII; E = EcoRI; R = EcoRV; H = HindIII; h = HpaII; N = NcoI; e = NheI; P = PstI; V = PvuII; s = SacI; S = SacII; L = SalI; a = ScaI; U = StuI; X = XbaI; O = XhoI.

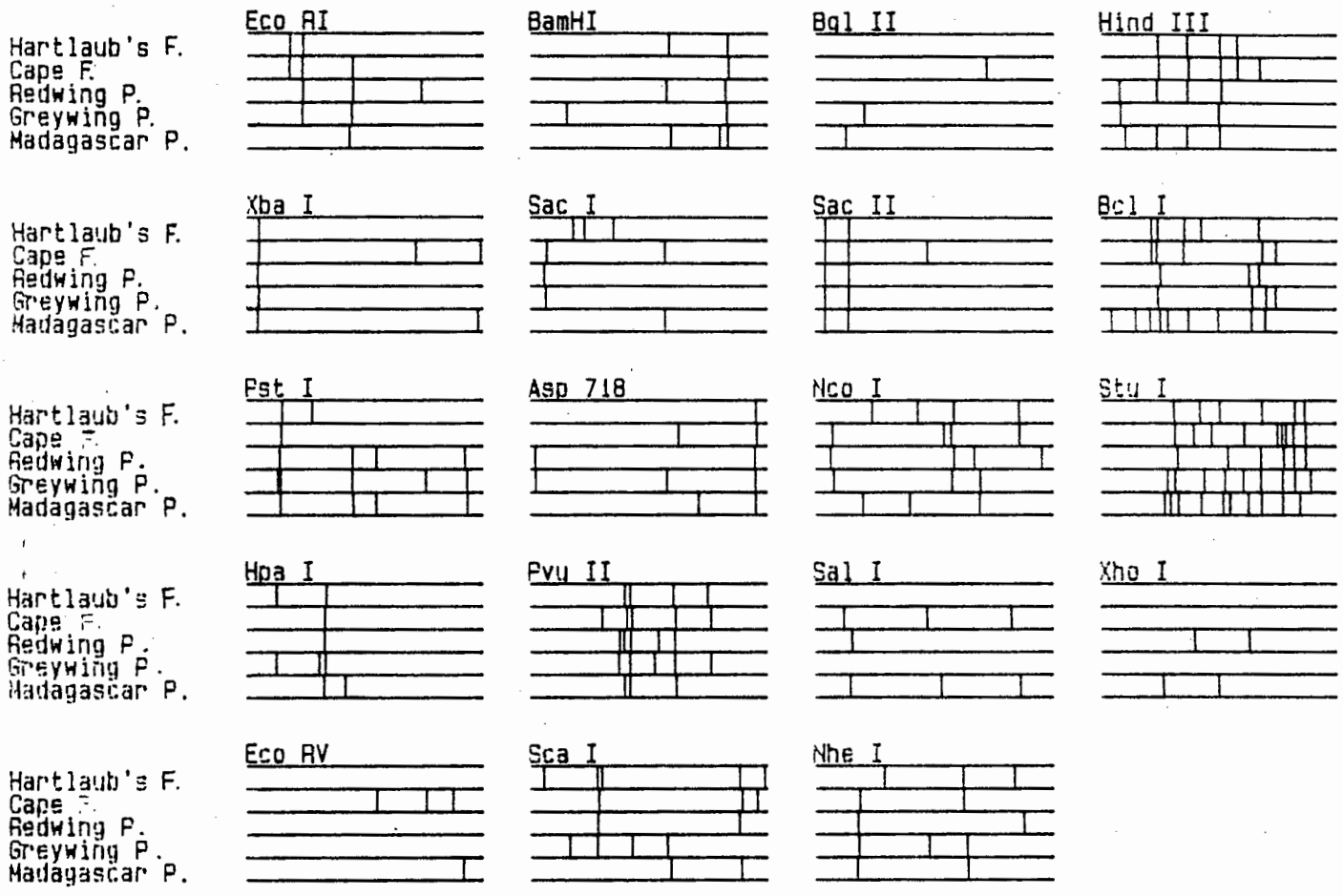


Fig. 17(b) Restriction maps of the mtDNA's of F.hartlaubi (Hartlaub's Francolin), F.capensis (Cape Francolin), F.levaillantii (Redwing Partridge) , F.africanus (Greywing Partridge) and Margaroperdix madagascariensis (Madagascar Partridge). Each map shows individual cleavage sites for each of the 19 restriction enzymes.

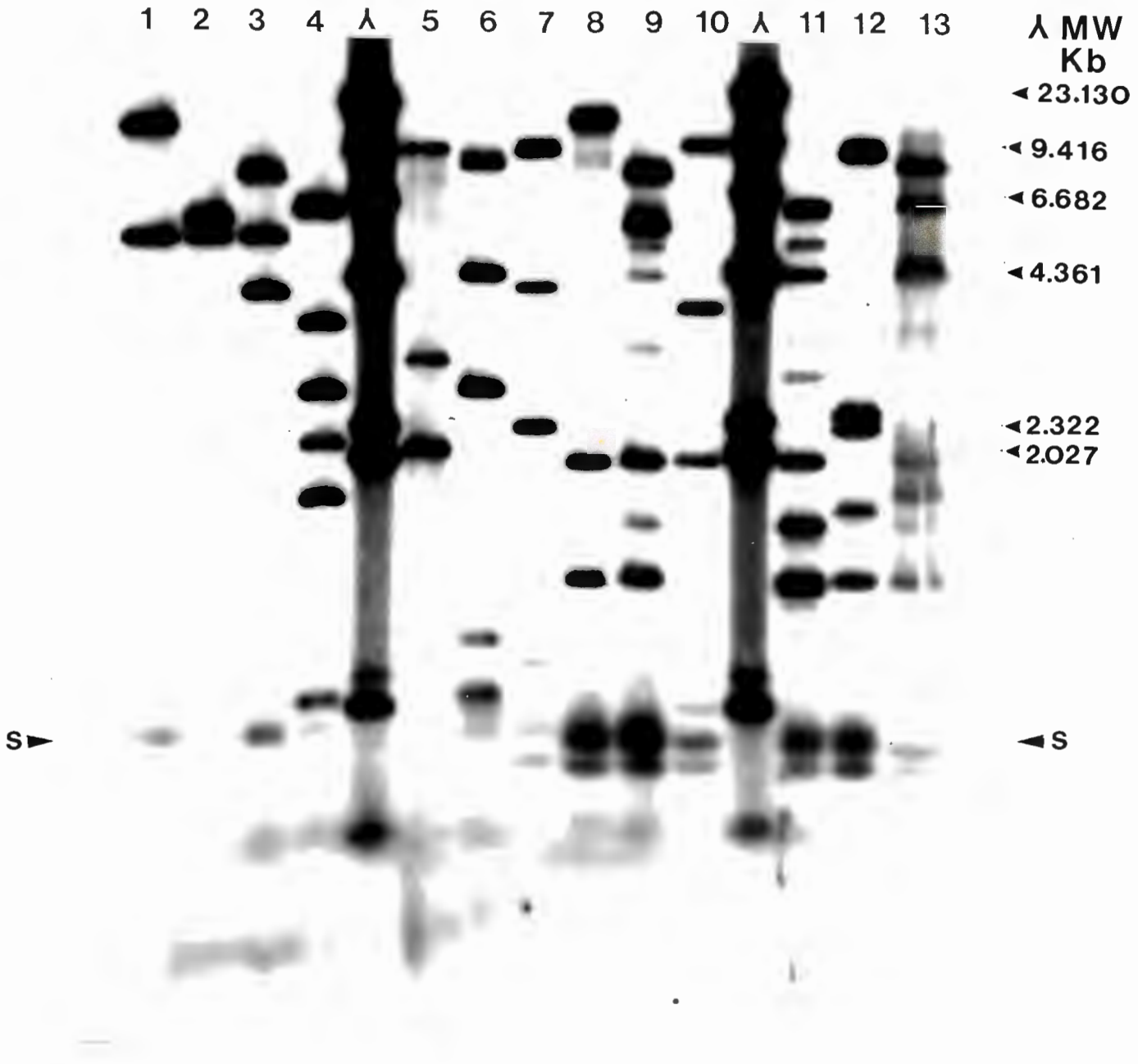


Fig. 18 Autoradiograph of various double-digests of Redwing Partridge mtDNA, end-labelled and subjected to agarose gel electrophoresis at approximately 2V/cm in a 1.2% gel. As a size marker, Lambda DNA (λ) was digested with HindIII and end-labelled. Lane 1 = XhoI; 2 = XhoI+XbaI; 3 = XhoI+SalI; 4 = XhoI+HindIII; 5 = XhoI+BamHI; 6 = XhoI+EcoRI; 7 = XhoI+BclI; 8 = PvuII; 9 = PvuII+XhoI; 10 = PvuII+BamHI; 11 = PvuII+Asp718; 12 = PvuII+EcoRI; 13 = PvuII+SacII. Satellite DNA's are indicated by arrowheads.

4.3 Genome Size

The sizes of Francolin and Madagascar Partridge mitochondrial genomes were obtained by summing estimated fragment sizes after double-digestion, and then taking averages for several double-digests per bird. Table 5 shows the number and relative sizes of fragments resulting from examples of double-digests, with three sets chosen randomly from the mapping of each bird. The average size of 16.47 ±0.12kb appears to be in good agreement with the size estimate based on a measurement of chicken circular mtDNA by electron microscopy (Glaus, pHD thesis, 1980). It is also consistent with Galliform genome size based on restriction mapping and restriction fragment studies (Shields and Helm-Bychowski, 1988). This size estimate can only be considered as an indirect one, since no avian mitochondrial genomes had actually been sequenced at the time of writing.

4.4 Functions of RESOLVE

The programme offers three main functions:

i) The construction of restriction maps of DNA molecules from double-digestion data. This is performed in a 2-step process to be elaborated upon below (Sections 4.4.1 and 4.4.2);

ii) The management of sets of mapped data from different

SPECIES	F.hartlaubi (HT)			F.capensis (CA)			F.africanus (GW)			F.levallantii (RP)			M.madagascariensis (MP)		
Restriction enzymes used	ScaI/ BamHI	NheI/ Asp718	BclI/ Asp718	HindIII/ Asp718	NheI/ SalI	SacII/ XbaI	NheI / Asp718	ScaI/ PstI	Asp718/ PstI	Hind III/ Bam HI	BclI/ PstI	StuI/ SalI	HindIII/ Asp718	XhoI/ Asp718	PstI/ Asp718
Sizes of fragments generated (Kb)	5.00	5.60	4.50	4.80	4.70	5.55	5.20	3.40	5.20	4.10	4.10	4.70	4.00	5.10	4.90
	4.00	5.40	4.30	4.60	4.70	4.50	5.00	2.90	2.85	4.00	3.45	3.60	3.40	4.10	3.70
	3.80	3.55	4.20	2.30	3.30	4.10	2.60	2.70	2.70	2.69	3.05	2.80	2.40	3.85	3.30
	1.63	1.80	1.85	2.15	2.60	1.45	1.70	2.50	2.35	2.30	1.85	2.40	2.30	3.45	2.70
	1.05		1.20	1.22	1.08	0.68	1.07	2.40	1.90	2.15	1.55	1.47	2.22		1.53
	0.92		0.45	0.96		0.20	0.94	1.94	1.10	1.33	1.55	0.80	2.15		0.20
	0.32			0.99				0.32	0.40		0.73	0.76			
								0.20	0.17						
Total of fragment sizes (Kb)	16.40	16.35	16.50	16.69	16.38	16.48	16.51	16.44	16.67	16.55	16.28	16.53	16.47	16.50	16.33
Average size of mtDNA (Kb) genome	16.42 ± .08			16.52 ± 0.16			16.54 ± 0.12			16.43 ± 0.15			16.43 ± 0.09		
Overall average size of Francolin and Margaroperdix mtDNA (Kb)							16.47 ± 0.12								

Table 5: Examples of double-digests and product fragments generated for each species in the restriction enzyme site comparison.

DNA's, after catalogueing, editing and manipulating them where necessary;

iii) The comparison of maps from related taxa and the construction of output files of either phylogenetically informative sites or sequence divergence matrices. These are then formatted for analysis using an appropriate phylogenetic package.

4.4.1. Construction of Restriction Maps with RESOLVE

After performing the selected double-digests with a particular mtDNA, the number of digestion fragments given by each enzyme separately was entered. Data was accepted with a requirement for a likely value of the percentage error (reasonable to expect from one's measurements).
The percentage error chosen may be too low, in which case no solutions are likely to be found - or if it is too high, there may be too many solutions (some of which are obviously not appropriate). However, several different error values can be tried sequentially. It must be noted that it is not always appropriate to use an error value which provides only one solution, as it is also possible for another solution to be found at a slightly higher error value, which is in fact the correct one.
After choosing a value for error, the fragment sizes for the single and for the double-digest results were entered. The

programme then searches for solutions, of which there may be several. These partial digest solutions are then stored in a temporary file. Redundant solutions occur, due to unfixed sites (Fig.19(a), A*). Such sites occur when two or more "A" fragments lie completely within a B fragment, thus resulting in alternative positions for the internal A* site. Thereafter, more rigorous strategies can be used to eliminate ambiguities such as unfixed sites. Where more than one unfixed site occurs, they are stored on temporary file, in decreasing order of size (Fig.19(b)).

The temporary data set for each of the 5 mapped bird mtDNA's, was then subject to the second step in RESOLVE's mapping procedure: the three enzyme consensus analysis.

4.4.2 Three Enzyme Consensus Analysis

This step of the programme has a triple function:
i) The resolution of the correct partial digest solution for an enzyme pair in temporary file;
ii) The resolution of unfixed sites;
iii) The checking of site positions a second time, to ensure accuracy of the final maps.
Each enzyme is registered in the final file only when it has been correctly mapped relative to two other mapped enzymes in final file, with all the sites compatible within given error limits.

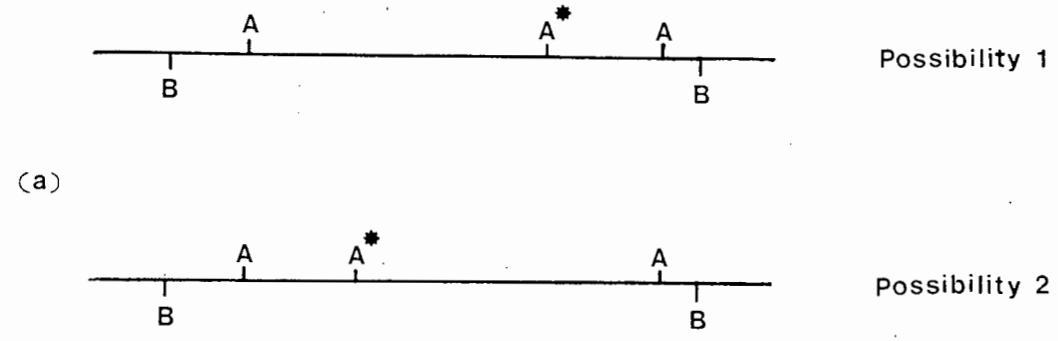


Fig. 19(a) Restriction maps of DNA showing unfixed sites (A*), where two or more "A" fragments lie completely within the "B" fragment. The two optional positions for A* are shown in Possibility 1 and Possibility 2. A,B = restriction enzymes A and B.

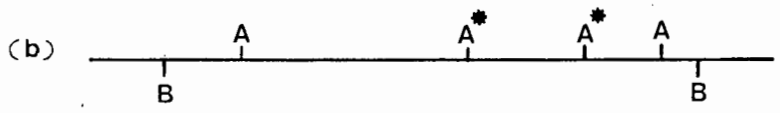


Fig. 19(b) Where there are two adjacent unfixed sites (A*), RESOLVE orders them in decreasing order of size to the right, as depicted in this temporary mapping solution.

Once again, error is taken into account and is in the site alignment - it is measured as a percentage of overall DNA length, as opposed to the percentage of individual fragment length, at the temporary file stage. In the 3-enzyme consensus analysis, a clean data set usually resolves at less than 5% error, and very rarely does consensus analysis provide two different solutions with significant ambiguity. In such cases, the fault is likely to lie in the original digestion fragment size data, even though it may have been successfully mapped to a temporary file.

The 5 birds were thus mapped with each of the chosen enzymes, and their final maps manipulated for phylogenetic analysis of RE site data.

4.4.3 Manipulation of Final Maps by RESOLVE

Using an editing option for final map data files, all 5 maps were realigned on a specified map position. This was usually done as the maps were being built up, and was necessary since the first map produced for each bird had an arbitrary starting point. As further map positions were built on this point, it was essential to select a site for alignment in order for comparisons to be made with the other bird maps.

To this end, the sites selected were those in invariant positions in vertebrate mtDNA - the two Sac II sites (at

position 676bp and 2356bp, which give an invariant Sac II fragment of 1680bp); all 5 maps were aligned on these. Orientation of the maps was further necessary in order to establish the correct positioning of sites on either side of the invariant Sac II sites. Thus, an Hpa I site in all 5 birds (also invariant in vertebrates and occurring at 5540bp, therefore lying 3184bp to the right of the second Sac II site), was selected as the site for map orientation. The initial maps were drawn up independently for each species, and then aligned with one another so as to maximize conservation of cleavage sites. Fig.17(a) shows the aligned 5 maps of Hartlaub's Francolin, Cape Francolin, Redwing Partridge, Greywing Partridge and the outgroup Madagascar Partridge.

4.4.4 Analysis of Final Maps using RESOLVE

This procedure offers three features for the analysis of enzyme sites in the final map file: The first provides a comparison of individual enzyme site alignments; the set of 5 maps was chosen, followed by each of the 19 RE's used. Individual RE sites were plotted under one another as seen in Fig.17(b). This is useful in finding alignments or orientation in new maps that may be added, and was useful in identifying sites that may have been misplaced.

The second and third features find phylogenetically informative sites, and measure pairwise sequence divergence, respectively. Where these were done manually for the fragment comparison, RESOLVE provided a very convenient and quick way of presenting the RE site data.

4.5 Phylogenetic Analysis

4.5.1 Cladistic Approach

RESOLVE was able to find all the phylogenetically informative RE sites in the 5 cleavage maps, for further application in the construction of phylogenies using cladistic approaches. The programme requires at least 4 maps, since one constraint of cladistic analysis is that it identifies phylogenetically informative sites only if sites are shared by at least 2 and not more than n-2 taxa (where n = total number of taxa).

An error value was set (% of total DNA length), such that sites in different maps within this value were presumed to be fully aligned.

Single sites with no alignments were treated as autapomorphies. RE sites shared by all, or all barring one of the taxa, were assumed to be shared ancestral characters (symplesiomorphies).

RESOLVE's assumptions as to which RE sites are autapomorphies and symplesiomorphies are usually, though not

always, correct. However, autapomorphies and symplesiomorphies are not used in the subsequent analyses.

Thus, a table of phylogenetically informative RE site positions was drawn up for the 5 maps in the comparison, as well as a table of informative character states (Table 6(a), 6(b)), where "1" is indicative of the presence of a shared site,, and "0" denotes the absence of such.

4.5.1.1 Cladistic Analysis of MtDNA RE Site Data

This table of phylogenetically informative sites served as input for analysis by each of the character-based programmes PAUP (D.L. Swofford, version 2.4) and Hennig 86 (J.S. Farris, version1.5), as well as DOLLO (in PHYLIP). The shortest possible cladograms for the 5 birds were produced using Hennig 86's I.E. (Imlicit Enumeration) option and the Branch and Bound option of PAUP.

Hennig 96 and PAUP analysis of the phylogenetically informative site characters produced the same single unrooted tree (Fig.20(a)).

Dollo analysis of the phylogenetically informative characters produced two equally parsimonious trees and attempted to place a root. These are shown in Fig.20(b).

INFORMATIVE SITE NUMBER	RESTRICTION ENZYME	MT GENOME POSITION (bp)	INFORMATIVE SITE NUMBER	RESTRICTION ENZYME	MT GENOME POSITION (bp)
1	a	14562	22	N	1141
2	h	2126	23	N	9511
3	P	2300	24	N	9669
4	P	7356	25	N	11443
5	P	8993	26	N	14115
6	P	15262	27	e	2951
7	s	1027	28	e	10283
8	s	9359	29	e	10573
9	E	3008	30	B	9628
10	E	7224	31	A	383
11	E	7360	32	U	5150
12	c	3516	33	U	5400
13	c	3976	34	U	7000
14	c	4194	35	U	8650
15	c	5845	36	U	8975
16	c	10469	37	U	10010
17	C	11059	38	U	14323
18	C	11533	39	V	6175
19	C	12248	40	V	6554
20	H	1264	41	V	12550
21	H	9510	42	L	2512

Table 6 (a): Informative mtDNA restriction site positions. Abbreviations for restriction enzymes are as follows:

a = ScaI; h = HpaI; P = PstI; s = SacI; E = EcoRI; c = BclI; H = HindIII;

N = NcoI; e = NheI; B = BamHI; A = Asp718; U = StuI; V = PvuII; L = SalI

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
F. hartlaubi	1	1	0	0	0	0	0	0	1	0	0	1	1	0	1	0	1	0	0	0	1
F. capensis	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1	0	0	0	1	0	1
F. levaillantii	1	0	1	1	1	1	1	0	0	0	1	0	0	1	0	1	1	0	0	1	0
F. africanus	0	1	1	1	0	1	1	0	0	1	0	0	1	0	0	0	0	1	1	1	0
M. madagascariensis	1	0	0	1	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0

	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
F. hartlaubi	0	0	1	0	1	0	1	0	1	0	1	0	1	0	0	0	1	0	1	0	0
F. capensis	1	1	0	0	1	0	1	0	0	0	1	0	0	0	0	1	1	0	0	1	0
F. levaillantii	1	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0	1	1	1	0	1
F. africanus	0	1	0	1	0	1	0	1	0	1	1	0	0	1	0	1	0	1	0	1	0
M. madagascariensis	0	0	0	1	0	1	0	1	1	0	0	1	1	1	1	0	0	0	1	0	1

Table 6 (b): Phylogenetically Informative mtDNA Restriction Enzyme Site Characters

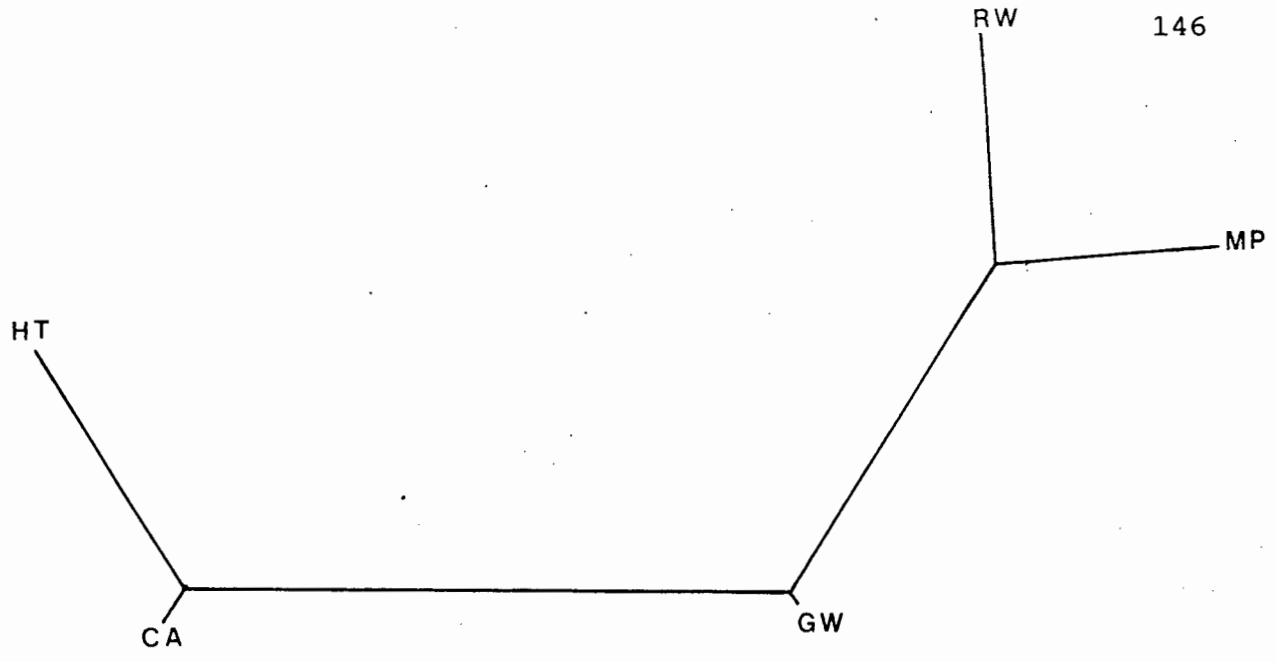


Fig. 20(a) Cladogram inferred by Wagner Parsimony (Hennig 86 and PAUP) analysis of phylogenetically informative Francolin mtDNA restriction enzyme site characters.

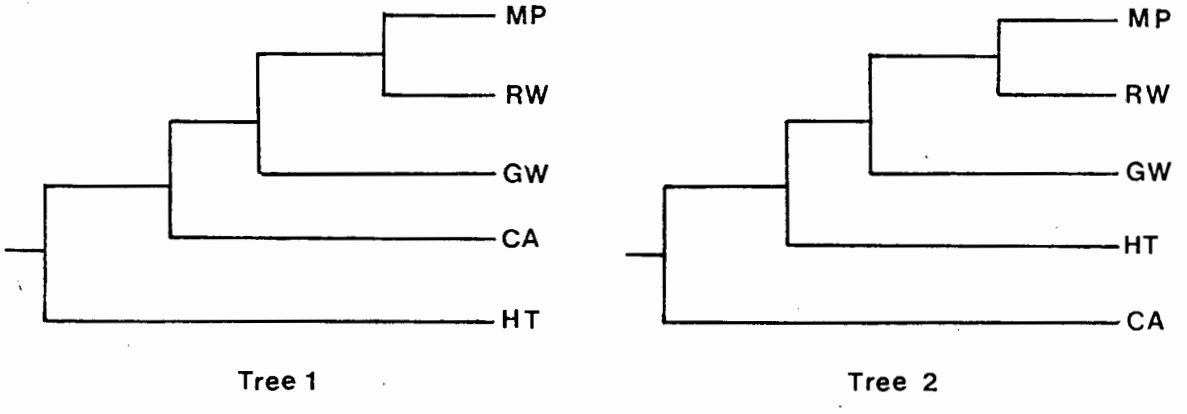


Fig. 20(b) Trees inferred by Dollo Parsimony analysis (DOLLO, in PHYLIP) of phylogenetically informative Francolin mtDNA restriction enzyme site characters. MP = Madagascar Partridge; RW = Redwing Partridge; GW = Greywing Partridge; CA = Cape Francolin; HT = Hartlaub's Francolin.

4.5.1.2 Bootstrap Analysis

Data was subjected to Bootstrapping with 100 replications, using BOOT (in PHYLIP).

Fig. 21 shows the tree generated by BOOT upon analysis of RE site data.

4.5.2 Distance Measure Approach

RESOLVE was used to estimate percentage sequence divergence by measuring the proportion of shared RE sites between the pairs of mapped birds. Recognition sites were considered shared if their positions on the cleavage maps coincided within an error value of 1% of the genome length. An overall distance matrix was produced (see Table 7).

4.5.2.1 Estimation of Percentage Sequence Divergence

The algorithm used, is based on equation 16 of Nei and Li (1979), which assumes there is heterogeneity among cleavage sites with respect to probability of base substitution. The mean number of substitutions per nucleotide site (δ) can therefore be estimated using either of the following equations:

$$\delta = (\ln s)/r$$

or

$$\delta = -(3/2)\ln[(4s^{\frac{1}{2}} - 1)/3]$$

where : r = length of recognition sequences for RE's and is equal to 6 in this study;
s is the proportion of ancestral RE sites that have remained unchanged, until present, in both taxa of the pair compared, and is taken from

$$s = 2n_{xy} (n_x + n_y)$$

where: n_{xy} = number of identical sites shared by two taxa
n_x, n_y = total number of RE sites in taxa x and y, respectively.

4.5.2.2 Distance-Based Analysis of MtDNA RE Site Data

Table 5 served as input data for analysis by the distance-based programmes of FITCH, KITSCH (in PHYLIP), and Saitou and Nei's (1987) Neighbour-Joining, each of which produced a single tree (see Figs. 22 (a), (b) and (c).

4.6 Outgroup

The quail-like Madagascar Partridge was initially chosen as a reasonable candidate for outgroup because of its morphological characteristics, geographical habitat, and the

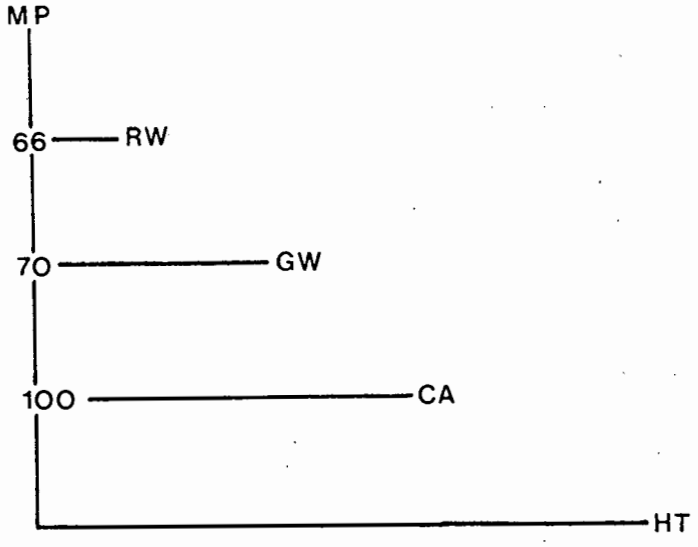
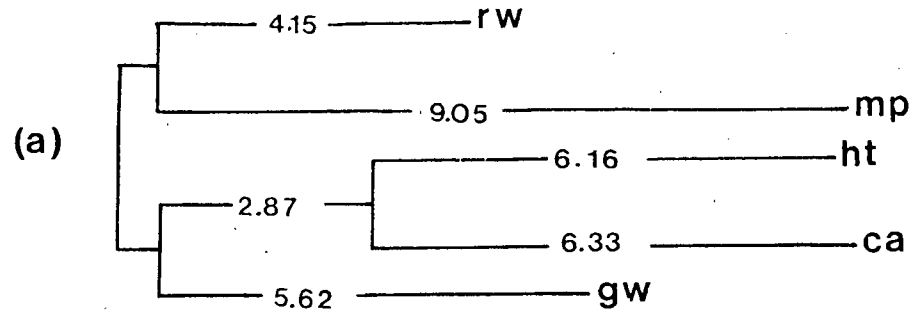


Fig. 21 Tree resulting from Bootstrap analysis (BOOT, in PHYLIP) of phylogenetically informative Francolin mtDNA restriction enzyme site characters. 100 replicates were made. Bootstrapping frequencies are indicated at the nodes. MP = Madagascar Partridge; RW = Redwing Partridge; GW = Greywing Partridge; CA = Cape Francolin; HT = Hartlaub's Francolin.

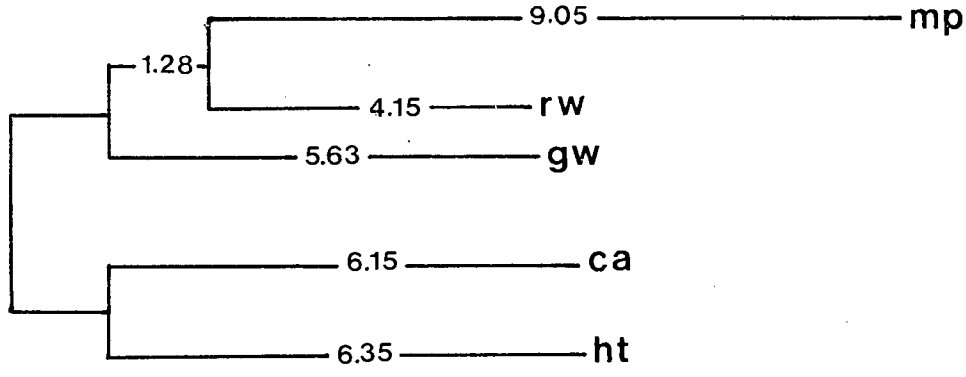
	har	cap	lvi	afr	mad
F.hartlaubi		12.500	13.600	16.300	18.400
F.capensis	12.500		14.600	13.200	21.000
F.levaillantii	13.600	14.600		11.300	13.200
F.africanus	16.300	13.200	11.300		15.500
M.madagascariensis	18.400	21.000	13.200	15.500	

Table 7: Matrix of percent nucleotide divergance (δ) estimates (lower half matrix) and proportion of shared mtDNA restriction enzyme sites (upper half matrix) (Nei and Li, 1979) for Francolinus spp. and MMadagascariensis

(a) Tree inferred by Neighbour-Joining analysis.



(b) Tree of best fit inferred by FITCH (in PHYLIP) analysis.



(c) Tree of best fit inferred by KITSCH (in PHYLIP) analysis.

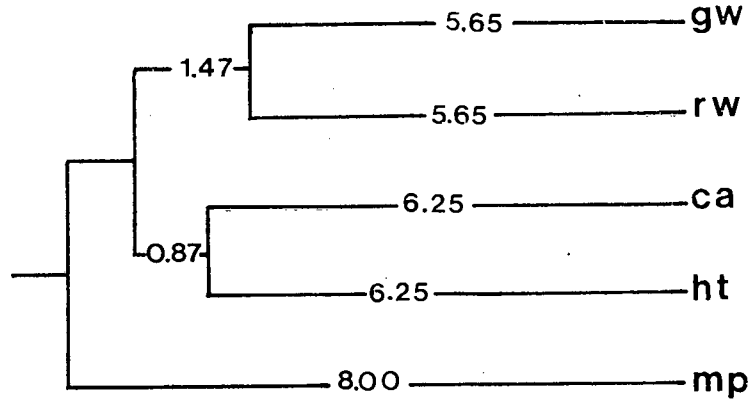


Fig. 22 Distance-based phylogenies of Francolin mtDNA restriction enzyme site data. MP = Madagascar Partridge; RW = Redwing Partridge; GW = Greywing Partridge; CA = Cape Francolin; HT = Hartlaub's Francolin. Branch lengths are shown.

fact that it is osteologically the nearest neighbour of F.hartlaubi (F.hartlaubi itself appearing in the fragment comparison to be a sister-taxon to the other Francolins; see Discussion, Chapter 3). However, δ values for Madagascar Partridge gave some indication that it shared more restriction sites with F.levaillantii and F.africanus (δ = 13.2 and 15.5 respectively; see Table 7) than expected. Percentage sequence divergence values between Margaroperdix and the other two "francolins", F.hartlaubi and F.capensis, on the other hand, were much higher (δ = 18.4 and 21.00 respectively; Table 7), relatively speaking. Thus, the use of Madagascar Partridge as an outgroup might be inappropriate, but there was at the time no suitable alternative.

DISCUSSION

The comparison of Francolin mtDNA site data provided a firmer phylogenetic identity for F.levaillantii and F.hartlaubi, although the results of phylogenetic analysis fell slightly short of expected topology where Margaroperdix was concerned.

4.1 Cladistic Analysis

The phylogeny inferred by Parsimony analysis grouped hartlaubi with capensis, but the chosen outgroup M.madagascariensis appeared to be more closely related to levaillantii than the latter was to africanus. If the "partridges" and the "francolins" are separate monophyletic groups, then this association makes the Madagascar Partridge appear as a recent offshoot of the genus, and more closely related to the Redwing Partridge than to the Greywing Partridge. This compounds the unexpected taxonomic status of the levaillantii.

Dollo analysis provided two alternative trees.
Tree 2 differed from Tree 1 in that hartlaubi became sister-taxon to the Greywing, Redwing and Madagascar Partridge assemblage (here, once again, levaillantii and Margaroperdix were grouped as sister-taxa), where capensis fulfilled this role in Tree 1. The major difference between the two Dollo

and the Wagner Parsimony-based (PAUP or Henning 86) tree was that the latter kept capensis monophyletic with hartlaubi, whereas Dollo has made the "francolins" paraphyletic. This feature could be a demonstration of overestimation by DOLLO of the number of evolutionary changes (see Chapter 1, Section C, 1.3.1.4), since distance-based phylogenies (see Section 4.2 below), like that of Maximum Parsimony, also support the monophyly of the "francolins".

Bootstrapping frequencies for RE site data can be seen in Fig.20, where the Madagascar Partridge was used as outgroup. This was not a very valuable exercise for these data since there were only 5 taxa and the use of Margaroperdix as outgroup was of very doubtful validity. The main point of value was the association of capensis with hartlaubi with a 70% confidence figure.

4.2 Distance-Based Analysis

Although representation of the "francolin" and "partridge" clades was limited in this mapping approach, several observations could be made about the phylogeny of the represented Francolins. Sequence divergence values appeared to be higher in RE site as opposed to fragment data (see Table 3 and Table 7). This is probably an example of the occurrence of convergence in the fragment approach, i.e. the fragment comparison has probably resulted in an

excessive number of homoplasic alignments in the more distantly related taxa.

Percentage sequence divergence values averaged within a range of 11.3 between levaillantii and africanus, to 16.3 between hartlaubi and africanus, but values between Margaroperdix and hartlaubi, and Margaroperdix and capensis were higher, being 18.4 and 21.0 respectively. The last two values were in accordance with those expected for an outgroup, relative to the rest, but when the δ values between Margaroperdix and levaillantii, and between Margaroperdix and africanus were considered (13.2 and 15.5 - resp.), they were too close to the inter-Francolin values (e.g., 12.5 for hartlaubi and capensis, and 13.2 for africanus and capensis) to justify its use as an outgroup. At this early stage, then, the δ values for Madagascar Partridge pointed to its potential alliance with levaillantii and africanus - if this holds true, then the results dispute Hall's (1963) allegation that Margaroperdix madagascariensis is not closely related to the Francolins, despite superficial similarities in plumage colouration to the Forest Francolin, F.lathami of West Central Africa.

The Neighbour-Joining (NJ) analysis of RE site data gave rise to a single tree, giving branch lengths indicative of evolutionary change under the asssumption of a molecular clock. The method utilizes midpoint-rooting by taking the

midpoint of the longest patristic distance - in this case the branch lengths linking Margaroperdix to capensis.
The "partridges" did not form a monophyletic group, although the "francolins" did (Fig.22(a)).
The use of the mid-patristic distance resulted in africanus apparently clustering with the "francolins", but since its placement was so near to the root, no confidence could be placed on this topology.

This point was emphasized in the FITCH analysis, which yielded a single unrooted tree having virtually identical branch lengths to the NJ-based tree, and with very similar topologies. The exception was that africanus now grouped with the "partridges", thus engendering a "francolin" - "partridge" dichotomy (Fig.22(b)). However, as with NJ analysis, africanus is placed very near to the root, thus allowing little confidence in the topology.

It was interesting, however, that both NJ and FITCH still maintained levaillantii as the closest relative of Margaroperdix. It was notable that the branch lengths of levaillantii and Margaroperdix from the node above were very dissimilar. This reflected either a different mutation rate in mtDNA, or more likely, was merely the result of stochasticity.

Fig.22(c) shows the rooted phylogeny inferred by KITSCH

analysis (which assumes the molecular clock, subject to stochastic variation). The approach resulted in Margaroperdix becoming the outgroup, and of all the trees based on RE site data, this phylogeny appeared to agree best with the classical expected order of branching, where both "partridges" (africanus and levaillantii) and "francolins" (capensis and hartlaubi) were clearly monophyletic.

4.3 General Discussion of Site Mapping Approach

Fragment Comparison data served as a useful basis for the construction of the restriction maps, which in turn provided a much more robust set of data for subsequent distance and cladistic analysis.

As in the fragment comparison, it should be noted that in the present study, the Nei and Li (1979) model used to calculate percentage sequence divergence is based on rather precise probabilities for the evolutionary change of RE sites: The model also ignores the fact that transitions occur more often than transversions thus possibly underestimating the extent of point-mutational divergence.

RE mapping does eliminate problems of convergent fragment lengths - however, the sites themselves may be convergent.

Even when sites are mapped, RE data might be less than ideal for phylogenetic analysis, due to asymmetry in the probabilities of gaining and losing sites: The probability

of convergent site losses is far greater than that of convergent site gains; a site loss may be caused by any point mutation within a cleavage site, whereas a site gain requires a specific base substitution at a specific base pair (Templeton, 1983b; Li, 1986). These inequalities should be considered when using RE site data for phylogenetic analysis.

Both distance and cladistic analyses of site data produced phylogenies more in keeping with expected results (as from traditional classification by morphology) than the fragment data; F.hartlaubi grouped with the representative of Hall's Vermiculated Group (i.e. Cape Francolin) and F.levaillantii grouped with the representative of Hall's Red-Winged Group (i.e. Greywing Partridge).

CHAPTER 5

GENERAL DISCUSSION

5.1 Evaluation of Methods of Phylogenetic Analysis

The applications of molecular approaches such as those utilized in this study, have provided some insight into the performance of distance-based methods, as opposed to those of cladistic analysis. Several sources of error arise when using either approach, which should be borne in mind.

Cladistic methods, apart from de-emphasizing measures of branch length (and therefore the time at which any pair diverged from a common ancestor), are prone to error in the form of convergence or homoplasy. Chance dictates that when a phylogeny contains two very divergent sequences, some of the mutational changes introduced will be the same, even though the evolution along the two lineages was independent.

Such homoplasies tend to bring the diverged sequences together artifactually, as sister-groups in an unrooted phylogenetic tree. This is often seen in the attraction of long branches, which have many homoplasies, a phenomenon well illustrated in the SEvoNA simulation. The effect is compounded for sequence data when there is a high transition bias, as is known for mtDNA.

Parsimony analysis in particular suffers from this effect (Felsenstein, 1978a). When long branches are separated in the tree by short edges, Parsimony tends to select trees with long edges linked together. This is because such trees can be shorter than the original, i.e. have fewer observed changes.

RE sites are less likely than sequence data to show this kind of convergence, however, since parallel losses are more likely to occur than parallel gains.

The phenomenon of convergence can be eliminated in sequence data if homoplasic characters are recognized as such, and given little weight in the analysis. However, this is not generally possible for molecular fragment and site data (except when there is avoidance of information from single restriction cuts such as in a fragment study; see Chapter 3, Results, Section 3.5.2).

Another source of error lies in the incorrect identification of shared derived states (synapomorphies) as opposed to ancestral states (symplesiomorphies). Avise and his co-workers have demonstrated that for recently diverged species, mtDNA phylogenies may be affected by retained ancestral polymorphism. Over time, stochastic lineage sorting eliminates mtDNA's with ancestral states, while

mtDNA's with derived states are created by mutation (Avise et al., 1983; Neigel and Avise, 1985; Avise, 1986).

Distance methods suffer from different problems. A major factor affecting measurements of nucleotide substitution is stochasticity - the variation due to the random nature of the process. Stochastic variation decreases however, as the size of the data set increases, e.g. 5000 sequenced base pairs offer much less stochastic variation than 500 sequenced base pairs. Any distance method for generating trees from sequence data has a high probability of producing an erroneous tree, unless the number of nucleotides examined is large (Saitou and Nei, 1986).

Again, the difference in mutation rates between taxa in a comparison is disadvantageous to distance-based methods. The molecular clock assumption, itself, is susceptible to a relatively large margin of error, as few serious attempts have been made to determine confidence limits for it. Potential errors originate at various stages of analysis (Hillis and Moritz, 1990). For example, errors of measurement in collecting data, errors associated with conversion of original data into a distance divergence measure, and errors arising in calibrating the clock. The error in clock calibration arises because it is difficult to establish the age of the last common ancestor of a group of extant species, particularly since palaeontological

estimates may be inaccurate (Carlson et al., 1978) and lack of fossil material poses a major problem.

The phenomenon of saturation provides another source of disturbance to distance methods; this is where most synonymous (neutral) positions within the mtDNA have mutated at least once already. At high levels of divergence, the rapidly evolving sequences are saturated with substitutions and multiple events have occurred at many sites - this effectively reduces the resolution of phylogenetic analysis by homoplasy.

These factors could be addressed in various ways: Using slowly evolving regions, such as rRNA genes, or second codon positions of protein genes, should increase resolution at higher levels of divergence. Also, where nucleotides are used as characters for moderately divergent taxa, the comparisons could be restricted to transversion events, to avoid problems with convergent transitions (Hasegawa et al., 1985).

Although Maximum Parsimony is widely used in cladistic analysis (Benton, 1990), the evolutionary change of a character does not necessarily occur parsimoniously, as DNA sequences are particularly subject to back- and parallel mutations (Sourdis and Nei, 1988).

Because of the conversion of original data into distances, the distance approach is inherently weaker than Maximum

Parsimony, since the original data cannot be recovered from distances (Tables 3 and 7; see also Penny, 1982). Similarly, Maximum Parsimony does not utilize all the information available because autapomorphies are excluded from phylogenetic analysis, as seen in the cladistic analyses in Chapter 3 and Chapter 4.

The performance of distance methods is also affected by parallel and back-mutations, but their effects are apparently smaller than those for Maximum Parsimony because of the use of autapomorphies as well as informative sites, as seen in the distance analyses in Chapter 3 and Chapter 4.

The Bootstrap analysis of Francolin mtDNA fragment data in this study appeared to be successful in assessing the accuracy of cladograms. However, bootstrapping of Francolin mtDNA site data in this study was not very informative, since there were only 5 taxa in the comparison and the validity of Margaroperdix as an outgroup was questionable. This resampling method should prove useful in a comparison of the maps of at least all the taxa examined in the fragment approach, i.e. 13 Francolin species (see Table 1).

Notwithstanding these observations, then, for molecular data, given a molecular clock with equal mutation rates in the Francolin species studied, and provided the stochastic effects are not too great, distance measures are almost as

effective as cladistic approaches for reconstructing the true phylogeny.

Perhaps the best practical approach is to use cladistic methods for defining the topology (preferably Maximum Parsimony), and the distance methods for providing information on the branch lengths and the timing of radiation events (preferably FITCH or KITSCH analysis; see Chapter 4, Discussion Section 4.2). It perhaps ironical in terms of this decision to use cladistic approaches for topology and distance methods for branch lengths, that the tree best agreeing with a preconception of expected topology in this study, was provided by the distance-based KITSCH analysis.

5.2 General Discussion of Francolin MtDNA Fragment and RE Site Comparisons

5.2.1 MtDNA Fragment Comparison

Despite caveats attached to the use of mtDNA fragment methods in phylogenetic analysis, the approach utilized in this study proved a simple and serviceable means of pointing out molecular relatedness between the individuals under comparison, particularly since it has addressed a closely related group such as the Francolins - i.e. the level of convergence appeared not to be high enough

to warrant the exclusion of the method from phylogenetic studies. The fragment comparison has been successful in helping to define a basic phylogeny comparable with that of the classic topology as proposed by Hall (1963; see Fig. 1), i.e. separate monophyletic groups (Vermiculated, Red-Winged and Bare-Throated Groups) could be identified from fragment data. The "partridge" - "francolin" dichotomy proposed by Milstein and Wolff (1987) was supported by fragment data, except for the placement of F.levaillantii and F.hartlaubi.

It must be emphasized, however, that the mtDNA fragment comparison, although it provides a good preliminary basis for the study of molecular phylogenetics, is not adequate on its own, but is best accompanied by a comparison of cleavage maps or ultimately, direct sequence data. This would give a much more accurate assessment of evolutionary interrelationship (e.g. Cann and Wilson, 1983; Cann et al., 1984).

5.2.2 MtDNA Site Comparison

The RE Site comparison was designed to:
a) resolve the quadrichotomy at the base of the trees resulting from phylogenetic analysis of fragment data, and
b) investigate the phylogenetic placement of Margaroperdix.

It has successfully resolved the quadrichotomy by showing that levaillantii does indeed cluster with the "partridges", and that hartlaubi clusters with the "francolins".
The distance analyses enabled the calculation of a time of divergence of "partridges" from "francolins" at approximately 7-8 x 10^6 y.b.p. (assuming for the present the mammalian calibration scale of 2% per million years; Brown et al., 1979).

From the site data to hand, the F.levaillantii genotype no longer appears to be as anomalous as it did in the fragment approach, now tending to side with the "partridges" as originally expected. Redwing Partridge is now seemingly more in step with the classical morphological phylogeny.
However, these findings are by no means unequivocal, and a comparison with cleavage maps of F.levaillantoides, F.shelleyi, or F.finschi (see Fig.1) would provide a sharper resolution for F.levaillantii and the members of Hall's Red-Winged Group.

F.hartlaubi is confirmed as a sister-taxon to the "francolins", as a result of its observed closer alliance with F.capensis than to levaillantii and africanus. Similarly, mapping other representatives of the Vermiculated Group, such as F.harwoodi, F.hildebrandti or F.bicalcaratus would add to this preliminary study by reinforcing the position of F.hartlaubi within the "francolin" clade.

Margaroperdix, surprisingly, did not appear in any of the cladistic or distance analyses (except in the KITSCH phylogeny) to be an outgroup, since it grouped with levaillantii. More data is required to resolve the issue and a fuller survey of putative sister-taxa is necessary. Plausible candidates are members of the Alectoris or Perdix genuses.

Clearly the data set for the mapping approach was limited, and a wider range of Francolins need to be addressed, with adequate representation by members of the Scaly, Montane and Spotted Groups, as well as by more individuals from the Red-Tailed and Striated Groups, and from the enigmatic Forest Francolin species, F.nahani and F.lathami.

5.3 Final Conclusion

It was expected that molecular data would give more useful additional data for resolution of Francolin evolutionary relationships, because they show a much more regular pattern of evolutionary change than morphological characters and more (unweighted) characters can be accumulated. Nonetheless, no single systematic data set can be expected to be informative at all phylogenetic levels simultaneously, and it was useful to compare Francolin molecular data with Francolin morphological characters.

The approaches presented in this dissertation combined as many sources of information as possible, to maximize the chances of a correct phylogenetic inference for the Francolins. An unfortunate drawback to the analysis of a combined macrocharacter and mtDNA character set is that it cannot be done using distance methods, and so the combined data set was limited to character-based analysis such as Hennig 86.

The use of restriction enzymes in the analysis of mtDNA allowed direct examination of perhaps the best understood piece of the vertebrate genome. The maternal inheritance, rapid rate of nucleotide substitution and easy analysis of mtDNA are all instrumental in expanding our knowledge of molecular evolution.

The full resolving power of mitochondrial DNA will be available only when mtDNA from a number of avian species can be completely sequenced both rapidly and efficiently. Nucleotide sequence data would facilitate more accurate estimates of sequence divergence among the Francolins. Nevertheless, this study, in indicating doubt as to the exact phylogenetic positioning of F.levaillantii and F.hartlaubi has provided a good basis for systematic orientation of the Francolins. Molecular studies such as the mtDNA fragment and RE site comparisons presented in this

dissertation have been effective in providing the groundwork for future molecular research on Francolinus.

APPENDIX I

MATERIALS

1. Isolation and Purification of Mitochondrial DNA

1.1 Buffers

(a) Extraction Buffer (1 M)

To make 1000 ml:

100 mM Tris.Cl (pH 8)
150 mM NACl
20 mM EDTA (pH 8)
10% w/v sucrose

Add distilled H_2O to 1000 ml. Autoclave.

(b) Tris EDTA (TE) Buffer (1 M, pH 8)

To make 1000 ml:

10 mM Tris.Cl (pH 8)
1 mM EDTA (pH 8)

Add distilled H_2O to 1000 ml. Autoclave.

(c) Saline Tris EDTA (STE) Buffer (1 M, pH 8)

To make 1000 ml:

10 mM Tris.Cl (pH 8)
1 mM EDTA (pH 8)

Add distilled H_2O to 1000 ml. Autoclave.

(d) Tris.HCl Buffer (1 M, pH 8 or pH 7.5)

To make 1000 ml: Dissolve 121.1 g Tris base in 800 ml distilled H_2O. Adjust pH to 8 with conc. HCl. Autoclave. If pH 7.5 required, add another 20 ml conc. HCl before making up to 1000 ml.

1.2 Miscellaneous Solutions

(a) Ethylene Diamino Tetra-Acetic Acid (EDTA)(0.5 M,pH 8)

To make 500 ml: Dissolve 93 g EDTA in 400 ml distilled H_2O. Adjust pH to 8 with 10 g NaOH and 5 M NaOH. Add distilled H_2O to 500 ml. Autoclave.

(b) Sodium Chloride (NaCl) (5 M Stock Soln.)

To make 500 ml: Dissolve 146 g NaCl in 400 ml distilled H_2O. Add H_2O to 500 ml.

(c) Sodium Dodecyl Sulphate (SDS) (10 % Stock Soln.)

To make 100 ml: Add 10 g SDS to 90 ml distilled H_2O. Warm to dissolve, and make up to 100 ml with H_2O.

(d) Sodium Hydroxide (NaOH) (5 M Stock Soln.)

To make 500 ml: Dissolve 100 g pellets NaOH in 400 ml distilled H_2O. Add dH_2O to 500 ml.

(e) Ethidium Bromide (EtBr) (10 mg/ml Stock Soln.)

To make 100 ml: Add 1 g EtBr to 100 ml distilled H_2O. Stir to dissolve. Wrap container in aluminium foil and store at 4°C.

2. Restriction Enzyme Digestion

2.1 Buffers

KGB Buffer (2 X Stock Soln.)

To make 1000 ml:	200 mM	Potassium Glutamate
	50 mM	Tris Acetate (pH 7.6)
	20 mM	Magnesium Acetate
	100µg/ml	Bovine Serum Albumin
	1 mM	2-ß Mercapto-Ethanol
To make 20 ml:	0.741 g	Pot. Glutamate
	1 ml	1M Tris Acetate(pH7.6)
	0.090 g	Mg. Acetate
	1 ml	2mg/ml BSA Soln.
	400 µl	50 mM 2-ß M EtOH

Add sterile distilled H_2O to 20 ml. Filter sterilize. If a 1 X conc. is required, dilute 1/2 before use.

2.2 Miscellaneous Solutions

(a) Tris Acetate (1 M, pH 7.6)

To make 1000 ml: Dissolve 121,1 g Tris base in 800 ml distilled H_2O; adjust pH to 7.6 with Acetic Acid. Make volume up to 1000ml. Autoclave.

(b) Bovine Serum Albumin (BSA) (2 mg/ml Stock Soln)

To make 20 ml: Dissolve 40 mg BSA powder in 20 ml distilled H_2O. Filter Sterilize.

(c) 2-ß Mercapto-Ethanol (2-ß M EtOH) (50 mM Stock Soln.)

To make 10 ml: Add 35 µl 14,4 M 2-ßM EtOH stock Soln. to 9.965 ml distilled H_2O. Store in a dark bottle at 4°C.

(d) Restriction Enzymes

Enzymes were diluted in 1 X KGB Buffer from laboratory stocks, to a working concentration of 2 units/μl per digest.

3. End-Labelling Reactions

3.1 Reagents Used

(a) Deoxynucleotides

Each of the three Deoxynucleotides, dATP, dTTP and dGTP, were diluted in sterile distilled H_2O from laboratory stock solutions of concentration 20 mM, to a final concentration of 2 mM per end-labelling reaction.

(b) ^{32} P DeoxyCytidine Phosphate

^{32} P dCTP was diluted in sterile distilled H_2O from a laboratory stock solution of 10 μCi/μl, to a working concentration of 1 μCi/μl per end-labelling reaction.

(c) Klenow Polymerase

Klenow Polymerase was diluted in sterile distilled H_2O from a laboratory stock solution of 6 units/μl, to a working concentration of 1 unit/μl, per end-labelling reaction.

4. Gel Electrophoresis

4.1 Buffers

(a) Tris Acetate EDTA (TAE) Buffer (50 X Stock Soln.)

To make 1000 ml:	242.0	g	Tris base
	57.1	ml	Glacial Acetic Acid
	100.0	ml	0,5 M EDTA (pH 8)
	0.05%	v/v	Sodium Pyrophosphate

Add distilled H_2O to 1000 ml. Autoclave. Store at Room Temperature. Dilute 1/50 before use.

(b) Gel Loading Buffer (6 X)

To make 30 ml:

0.25% w/v Bromophenol Blue
40.0% w/v Sucrose
20 mM EDTA (pH 8)

Add distilled H_2O to 30 ml. Store at 4°C.

4.2 Miscellaneous Solutions

(a) Sodium Pyrophosphate (NaPP) (10% Stock Soln.)

To make 1000 ml: Dissolve 100 g Sodium Pyrophosphate in 1000ml distilled H_2O.

(b) Preparation of Large Agarose Gels for Electrophoresis

For a typical 1.2% gel, dissolve 1,8 g dry agarose powder in 150 ml 1 X TAE Buffer, by heating in microwave oven. When cooled to about 50°C, pour onto gel apparatus and allow to set.

5. Southern Transfer

5.1 Solutions Used

(a) Denaturing Solution

To make 1000 ml:

0.5 M NaOH
1.5 M NaCl

Add distilled H_2O to 1000 ml.

(b) Neutralization Solution

To make 1000 ml:

0.5 M Tris.Cl (pH 7.5)
20X SSC (pH 7.5)

Add distilled H_2O to 1000 ml.

(c) Prehybridization Solution

To make 1000 ml:

6 X	SSC
0.1%	SDS
0.06%	NaPP
0.25% w/v	Protea Milk Powder

Make up to 1000 ml with distilled H_2O.

(d) SSC (20 X Stock Soln., pH 7.5)

To make 1000 ml: Dissolve 175.3 g NaCl and 88.2 g Sodium Citrate in 800 ml distilled H_2O. Adjust pH to 7.5 with a few drops of a 10 M solution of NaOH. Adjust volume to 1000 ml. Autoclave.

(e) Stringency Washes

I To make 1000 ml:

6 X	SSC
0.1%	SDS
0.06%	NaPP

II To make 1000 ml:

3 X	SSC
0.1%	SDS

III To make 1000 ml:

0.1 X	SSC
0.1%	SDS

APPENDIX II

Restriction fragment sizes (kb) of Lambda DNA digested with HindIII:

23.130

9.416

6.682

4.361

2.322

2.027

1.564

0.125

BIBLIOGRAPHY

Adams, E.N.III. 1972. Syst. Zool. 21: 390-397.

Adams, J. and Rothman, E.D. 1982. Proc. Natl. Acad. Sci. USA 79: 3560-3564.

Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. 1981. Nature (London) 290: 457-465.

Aquadro, C.F. and Greenberg, B.D. 1983. Genetics 103: 287-312.

Arctander, P. 1988. J.Ornithol. 29: 205-216.

Ashley, M.V., Laipis, P.J. and Hauswirth, W.W. 1989. Nucl. Acids Res. 17(18):7325-7331.

Avise, J.C. 1986. Phil. Trans. Roy. Soc. London Ser. B 312:325-342.

Avise, J.C., Giblin-Davidson, C., Laerm, J., Patton. J.C., and Lansman, R.A. 1979a. Proc. Natl. Acad. Sci. USA 76: 6694-6698.

Avise, J.C., Lansman, R.A. and Shade R.O. 1979b. Genetics 92:279-295.

Avise, J.C., Patton, J.C. and Aquadro, C.F. 1980. J. Heredity 71:303-310.

Avise, J.C. and Aquadro, C.F. 1982. Evol. Biol. 15: 151-158.

Avise, J.C. and Lansman, R.A. 1983. Polymorphism of Mitochondrial DNA in Populations of Higher Animals., pp. 165-190. In M. Nei and R.K. Koehn (eds.), *Evolution of Genes and Proteins*. Sunderland, MA.: Sinauer Associates.

Avise, J.C., Arnold, J., Ball, R.M., Bermingham, E., Lamb, T., Neigel, J.E., Reeb, C.A. and Saunders, N.C. 1987. Ann. Rev. Ecol. Syst. 18:489-522.

Avise, J.C. and Zink, R.M. 1988. Auk 105:516-528.

Barrowclough, G.F. and Corbin, K.W. 1978. Auk 95:691-702.

Baverstock, P.R., Cole, S.R., Richardson, B.J. and Watts, C.H.S. 1979. Syst. Zool. 28:214-219.

Benjamin, D.C., Berzofsky, J.A., East, I.J., Gurd, F.R.N., Hannum, C., Leach, S.J., Margoliash, S., Michael, J.G., Miller, A., Prager, E.M., Reichlin, M., Sercarz, E.E., Smith-Gill, S.J., Todd, P.E. and Wilson, A.C. 1989. Annu. Rev. Immunol. 2:67-101.

Benton, M. 1990. Tree 5:393-394.

Bock, W.J. and Farrand, J. 1980. Amer. Mus. Novit. No.2703:1-29.

Bonell, M.J. and Selander, R.K. 1974. Science 184:908-909.

Britten, R.J. 1986. Science 231:1393-1398.

Britten, R.J., Cetta, A. and Davidson, E.H. 1978. Cell 15: 1175-1186.

Brown, W.M. 1980. Proc. Natl. Acad. Sci. USA 77:3605-3609.

Brown, W.M. and Vinograd, J. 1974. Proc. Natl. Acad. Sci. USA 71:4617-4621.

Brown, W.M., George, M. and Wilson, A.C. 1979. Proc. Natl. Acad. Sci. USA 76:1967-1971.

Brown, G.G. and Simpson, M.V. 1982. Proc. Natl. Acad. Sci. USA 79:3246-3250.

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. 1982. J. Mol. Evol. 18:225-239.

Caccone, A. and Powell, J.R. 1987. Evolution 41:1215-1238.

Cann, R.L. and Wilson, A.C. 1983. Genetics 104:699-711.

Cann, R.L., Brown, W.M. and Wilson, A.C. 1984. Genetics 106:479-499.

Cann, R.L., Stoneking, M. and Wilson, A.C. 1987. Nature 325:31-36.

Carlson, S.S., Wilson, A.C. and Maxson, R.D. 1978. Science 200:1183-1185.

Carr, S.M. and Brothers, A.J. and Wilson, A.C. 1987. Evolution 41:176-190.

Cavalli-Sforza, L.L. and Edwards, A.W.F. 1967. Evolution 21:550-570.

Collier, G. and O'Brien, S. 1985. Evolution 39:473-487.

Cracraft, J. 1981. Auk 98:681-714.

Crowe, T.M. 1988. Trans. Roy. Soc. S. Afr. 46:317-334.

Crowe, T.M. and Crowe, A.A. 1985. The genus Francolinus as a model for avian evolution and biogeography in Africa: I. Relationships among Species, pp. 207-231. In K.L. Schumann (ed.), *Proceedings of the International Symposium on African Vertebrates*. Bonn, Museum Alexander Koenig.

Crowe, T.M., Harley, E.H.H., Jakutowicz, M., Komen, J. and Crowe, A.A. 1990. Phylogenetic, Taxonomic, and Biogeographical Implications of Variation in Mitochondrial DNA, Morphology, and Behaviour of Francolins (Galliformes: Phasianidae, Genus Francolinus). Submitted.

Dawid, I.B. 1972. Dev. Biol. 29:139-151.

Debry, R.W. and Slade, N.A. 1985. Syst. Zool. 34:21-34.

Densmore, L.D., Wright, J.M. and Brown, W.M. 1985. Genetics 110:689-707.

Eck, R.V. and Dayhoff, M.O. 1966. *Atlas of Protein Sequence and Structure*. Silver Springs, MD.: National Biomedical Research Foundation.

Efron, B. and Gong, G. 1983. Am. Stat. 37:36-48.

Estabrook, G.F. 1983. The Causes of Character Incompatibility., pp. 279-295. In J. Felsenstein (ed.), *Numerical Taxonomy*. NATO ASI Series, Vol. G1. Berlin: Springer-Verlag.

Farris, J.S. 1970. Syst. Zool. 34:21-34.

Farris, J.S. 1971. Ann. Rev. Ecol. Syst. 2:277-302.

Farris, J.S. 1972. Am. Natur. 106:645-668.

Farris, J.S. 1977. Syst. Zool. 26:77-88.

Fauron, C. M.-R. and Wolstenholme, D.R. 1980. Nucl. Acids Res. 11:2439-2453.

Ferris, S.D., Brown, W.M., Davidson, W.S. and Wilson, A.C. 1981. Proc. Natl. Acad. Sci. USA 78:6319-6323.

Ferris, S.D., Sage, R.D., Huang, C.-M., Nielsen, J.T., Ritte, U. and Wilson, A.C. 1983a. Proc. Natl. Acad. Sci. USA 80:2290-2294.

Ferris, S.D., Sage, R.D., Prager, E.M., Ritte, U. and Wilson, A.C. 1983b. Genetics 105:681-721.

Felsenstein, J. 1978a. Syst. Zool. 27:401-410.

Felsenstein, J. 1978b. Syst. Zool. 27:27-33.

Felsenstein, J. 1981a. J. Mol. Evol. 17:368-376.

Felsenstein, J. 1981b. Evolution 35:1229-1242.

Felsenstein, J. 1983. Ann. Rev. Ecol. Syst. 14:313-333.

Felsenstein, J. 1988. Ann. Rev. Genet. 22:521-565.

Fitch, W.M. 1971. Syst. Zool. 20:406-416.

Fitch, W.M. and Margoliash, E. 1967. Science 155:279-284.

Freifelder, D. 1983. *Molecular Biology*: A Comprehensive Introduction to Prokaryotes and Eukaryotes. Boston, Portola Valley: Jones and Bartlett Publishers, Inc.

Frost, P.G.F. 1975. Bull. Br. Orn. Club 95:64-68.

Gilbert, D.A., Lehman, N., O'Brien, S.J. and Wayne, R.K. 1990. Nature 344:764-766.

Gillespie, J.H. 1984. Proc. Natl. Acad. Sci. USA 81: 8009-8013.

Gillespie, J.H. 1986a. Mol. Biol. Evol. 3:138-155.

Gillespie, J.H. 1986b. Genetics 113:1077-1091.

Glaus, K.R. 1980. The Structure, Organization and Evolution of Avian Mitochondrial DNA, Ph.D. Thesis, Ohio State University, Columbus.

Goodman, M. 1961. Hum. Biol. 33:131-162.

Goodman, M. 1985. Bioessays 3:9-14.

Goodman, M., Czeluszniak, J., Moore, G.W., Romero-Herrera, A.E. and Matsuda, G. 1979. Syst. Zool. 28:132-163.

Goodman, M., Miyamoto, M.M. and Czeluszniak, J. 1987. Pattern and Process in Vertebrate Phylogeny Revealed by Coevolution of Molecules and Morphologies., pp. 141-176. In C. Patterson (ed.), *Molecules and Morphology in Evolution: Conflict or Compromise?* Cambridge: Cambridge University Press.

Greenberg, B.D., Newbold, N.E. and Sugino, A. 1983. Gene 21:33-49.

Grivell, L.A. 1983. Sci. Am. 248:60-73.

Gutierrez, R.J., Zink, R.M. and Yang, S.Y. 1983. Auk 100: 33-40.

Gyllensten, U., Wharton, D. and Wilson, A.C. 1985. J. Hered. 76:321-324.

Hall, B.P. 1963. Bull. Brit. Mus (Nat. Hist.) 10:8-204.

Harley, E.H. 1988. S. Afr. J. Science. 84:158-159.

Harris, H. 1966. Proc. Roy. Soc. London Ser. B 164: 298-310.

Hasegawa, M., Kishino, H. and Yano, T. 1985. J. Mol. Evol. 22:160-174.

Helm-Bychowski, K.M. 1984. Evolution of Nuclear and Mitochondrial DNA in Gallinaceous Birds, Ph.D. Thesis, University of California, Berkeley.

Hendy, M.D. and Penny, D. 1982. Math. Biosci. 59:277-290.

Higuchi, R.G., Wrischnik, L.A., Oakes, E., George, M., Tong, B. and Wilson, A.C. 1987. J. Mol. Evol. 25:283-287.

Hillis, D.M. 1985. Syst. Zool. 34:109-126.

Hillis, D.M. 1987. Ann. Rev. Ecol. Syst. 18:23-42.

Hillis, D.M. and Moritz, C (eds.). 1990. *Molecular Systematics*. Sunderland, MA.: Sinauer Associates, Inc.

Hixson, J.E. and Brown, W.M. 1986. Mol. Evol. Biol. 3: 1-18.

Hubby, J.L. and Lewontin, R.C. 1966. Genetics 52:203-215.

Jeffreys, A.J., Wilson, V. and Thein, S.L. 1985. Nature 316:76-79.

Johnsgard, P.A. 1986. *The Pheasants of the World*. Oxford: Oxford University.

Johnson, N.K., Zink, R.M., Barrowclough, G.F. and Marten, J.A.. 1984. Wilson Bull. 96:543-560.

Jukes, T.H. and Cantor, C.R. 1969. Evolution of Protein Molecules., pp. 21-132. In H.N. Munro (ed.), *Mammalian Protein Metabolism*. New York: Academic Press.

Kaplan, N. and Langley, C.H. 1979. J. Mol. Evol. 13: 295-304.

Kasamatsu, H., Robberson, D.L. and Vinograd, J. 1971. Proc. Natl. Acad. Sci. USA 68:2252-2257.

Kessler, L.G. and Avise, J.C. 1985. Mol. Biol. Evol. 2: 109-125.

Kimura, M. 1968. Nature 217:624-626.

Kimura, M. 1980. J. Mol. Evol. 16:111-120.

Kimura, M. 1983. *The Neutral Theory of Evolution.* Cambridge: Cambridge University Press.

Kluge, A.G. 1983. Cladistics and the Classification of the Great Apes, pp. 151-177. In R.L. Ciochan and R.S. Corrucini (eds.), *New Interpretations of Ape and Human Ancestry..* New York: Plenum.

Kluge, A.G. and Farris, J.S. 1969. Syst. Zool. 18:1-32.

Kohne, D.E. 1970. Q. Rev. Biophys. 33:327-375.

Komen, J. 1987. S. Afr. J. Wildl. Res. Suppl. 1:82-86.

Lansman, R.A. and Clayton, D.A. 1975. J. Mol. Biol. 99: 761-776.

Lansman, R.A., Shade, R.O., Shapira, J.A. and Avise, J.C. 1981. J. Mol. Evol. 17:214-226.

Lansman, R.A., Avise, J.C. and Huettel, M.D. 1983a. Proc. Natl. Acad. Sci. USA 80:1969-1971.

Lansman, R.A., Avise, J.C., Aquadro, C.F., Shapira, J.F. and Daniel, S.W. 1983b. Evolution 37:1-16.

Le Quesne, W.J. 1974. Syst. Zool. 23:513-517.

Le Quesne, W.J. 1982. Zool. J. Linn. Soc. 74:267-275.

Lewin, R. 1988a. Science 241:1598-1600.

Lewin, R. 1988b. Science 241:1756-1759.

Lewontin, R.C. 1974. *The Genetic Basis of Evolutionary Change.* New York: Columbia University Press.

Li, W.-H. 1981. Proc. Natl. Acad. Sci. USA 78:1085-1089.

Li, W.-H. 1986. Genetics 113:187-213.
Lynch, M. 1988. Mol. Biol. Evol. 5:584-599.

Margush, T. and McMorris, F.R. 1981. Bull. Math. Biol. 43: 239-244.

Matsumoto, L., Kasamatsu, H., Piko, L. and Vinograd, J. 1974. J. Cell. Biol. 63:146-149.

Maxson, R.D. and Maxson, L.R. 1986. Mol. Biol. Evol. 3: 375-388.

McLelland, M., Hamish, J., Nelson, M. and Patel, Y. 1988. Nucl. Acids Res. 16:364.

Mihok, S., Fuller, W.A., Canham, R.P. and McPhee, E.C. 1983. Evolution 37:332-340.

Milne-Edwards, A. and Grandidier, A. 1885. Histoire Physique, Naturelle et Politique de Madagascar. Vol. 12, Histoire Naturelle des Oiseaux Tome 1:377-779.

Milstein, P. le S. and Wolff, S.W. 1987. S. Afr. J. Wildl. Res. Suppl. 1:58-65.

Miyamoto, M.M. 1985. Syst. Zool. 32:109-124.

Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M. and Kobayashi, M. 1982. J. Mol. Evol. 19:28-35.

Moritz, C. and Brown, W.M. 1987. Proc. Natl. Acad. Sci. USA 84:7183-7187.

Moritz, C., Dowling, T.E. and Brown, W.M. 1987. Ann. Rev. Ecol. Syst. 18:269-292.

Mueller, L.D. and Ayala, F.J. 1982. Genet. Res. 40: 127-137.

Mullis, K.B. and Faloona, F.A. 1987. Methods Enzymol. 155: 335-350.

Nei, M. and Li, W.-H. 1979. Proc. Natl. Acad. Sci. USA 76: 5269-5273.

Nei, M. and Tajima, F. 1981. Genetics 97:145-163.

Nei, M. anf Tajima, F. 1985. Mol. Biol. Evol. 2:189-205.

Nei,M. *Molecular Evolutionary Genetics*. 1987. New York: Columbia University Press.

Neigel, J.E. and Avise, J.C. 1985. Phylogenetic Relationships of Mitochondrial DNA under Various Demographic Models of Speciation., pp. 515-534. In E. Nevo and S. Karlin (eds.), *Evolutionary Processes and Theory*. New York: Academic Press.

Nelson, G. 1979. Syst. Zool. 28:1-21.

Neyman, J. 1971. Molecular Studies of Evolution: A Source of Novel Statistical Problems., pp. 1-27. In S.S. Gupta and J. Yackel (eds.), *Statistical Decision Theory and Related Topics*. New York: Academic Press.

Nuttal, G.H.F. 1904. *Blood Immunity and Blood Relationship*. Cambridge: Cambridge University Press.

Pääbo, S. 1989. Proc. Natl. Acad. Sci. USA 86:1939-1943.

Palmer, J.D., Jorgensen, R.A. and Thompson, W.F. 1985. Genetics 109:195-213.

Penny, D. 1982. J. Theor. Biol. 96:129-142.

Penny, D. and Hendy, M.D. 1985. Cladistics 1:266-272.

Penny, D. and Hendy, M.D. 1986. Mol. Evol. Biol. 3: 403-417.

Plante, Y., Boag, P.T. and White, B.N. 1987. Can. J. Zool. 65:175-180.

Platnick, N.I. 1988. Nature 335:310.

Potter, S.S., Newbold, C.A., Hutchison, C.A. and Edgell, M.H. 1975. Proc. Natl. Acad. Sci. USA 72: 4494-4500.

Powell, J.R., Caccone, A., Amato, G. and Yoon, D. 1986. Proc. Natl. Acad. Sci. USA 83:9090-9093.

Prager, E.M., Brush, A.H., Nolan, R.A., Nakanishi, M. and Wilson, A.C. 1974. J. Mol. Evol. 3:243-262.

Quinn, T.W. and White, B.N. 1987. Analysis of DNA Sequence Variation., pp. 163-198. In F. Cooke and P.A. Buckley (eds.), *Avian Genetics*. London: Academic Press.

Rabinowitz, M. and Swift, H. 1970. Physiol. Rev. 50: 376-427.

Rand, D.M. and Harrison, R.G. 1956. Genetics 114:955-970.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. 1977. J. Mol. Biol. 113:237-251.

Ryman, N., and Utter, F. (eds.). 1987. *Population Genetics and Fishery Management*. Seattle: University of Washington Press.

Saitou, N. and Nei, M. 1986. J. Mol. Evol. 24:189-204.

Saitou, N, and Nei, M. 1987. Mol. Biol. Evol. 4:406-425.

Sarich, V.M. 1969. Syst. Zool. 18:286-295.

Sarich, V.M. and Wilson, A.C. 1966. Science 154:1563-1566.

Scharf, S.J., Horn, G.T. and Erlich, H.A. 1986. Science 233:1076-1078.

Sharkey, M.J. 1989. Cladistics 5:63-86.

Sheldon, F.H. 1987. Mol. Biol. Evol. 4:56-69.

Shields, G.F. and Helm-Bychowski, K.M. 1988. Mitochondrial DNA of Birds., pp. 273-295. In R.F. Johnston (ed.), *Current Ornithology*, Vol. 5. New York: Plenum.

Shoshani, J. 1986. Mol. Biol. Evol. 3:222-242.

Sibley, C.G. and Ahlquist, J.E. 1983. The Phylogeny and Classification of Birds based on the Data of DNA-DNA Hybridization., pp. 245-292. In R.F. Johnston (ed.), *Current Ornithology*, Vol. 1. New York: Plenum.

Sibley, C.G. and Ahlquist. J.E. 1984. J. Mol. Evol. 20: 2-15.

Sibley, C.G. and Ahlquist, J.E. 1985. The Relationships of some Groups of African Birds, based on comparisons of the Genetic Material, DNA, pp. 115-170. In K.-L. Schumann (ed.), *Proceedings of the International Symposium of African Vertebrates*. Bonn, Museum Alexander Koenig.

Sibley, C.G. and Ahlquist, J.E. 1987a. Avian Phylogeny Reconstructed from Comparisons of the Genetic Material, DNA., pp. 95-121. In C. Patterson (ed.), *Molecules and Morphology in Evolution: Conflict of Compromise?* Cambridge: Cambridge University Press.

Sibley, C.G. and Ahlquist, J.E. 1987b. J. Mol. Evol. 26:99-121.

Smith, C.A., Jordan, J.M. and Vinograd, J. 1971. J. Mol. Biol. 59:255-272.

Smithies, O. 1955. Biochem. J. 61:629-641.
Smouse, P.E. and Li, W.-H. 1987. Evolution 41:1162-1176.
Sneath, P.H.A. and Sokal, R.R. 1973. *Numerical Taxonomy*. San Fransisco: W.H. Freeman.

Sourdis, J. and Nei, M. 1988. Mol. Biol. Evol. 5:298-311.

Southern, E.M. 1975. J. Mol. Evol. 98:503-517.

Springer, M.S. and Krajewski, C. 1989. Q. Rev. Biol. 64: 291-318.

Swofford, D.L. and Olsen, G.J. 1990. Phylogeny Reconstruction., pp. 411-501. In D.M. Hillis and C. Moritz (eds.), *Molecular Systematics*. Sunderland, MA: Sinauer Associates, Inc.

Tateno, Y., Nei, M. and Tajima, F. 1982. J. Mol. Evol. 18: 387-404.

Templeton, A.R. 1983a. Convergent Evolution and Non-Parametric Inferences from Restriction Fragment and DNA Sequence Data., pp. 151-179. In B. Weir (ed.), *Statistical Analysis of DNA Sequence Data*. New York: Marcel Dekker.

Templeton, A.R. 1983b. Evolution 37:221-244.

Thorpe, J.P. 1982. Am. Rev. Ecol. Syst. 13: 139-168.

Tilley, S.G. and Hansman, J.S. 1976. Copeia 1976:734-741.

Upholt, W.B. 1977. Nucl. Acids Res. 4:1257-1265.

Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A.S. and Christophe, D. 1987. Science 235: 683-684.

Vawter, L. and Brown, A.C. 1986. Science 234:194-246.

Verheyen, R. 1956. Inst. R. Sci. Nat. Belg. 32:1-24.

Wake, D.B. and Larson, A. 1987. Science 238:42-48.

Wetmur, J.G. and Davidson, N. 1968. J. Mol. Evol. 31: 349-370.

Wetton, J.H., Royston, E.C., Parkin, D.T. and Walters, D. 1987. Nature 327:147-149.

Wilson, A.C., Sarich, V.M. and Maxson, L.R. 1974. Proc. Natl. Acad. Sci. USA 71:3028-3030.

Wilson, A.C., Carlson, S.S. and White, T.J. 1977. Ann. Rev. Biochem. 46:573-639.

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Helm-Bychowski, K.M., Higuchi, R.C., Palumbo, S.R., Prager, E.M., Sage, R.D. and Stoneking, M. 1985. Biol. J. Linn. Soc. 26:375-400.

Wolters, H.E. 1975. *Die Vogelarten der Erde*. Berlin: Paul Parey.

Wrischnik, L.A., Higuchi, R.A., Stoneking, M., Erlich, H.A., Arnheim, N.and Wilson, A.C. 1987. Nucl. Acids Res. 15: 529-542.

Zückerkandl, E. and Pauling, L. 1962. Molecular Disease, Evolution and Genic Heterogeneity., pp. 189-225. In M. Kasha and B. Pullman (eds.), *Horizons in Biochemistry*. New York: Academic Press.