Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

SEQUENCES AND SIGNALS: EVOLUTIONARY HISTORIES OF NEW ZEALAND SKINKS.

A thesis presented in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy in Genetics
at Massey University
New Zealand

Robert Eric Hickson

May 1993

Abstract

The application of DNA sequencing to studies of the New Zealand biota is illustrated by investigations into the evolutionary relationships of skinks in the genus *Leiolopisma*. DNA sequences from a region of the mitochondrial 12S rRNA gene were determined for 20 taxa by use of the polymerase chain reaction. A vertebrate secondary structure model for this part of the gene was developed using comparative sequence analysis and calculations of RNA folding energies. Approximately one third of the molecule does not vary between the vertebrates examined, and there are similar patterns of sequence variability among the vertebrates. The secondary structure model was subsequently used to assist phylogenetic analyses of the skink sequence data set.

Analyses of the mitochondrial DNA sequence information, using newly developed and more sophisticated algorithms, did not produce a fully resolved phylogenetic tree for all the skinks, though relationships between some taxa are less ambiguous. The lack of resolution does not appear to be due to limitations in the analytical methods, nor to the patterns of nucleotide substitutions in the skink 12S rRNA sequences. The skink sequence data set is unusual in that most of the taxa have similar numbers of nucleotide substitutions when compared to each other. These results are interpreted as reflecting a rapid divergence of the *Leiolopisma* group of skinks. Simulation studies support this interpretation.

Three hypotheses are presented to account for the patterns of sequence differences between the skinks. One proposes a Gondwanan divergence of the *Leiolopisma* group, about 80 million years ago. Under this hypothesis the distribution of the skinks on islands in the Pacific and Indian oceans can be explained, in part, by continental drift. A second hypothesis suggests that New Zealand *Leiolopisma* are derived from a Miocene (15-25 million years ago) evolutionary radiation in New Zealand. This hypothesis however is inconsistent with observations of the sequence similarities between Mauritian, Australian, and New Zealand skinks. A third hypothesis, proposing several independent colonizations of New Zealand by *Leiolopisma*, is also not as well supported by the available sequence data. However, a close relationship between the 12S rRNA sequences of one New Zealand species *L. infrapunctatum* and the Australian *Lampropholis guichenoti* suggests that at least two skink immigrations to, or emigrations from, New Zealand may have occurred. Predictions of the three hypotheses and strategies to test them are discussed.

Some of the conclusions derived from analyses of the mitochondrial DNA sequences conflict with those obtained from allozyme information, though there are points of agreement. Comparison of the allozyme and sequence data do also revealed a case of hybridization between two sympatric species, *L. n. polychroma* and *L. maccanni*, at a site in Southland. Analyses of both data sets indicate that the morphological similarity of *Leiolopisma* species obscures a large amount of genetic diversity, and the evolutionary histories of New Zealand *Leiolopisma* are older and more complex than previously considered. Further genetical and ecological studies of *Leiolopisma* are required, but this thesis emphasizes both the suitability and necessity of molecular genetic approaches for evolutionary investigations in New Zealand.

Acknowledgments

Barry Scott and Garry Latch are penultimately to blame for this thesis. None of us knew what we were really in for.

Financial support was provided by the University Grants Committee, the Nga Manu Trust, Massey University, the Molecular Genetics Unit and Department of Microbiology and Genetics. Additional support came from the Dean of Science, the Royal Society of New Zealand, the Alfred P. Sloan Foundation (not once but twice!), and my parents. To all I am very grateful.

Iain Lamont and Liz Poole of Otago University provided hospitality and Liz showed me how to double-strand sequence. Brad Shaffer, Centre for Population Biology, University of California Davis, permitted me time, space and consumables in his lab. Ross Sadlier of the Australian Museum, Sydney, provided tissue samples of New Caledonian skinks.

With Robert Fisher, Centre for Population Biology, University of California Davis, I enjoyed many hours of entertaining and interesting discussions, and he let me share a room with his lizards. Thanks Robert, but get on with your thesis! Mike Hendy, Mike Steel, Mike Charleston, Peter Waddell, and Chris Simon are also thanked for stimulating discussions and ideas.

Charlie Daugherty, Victoria University, provided the initial inspiration for this study. He gave me access to the skink tissues and was ever willing to fill me in on the details of New Zealand skinks.

Barry Scott, Tim Brown, and Geoff Malcolm provided a stimulating and friendly atmosphere in which to work, freedom to do as I pleased and to go where I wanted, and money to make it possible. Mister Tucker, Paul, Laura and Terri covered administration and clerical things.

Thanks also to my travel agent Sue. But absolutely no thanks to Guido and the Linguine Gang, Milano Station, Italy, even though they too relieved me of a burden.

Lots of people have been and gone, or are still around. Thanks to them for all sorts. Especially Carolyn and Miranda.

To the following, words are no recompense:

David and Eric, both much more than superb and challenging supervisors and friends.

Alan, Gina, and particularly Mary of "Capital City Campus".

Stan, Pete, Trish and Liz of the Farside (and other places).

But most of all I thank my parents, with much love.

Table of Contents

Abstract i
Acknowledgements ii
Table of Contents iii
List of Figures vii
List of Tables x

Chapter One: Introduction 1
New Zealand in an Evolutionary Context 1
The Oligocene 1
The Pliocene 2
The Pleistocene 2
The Holocene 2
Evolutionary Studies in New Zealand 2
Biogeographic Studies in New Zealand 3
Assessing Taxonomic Relationships 4
Secondary Metabolites 4
Immunological Techniques 4
Chromosomal Studies 4
Allozyme Studies 5
DNA Fragments 5
DNA Sequences 6
Mitochondrial DNA and Evolution 6
Rates of Evolution in Cytoplasmic Organelles 7
Limitations of Sequence Data 8
Rates of Sequence Evolution 9
Skinks as a Model Group for Evolutionary Studies 9
Views on the Origin of New Zealand Lizards 9
Ecological Aspects of Some *Leiolopisma* 10
Genetical Investigations of *Leiolopisma* 11

Chapter Two: Methods 13
The Polymerase Chain Reaction 13
Mitochondrial DNA Isolation 14
Total DNA Isolation 14
Amplification of DNA by the Polymerase Chain Reaction 16
Direct Sequencing of PCR Products 17
RESULTS 18

Chapter Three: A Refined Secondary Structure Model
for Domain III of Vertebrate 12S rRNA 22
Refining a Secondary Structure Model 22
The Skink and Other Vertebrate 12S rRNA Sequences 23
Approaches for Refining a Vertebrate Domain III Model 23
Energetics of the Folding of the rRNA Sequences 23
Identifying Individual Helices 27
Helix 35 27
Helix 36 27
Helix 37 27
Helix 38 29
Helix 39 29
Helix 40 29
The Refined Secondary Structure Vertebrate Model 29
Conserved Positions in the Secondary Structure Model 30
Variability within Domain III 30
Nucleotide Substitutions in the Skink Structure 30
DISCUSSION 32
Interactions Between Domain III and Ribosomal Proteins 32
Tetra-loops 36
Comparison of Secondary Structure Models 36
Limitations of the RNA Folding Algorithm 37
Variability between Regions 38
Secondary Structure Models and Sequencing Precision 38
Bias in Types of Substitutions 39

Chapter Four: Transitions and Transversions 40
Transitions and Transversions 40
Saturation of Transitions? 40
Simulations 45
Estimating Divergence from Transversions 47
Comparisons to Other Data Sets 47
Changes in the Context of Secondary Structure 51
DISCUSSION 52
Rates of mtDNA Evolution 53
Estimating Degree of Divergence with Transversions 54
Regions of Variability 54

Chapter Five: Signals in the Data and Phylogenetic Analyses 56
Spectral Analysis of DNA Sequences 56
What is a Bipartition and How is it Represented? 57
Lack of Resolution in the Skink 12S rRNA Data Set 60
Spectral Analysis of other Vertebrate 12S rRNA Data Sets 60
Parsimony and Neighbor-Joining Analyses of the Skink Data Set 60
Partially Resolved Relationships Using the Hadamard Conjugation 61
Weighting of Characters 61
Weighting of Transversions 62
Weighting of Paired Regions 62
Weighting of More Variable Sites 64
The Effect of Constant Columns in Sequence Analysis 64
The Effect on Skink Phylogeny of Weighting of Constant Columns 65
Summary of the Effect of Weighting Schemes on the Skink Phylogeny 65
Spectral Analysis of Simulated Sequences 66
Relationships of Northern & Southern New Zealand *Leiolopisma* 66
Spectral Analysis of Skinks from Northern New Zealand 68
The Complex Relationships of *C. aenea* 68
Spectral Analysis of Southern New Zealand Skinks 69
Mitochondrial Lineages of the Skinks 69
Addition of a New Caledonian Skink 69
Cytochrome *b* Sequence Data 71
DISCUSSION 71
"Rapid" Divergence of *Leiolopisma* 74
Estimating Times of Divergence for the Skinks 74
Estimates from Immunological Data 75
Estimates from Allozyme Data 75
Estimates from Biogeographic Information 76
Hypothesis One: A Gondwanan Origin and Diversification 76
The Mauritian Leiolopisma are Still Enigmatic 77
Origins of the New Zealand Skinks 77
Diversification Within New Zealand 78
Hypothesis Two: Post-Oligocene Diversification 78
Phylogenetic Limitations of the Skink Allozyme Data 79
Hypothesis Three: Multiple *Leiolopisma* Colonizations of NZ 79
Hypotheses for the testing 80

Chapter Six: Comparing Sequence and Allozyme Data 81
Conflict Between Allozyme and 12S rRNA Sequence Data 81
Congruence Between Allozyme and 12S rRNA Sequence Data 81
Hybridization 83
DISCUSSION 84
Hybridization - Questions to Address 85
Species Concepts 87
Origins of Some Specific New Zealand *Leiolopisma* 87

Chapter Seven: Discussion 89
Secondary Structure and Phylogenetic Analyses 89
Evolution of the 12S rRNA Molecule 90
Simulation Studies 91
The Hadamard Conjugation and Spectral Analysis 92
Incompletely Resolved Skink Phylogeny 92
Hypotheses for the Origins of New Zealand *Leiolopisma* 93
Predictions of the Three Hypotheses 94
Hypothesis One: Gondwanan Origins 94
Hypothesis Two: Oligocene Drowning 94
Hypothesis Three: Multiple Colonists 94
Testing the Hypotheses: What Other Skinks to Sample? 95
Testing the Hypotheses: What Other Genes to Sample? 95
Investigation of Nuclear Gene Sequences 96
Hybridization between *L. n. polychroma* and *L. maccanni* 97
The Potential of *L. n. polychroma* For Population Studies 98
Rates of Molecular and Morphological Evolution in Skinks 98
Evolutionary Investigations of New Zealand Biota 99

References 101

Appendix 1 ..... *on diskette*
a). **Skink 12S rRNA sequences**
b). **Skink cytochrome *b* sequences**

Appendix 2 ..... *on diskette*
a). **Xantusiid 12S rRNA sequences**
b). **Ratite 12S rRNA sequences**
c). **Bovid 12S rRNA sequences**
d). **Great Ape 12S rRNA sequences**

List of Figures

		Following page
Figure 2.1	Map of New Zealand.	15
Figure 2.2	Inferred phylogenetic relationships of vertebrates based on 12S rRNA sequences.	19
Figure 3.1	Secondary structure of prokaryote and mitochondrial small subunit rRNA.	22
Figure 3.2	Potential pairings for helix 36 in vertebrate 12S rRNA.	28
Figure 3.3	Secondary structure model for domain III of skink 12S rRNA.	29
Figure 3.4	Potential alternative secondary structures for skink domain III.	29
Figure 3.5	Conserved positions in vertebrate 12S rRNA domain III sequences.	30
Figure 3.6	Distribution of variable sites in domain III of vertebrate 12S rRNAs.	31
Figure 3.7	Regions of variability in the skink 12S rRNA sequence.	31
Figure 4.1	Frequency distributions for total nucleotide differences, transitions and transversions in skink 12S rRNA sequences.	42
Figure 4.2	Numbers of transversions versus numbers of transitions for skink 12S rRNA sequences.	44
Figure 4.3	Total number of differences versus numbers of transversions for skink 12S rRNA sequences.	44
Figure 4.4	Relationship between numbers of transitions and numbers of transversions for ratites and pecoran bovids.	44
Figure 4.5	Proportion of transition substitutions in skink 12S rRNA sequences.	44
Figure 4.6	Relationship between number of transversions and the transition/transversion ratio for the skink data set.	45
Figure 4.7	Relationship between number of transversions and the transition/transversion ratio for other data sets.	45
Figure 4.8	Frequency distribution of transitions and transversions in simulated sequence data.	46
Figure 4.9	Relationship between numbers of transitions and numbers of transversions in simulated sequence data.	46
Figure 4.10	Relationship between numbers of transitions and numbers of transversions in simulated sequence data using 40 and 55 expected substitutions.	46
Figure 4.11	Relationship between numbers of transitions and numbers of transversions for other simulations using Ts/Tv ratios of 2/1 and 8/1.	46
Figure 4.12	Frequency distributions of transitions and transversions in simulated sequence data for a large number of iterations.	46
Figure 4.13	Relationship between numbers of transitions and numbers of transversions in simulated sequence data when 500 iterations are performed.	46

Figure 4.14	Means and standard deviations of estimates for the real number of substitutions.	47
Figure 4.15	Means and standard deviations of estimates for the real numbers of substitutions for other simulations using 40 and 55 expected substitutions.	47
Figure 4.16	Means and standard deviations of estimates for the real numbers of substitutions for other simulations using Ts/Tv ratios of 2/1 and 8/1.	47
Figure 4.17	Variable sites in the skink 12S rRNA domain III.	51
Figure 4.18	Location of nucleotide substitutions in 12S rRNA for pairs of skink taxa with a small number of differences between them.	51
Figure 4.19	Comparisons of regions of variability in the 12S rRNA sequence for pairs of closely and more distantly related skink taxa.	51
Figure 4.20	Comparisons of regions of variability in the 12S rRNA sequence for closely and more distantly related ratites, and great apes.	51
Figure 5.1	Spectral analysis of skink 12S rRNA sequences.	60
Figure 5.2	Spectral analyses of ratite and bovid 12S rRNA sequences.	60
Figure 5.3	Maximum parsimony tree for skink 12S rRNA sequences.	61
Figure 5.4	Neighbor-joining tree for skink 12S rRNA sequences.	61
Figure 5.5	Inferred phylogenetic relationships of 20 skink taxa based on spectral analysis of 12S rRNA sequences.	61
Figure 5.6	Effect of weighting transversions in skink data set.	62
Figure 5.7	Inferred phylogenetic relationships of 20 skink taxa based on spectral analysis of 12S rRNA sequences (transversions weighted four times higher than transitions).	62
Figure 5.8	Effect of weighting upaired regions three-fold higher than unpaired regions in skink 12S rRNA.	63
Figure 5.9	Effect of down-weighting the eight most variable sites in the skink 12S rRNA.	64
Figure 5.10	Effect of removing the eight most variable sites.	64
Figure 5.11	A resolved tree of 10 skink taxa, in which the two strongest signals are not in the tree.	65
Figure 5.12	Effect on the 10 taxa spectrum of removing constant sites from the sequence.	65
Figure 5.13	Spectral analysis of simulated sequences.	66
Figure 5.14	Spectral analysis of northern and southern skink 12S rRNA sequences.	66
Figure 5.15	Bipartitions in the optimal tree for 11 northern New Zealand skinks.	68
Figure 5.16	Bipartitions in the optimal tree for 10 northern New Zealand skinks.	68

Figure 5.17	Bipartitions in the optimal tree for 10 southern New Zealand skinks.	69
Figure 5.18	Inferred phylogenetic relationships for northern and southern New Zealand skinks based on 12S rRNA sequences.	69
Figure 5.19	Influence of the first 80 bases on spectral signals in the skink data set.	70
Figure 5.20	Spectral analysis of 19 *Leiolopisma* skinks and the New Caledonian *Tropidoscincus rohssii*.	71
Figure 5.21	Inferred phylogenetic relationships of four skink taxa from Southland using 12S rRNA and cytochrome *b* sequences.	72
Figure 5.22	Three hypotheses for the origin of *Leiolopisma* skinks.	76
Figure 6.1	Evidence for sexual hybridization in a skink population.	83
Figure 6.2	Potential offspring from sexual hybridization.	84

List of Tables

		Page
Table 2.1	Skink taxa from which DNA sequence information was obtained.	15
Table 2.2	Base composition of the skink 12S rRNA sequences.	21
Table 3.1	Alignment of 20 skink 12S rRNA sequences.	24
Table 3.2	Species and sources of 12S rRNA sequences used in Table 3.3	24
Table 3.3	Alignment of vertebrate 12S rRNA sequences.	26
Table 3.4	Local minimal free energies for folding of helices.	28
Table 3.5	Distribution of conserved nucleotides in vertebrate 12S rRNA.	31
Table 3.6	Changes occurring in paired and unpaired regions of skink 12S rRNA.	31
Table 3.7	Substitutions among paired nucleotides in skink 12S rRNA.	33
Table 3.8	Observed and expected numbers of transitions and transversions in the skink data set.	34
Table 3.9	Base composition of conserved residues in skink 12S rRNA.	35
Table 4.1	Distance matrices for skink 12S rRNA sequences.	41
Table 4.2	Mean numbers of total differences, transitions, transversions, and transition/transversion ratios for each taxon.	43
Table 4.3	Variability in total substitutions, transitions and Ts/Tv ratio in relation to numbers of transversions.	44
Table 4.4	Parameters and results of simulations.	46
Table 4.5	Distance matrices for 12S rRNA from other vertebrates.	48
Table 4.6	Comparison of estimated times of divergence among the skinks.	50
Table 5.1	Determining a relative weighting factor for paired regions.	63
Table 5.2	New Zealand *Leiolopisma* divided into northern and southern groups.	67
Table 5.3	Mitochondrial lineages in skinks.	70
Table 5.4	Distance matrices for cytochrome *b* and 12S rRNA sequences from four skink taxa.	72
Table 6.1	Comparisons of genetic distances inferred from 12S rRNA sequences and allozymes.	82