UNIVERSITY OF ADELAIDE
26. 3. 98
LIBRARY

MOLECULAR BIOLOGY AND EVOLUTION OF β-GLOBIN GENES IN MONOTREMES

Mi-Hye Lee M. Sc.

Thesis submitted for the Degree of

Doctor of Philosophy

December, 1997

Department of Genetics

University of Adelaide

MONOTREMES



Tachyglossus aculeatus

Australian Echidna

Zaglossus bruijni

New Guinea Echidna





Ornithorhynchus anatinus

Platypus

TABLE OF CONTENTS

	Page
SUMMARY	
DECLARATION	
ACKNOWLEDGEMENTS	
TERMINOLOGY	
ABBREVIATIONS	
INTRODUCTION	1
AIMS AND OBJECTIVES	5
CHAPTER 1 LITERATURE REVIEW	
1.1 HAEMOGLOBIN GENES	6
1.1.1 The globins of plants, invertebrates and vertebrates	8
1.1.2 Haemoglobin function	11
1.1.3 The evolution of globin gene structure	12
1.1.4 Expression of haemoglobin genes in vertebrate	15
1.1.5 Regulation of the β-globin switches	16
1.1.6 The structure of β-globin genes	18
1.1.6.1 Non-coding regions	18
1.1.6.2 Pseudogenes	19
1.1.7 Arrangement of β-globin gene families	20
1.1.8 Evolution of the β-globin gene family in mammals	23
1.2 MONOTREMES	27
1.2.1 Morphology and reproduction	29
1.2.2 Fossil data	31
1.2.3 Evolution of monotremes and their phylogenetic relations to the marsupials and eutherians	33
1.2.4 Biochemical and haematological studies	38
1.2.5 Chromosome studies	39
1.2.6 Monotreme globins	42
1.3 MOLECULAR EVOLUTION AND PHYLOGENY	45
1.3.1 Molecular evolution and evolutionary clocks	45
1.3.2 Molecular phylogeny	48
CHAPTER 2 MATERIALS AND METHODS	
2.1 MATERIALS	53
2.1.1 Tissue samples	53
2.1.2 Chemical reagents and other materials	53
2.1.3 Enzymes	54
2.1.4 Kits and special reagents	54
2.1.5 DNA	54
2.1.6 Nucleotides and radio-isotopes	54
2.1.7 *E. coli* strains	55

2.1.8 β-globin gene probes used in this project 55
2.1.9 Buffers and solutions 55
2.1.10 Nucleotide sequences 58
2.2 METHODS 59
2.2.1 Constructing monotreme genomic libraries 59
2.2.1.1 Preparation of high molecular weight genomic DNA 59
2.2.1.2 Partial digestion of genomic DNA with *Sau3A* 59
2.2.1.3 Separation and recovery of partially digested DNA using a glycerol gradient 59
2.2.1.4 Library construction 60
2.2.1.5 Plaque lifting 61
2.2.1.6 Oligo-labelling of DNA restriction fragments 61
2.2.1.7 Hybridisation of Hybond N+ filters with radio-labelled DNA probes 61
2.2.1.8 Column separation of unincorporated radionucleotides 62
2.2.1.9 Preparation of lambda phage DNA 62
2.2.2 Characterisation of lambda clones and recombinant plasmids 63
2.2.2.1 DNA digestion with restriction endonuclease 63
2.2.2.2 Agarose gel electrophoresis 64
2.2.2.3 Southern transfer using the alkaline blot procedure 64
2.2.2.4 Purification of DNA fragments 64
LMT Agarose gels 65
Freeze-squeeze 65
2.2.3 Preparation of plasmid subclones 65
2.2.3.1 Ligation of DNA fragment into plasmid *PGM3zf* 65
2.2.3.2 Preparation of competent *E. coli* JM 109 cells 66
2.2.3.3 Transformation of *E. coli* with plasmid DNA 66
2.2.3.4 Preparation of plasmid DNA 67
2.2.3.5 Rapid boiling plasmid mini preparations 67
2.2.4 DNA sequencing 68
2.2.4.1 Purification and denaturation of plasmid DNA 68
2.2.4.2 DNA sequencing with *Sequenase* kit 68
2.2.4.3 DNA sequencing with the *fmol* sequencing kit 69
2.2.5 Polymerase Chain Reaction 69
2.2.5.1 Synthesis and processing of oligonucleotide PCR primers 69
2.2.5.2 PCR primer design 69
2.2.5.3 PCR conditions 70
2.2.5.4 Purification of PCR products 70
2.2.6 Chromosome preparation and *in situ* hybridisation 71
2.2.6.1 Echidna blood cell culture and chromosome preparation 71
2.2.6.2 Preparation of probe DNA for fluorescent *in situ* hybridisation 71
2.2.6.3 Estimation of labelling efficiency 72
2.2.6.4 Chromosome pre-treatment 72
2.2.6.5 Denaturation of chromosomal DNA and probe hybridisation 72
2.2.6.6 Post-hybridisation washing 73
2.2.6.7 Detection and amplification of hybridised biotinylated probe 73
2.2.6.8 Chromosome staining 73
2.2.6.9 Chromosome G-banding after *in situ* hybridisation 74
2.2.7 DNA sequence analysis 74

2.2.8 Phylogenetic analysis 74
2.2.8.1 Relative rate test and synonymous/non-synonymous substitutions 74
2.2.8.2 Phylogenetic tree construction 75

CHAPTER 3 RESULTS
3.1 CHARACTERISATION OF β-GLOBIN GENES IN MONOTREMES 76
3.1.1 Identification of globin clones in echidna genomic DNA libraries 76
3.1.2 Southern analysis and restriction mapping of the echidna clones λEL.1.1 and λEPL.1 76
3.1.3 Construction of a platypus genomic library and identification of globin clones 77
3.1.4 Southern analysis and restriction mapping of the platypus clones λPL.1.1 and λPL.1.2 82
3.1.5 Sequencing of the putative β-globin gene from λEL.1.1 90
3.1.6 Sequencing of the platypus β-globin gene from λPL.1.1 92
3.1.7 Sequencing of the platypus β-globin gene from λPL.1.2 95
3.1.8 Is *platy* β2 an authentic gene in the platypus ? 98
3.1.8.1 PCR amplification of the putative *platy* β2 gene from platypus genomic DNA 98
3.1.8.2 *Eco*RI restriction digestion of PCR products 102
3.1.8.3 Direct sequencing of the Lee F2/Lee R PCR product 104
3.1.8.4 Alignments of "TGAA" repetitive boundary sequences 104
3.1.9 PCR amplification and identification of a β-globin gene from the New Guinea echidna 108
3.1.10 Comparisons involving the four monotreme β-globin genes 113
3.1.10.1 DNA sequence comparisons 113
3.1.10.2 Intron lengths and sequences 116
3.1.10.3 5' promoter regions 118
3.1.10.4 Amino acid sequence comparisons 120
3.1.10.5 G+C content 122
3.1.11 Southern analyses of monotreme genomic DNA 122
3.2 LOCALISATION OF A β-GLOBIN GENE ON ECHIDNA CHROMOSOMES USING *IN SITU* HYBRIDISATION 126
3.2.1 G-banded karyotype of the Australian echidna 126
3.2.2 Fluorescent *in situ* hybridisation (FISH) of Australian echidna chromosomes using a β-globin probe 126
3.3 PHYLOGENETIC ANALYSIS USING β-GLOBIN GENE SEQUENCES 130
3.3.1 Maximum parsimony analysis 131
3.3.2 Distance tree 133
3.3.3 Maximum likelihood analysis 136
3.3.4 Transitions and transversions 138
3.3.5 Relative rates tests 141
3.3.6 Synonymous and non-synonymous comparisons 143
3.3.7 Estimation of the separation time of the three monotreme species 144

CHAPTER 4 DISCUSSION
4.1 Structure of monotreme β-globin genes 151

4.1.1 Features of the monotreme β-globin gene introns and flanking sequences 153
4.1.2 G+C content of the monotreme β-globin genes 154
4.2 The *platy* β2-globin gene 155
4.3 Existence of ε-like globin genes in monotremes 159
4.4 Chromosomal location of β-globin genes in monotremes 161
4.5 Rates of evolution of β-globin genes in mammals 162
4.6 Evolutionary relationships of the monotreme β-globin genes 164
4.7 Time of separation of the monotreme species 165
4.8 Evolution of the β-globin gene family in mammals 166
4.8.1 Evolutionary relationships among monotreme, marsupial and eutherian β-like globin genes 168

CONCLUDING REMARKS 177

BIBLIOGRAPHY 180

APPENDIX

Appendix 1 Sequence of primer used for sequencing and PCR 203
Appendix 2 Partial sequences of λEL.1.1 subclones 204
Appendix 3 Partial sequences of a subclone, pobg 1-3 of λPL.1.2 205
Appendix 4 Sequences of Lee F2/Lee R PCR product from platypus genomic DNA 205
Appendix 5 Multiple sequence alignment of four monotreme β-globin genes 206
Appendix 6 Intron sizes of β-like globin genes 210
Appendix 7 Dot plots analyses 211
Appendix 8 Dot plot comparisons of intron I sequences of β-like globin genes 215
Appendix 9 The A+T and G+C contents of β-like globin genes 216
Appendix 10 DNA sequences of coding regions of birds, monotreme, marsupial and eutherian β-like globin genes 217

LIST of FIGURES

		Page
Fig. 1.1.1	The organization of α- and β-globin genes in *X. laevis* and chicken	10
Fig. 1.1.2	Two evolutionary models of globin exon / intron structure	14
Fig. 1.1.3	Organization of the β-globin gene cluster of marsupial and eutherian	22
Fig. 1.1.4	An evolutionary tree of globin genes	25
Fig. 1.2.1	The classification of mammals	29
Fig. 1.2.2	The distribution of monotremes	30
Fig. 1.2.3	The geological fossil record of monotremes, marsupials and eutherians in Australia and America continents	32
Fig. 1.2.4	Hypotheses for the phylogenetic relationship of monotremes, marsupials and eutherians	34
Fig. 3.1.1	Autoradiographs showing hybridisation of the probe pTAG-2 to clone λEPL.1 in the echidna library EGL.1.2	78
Fig. 3.1.2	Restriction digestion and Southern analysis of λEPL.1	79
Fig. 3.1.3	Restriction map of λEL.1.1 and the location of plasmid subclones used for DNA sequence determination	80
Fig. 3.1.4	Restriction map of λEPL.1 showing the region of overlap between λEPL.1 and λEL.1.1	81
Fig. 3.1.5	Partial digestion with *Sau*3A, and size separation on 0.8% (w/v) agarose gel of platypus genomic DNA	83
Fig. 3.1.6	Autoradiographs of plaque lift filters from platypus genomic DNA library, PGL.1.1	84
Fig. 3.1.7	Restriction digestion and Southern analysis of λPL.1.1	86
Fig. 3.1.8	Restriction map of λPL.1.1 and location of the plasmid subclones used for sequence analysis	87
Fig. 3.1.9	Restriction digestion and Southern analysis of λPL.1.2	88
Fig. 3.1.10	Restriction map of the λPL.1.2 insert and the position of fragments that were subcloned and used for sequencing of the *platy* β2-globin gene	89
Fig. 3.1.11	Complete nucleotide sequence and the corresponding amino acid sequence of the β-globin gene (*echid* β1) from the Australian echidna	91
Fig. 3.1.12	Complete nucleotide sequence and the corresponding amino acid sequence of the *platy* β1-globin gene	94
Fig. 3.1.13	Complete nucleotide sequence and the corresponding amino acid sequence of the *platy* β2-globin gene	96
Fig. 3.1.14	DNA sequence comparisions of *echid* β1-, *platy* β2- and *platy* β1-globin gene sequences	99
Fig. 3.1.15	PCR amplification of *platy* β1 and *platy* β2 genes from monotreme genomic DNA with Lee F2/Lee R and B F1/Lee R primers	101
Fig. 3.1.16	*Eco*RI digestions of PCR products amplified from lambda clones and monotreme genomic DNAs using primer sets, Lee F2/Lee R and B F1/Lee R	103

Fig. 3.1.17	Alignments of the partial DNA sequences obtained from the Lee F2/Lee R PCR products and *platy* β1 and *platy* β2 sequences	106
Fig. 3.1.18	Comparison of the restriction maps of λEL.1.1, λEPL.1 and λPL.1.2, and alignment of "TGAA" repeats and their boundary sequences in the *platy* β2 and partial sequences from a plasmid subclone pTAG-5 of λEL.1.1	107
Fig. 3.1.19	PCR amplification of a putative β-globin gene from the New Guinea echidna	109
Fig. 3.1.20	Complete nucleotide sequence and corresponding amino acid sequence of the New Guinea echidna β1-globin gene	111
Fig. 3.1.21	Alignment of the 146 amino acid residues of monotreme β-globin putative gene products	112
Fig. 3.1.22	Mutiple sequences alignment of four monotreme β-globin genes using the CLUSTALW computer program	115
Fig. 3.1.23	Length distributions of β-like globin gene introns	117
Fig. 3.1.24	Comparison of 5' promoter elements in monotreme β-globin genes and various marsupial and eutherian β-like globin genes	119
Fig. 3.1.25	A comparison of A+T and G+C contents of β-like globin genes from mammals and birds	123
Fig. 3.1.26	Southern analysis on monotreme genomic DNA using β-globin gene probes	125
Fig. 3.2.1	G-banded chromosomes of a female Australian echidna	128
Fig. 3.2.2	*In situ* hybridised cells from an Australian echidna	129
Fig. 3.3.1	Maximum parsimony tree derived from 30 β-like globin genes using the heuristic algorithm	132
Fig. 3.3.2	Cladogram representing the Kimura-2-parameter\Neighbor-Joining tree of the 30 β-like globin genes	134
Fig. 3.3.3	Neighbor-Joining tree constructed using PAUP (*)	135
Fig. 3.3.4	Maximum likelihood tree constructed from the 30 β-like globin nucleotide sequences with HKY 85 model	137
Fig. 3.3.5	Maximum likelihood tree constructed from the 30 β-globin nucleotide sequences with HKY 85 and gamma distribution models	139
Fig. 3.3.6	A plots of the propotion of sites with transitional (A) and transversional (B) nucleotide substitutions against the Jukes-Cantor genetic distances	140
Fig. 4.1	An hypothetical phylogenetic tree of monotreme β-globin genes that could account for the existence of *platy* β2 in the platypus	157
Fig. 4.2	Phylogenetic trees illustrating four possible evolutionary models	173

LIST of TABLES

Table 1.2.1	The estimated separation dates between different mammalian groups	36
Table 2.2.1	PCR primer sequences and conditions for PCR amplification of monotreme β-globin genes	70
Table 3.1.1	The predicted sizes of PCR products with Lee F2/Lee R and B F1/Lee R primers and *Eco*RI digestion of the PCR products	102
Table 3.1.2	Sequence identity (%) in the three exons and two introns from the four monotreme β-globin genes	113
Table 3.1.3	Sequence identity (%) in 5' and 3' flanking regions of three monotreme β-globin genes	113
Table 3.1.4	Intron lengths of β-globin genes in monotremes	116
Table 3.1.5	β-like globin polypeptide sequence comparisons	121
Table 3.3.1	Relative rates tests of bird, marsupial, eutherian and monotreme β-like globin genes	145
Table 3.3.2	Synonymous and non-synonymous divergence values for pairs of coding regions from β-like globin genes	150
Table 4.1	The estimated separation dates among monotreme species	166

SUMMARY

The β-globin genes have been the focus of many studies concerned with molecular evolution and have been used as model systems for investigating the role of gene duplication in the evolution of coordinately regulated gene families. As a result of these studies, models depicting the evolution of the β-globin genes have been developed. These models attempt to account for the complexity (and indirectly, the function) of the β-globin genes, in terms of gene duplications and sequence divergence events.

Despite the large amount of data on the structure, function and evolution of the β-globins, no molecular data at the nucleotidetide level on the β-globin genes have previously been obtained from a monotreme. The Monotremata, a group of egg laying mammals, have attracted particular interest from evolutionists because of their reptilian / bird-like as well as mammalian-like characteristics. The monotreme lineage is thought to have separated over 150 MYA from the lineage that subsequently gave rise to the marsupials and eutherians. Recent analyses based on mitochondrial DNA sequences from the platypus have suggested that the monotremes may form a monophyletic group with the marsupials to the exclusion of the eutherians (Janke *et al.* 1996, 1997). While the true phylogenetic relationships of the monotremes remains unresolved, it is clear that monotremes represent a very ancient mammalian branch, and that molecular studies on their β-globin genes are likely to be especially informative in the context of β-globin gene family evolution.

This thesis contains the results of molecular and evolutionary studies of β-globin genes in three monotreme species: the Short-beaked Echidna, *Tachyglossus aculeatus* (sometimes called the Australian Echidna), the Long-beaked Echidna, *Zaglossus bruijni* (New Guinea Echidna) and the Platypus, *Ornithorhynchus anatinus*. The basic aims of this research project are to investigate i) the evolution of β-globin gene

families in mammals, and ii) the phylogenetic relationships among the monotreme species, and between the monotremes, marsupials and eutherians.

In this project, a β-like globin gene was isolated and characterised from each of the three monotreme species. *Echid* β1 was isolated from an Australian echidna genomic library, using a marsupial ε-globin gene probe. *Platy* β1 was isolated from a platypus genomic library using *echid* β1 as a probe, and *NGechid* β1 was successively amplified using PCR from New Guinea echidna genomic DNA using primers designed on the basis of the *echid* β1 sequence.

In their overall structure the monotreme β-globin genes are typical of β-globins from other mammals. They comprise three exons and two introns, the latter being located in the positions characteristic of introns in all vertebrate β-globin genes. The amino acid sequence encoded by *echid* β1 proved to be identical to the published amino acid sequence of the β-globin chain isolated from an adult Australian echidna (Whittaker *et al.* 1972). Likewise, the amino acid sequence encoded by *platy* β1 was identical to the published sequence for adult platypus β-globin (Whittaker and Thompson, 1975). Comparisons of the 5' promoter regions show that in their sequence organisation, the monotreme β-globin genes more closely resemble the mammalian adult β-globin genes, than the early expressed ε-, γ- and η-globin genes.

Compared to other mammalian β-like globin genes, the monotreme β-globins have some interesting features. They have the longest first intron (146-153 bp) and the shortest second intron (440-465 bp) of any described functional adult mammalian β-globin gene. The (G + C) content of the coding regions of the monotreme β-globins is high compared to other mammalian β-globins, but is similar to that of the chicken.

Southern analysis (using various β-globin gene probes) of DNA from the Australian echidna and platypus, revealed three hybridising *Eco*RI bands suggesting the existence of more than one β-like globin gene in the genomes of these species.

Echid β1 was localised by fluorescent *in situ* hybridization to one of the small chromosomes (Group 14-22) in the Australian echidna. Previously, a β-like globin gene had been localised to chromosome 4 in the platypus. This difference between the echidna and platypus in the chromosomal location of the β-globin locus provides evidence that the chromosomes concerned (which are very different in size and general morphology) may be derived from a common ancestral chromosome and that divergence in chromosomal morphology occurred either as part of, or as a consequence of the speciation event that separated the platypus and echidna lineages.

Amino acid sequence comparisons between β-globins from various mammalian species showed that the monotreme β-globins share about the same degree of sequence identity with the eutherian adult β-globins as they do with the eutherian and marsupial embryonic-like β-globins. However, the monotreme β-globins have a lower degree of sequence identity with the marsupial adult-like β-globins.

The β-globins of the Australian echidna and New Guinea echidna differ by 6 amino acid residues. *Echid* β1 and *NGechid* β1 show a far greater degree of nucleotide sequence identity to one another than either does to *platy* β1. If it is assumed that the rates of amino acid and nucleotide substitution in these genes are approximately constant over time (tests revealed no significant differences in the rates of nucleotide substitution within the monotremes), the estimated divergence time of the echidna and platypus lineages is 41-57 MYA, which falls within the generally accepted time range of 30-60 MYA. However, the divergence time of the Australian echidna and New Guinea echidna is estimated to be 11-20 MYA, which is far older than the estimates (1-2 MYA) obtained from DNA-DNA hybridisation and mitochondrial 12S rDNA sequence comparisons.

Sequence comparisons indicate that at non-synonymous and synonymous sites, monotreme β-globins show no closer resemblance to eutherian β-globins than to eutherian ε-globins. Furthermore, at non-synonymous sites, monotreme β-globins

show no closer resemblance to marsupial β-like globins than to eutherian β-like globins.

The evolutionary relationships of the β-like globin genes were studied by applying maximum parsimony methods to aligned DNA sequences. The analyses, which included β-like globin sequences from representative species of birds, marsupials and eutherians, were consistent with a hypothesis that the gene duplication that gave rise to the proto ε-globin and proto β-globin occurred after the monotreme / therian separation, but before the marsupial / eutherian separation. The analysis also provided some support for an evolutionary model that places the monotremes outside a monophyletic group comprising marsupials and eutherians.

DECLARATION

I declare that this work does not contain any material published elsewhere, except where due reference is given in the text. Nor has it been presented previously as a component of any other degree or diploma.

December, 1997

NAME: Mi-Hye Lee COURSE: PhD

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Libraries, being available for photocopying and loan.

DATE: 5/12/97

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Dr. Rory Hope for his encouragement and advice throughout this project, and his valuable assistance. To Dr. Steven Cooper, thank you for your help, encouragement and advice, and for your time. To Rob Shroff, thank you for your support and guidance and for allowing me to use the clones you derived and some of your data. To Guy Webber, my thanks for your support in the early stages of this project.

I would like to thank the following people who helped me with different aspects of the project: Prof. Jenny Graves for giving me platypus DNA samples and slides containing platypus chromosome spreads, Dr. Steve Donnellan for donating New Guinea echidna tissue samples, Dr. D. Schultz for providing an Australian echidna blood sample for chromosome studies, Dr. P. Rismiller for the provision of echidna tissue, Dr. N. Gemmell for giving me platypus liver samples, Iole Barbier for sending me platypus tissue samples and chemicals for chromosome studies, Chris Leigh for organising the collection of echidna blood samples, Dr. Eugene Diatloff for proof-reading, Dr. A. Houben for proof-reading some sections of my thesis and for advice, encouragement and assistance with the *in situ* work. Also, I would like to thank all the lab. 108 members who have been working with me over the last few years, for their friendship, and Shiro Akiyama for discussions on microsatellites in monotremes.

I would like to thank the members of the Genetics Department including Prof. Robert Saint for all their support and interest throughout my PhD years, and for enabling me to carry out my PhD project.

Thank you all my friends in KLC for your friendship throughout my PhD years. I would especially like to thank one of my best friends in KLC, Dr. Burkard Poster for help with proof-reading and statistical tests, and for his friendship.

Finally, I would like to thank my mum and family in Korea for their support and for believing in me.

TERMINOLOGY

Adult-like β-globins	β globin genes that are expressed at post natal stages and are orthologous to human δ- and β-globin genes.
β-like globin gene	Any globin gene that is descended from the hypothesised proto β-like globin gene (see Fig. 1.1.4 page 25).
Embryonic-like β-globins	β globin genes that are expressed in embryonic and foetal stages and are orthologous to human ε-, γ- and η-globin genes.
Gene tree	A phylogenetic tree that has been constructed from homologous genes from each species.
Homology	Similarity by common ancestry or genetic relatedness.
Homoplasy	A collection of phenomena that leads to similarities in character states for reasons other than inheritance from a common ancestor. These include convergence, parallelism and reversal.
Concerted evolution	Maintenance of homogeneity of nucleotide sequences among members of a gene family in a species, although the nucleotide sequences change over time.
Isochore	A genomic DNA segment that is homogeneous in base composition.
Non-synonymous	A nucleotide substitution that alters a codon to that of another amino acid.
Monophyletic	A group that contains all the descendants of the most recent common ancestor of the constituent species.
Orthologous	Descendants of gene-lineages whose divergences from one another coincide with the divergence of the species-lineages within which they occur.
Paralogous	Descendants of gene-lineages whose divergences occurred by gene duplication prior to divergence of the species-lineages.
Plesiomorphy	An ancestral character state.

Proto ε-globin gene (see proto β-globin gene).

Proto β-globin gene One of the two putative globin genes formed by gene duplication of a proto β-like globin gene. This gene became developmentally delayed in its expression. The other duplicate gene so formed was the proto ε-globin which became expressed in embryonic and foetal stages (see Fig. 1.1.4 page 25).

Proto β-like globin gene One of the two putative globin genes that was formed by gene duplication. The other gene so formed (proto α-like globin) gave rise to the α-globins (see Fig. 1.1.4 page 25).

Species tree A phylogenetic tree that represents the evolutionary relationships of a group of species.

Synapomorphy The possession by two or more related lineages of the same character state derived from a different but homologous character state in the ancestral lineages.

Synonymous A nucleotide substitution resulting in a codon specifying the same amino acid as before.

Nomenclature of Monotreme β-globin genes

***Echid* β1** An adult-like β-globin gene that was isolated from a Australian echidna genomic library.

***NGechid* β1** An adult-like β-globin gene that was amplified from New Guinea echidna genomic DNA.

***Platy* β1** An adult-like β-globin gene that was isolated from a platypus genomic library.

***Platy* β2** A putative β-like globin gene that was isolated from a platypus genomic library, and that is very similar in its DNA sequence to *echid* β1.

ABBREVIATIONS

A	Adenine
ANGIS	Australian National Genomic Information Service
α	*alpha*
α-^{32}P-dATP	α-labelled deoxyadenosine 5'-triphosphate
BCIP	5-Bromo-4-chloro-3-indolyphosphate
Biotin-11-dUTP	biotin-labelled deoxyadenosine 5'-triphosphate
bp	base pair (s)
BSA	bovine serum albumin
C	cytosine
°C	degrees Celsius
DAPI	4,6 -diamidino-2-phenylinedole
dATP	2'-deoxyadenosine-5' triphosphate
dCTP	2'-deoxycytosine-5' triphosphate
dGTP	2'-deoxyguanosine-5' triphosphate
dTTP	2'-deoxy thymidine-5' triphosphate
DME	Dulbecco's modification of Eagle's medium
DNase	deoxyribonuclease
DTT	dithiothreitol
dTTP	2'-deoxythymidine-5' triphosphate
dUTP	2'-deoxyuridine-5' triphosphate
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetra-acetic acid
EtBr	ethidium bromide
ε	epsilon
FITC	fluoro iso thio cyanate
F	forward sequencing primer
G	guanine
g	force of gravity
g, mg, μg, ng, pg	gram, milligram, microgram, nanogram, picogram
HB	haemoglobin
IPTG	isopropyl β-d-thiogalactopyranoside
kb	kilo bases
λ	lambda
LB	luria broth
MEGA	molecular evolution genetical analysis
min	minute (s)
m, cm, μm	metre, centimetre, micrometre
M, mM, μM	moles per litre, millimoles per litre, micromoles per litre
mA	milliampere
ml, μl	millilitre, microlitre
MW	molecular weight
MYA	million years ago
NBT	Nitro-blue tetrazolium salt

dNTP	deoxy nucleoside triphosphate
NOR	nucleolar organising region
nt	nucleotide (s)
ODn	optical density at wavelength of n-nanometres
PAUP	phylogenetic analysis using parsimony
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG 8000	polyethylene glycol, molecular weight 8000
pH	hydrogen ion concentration
pfu	plaque forming units
®	registered
γ	gamma
rDNA	ribosomal DNA
RNA	ribonucleic acid
RNase A	ribonuclease A
rRNA	ribosomal RNA
R	reverse sequencing primer
rpm	revolutions per minute
RT	room temperature
SDS	sodium dodecyl sulphate
SSC	saline sodium citrate
SSPE	saline sodium phosphate EDTA
T	thymidine
TBE	tris-borate EDTA
Taq	*Thermus aquaticus*
TEMED	N N N N-tetramethylethylene-diamine
$\mathbf{T_M}$	melting temperature
TM	Trademark
Tris	tri-hydroxymethyl amino methane
μCi	micro Curies
U	unit of enzyme activity
UV	ultraviolet light
V	volts
v/v	volume per volume
W	watts
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-B-D-galactoside

INTRODUCTION

Globin proteins, whose principal function is oxygen binding and transportation, are found in a wide variety of organisms including bacteria, fungi, plants, invertebrates and vertebrates. Vertebrate haemoglobin molecules are composed of four polypeptides, two α-like and two β-like chains. The different globin chains are encoded by related families of genes in which the expression of each gene is developmentally regulated (Slightom *et al.* 1980; Goodman *et al.* 1984; Koop *et al.* 1986). The abundance of globin protein in red blood cells, and the involvement of globin gene mutations in several widespread and debilitating human diseases (e.g. sickle cell disease and thalassemia) have made the globin genes an important and popular model system for investigating gene structure, regulation, function and evolution (Bashford *et al.* 1987; Gumucio *et al.* 1996). Information on these various aspects of the globins has been accumulated in a wide range of species, making the globin genes one of the most extensively studied and well characterised group of genes.

The β-globin genes have been the focus of many studies concerned with molecular evolution and a number of models depicting the evolutionary history of these genes have been developed. These models attempt to account for the numerical and DNA sequence complexity of the β-globin genes, in terms of events such as gene duplications followed by sequence divergence.

Phylogenetic analyses based on DNA and protein sequence comparisons provide evidence that the α- and β-like globin genes arose from a common ancestral gene that duplicated during the evolution of higher vertebrates approximately 450 MYA (Czeluzniak *et al.* 1982). The pair of globin genes so formed underwent a further series of gene duplications to give rise to the various globin genes found in extant birds, reptiles and mammals.

The "original" β-globin gene (proto β-like globin) duplicated during the evolution of early mammals (probably after the separation of the avian and mammalian lineages, but before the separation of marsupials and eutherians) about 150-200 MYA, and gave rise to two β-like globin genes, one of which became developmentally delayed in its expression (proto β-globin) and the other became expressed in embryonic tissues only (proto ε-globin) (Goodman *et al.* 1987a; Koop and Goodman, 1988; Hope *et al.* 1990; Cooper and Hope, 1993; Cooper *et al.* 1996). The proto ε-globin was the progenitor of the so called "early expressed" β-globins, i.e. those genes whose timing of expression is now limited to embryonic and foetal stages, presently exemplified by ε-, γ-, and η-globin in eutherians, and ε-globin in marsupials. The proto β-globin was the progenitor of the "late expressed" or so called "adult" β-globins, presently exemplified by β- and δ-globin in eutherians and β-globin in marsupials. The two genes, ε- and β-globin, that make up the β-globin gene clusters of marsupials, or at least of *Sminthopsis crassicaudata* and *Didelphis virginiana*, are thought to be the direct descendants of proto ε-globin and proto β-globin, respectively.

Despite the large amount of data on the structure, function and evolution of the β-globins, no molecular data at the nucleotide level on the β-globin genes have been obtained previously from a monotreme. The rationale underlying this research project is that studies on monotreme β-globin genes are potentially informative in an evolutionary context. Such studies can fill gaps in our knowledge of the timing of the various gene duplication events, and can also provide data on the evolution of the monotremes themselves. For example, an unanswered question in the current models for β-globin evolution is whether the gene duplication that gave rise to the proto ε- and proto β-globins occurred before or after the separation of the stem of monotremes from the stem of marsupials / eutherians. Potentially, studies of monotreme β-globins can provide key information on the timing of this duplication event. Parsimony methods permit likely ancestral states to be determined, and data on monotreme globins will lead to a greater understanding of the structure and organisation of the globin genes in the stem mammals.

The β-globin gene sequences, and the structure and organisation of the β-globin gene families, have been determined at the molecular level in many eutherian species, but in only three species of marsupials (Wainwright and Hope, 1985; Goodman *et al.* 1987a; Koop and Goodman, 1988; Hope *et al.* 1990; Cooper and Hope, 1993; Cooper *et al.* 1996). The abundance of β-globin nucleotide and protein sequence data from a variety of mammal species means that as new data become available, e.g. from monotremes, comparative studies involving the relatively well characterised marsupials and eutherians will be immediately possible.

Studies on monotreme globins take on an added interest and significance because the structure and function of these genes may well have evolved in concert with unique features of monotreme reproduction and lifestyle that impinge on respiration. These features include the delivery of oxygen to the embryo after birth by diffusion through the egg membrane; the requirement, at a very early stage of development, for lung function in a pouch where non-ambient levels of oxygen and carbon dioxide are likely to occur; and the oxygen stress placed on tissues during hibernation, burrowing and diving (Parsons *et al.* 1971a,b; Bentley *et al.* 1967; Whittington and Grant, 1983; Grigg *et al.* 1992; Bech *et al.* 1992). Comparative studies that include marsupial and monotreme haemoglobins are likely, therefore, to provide useful information on the relationships between haemoglobin structure, function and evolution.

The monotremes, a group of egg laying mammals, have attracted particular interest from evolutionists because of their reptilian / bird-like as well as mammalian-like characteristics. The lineage that gave rise to the extant monotremes is thought to have diverged from the marsupial and eutherian lineage some time in the Mesozoic, about 150-200 million years ago (MYA), well before the marsupials and eutherians diverged from one another in the early Cretaceous, about 135 MYA (Clemens, 1989). However, the timing of these evolutionary events remains problematical, and has recently been questioned on the basis of the analysis of mtDNA sequences from the platypus (Janke *et al.* 1997) which suggest that the monotremes may form a monophyletic group with the marsupials to the exclusion of the eutherians (Janke *et al.* 1996, 1997). While the true phylogenetic relationships of the monotremes remains

unresolved (the evolution and radiation of the monotremes have been difficult to trace, partly because of an inadequate fossil record) it is clear that monotremes represent a very ancient mammalian branch. It is the age of this branch that makes comparative studies of genome organisation and function so interesting.

The amino acid sequences of α- and β-globin and myoglobin of the Australian echidna and platypus were determined during the 1970s (Whittaker *et al.* 1972, 1973, 1978; Dodgson *et al.* 1974; Whittaker and Thompson, 1974, 1975; Fisher and Thompson, 1976; Castillo *et al.* 1978). Amino acid sequence data from α- and β- haemoglobin, and myoglobin (Goodman *et al.* 1985), and DNA sequence analysis of the protamine p1 (Retief *et al.* 1993), mitochondrial cytochrome-b and 12S ribosomal RNA genes (Gemmell and Westerman, 1994) and mitochondrial protein-coding genes (Janke *et al.* 1996, 1997) have provided some information on the interrelationships of the three extant monotreme species, and of monotremes, marsupials and eutherians (Westerman and Edwards, 1991, 1992; Retief *et al.* 1993; Gemmell and Westerman, 1994). Unfortunately, these approaches have not enabled firm conclusions to be drawn. Therefore, one of the aims of this project was to carry out a comparative study based on β-globin amino acid and nucleotide sequences, to help clarify these evolutionary relationships.

Comparative studies of chromosomal banding patterns and chromosomal mapping of homologous genes have provided information about the evolution of mammalian chromosomes (O'Brien and Graves, 1990). Wrigley and Graves (1984) have suggested that a comparative study of gene arrangements in eutherians, marsupials and, perhaps most importantly, monotremes, would provide an insight into the evolution of mammalian genome organisation and may even lead to an identification of the ancestral karyotype and gene map of mammals (Wrigley and Graves, 1984; Watson, 1990; Maccarone *et al.* 1992; Watson and Graves, 1992). Therefore, mapping the β-globin genes to monotreme chromosomes, may provide additional information on mammalian genome evolution, and may help describe the chromosome changes and their possible role in mammalian evolution.

AIMS AND OBJECTIVES

The basic aims of this research project are to investigate the evolution of β-globin gene families in mammals and to elucidate the phylogenetic relationships of the three monotreme species and of monotremes and other mammals.

The specific objectives are to :

- isolate and characterise the β-globin genes from 3 species of Monotreme, *Tachyglossus aculeatus* (Australian echidna), *Zaglossus bruijni* (New Guinea echidna) and *Ornithorhynchus anatinus* (platypus);
- carry out a phylogenetic and statistical analysis of β-globin protein and nucleotide sequences;
- estimate the divergence times between the three monotreme species by the pair-wise comparison of β-globin nucleotide substitutions, assuming the existence of a molecular clock;
- determine the chromosomal location of these genes in the Australian echidna using *in situ* hybridisation.

CHAPTER 1 LITERATURE REVIEW

LIBRARY
UNIVERSITY OF ADELAIDE

1.1 HAEMOGLOBIN GENES

Multigene families are a common feature of mammalian genomes. Such families often contain clustered genes which are structurally and evolutionarily related to each other and are expressed at different times during development. It is likely that some regulatory elements reside within or near each gene cluster that control the temporal and tissue-specific expression of its members (Richard *et al.* 1987). A considerable amount of information has been accumulated on the organization, expression and evolution of the globin genes (Brown *et al.* 1984; Loomis and Gilpin, 1986). Globin genes are found in a wide variety of organisms, ranging from bacteria (Wakabayashi *et al.* 1986), fungi (Iwaasa *et al.* 1991), higher plants (Hyldig-Nielsen *et al.* 1982; Jensen *et al.* 1981; Lee *et al.* 1983), many invertebrate phyla (Antoine and Niessing, 1984; Dixon *et al.* 1991; Kloek *et al.* 1993; Blaxter, 1993; Nassif *et al.* 1994) and vertebrate (Maniatis *et al.* 1980). Amino acid sequences are now known for several hundred globin chains and the DNA sequences of over 250 different globin genes have been published (Bashford *et al.* 1987).

The principal function of vertebrate globins is the transport of oxygen to tissues and cells for respiration. A second function of globins (myoglobin) is oxygen storage in muscles. However, in bacteria and nematodes, the function of globin is still subject to debate. Evidence suggests that in these species globins may have a role as oxygen sinks or in enzymatic reactions rather than as oxygen transport mediators (Blaxter, 1993).

Comparisons of globin gene clusters from different species reveal that there is considerable variation in the size and number of genes, but the three-dimensional structure of globin proteins and the structure of globin genes are superficially similar (Bashford *et al.* 1987). There has been a rapid growth of amino acid and nucleotide sequence data on haemoglobin genes and pseudogenes, with cladistic analysis continuing to reveal new facts about globin evolution (Czelusniak *et al.* 1982; Richard

et al. 1987; Bashfore *et al.* 1987; Blaxter, 1993). Comprehensive models explaining the evolutionary history of the globin genes in vertebrate mammals, invertebrates and plants and how the various genes are related to each other have emerged (Shapiro *et al.* 1983; Hardies *et al.* 1984; Hardison, 1984; Hill *et al.* 1984).

Several mechanisms to explain the evolution of the globin genes have been proposed:

1) Intron loss. Bacterial globins and some invertebrate globins lack introns altogether (Craik *et al.* 1981; Antoine and Niessing, 1984; Trewitt *et al.* 1995) and the globin of a nematode, *Caenorhabditis elegans,* contains a single intron (Kloek *et al.* 1993). The absence of introns in the globin genes of *Chironomus thummi thummi* was explained by suggesting that the gene arose by reverse transcription from spliced mRNA in germ-line cells to produce an intron-less cDNA that subsequently became integrated back into the genome (Antoine and Niessing, 1984).
2) Tandem duplication of one or more genes. This is suggested to be a major mechanism to increase the number of β-globin genes. For examples, the β-haemoglobin gene clusters of ruminants (goats, cattle, sheep) contain a large number of genes compared to other eutherian mammals; cattle contain 9 β-globin genes (Schimenti and Duncans, 1985a,b) and goats 12 (Townes *et al.* 1984). These multiple genes are related through tandem duplications of an ancestral four gene set.
3) Gene conversion by unequal crossing over. A large body of evidence demonstrates that gene conversion has played a major role in the evolution of globin gene families (Slightom *et al.* 1980; Roninson and Ingram, 1982; Alan *et al.* 1984). Sequence comparisons between various globin genes show that some regions have high sequence similarity which could be explained most readily by gene conversion events, the nonreciprocal transfer of DNA from one gene to another. Gene conversion events are proposed to explain the high degree of similarity in DNA sequences between the human $^{A}\gamma$ and $^{G}\gamma$ -globin genes (Efstratiadis *et al.* 1980). A fusion δ/β haemoglobin, "Lepore" also resulted from the unequal crossing-over between δ- and β-globin genes (Jeffreys, 1982). It has been suggested that gene conversion occurs frequently in higher eukaryote globin genes (Slightom *et al.*

1980; Shen *et al.* 1981) and that it serves as a mechanism for the co-evolution of genes in multigene families (Hood *et al.* 1975).

1.1.1 The globins of plants, invertebrates and vertebrates

Globins are found universally, with few variations in quaternary structure, throughout vertebrates, but are only sporadically found in the invertebrates (Goodman *et al.* 1988). The molecular structure of myoglobin, leghaemoglobin and animal haemoglobin demonstrate that very different amino acid sequences may produce similar three-dimensional structures (Lee *et al.* 1983). Comparisons of globin nucleotide sequences showed significant sequence similarity in the regions surrounding the gene including CAP, ATA, and CCAAT boxes in the promoter region (Lee *et al.* 1983; Gerstein *et al.* 1994).

Invertebrate intracellular haemoglobins are generally monomeric, dimeric or tetrameric, although highly variable forms can also be found (Terwilliger, 1980; Proundfoot *et al.* 1982; Goodman *et al.* 1988). Plant leghaemoglobin and myoglobin are monomeric (Lee *et al.* 1983). Vertebrate haemoglobin is heterotetrameric, consisting of two types of chains, one encoded by a gene in the α-family and the other by a β-family gene.

In addition to haemoglobins, myoglobin-like-molecules are also found in nitrogen-fixing cyanobacterum (algae) and lower prokaryotes, for example, the ciliated protozoa (Potts *et al.* 1992). Vertebrate myoglobin is the principal muscle globin protein which facilitates the diffusion of oxygen from the vascular system to muscle mitochondria (Weller *et al.* 1984). In mammals, myoglobin accumulates in cardiac muscle early in foetal development, whereas the appearance of myoglobin in skeletal muscle is delayed until birth, correlating with the relative activities of these two muscle types during development (Tipler *et al.* 1978). Elevated levels of myoglobin in skeletal and/or cardiac muscle are found in diving eutherian mammals (e.g. whales and seals) and birds. It is assumed that this high level of myoglobin is an adaptation to

hypoxic subterranean or high altitude environments (Wittenberg, 1970; Butler and Jones, 1982; Blachetot *et al.* 1983; Weller *et al.* 1984).

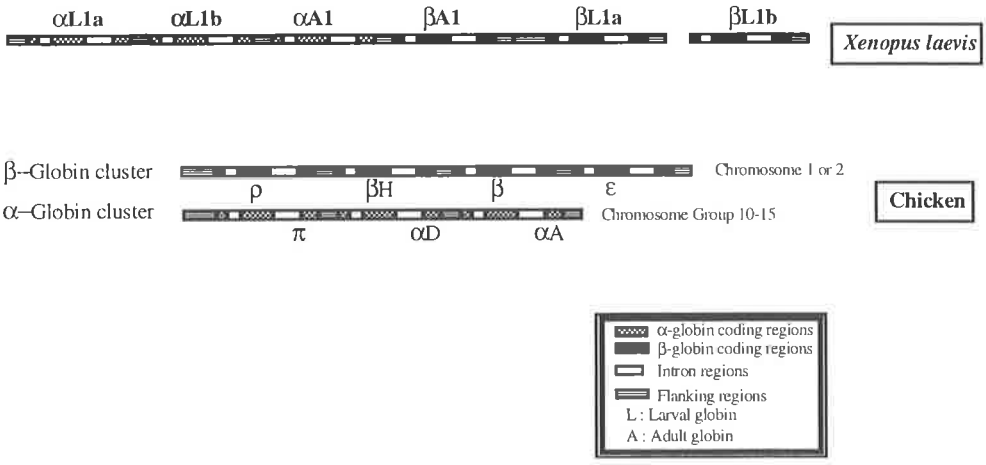
Leghaemoglobins are the most abundant gene products in nitrogen fixing legume root nodules, formed as a result of a symbiotic association of plants with the soil bacteria, *Rhizobium*. Leghaemoglobins are localised in the cytoplasm of infected cells and function in providing an adequate supply of oxygen for bacterial respiration whilst protecting the oxygen sensitive nitrogenase enzyme. Leghaemoglobins have also been found in the non-legumes, *Trema tomentosa* and *Parasponia andersonii*, which indicate that globin genes may be widespread in the plant kingdom (Bogusz *et al.* 1984, 1988; Peacock, 1988), and imply that globin genes existed in a common ancester of plants and animals.

Analysis of cDNA and genomic clones from the amphibian *Xenopus laevis* indicated that this species contains 12 globin genes arranged in two almost identical clusters, each of six genes (Jeffreys *et al.* 1980; Hosbach *et al.* 1983) (Fig. 1.1.1). Each cluster contains two adult expressed genes (αA and βA) and four genes expressed in larval tissues (αL1a, αL1b, βL1a and βL1b). Each cluster is assumed to be the result of chromosome duplication and within each cluster, the α- and β-globin genes are closely linked on the same chromosome (Jeffreys *et al.* 1980). In contrast to amphibians, the α- and β-globin gene families in birds and mammals are located on separate chromosomes (Hughes *et al.* 1979; Maniatis *et al.* 1980) (Fig. 1.1.1). These observations have been interpreted as indicating that: i) the α and β globin genes arose by a tandem gene duplication of a single primordial globin gene present in the common ancestor of amphibians, birds, and mammals; ii) the ancestral linked arrangement of α and β globins was retained in *X. laevis*; iii) the α and β globin genes were separated through a translocation event in a common ancestor of birds and mammals (Jeffreys *et al.* 1980; Hosbach *et al.* 1983; Knochel *et al.* 1986). In humans, the myoglobin gene is located on chromosome 22q11.2, the α-globin on chromosome 16p13 and the β-globin cluster on chromosome 11p15.5 (Hardison, 1984; Klinger, 1987).

Monomeric myoglobins and tetrameric haemoglobins of vertebrates form part of the same globin superfamily and have arisen by successive duplications of a primordial globin gene. The duplication that resulted in myoglobin and haemoglobin genes is estimated to have occurred about 600-800 MYA and each of the resultant proteins evolved specialized roles: myoglobin became the oxygen-storage protein in muscles, and haemoglobin became the oxygen carrier in blood (Brown *et al.* 1984; Antoine and Niessing, 1984; Li and Graur, 1991) (Fig. 1.1.4). Prior to this gene duplication, the ancestral globin gene probably encoded a monomeric protein, similar to the structure of myoglobin but after the gene duplications, the genes evolved to encode the α- and β- polypeptide chain that functioned in a tetrameric conformation (Jeffreys, 1982; Czelusniak *et al.* 1990; Hardison, 1991).

The duplication that formed the α and β vertebrate haemoglobin genes occurred about 450-500 MYA. Both the α-like and β-like genes duplicated several more times in their subsequent evolution, leading to families of genes expressed at different developmental stages (Goodman and Moore, 1975; Czelusniak *et al.* 1982; Proundfoot *et al.* 1982; Goodman *et al.* 1987a).

Figure 1.1.1 The organization of α- and β-globin genes in *X. laevis* and chicken. Only one of the 2 gene clusters in *X. laevis* is shown.



1.1.2 Haemoglobin function

Haemoglobin performs its function effectively by binding oxygen cooperatively, and with an overall affinity appropriate to the physiological conditions. In vertebrates, the α and β subunits of haemoglobin are connected by salt bridges and each carries one haem group. Thus a haemoglobin molecule carries four haem groups and can bind four molecules of oxygen. In contrast to haemoglobin, myoglobin is a single polypeptide, but its primary structure is closely related to that of haemoglobin. Each myoglobin molecule binds one molecule of oxygen, but binding is much stronger than that of haemoglobin (Gerstein *et al.* 1994).

Regulation of the oxygen affinity of haemoglobin is achieved by the preferential binding of hydrogen atoms, carbon dioxide, chloride ions and 2,3-diphosphoglycerate (DPG). These factors bind to the haemoglobin active site, or in the case of DPG, bind to the central cavity where the four subunits are in contact with each other, and induce a conformational change in the haemoglobin molecule. Therefore, any increase in the concentration of these factors effectively lowers the affinity of haemoglobin for oxygen (Shulman *et al.* 1975; Perutz, 1978).

Based on the three-dimensional structure of human haemoglobin, Eaton (1980) classified the amino acid residues into five functional categories: i) residues involved with haem binding (these are of critical importance to the function of haemoglobin as an oxygen transporter protein), ii) α1-β2 contact residues[1] (these residues are involved in cooperative alterations of the quaternary structure of haemoglobin), iii) α1β1 contact residues, iv) residues of the β chain that bind DPG (DPG binds preferentially to deoxyhaemoglobin and also has the effect of reducing the oxygen affinity of haemoglobin), and v) residues involved in the Bohr effect (an effect whereby H^+ and CO reduce the oxygen affinity of haemoglobin and thereby enhance the release of oxygen in metabolically active tissue). Differences in the tertiary conformation

[1] Contact residues are those that are sufficiently close to one another to be of special significance in maintaining the correct three dimensional conformation of the molecule.

associated with the quaternary structure of haemoglobin produce a difference in affinity of the haem groups for oxygen. Selective maintenance of the functional molecular configuration would be expected to be conserved during evolution, particularly at the five types of contact residues referred to above (Air and Thompson, 1969; Perutz *et al.* 1968; Eaton, 1980).

Air and Thompson (1969) studied the structure-function relationship of haemoglobins in marsupials and monotremes. Stenzel *et al.* (1979) also emphasised the importance of the structure-function relationship in their analysis of opossum α- and β-globin chains. Hope *et al.* (1990) summarised the data on conservation of residues at functionally important sites from marsupial and monotreme globins. They found i) no evidence of significant conservation at α1-β1 contact sites compared with all other sites, ii) evidence for significant conservation at globin-haem, and α1-β2 contact sites.

Eaton (1980) suggested a possible pathway for the evolution of haemoglobins in respect of their function. The existence of monomeric haemoglobin in lower vertebrates and invertebrates led him to assume that primitive haemoglobin was monomeric and capable of reversibly binding oxygen. Gene duplication and mutation of the gene encoding this monomeric protein led to the development of a dimeric protein (α/β dimer) that was capable of more efficient and cooperative oxygen binding. The Bohr effect and 2,3-diphosphoglycerate (DPG) binding effect subsequently evolved and these were responsible for fine-tuning the function of haemoglobin.

1.1.3 The evolution of globin gene structure

Based on protein structure analysis, Gö (1981) and Blake (1978, 1981) proposed that modern globins arose from an ancestral four-exon / three-intron gene from which the central intron was lost during the course of animal evolution. They found that in many globular proteins for which the internal division into structural domains has been determined, a more or less exact correspondence exists between the exons of the gene and structural domains of the protein. Furthermore, the positions of introns in the

genes are closely correlated with substructures within a structural domain (Gö and Nosaka, 1987).

Vertebrate haemoglobin α and β chains consist of four functional domains while their genes consist of only three exons. The central exon of globin genes encode the haem-binding pocket, but the analysis of folded domains in globins showed that this haem-binding region is separated into two compact structures (Eaton, 1980; Craik *et al.* 1981; Jensen *et al.* 1981: Jhiang *et al.* 1988). This apparent discrepancy between the four conformational structures and the three exons was "resolved" by the analysis of leghemoglobin genes (Jensen *et al.* 1981; Brisson and Verma, 1982; Hyldig-Nielsen *et al.* 1982). These genes have four exons and three introns. The position of the two outer introns show precise homology to the position of vertebrate haemoglobin gene introns, and the middle intron separates the homologue of the vertebrate central exon into two exons (Jensen *et al.* 1981; Czelusniak *et al.* 1982; Blanchetot *et al.* 1983; Weller *et al.* 1984; Gilbert, 1978, 1987).

Some invertebrate globin genes (e.g. in the nematode *Asuris)* also possess the four exon - three intron structure found in plants (Dixon *et al.* 1992). Surprisingly, the globin genes from the earthworm (Jhiang *et al.* 1988) possess the three exon - two intron structure exhibited by vertebrate globins, while the insect (*C. thummi thummi*) has intronless globin genes (Antoine and Niessing, 1984; Kloek *et al.* 1993). The latter arrangement appears to be the product of an unusual conversion event of intron-less cDNA brought about by reverse transcription and re-insertion into the genome, from spliced mRNA in germ-line cells (Antoine and Niessing, 1984; Lewin, 1984; Jhiang *et al.* 1988). Intron loss is thought to be widespread in insects, perhaps the result of selective pressure to maintain a smaller genome size (Hawkins, 1988). Bacteria (Wakabayashi *et al.* 1986) and yeast (Zhu and Riggs, 1992) globin genes also appear either to have lost all three introns or to have never possessed introns.

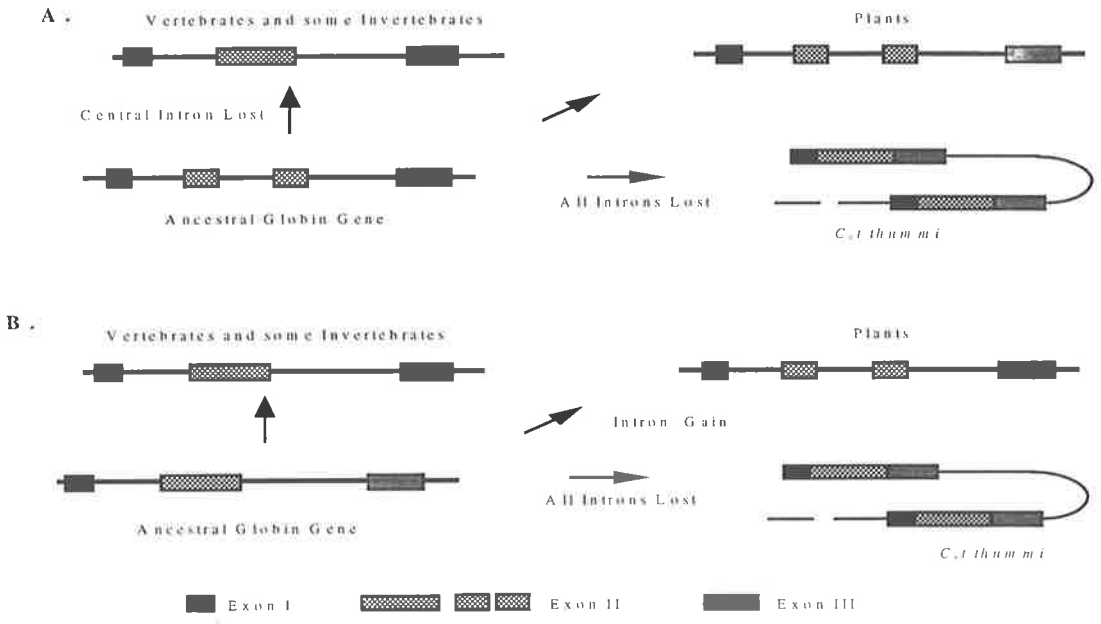
Several investigators have concluded that the plant four-exon / three-intron structure reflects the structure of the primordial globin gene that is a common ancestor of all animal globin genes with a three-exon / two-intron configuration. During vertebrate

evolution, the central intron was lost and in an invertebrate (*C. t. thummi*) lineage all introns were lost (Efstratiadis *et al.* 1980; Lacy and Maniatis, 1980; Li and Graur, 1991; Jellie *et al.* 1996) (Fig. 1.1.2 A). However, the absence of a central intron in some invertebrate globin genes suggests that this model is incorrect. Dixon and Pohajdak (1992) proposed that the typical vertebrate three-exon / two-intron gene structure is the primordial eukaryotic form, and that following the divergence of plants and animals, a common ancestor of plants gained a central intron in the globin gene (Fig. 1.1.2 B). At present, it is not clear whether leghaemoglobin genes acquired their introns during evolution or whether the split gene organisation was already present in the common ancestor of globin genes. The two-intron / three-exon gene organisation has been strictly conserved in vertebrate haemoglobins and myoglobins.

Figure 1.1.2 Two evolutionary models of globin exon / intron structure (adapted from Marchant, 1994).

A: The ancestral globins contain four / exons and three / introns as found in plant leghaemoglobin genes. During vertebrate evolution the central intron was lost and in the *C. t. thummi* lineage all introns were lost (Li and Graur, 1991).

B: The Dixon and Pohajdak (1992) model. The ancestral globin gene contained three-exons and two-introns. During evolution plants gained the central intron and *C. t. thummi* lost all introns.



1.1.4 Expression of haemoglobin genes in vertebrate

Most of the functional α- and β-globin genes of vertebrates such as amphibians, birds and mammals occur as part of multiple gene clusters. The individual member genes of each cluster often show differential expression during development and code for globin proteins with different functional properties e.g. oxygen binding capacities. For example, the embryonic globins of eutherian mammals have a higher oxygen affinity than adult globins, and this likely enhances placental exchange of oxygen from the mother to the embryo (Honig and Adams III, 1986).

Erythropoisis in mammals and many other vertebrates is first observed in the blood of the yolk sac. The early erythrocytes are large nucleated cells that contain embryo-specific haemoglobins. Later in development the site of erythropoiesis shifts to the foetal liver. This shift is accompanied by the expression of different haemoglobin genes in humans (Maniatis *et al.* 1980). During human development there is a series of changes in the major site of erythropoiesis which more or less parallels the changes in the pattern of haemoglobin production. The earliest embryonic haemoglobin tetramer found in primitive nucleated erythrocytes in the placental yolk sac consists of two ζ (α-like) and two ε (β-like) polypeptide chains ($\xi_2\varepsilon_2$). At approximately six weeks gestation the embryonic chains are gradually replaced by the adult α-globin chains and two different foetal chains, designated $^{G}\gamma$ and $^{A}\gamma$. The site of erythropoiesis moves to the liver and spleen, and coincides with the production of foetal haemoglobin $\alpha_2\gamma_2$. At about 39 weeks, the γ chains are gradually replaced by the adult β- and δ-globin polypeptides. At six months after birth, the haemoglobin consist of two adult types Hb A ($\alpha_2\beta_2$) and Hb A2 ($\alpha_2\delta_2$) and erythropoiesis only occurs in bone marrow cells (Maniatis *et al.* 1980; Hosbach *et al.* 1983; Karlsson and Nienhuis, 1985; Zitnik, *et al.* 1993). This pattern of expression found in human is not found in all mammals. For example, the marsupial *S. crassicaudata* contains single ε- and β-globin genes (Cooper and Hope, 1993; Cooper *et al.* 1996).

1.1.5 Regulation of the β-globin switches

The switches in haemoglobin production as development progresses involve regulatory control at the transcription level including the action of local (or proximal) promoters, distal regulatory elements, and possibly environmental factors, all operating upon a dynamic chromatin domain (Fritsch *et al.* 1980; Hill *et al.* 1984).

Magram *et al.* (1985) detected *trans*-acting regulatory factors and several *cis*-regulatory factors in transgenic mice. In addition, they identified several conserved *cis*-acting elements that may be involved in differential expression of the β-like globin genes during development. Based on analyses of deletion and substitution mutants in chicken globin genes, Choi and Engel (1988) suggested that temporal regulation of β-like globin genes results from competition between embryonic and adult genes for interaction with a mininum of two separate genetic elements; one responsible for erythroid tissue specificity (enhancer) and another responsible for definitive cell stage specificity (the β-globin developmental stage selector element) (DeBoer *et al.* 1988; Kotkow and Orkin, 1995; Liu and Mertz, 1995).

The Locus Control Region (LCR) is a large segment of DNA that exerts profound effects on the expression of the β-globin genes in mammals. The β-globin LCR can be divided into four small subdomains (each about 300 bp) distributed over nearly 20 kb, coinciding with chromatin regions that exhibit hypersensitivity (HS) to DNase I digestion in erythroid, but not non-erythroid, cells. The five HS sites (5' HS1-HS5) were located 6.1, 10.9, 14.7, 18, and 21.5 kb upstream of the ε-globin gene cap site. The other major HS site was located 19 kb downstream of the β-globin gene and is referred to as 3' HS1 (Tuan *et al.* 1985, 1989; Townes and Behringer, 1990; Crossley and Orkin, 1993). Deletion experiments of 5' flanking regions of β-like globin gene family demonstrated that a single upstream HS site is sufficient for high level erythroid specific expression in transgenic mice and in erythroid cell cultures (Ryan *et al.* 1989; Collis *et al.* 1990). Studies of individuals with deletion forms of γδβ°-thalassemia have shown that deletion of the LCR region results in failure to express any of the

downstream globin genes even though all of these genes themseves remain intact (Drisocoll *et al.* 1989). Constructs containing the LCR linked directly to the human β gene are expressed at high levels in the mouse but proper developmental regulation is lost (Enver *et al.* 1990). These findings confirm that the LCR has an enhancer function and the LCR elements are important for the expression of each of the β-like globin genes (Gumucio *et al.* 1996).

It has been suggested that LCR sequences perform two distinct functions. First they organise the entire β-globin locus into an "'open" or DNase I sensitive domain. Second, they serve as a powerful enhancer of ε-, γ- and β-globin gene transcription. DNase I footprinting and gel shift analyses have revealed that each 5'HS core element contains binding sites for transcription factors including GATA, AP-1 and CACC. The binding of proteins to the 5'HS sites is likely to cause a change in the tertiary structure of the local DNA. As a consequence, specific bases become more exposed to DNase I digestion (Pruzina *et al.* 1991; Moon and Ley, 1991; Reddy *et al.* 1994).

Two models have been proposed to account for the role of the 5'LCR in human β-globin gene switching: the DNA looping model (Engel, 1993) and the holocomplex model (Bresnick and Felsenfeld, 1994; Ellis *et al.* 1996).

i) In the DNA looping model, the LCR interacts with β-globin genes *via* DNA looping formed by DNA-protein and protein-protein interaction. The sequential interactions between distal (HS sites) and proximal (promoter and negative regulatory sequence) *cis*-regulatory elements of the locus form stable complexes. The β-globin LCR appears to interact with all of the genes in the β-globin cluster but some variation in preference among active promoters might be expected, depending on the relative strength of interaction between promoter- and LCR-bound factors and upon the size of the loop.

ii) The holocomplex model suggests that the LCR forms a holocomplex by protein binding to individual 5'HS sites. This interaction opens the chromatin domain of β-globin. Once the open domain is established, the LCR functions as an enhancer to direct high level globin gene expression. This model accounts for all the 5'HS sites being necessary for full β-globin gene

expression but is not dependent on any preferential interaction between individual 5'HS sites and β-globin genes.

The mechanisms by which the LCR acts in the switching process is still under investigation.

Sequence comparisons as well as functional analyses have shown that the LCR is conserved between humans, mice, goats and rabbits. The 5'HS sites are conserved in size spacing and sequence, and the binding sites for nuclear transcription factors GATA-1 and NF-E2 are also highly conserved (Li *et al.* 1991; Jimenez *et al.* 1992; Hardison *et al.* 1993). It is possible, therefore, that a careful analysis of sequences within and surrounding the 5'HS in a wide range of mammals may reveal conserved functional elements that have been overlooked in expression studies.

1.1.6 The structure of β-globin genes

1.1.6.1 Non-coding regions

The sequences of the noncoding regions are of interest because these regions may play a role in transcription and translation control of gene expression. Alignments of sequences in the promoter regions of β-globin genes demonstrate a high level of sequence identity in all mammalian globin gene families (Efstratiadis *et al.* 1980; Jahn *et al.* 1980). The region at approximately -90 relative to the CAP site contains the sequence GCCACACCC, a common feature of β-globin gene promoters from many species. The CCAAT box at approximately -75 relative to the CAP site is a common element in many eukaryotic promoters transcribed by RNA polymerase II. The TATA box is located at about -30 in most eukaryotic genes where it is involved in determining the site of transcription initiation (Hartzog and Myers, 1993). In addition, erythroid-specific enhancer elements have been found in the 5' flanking regions of β-like globin gene domains, and at 10.2-11.0 kb upstream of the embryonic β-globin gene of several mammals (Tuan *et al.* 1989).

All β-globin genes in mammals contain two intervening sequences (IVS), the second 3' being larger than the first. Both introns are variable in size. IVS1 ranges from 90 bp in mouse ε-globin gene to 327 bp in mouse h0-globin gene and IVS2 from 566 bp in *Lepus. europaeus* (Lagomorpha) β-globin to 1,858 bp in the *Didelphis. virginiana* (marsupial) ε-globin (Koop and Goodman, 1988) (Appendix 6). Both the size variability and sequence alignments of the introns indicate that they have evolved by a combination of base substitution and deletion/insertion events (Konkel *et al.* 1979). The splice junction sites in both introns have been higly conserved in β-globin genes (Gö and Nosaka, 1987; Doolitle, 1987). Nearly all 5' sites have the dinucleotide GT, whereas the 3' junction has AG. The conserved consensus sequence of splice sites suggest their importance in the processing mechanism which removes intron sequences from nuclear precursor RNAs during the production of mature mRNA (Nishioka *et al.* 1980).

Repetitive sequences have been found interspersed within the β-globin gene clusters of rabbits (Shen *et al.* 1981), humans (Malcom *et al.* 1981), mice (Haigwood *et al.* 1981) and chickens (Wood *et al.* 1981). They are located in the flanking sequences and intervening sequences (Young *et al.* 1982). Middle repetitive sequences within introns may participate in genetic recombination resulting in an increase of the exon assortment rate, but the function of short interspersed repeats in general has not been determined (Gilbert, 1978, 1987; Blake, 1981; Fuhrman *et al.* 1981; Disegni *et al.* 1981; Schimid and Jelinek, 1982; Kao *et al.* 1994).

1.1.6.2 Pseudogenes

Two types of globin pseudogene have been identified. One type is usually closely linked to the functional gene from which it was presumably derived. Tandem gene duplication followed by mutational inactivation are thought to be the mechanisms of origin of these pseudogenes (Leder *et al.* 1981). The other type of pseudogene has lost intervening and flanking sequences and frequently has been transposed away from the parent gene family, possibly by unequal crossing-over between homologous sequences (Little, 1982; Hardison and Margot, 1984; Ghazal *et al.* 1985). These types

of pseudogene are thought to arise through reverse transcription of mRNA and re-insertion into the genome (Patterson, 1987).

At least one pseudogene of the former type has been found in the β-globin gene families of primates (ψβ), mouse (βh2 and βh3), rabbit (ψδ) and goat (ε^{III}, ε^{VI}, $\psi\beta^{X}$, $\psi\beta^{Z}$ and $\psi\beta^{Y}$) (Lacy and Maniatis, 1980; Cleary *et al.* 1981; Jeffreys *et al.* 1982; Hardies *et al.* 1984; Hardison and Margot, 1984; Harris *et al.* 1986). These pseudogenes are located between the embryonic or foetal genes and the adult globin genes. However, no pseudogenes have been reported in either the α- or β-globin gene families of chicken (Dolan *et al.* 1981). In non-primate mammals the known β-globin pseudogenes such as those in mouse (βh2 and βh3), rabbit (ψδ) and goat ($\psi\beta^{X}$, $\psi\beta^{Z}$ and $\psi\beta^{Y}$) show high sequence similarity to the adult expressed gene (β). In contrast, the higher primate pseudogenes, such as those present in lemur and human (ψβ) and goat (ε^{III} and ε^{VI}), show high sequence similarity to the embryonically expressed gene (ε) (Chang and Slightom, 1984; Hardison, 1991). The widespread occurrences of globin pseudogenes and the high conservation of the nucleotide sequence in and around the hominoid (human and gorilla) pseudogenes (ψβ) suggest that pseudogenes may be involved in the regulation of other β-globin genes (Chang and Slightom, 1984). Collins and Weissman (1984) have suggested that the location of pseudogenes between the γ- and δ-genes may be important in maintaining a *cis*-acting regulatory signal necessary for the switch from foetal to adult globin gene expression. Alternatively, and perhaps more likely, these pseudogenes may play no functional role.

1.1.7 Arrangement of β-globin gene families

Expression of the α-like and β-like globin genes is coordinately regulated so that equal amounts of both types of polypeptide are synthesized (Hardison, 1984; Roninson and Ingram, 1982). In the mouse, rabbit, primates and human the α- and β-globin genes are arranged on the chromosome in the order of their expression (5'→3') during development (Lacy *et al.* 1979; Leder *et al.* 1981). However, in the chicken, sheep, cow and goat, the arrangement of β-globin genes does not correspond to the

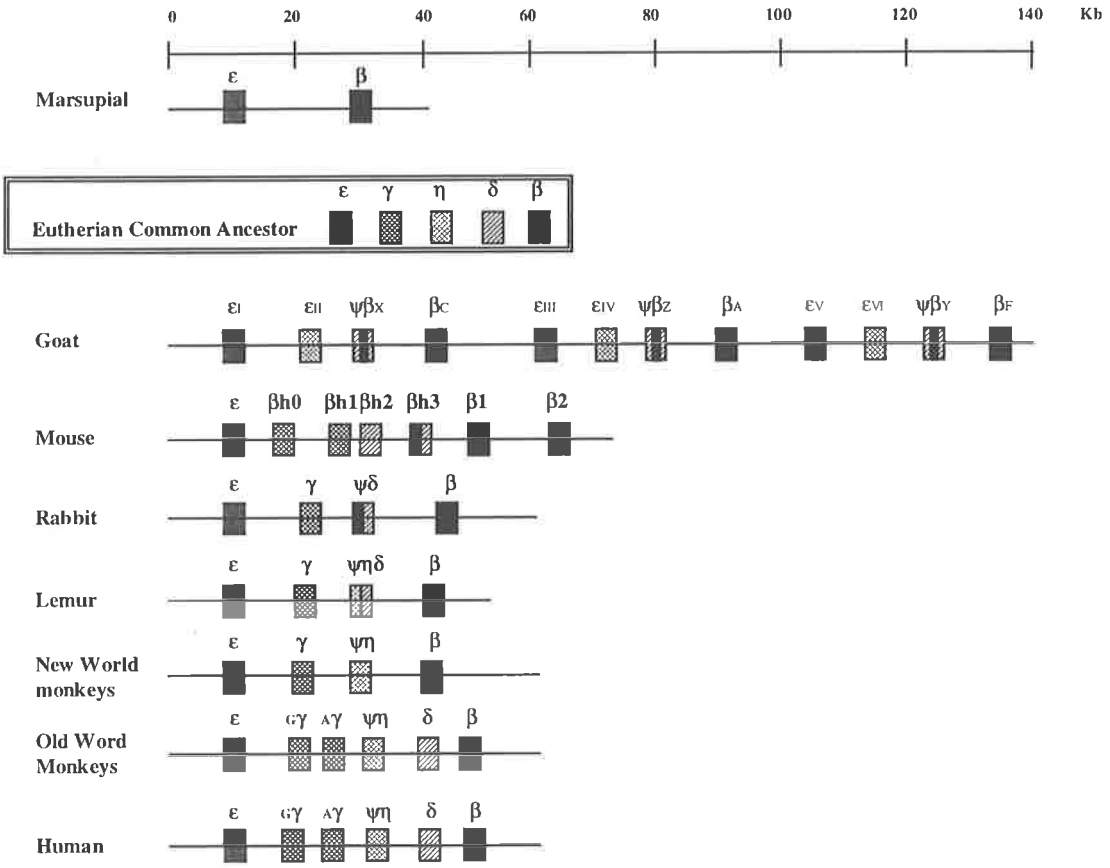
expressional order (Dolan *et al.* 1981; Townes *et al.* 1984; Schimenti and Duncan, 1985a,b).

The chicken β-globin genes are arranged over 14 kb of DNA in the order: 5'-ρ-βH-βA-ε-3' (Dolan *et al.* 1981) (Fig. 1.1.1). The ρ- and ε-globin genes are expressed during early embryonic development (days 2-5) and located on the outside of the cluster, while the βA- and βH-globin genes are transcribed after embryonic day 5 and located on the inside of the cluster (Dolan *et al.* 1981; Reitman *et al.* 1993). Roninson and Ingram (1982) compared the DNA sequences of the ρ- and ε-globin genes and deduced that each arose by independent duplication of the adult β-globin gene, and not by tandem duplication of an embryonic-like globin gene. This could therefore explain why these genes are located at opposite ends of the cluster.

The arrangement of the β-globin gene family among the three ruminant species (goat, sheep and cow) is very similar (Schon *et al.* 1981; Townes *et al.* 1984; Schimenti and Duncan, 1985a,b). The largest known β-globin gene cluster studied to date occurs in the goat and consists of twelve genes (Fig. 1.1.3). Comparisons of the nucleotide sequences of these genes suggests that the goat β-globin gene cluster evolved by the triplication of a common ancestral four-gene set ε-ε-ψβ-β (Townes *et al.* 1984; Garner and Lingrel, 1988). In goat, the ε-globin genes are expressed in the embryo while βF-, βC-, and βA-globin genes are expressed in foetal, pre-adult (after birth for approximately 3 months) and adult tissues, respectively. Interestingly, the ruminant β-globin genes, which are expressed after the embryonic stage, are located at the ends of the four-gene sets: βC pre-adult (juvenile), βA adult, βF foetal in goat, βB adult, βF foetal in sheep and β adult, γ foetal in cow. Nucleotide sequence comparisons between the βF- and βC-globin genes suggested that these genes originated from an adult β-globin gene and were then recruited into non-adult developmental regimes: βC became expressed in juveniles and βF at the foetal stage (Townes *et al.* 1984). In the cow (*Bos taurus*) there is a duplicated set of foetal and adult clusters, but no juvenile cluster (Schimenti and Duncan, 1985a,b).

The β-globin gene families of the mouse (Jahn *et al.* 1980; Hill *et al.* 1984; Leder *et al.* 1981) and rabbit (Lacy *et al.* 1979; Lacy and Maniatis, 1980) have a similar arrangement to that found in primates. The mouse has three embryonic-like (y, βh0, βh1), two pseudogenes (βh2 and βh3) and two adult β-globin genes (βmaj, βmin). The rabbit has two embryonic-like (ε and γ), one pseudogene (ψδ) and one adult β-globin gene.

Figure 1.1.3 Organization of the β-globin gene clusters of marsupial and eutherian: This diagram is adapted from Jeffreys, (1982) and Hardison, (1991).



The arrangement of β-globin gene clusters of Great apes and Old World monkeys are very similar, but the New World monkeys and lemur have fewer β-globin genes in their clusters (Barrie *et al.* 1981; Jeffreys *et al.* 1982; Harris *et al.* 1984; Bailey *et al.* 1992) (See Fig. 1.1.3).

The β-globin gene family of humans consists of five functional genes (ε, $^{G}\gamma$, $^{A}\gamma$, δ and β) and one pseudogene (ψη) arranged over a region spanning approximately 60 kb of DNA in the (5'→3') order of their developmental expression (Efstratiadis *et al.* 1980) (Fig. 1.1.3).

1.1.8 Evolution of the β-globin gene family in mammals

Extensive analyses of β-like globin genes in four eutherian orders (Artiodactya, Rodentia, Lagomorpha and Primates) have provided evidence that the progenitor cluster, i.e. the group of β-globin genes that existed prior to the major eutherian radiation 80 MYA, consisted of five genes, 5'-ε-γ-η-δ-β-3'. The extant β-globin clusters of eutherian mammals bear some resemblance to this five-member ancestral cluster (Hardies *et al.* 1984; Hardison, 1984; Harris *et al.* 1984; Goodman *et al.* 1984; Hardison and Gelinas, 1986; Tagle *et al.* 1988, Goodman, 1989). It has been proposed that during or just prior to the separation of the Rodentia and Lagomorpha lineages, the η-globin locus was deleted and the δ gene was involved in multiple gene conversions with the β-globin locus during the evolution of Lagomorphs, Primates and Artiodactyls (Jeffreys *et al.* 1982; Martin *et al.* 1983; Hardies *et al.* 1984) (Fig. 1.1.3). During the evolution of the Artiodactyla β-globin gene cluster, the η-globin gene was retained and the γ-globin gene was deleted. In fact, only in the goat is η known to be active. In this mammal it codes the embryonic ε^{II} globin (Lingrel *et al.* 1985; Goodman *et al.* 1984, 1985) (Fig. 1.1.3).

The early expressed human globin genes arose by two duplication events. First, a proto ε-globin gene duplicated to give rise to ε- and γ-globin genes about 80-140 MYA (Fig. 1.1.4). Secondly, γ-globin duplicated to give two γ-globin genes (γ^{G} and γ^{A}) about 20 to 40 MYA after the separation of the Catarrhines (New world monkey) and Prosimian (lemur) lineages (Catarrhines and Prosimians have only one γ-globin) (Chang *et al.* 1978; Jeffreys, 1979; Slightom *et al.* 1980; Barrie *et al.* 1981; Shen *et al.* 1981; Jeffreys *et al.* 1983; Goodman *et al.* 1987a) (Fig. 1.1.4). The orthologous embryonic-like globin genes of eutherians are not all expressed at the same time of

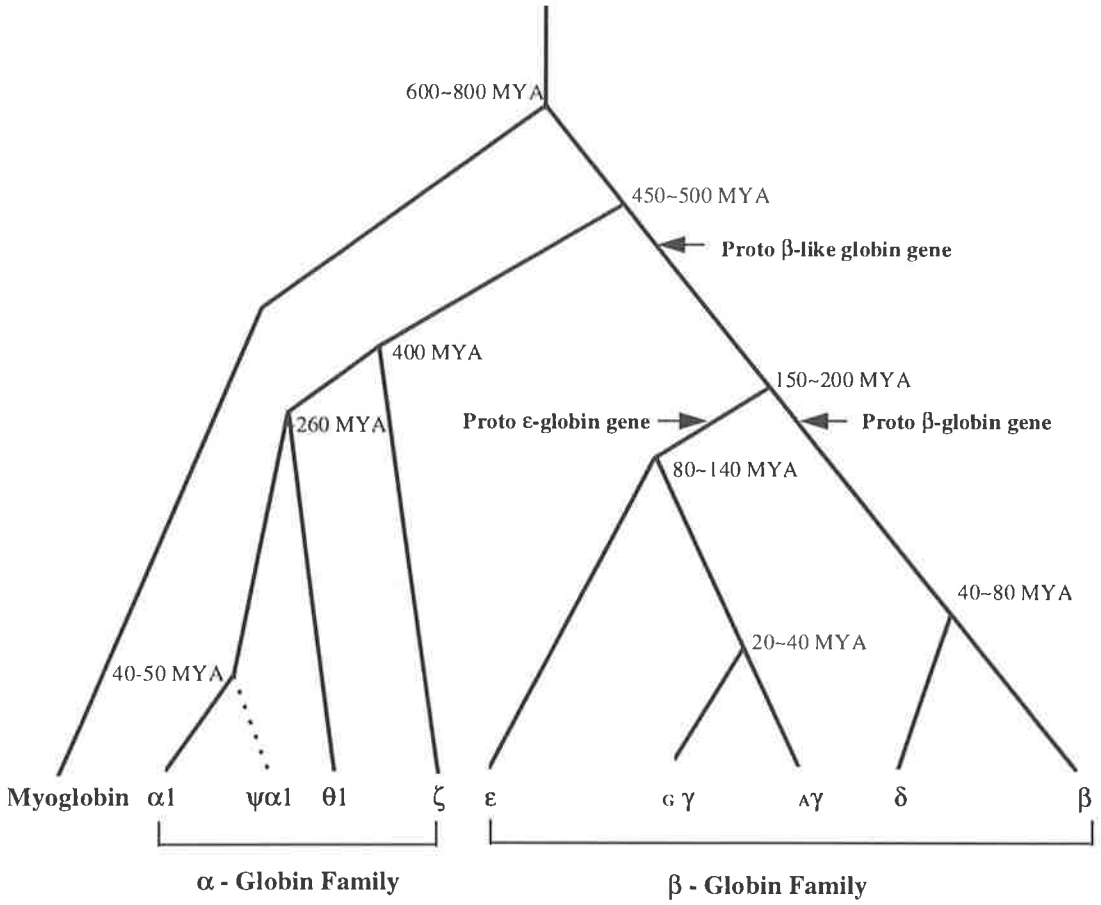
development. For example, the γ-like globin genes in the mouse (βh0 and βh1) and rabbit (γ) are normally expressed during embryonic life, but in higher primates, including New World and Old World monkeys, apes and humans, the γ-globin genes are expressed in foetal life. These data imply that after the duplication of an ancestral ε-globin gene, the γ-globin gene was recruited into expression during foetal development prior to the radiation of primates (Hardison and Margot, 1984; Hardison, 1984, 1991).

The separation of the δ- and β-globin genes was originally estimated to have taken place 40 MYA (Dayhoff, 1976; Efstratiadis *et al.* 1980) but DNA sequence comparison of primate β-globin genes suggest that it may have occurred before the eutherian radiation, about 80 MYA (Hardies *et al.* 1984; Hardison, 1991) (Fig. 1.1.4). However, the 5' region of the human δ- and β-globin genes show a remarkably similar nucleotide sequence suggesting that a conversion of this region between δ and β occurred recently during the evolution of the human lineage (Hardison and Margot 1984; Hardies *et al.* 1984; Hardison, 1991).

Two β-like globin genes have been isolated and sequenced from each of two marsupials, the North American opossum (*Didelphis virgiana*) and the Australian Fat-tailed Dunnart (*Smithopsis crassicaudata*). One of the genes in each species is similar in its coding and promoter sequences to the eutherian embryonic-like globin genes, while the other gene is similar to the adult β-like globin genes (Koop and Goodman, 1988; Cooper and Hope, 1993; Cooper *et al.* 1996) (Fig. 1.1.4). Based on genomic Southern data from the opossum, Koop and Goodman (1988) concluded that two functional β-like globin genes exist in this species. This finding was taken to indicate that the last common ancestor of eutherian mammals and marsupials had only two progenitors (ε-like and β-like globins) of the five linked β-related loci found in eutherians, and that these two ancestral genes were already differentiated with respect to promoters and developmental expression (Koop and Goodman, 1988). These studies also supported the view that the γ-, η-, and δ-globin genes of the ancestral

eutherian β-globin cluster arose by a series of gene duplications, in a stem species of eutherian, after the divergence of the marsupial and eutherian lineages.

Figure 1.1.4 An evolutionary tree of globin genes showing the proposed duplication events which led to the evolution of the human α- and β-globin gene clusters. The estimated duplication times are shown on a maximum parsimony tree derived from globin DNA sequence comparisons (Adapted from: Li and Graur, 1991).



Evidence for the existence of a third β-like globin gene in marsupials has recently been found in *Macropus. eugenii* (Tammar Wallaby). Using iso-electric focusing, Holland and Gooley (1997) reported the presence of four distinct haemoglobin types in neonatal Tammar Wallaby blood. All four types were different from the adult haemoglobin which made its appearance at about 2-3 days after birth. Purification and partial amino acid sequencing of the β-like globin chains present in these early

expressed globins revealed the presence of two β-like globin chains. One of these appeared to be orthologous to *S. crassicaudata* and *D. virginiana* ε-globin. However, the other β-like chain (named ω-globin) contained a number of unexpected features. Comparisons of its amino acid sequence to β-like chains from a range of vertebrate species showed that it most closely resembled avian β-globin chains (Hollen, Gooly, Hope, Cooper, Wheeler and Goodman-unpublished). Interestingly, ω-globin shows a greater degree of sequence identity to the avian β-like chains than it does to any mammalian β-globin chain. These results in the Tammar Wallaby clearly indicate the presence, in this marsupial species at least, of a third β-like globin gene. They also suggest that there is a need to re-examine current hypotheses concerning the evolution of the β-globin genes.

The first duplication of an ancestral proto β-like globin gene is estimated to have occured approximately 150-200 MYA. This time coincides approximately with the estimated divergence time (200 MYA) of the monotreme lineage from the lineage that subsequently gave rise to the marsupials and eutherians. Despite the wealth of information on the structure, function and evolution of marsupial and eutherian β-globin genes, no DNA data on the β-globin genes is available from monotremes. Because of this ancient separation from the (marsupial + eutherian) stem, molecular studies of the β-globin genes in monotremes should provide a greater understanding of β-globin gene family evolution in the mammals. It is noted that the maximum parsimony approach for deriving phylogenies from DNA sequence comparisons, enables ancestral states to be derived i.e. sequences that are presumed to have existed at the nodes on evolutionary trees.

1.2. MONOTREMES*

In most taxonomic works, monotremes are regarded as a very distinct group of mammals and are placed in a separate subclass, Prototheria (Fig. 1.2.1). It has been proposed that Prototherians diverged from therian mammals (marsupials and eutherians) some time in the Mesozoic, about 150-200 million years ago (MYA), well before the marsupials and eutherians diverged from one another in the early Cretaceous, about 135 MYA (Clemens, 1989). However, the timing of the evolutionary separation of the prototheria and theria is not well established. An exhaustive review of the morphological anatomy has shown that monotremes have many detailed parallelisms with therian mammals such as the brain case, ear ossicles, expansion of neopallium and epiphyses. It has been suggested that the prototheria/theria branch occurred close in time to (and may have even post-dated) the marsupial/eutherian branch (Kemp, 1983; Clemens, 1989; Benton, 1990). In other words, the generally accepted branching arrangement may well be wrong. Therefore, it would be wise to discontinue use of the terminology prototheria and theria until the order and timing of these evolutionary events is clarified.

Monotremes consist of two families, *Tachyglossus* (Echidna family) and *Ornithorhynchus* (Platypus family) (Fig. 1.2.1). The echidna family contains two extant species: the short-beaked echidna, *Tachyglossus aculeatus* (Australian echidna or echidna[1]), widespread in Australia (including Tasmania) (Fig. 1.2.2) and Papua New Guinea, and the long-beaked echidna, *Zaglossus bruijni* (New Guinea echidna) which formerly occurred in Australia, but now inhabits only Papua New Guinea (Fig. 1.2.2). The platypus, *Ornithorhynchus anatinus*, is the only extant species in the *Ornithorhynchus* family, and it's habitat is restricted to Tasmania and eastern coastal streams of Australia (Fig. 1.2.2) (Bentley *et al.* 1967; Augee, 1984; Griffiths, 1989; Grant, 1989b; Griffiths *et al.* 1991).

---

*Colour photographs of the three monotreme species are shown on the front page of this thesis.

[1] Throughout this thesis, the term "echidna" used alone should be taken to describe the Australian echidna. The other species of echidna will always be termed the New Guinea echidna.

Monotremes are specialised with regard to their habitats. The platypus is semi-aquatic and catches its food primarily under water in rivers and streams, whereas the echidna is terrestrial and lives on termites which it obtains by burrowing into their nests (Bentley *et al.* 1967). Morphologically, both animals give the immediate impression of being uniquely adapted to their specialised environments. The streamlined platypus is well adapted to its semi-aquatic life, while the echidna is well adapted to digging and burrowing (Hosken, 1978).

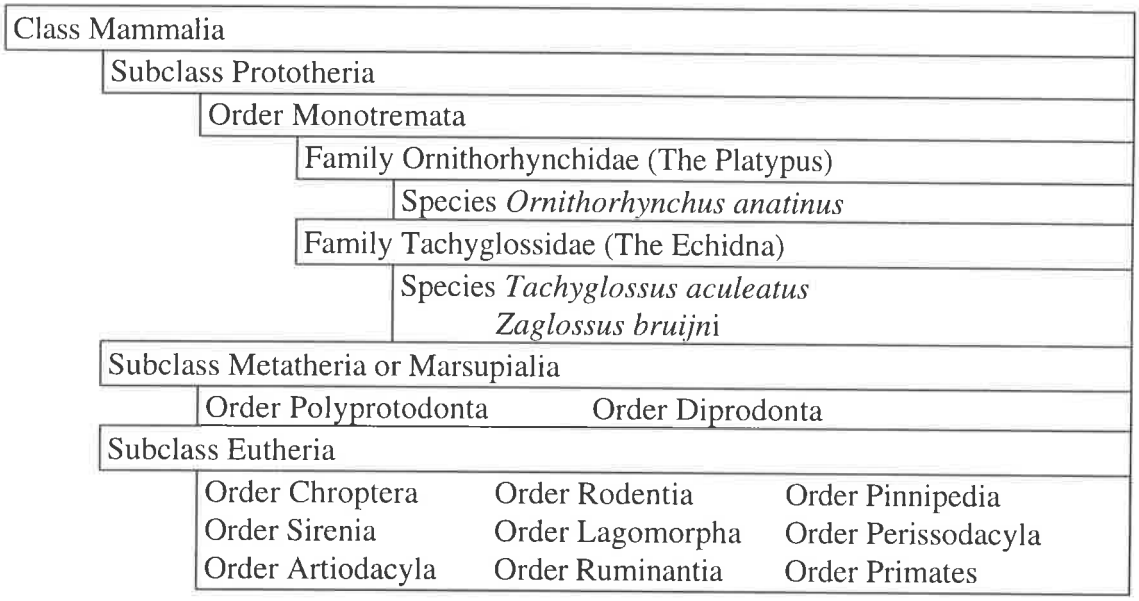
Echidnas show considerable variation in hair and spine characteristics. The hair colour ranges from light brown to almost black, depending on locality. The six subspecies are classified by the length of their spines, their hairiness and the relative length of the third hind claw as compared with the second (Rismiller and Seymour, 1991). Based on the differences in haemoglobin electrophoretic phenotypes in different locations, echidnas have been classified into four groups (Cooper *et al.* 1973; Murtagh, 1977). However, such classifications have been based on few animals and no studies of these haemoglobin electrophoretic phenotypes have been carried out at the molecular level. Therefore, these findings are insufficient to provide an accurate classification or to reflect genetic relationships in echidna.

The New Guinea echidna differs from the Australian echidna by its much greater size, and relatively long snout. New Guinea echidna once inhabited the Australian mainland but became extinct from this area in the late Pleistocene (Griffiths, 1989). The New Guinea echidna is common in Australian fossil deposits (Murtagh, 1977; Murray, 1984; Griffiths, 1989). Three species of New Guinea echidna have been recognised but insufficient specimens have been studied to exclude the possibility of extensive clinical variation and all may be attributable to a single highly variable species *Zaglossus bruijni* (Murtagh, 1977).

The most conspicuous feature of the platypus is its duck-like bill. This is covered with thick, but soft and flexible, naked skin which has blue-grey colour on its upper surface; it has such an improbable appearance that when specimens first reached Europe, they

were thought to be scientific hoaxes made up from parts of birds and parts of mammals (Hopson, 1970; Murtagh, 1977; Grant, 1989a,b).

Figure 1.2.1 The classification of mammals by Strahan (1991)



1.2.1 Morphology and reproduction

Monotremes possess a remarkable mixture of mammalian and reptilian/bird characters in their physiology, reproduction and osteology. They are more closely related to marsupial and eutherian mammals than to any group of reptiles, even though they have retained a surprising number of ancestral reptilian traits during the course of evolution (Bentley *et al* . 1967; Wrigley and Graves, 1984; Griffiths, 1978, 1989; Strahan, 1991; Augee and Gooden, 1993).

The name Monotremata refers to the single external opening, or cloaca, for both excretion and reproduction. The echidna female lays a single leathery shelled egg into its pouch where it is incubated for 7-10 days. After hatching, the bristle-less pouch young remains in the pouch until its spines develops (Rismiller and Seymour, 1991). The platypus lays 1-3 eggs in a nest within a burrow and stays with them until they hatch (Grant, 1989a,b). Unlike other mammals, monotremes have no nipples, but secrete milk from the ducts of the mammary glands to feed their young. Detailed

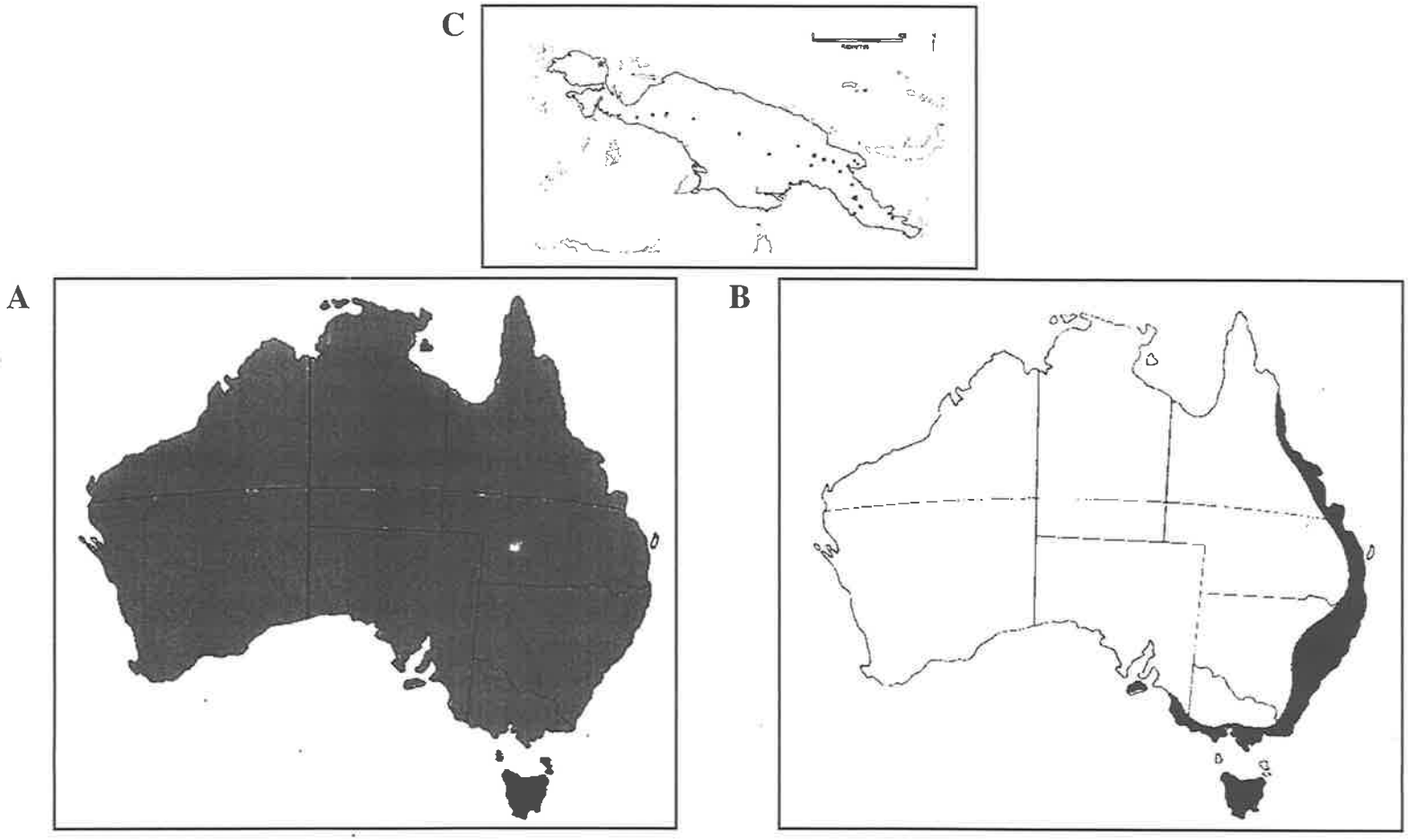


Figure 1.2.2 The distribution of Monotremes : **A**, theAustralian Echidna and **B**, the Platypus (from Strahan, 1991). **C**, the New Guinea Echidna is found in mountain forest and alpine meadows up to 4,000 m in New Guinea and on the island of Salawatti (Flannery, 1990).

studies of the parchment-like shell and the membranes of the egg have shown that it is fundamentally the same as the egg laid by reptiles and birds (Hughes, 1993; Augee and Gooden, 1993). Oviparity is one of the major characters which clearly separates the monotremes from the eutherians and marsupials (Augee, 1984; Augee and Gooden, 1993). Both the male platypus and the echidna have a hollow spur on the inside of the ankle which is connected with a gland situated higher in the leg (Murray, 1984). The monotremes have a body temperature of 32-33°C, significantly lower than that of most eutherian mammals. Monotremes are able to maintain a relatively constant body temperature when they are exposed to cold conditions (Rismiller and Seymour, 1991).

Some obvious characters that were immediately recognised were used to place monotremes together with marsupials and eutherians in the class Mammalia. Monotremes have a common pattern of mammary development such as areolar patches and functional mammary glands. Monotremes and marsupials share certain common structures which they do not share with eutherians: pelvis bones, tripartite structure of the spleen; a hammer type of ear ossicles, large open tympanic and the corpus calamus in the brain. Marsupials and eutherians share the characteristic of having the galactophores gathered together in a cluster opening at the ends of teats, which are lacking in monotremes (Griffiths, 1968, 1978; Gregory, 1974; Augee, 1984; Renfree, 1993).

The studies of comparative biochemistry and physiology have added many more characters that further establish the affinity of monotremes with marsupials and eutherians. However, only the monotremes appear to retain some reptilian characters (Augee and Gooden, 1993).

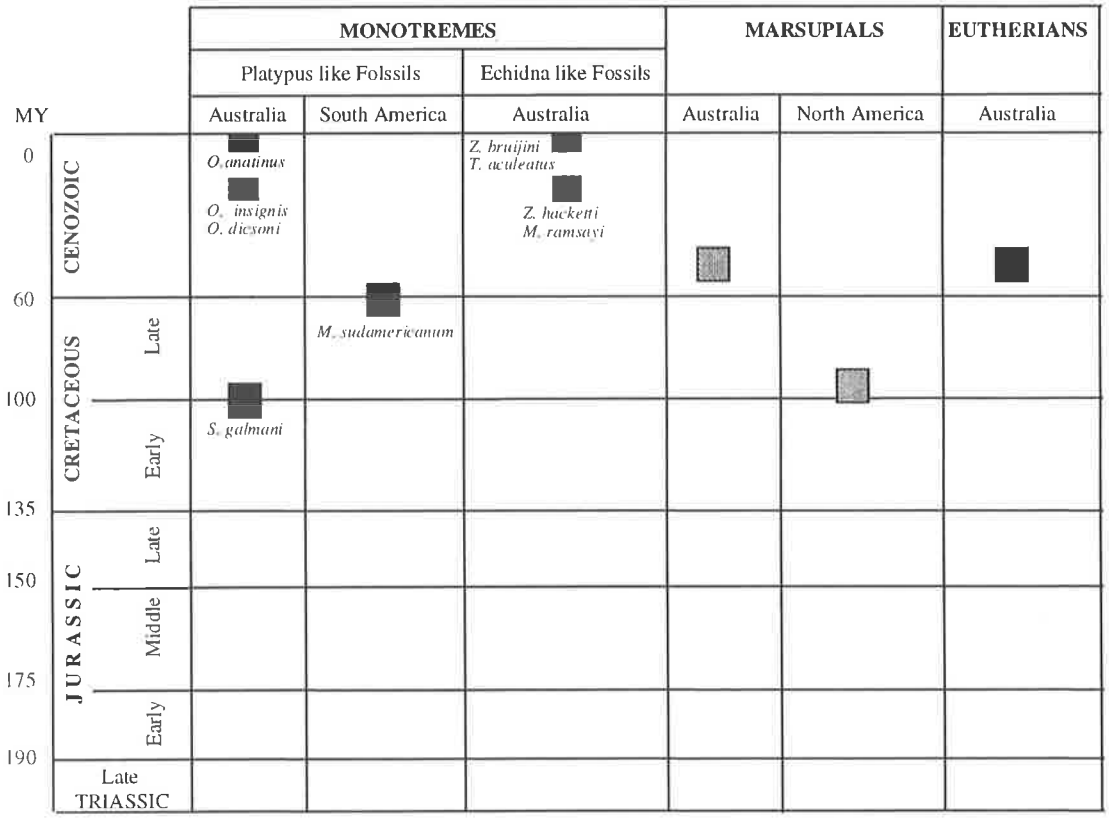
1.2.2 Fossil data

Today, Australia is the only continent to have representatives of all three living subclasses of mammals. The oldest fossil record for each of the subclasses is 110 MYA for Monotremata and 55 MYA for Marsupialia and Eutheria on the Australian continent (Rich, 1991; Archer *et al.* 1993), and 65 MYA for Monotremata and 95

MYA for Marsupialia on the American continent (Pascual *et al.* 1992a,b; Cifelli, 1993) (Fig. 1.2.3).

Few platypus-like fossils have been found in Australia and South America. In Australia the oldest fossil, *Steropodon galmani,* was found in early Cretaceous sediments dated at 110 MYA; *Obdurodon insignis* and *Obdurodon dicsoni* are from the old Miocene and dated at 25-15 MYA; and *Ornithorhynchus anatinus* is found in the Pleistocene and Holocene periods (5-0 MYA). In South America *Monotrematum subamericanum* has been found in early Paleocene sediments about 63-61 MYA (Murray, 1984; Archer *et al.* 1985, 1992, 1993; Kielan-Jaworowska *et al.* 1987; Pascual *et al.* 1992a,b) (Fig. 1.2.3).

Figure 1.2.3 The geological fossil record of monotremes, marsupials and eutherians on the Australian and American continents. The approximate geological ages are indicated by boxes.



There is no record of echidna prior to the late Miocene (about 20 MYA) where New Guinea echidna-like forms were found (Archer *et al.* 1992). Fossil data indicate that New Guinea echidna was once widespread throughout Australia, and it is assumed that increasing aridity led to the present contracted distribution (Sharman, 1978; Murray, 1984; Augee, 1984). The echidna fossils range in age from the late Miocene (20 MYA), *Megalibgwilia. ramsayi* and *Zaglossus hacketti,* to Holocene, *Zaglossus bruijni*, and *Tachyglossus aculeatus* (5-0 MYA) (Griffiths *et al.* 1991; Archer *et al.* 1992) (Fig. 1.2.3).

1.2.3 Evolution of monotremes and their phylogenetic relationships to the marsupials and eutherians

Eutheria is presently represented by approximately 3,000 species; marsupials, by 250 species (171 in Australia, 78 in South America and 1 in North America), and monotremes, by 3 species (Griffiths *et al.* 1991). The fragmentary fossil records and complete loss of adult teeth in extant monotreme species have been major problems to understanding the precise taxonomic relationship of monotremes to the other two extant mammal groups (Pascual *et al.* 1992a,b; Archer *et al.* 1992).

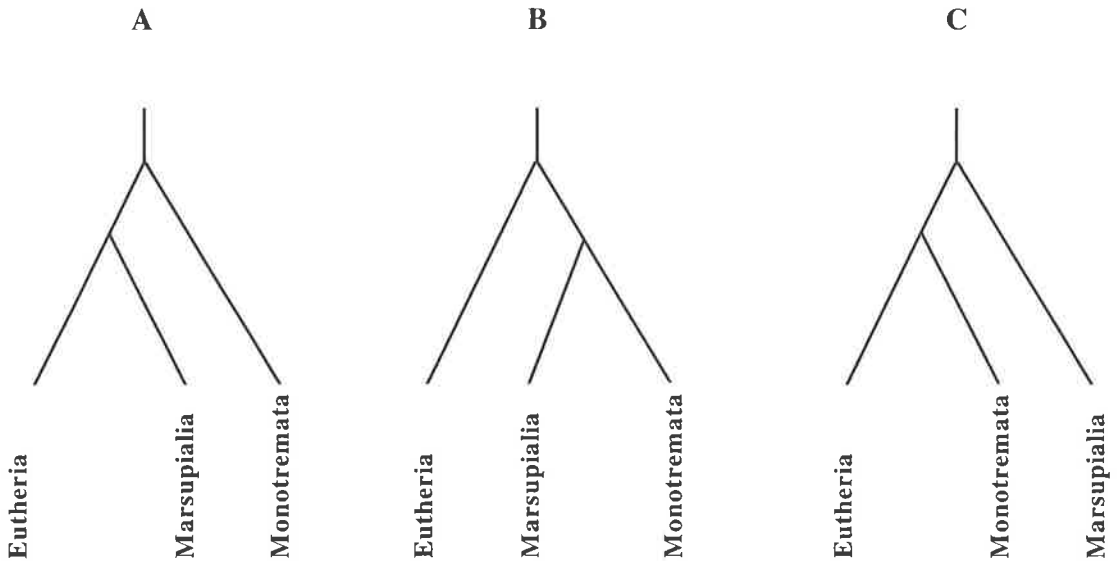
In considering the three extant groups of mammals, there are 3 possible bifurcated, rooted, phylogenetic trees (A-C) (Fig. 1.2.4). The generally accepted tree is "A" (monotremes, (marsupials, eutherians)) but recently it has been suggested that "B" (eutherians, (marsupials, monotremes)) may be the correct conformation (Janke *et al.* 1996, 1997). Based on an analysis of combined protein sequences (Goodman *et al.* 1987b) the phylogenetic relationships of mammalian subclasses were as shown in tree "C" (marsupials, (eutherians, monotremes)). However, they suggested that such an arrangement could be a result of fortuitous homoplasy between monotreme and eutherian β-globin sequences (Czelusniak *et al.* 1990).

Abundant morphological evidence supports the view that marsupials group with eutherian mammals, with monotremes more distantly related to these two groups i.e. tree "A" (Kermack and Kermack, 1984). Furthermore, phylogenetic studies with

molecular data such as DNA sequences of mitochondrial 12S rDNA and protamine p1 genes, as well as DNA-DNA hybridisation studies support a monotreme-marsupial/eutherian dichotomy (Westerman and Edwards, 1992; Retief *et al.* 1993; Gemmell and Westerman, 1994). Tree "A" is also supported by an analysis of synapomorphic characters and the dental structure of monotreme fossils (Kemp, 1983, 1988; Archer *et al.* 1985; Clemens, 1989; Rowe, 1993; Zeller, 1993).

In contrast, an alternative but widely disregarded hypothesis is that monotremes and marsupials, the so-called "Marsupionta", are sister groups (Fig. 1.2.4 B) (Gregory, 1974). Recently, Janke *et al.* (1996, 1997) lent support to this hypothesis using phylogenetic analysis of whole mitochondrial DNA genomes from representatives of each mammalian subclass.

Figure 1.2.4 Hypotheses for the phylogenetic relationships of monotremes, marsupials and eutherians.



Despite palaeontological evidence supporting the hypothesis that monotremes are a sister group to the marsupial and eutherian lineages (tree A), and use of abundant data on modern Ornithorhynchids and Tachyglossids to reconstruct the biology of Mesozoic mammals (Jenkins, 1990), the origin of the monotremes still remains a

central (yet unanswered) question to understanding mammalian evolution (Archer *et al.* 1992).

Estimates of mammalian divergence times are very varied and depend on the methods and data used for analysis. The date of the marsupial-eutherian evolutionary separation was estimated at 85-170 MYA (Air *et al.* 1971) and 81-144 MYA (Whittaker *et al.* 1978) using comparisons of amino acid residues from three globins (α-globin, β-globin and myoglobin) (Table 1.2.1). Using amino acid sequence comparisons from each of the three globins, Whittaker *et al.* (1978) estimated the dates of separation of the eutherian and monotreme lineages, and the platypus and echidna lineages, as 171 and 52 MYA from α-globin, 112 and 38 MYA from β-globin, and 93 and 28 MYA from myoglobin, respectively (Table 1.2.1). Based on amino acid sequence comparisons of the three combined globins, the separation time of the eutherian and marsupial lineages was estimated to be about 135 MYA, and the separation of the platypus and echidna lineages was estimated to be 54 MYA (Richardson, 1987) (Table 1.2.1). Hope *et al.* (1990) have summarised these dates and emphasised the large discrepancies between them.

Analysis of DNA-DNA hybridisation data gave an estimated time for the separation of the platypus and echidna lineages of 62-78 MYA and the separation of the monotreme-marsupial lineages of 110-137 MYA (Westerman and Edwards, 1991, 1992) (Table 1.2.1). Based on the protamine p1 DNA sequences, the divergence time of the platypus and echidna lineages was estimated to be around 18-27 MYA and the divergence of the monotreme and two other mammalian groups at approximately 160 MYA (Retief *et al.* 1993) (Table 1.2.1). Recently Janke *et al.* (1997), using nucleotide sequence comparisons of the mitochondrial coding regions from the platypus and representative marsupial and eutherians, suggested that monotremes and marsupials diverged from a common ancestor 115 MYA and that most extant eutherian orders had their origin in the middle to late Cretaceous, 115-65 MYA (Table 1.2.1).

Analysis of data from DNA sequences (e.g. protamine p1 and mitochondrial 12S) and from DNA-DNA hybridisation suggest that the monotreme-marsupial/eutherian divergence occurred not much more than 5-10 MYA before the eutherian and marsupials lineages split at some time in the late Jurassic-early Cretaceous period (Clemens, 1989; Westerman and Edwards, 1991, 1992; Retief *et al.* 1993; Gemmell and Westerman, 1994; Gemmell *et al.* 1994).

Table 1.2.1 The estimated separation dates (MYA) between different mammalian groups based on macromolecular sequence comparisons and DNA-DNA hybridisation.

Taxa compared				Basis of Comparisons	Reference
Marsupial/ Eutherian	Monotreme /other mammals	Monotreme/ Eutherian (or Marsupial)	Echidna/ Platypus		
124-154 104-170 85-139				α-globin β-globin Myoglobin	Air *et al.* 1971
144 136 81	186 114 90	171 112 93	52 38 28	α-globin β-globin Myoglobin	Whittaker *et al.* 1978
135			54	α-globin+ β-globin + myoglobin	Richardson, 1987
	110-137		62-78	DNA-DNA hybridisation	Westerman & Edwards, 1991
104-130	160		18-27	protamine p1	Retief *et al.* 1993
			25-30	mitochondrial 12S rDNA	Gemmell & Westerman, 1994
		115 (Marsupial)		mitochondrial coding sequences	Janke *et al.* 1997

Monotremes have been phylogenetically classified as a monophyletic group on the basis of synapomorphic characters such as the bill the spur on the inside of the hind foot, the detrahens muscle as the jaw adductor, and electroreceptor on the rhinarial

surface (Murray, 1984; Archer *et al.* 1993). However, the absence of echidna fossils make it difficult to deduce the time of divergence of modern monotremes from a common ancestor. The oldest known monotreme fossils are platypus-like rather than echidna-like in morphology. Of course, it can not be concluded from this observation that the extant platypus is more similar morphologically than the extant echidna to a common monotreme ancestor (Gregory, 1974; Murray, 1984; Archer *et al.* 1985, 1992, 1993; Pascual *et al.* 1992a,b; Augee and Gooden, 1993). Furthermore, no conclusions about the phylogenetic relationships of the monotremes can be drawn from this observation.

The separation of the Australian echidna and New Guinea echidna lineages is thought to have occurred in the mid to late Tertiary (1 MYA) (Griffiths *et al.* 1991), even though the oldest Australian echidna fossils are placed in the Pliocene (5 MYA) (Murray, 1984). Based on the fossil skulls of echidna and the molecular studies such as DNA-DNA hybridisation and parsimony analysis of mitochondrial 12S RNA sequences, the divergence of Australian and New Guinea echidna lineages may be as recent as 1-2 MYA (Watson, 1990; Griffiths *et al.* 1991; Westerman and Edwards, 1992; Gemmell *et al.* 1994).

The time of separation between the platypus and echidna lineages remains uncertain. Based on the fossil and DNA-DNA hybridisation data, the separation time is likely to be late Cretaceous or early Tertiary (Pascual *et al.* 1992a,b; Westerman and Edwards, 1992). However, analysis of the molecular data, derived from the protamine p1 gene and mitochondrial 12S RNA gene, suggest that the divergence could have been much more recent (Table 1.2.1) (Retief *et al.* 1993; Gemmell and Westerman, 1994). Most studies suggest that the Australian echidna and New Guinea echidna are last shared a common ancestor with the platypus 30-60 MYA (Retief *et al.* 1993; Gemmell and Westerman, 1994; Janke *et al.* 1996). Few of these studies, however, have included the platypus as well as both living echidnas.

Interestingly, based on the number of transversional differences in 12S mitochondrial DNA, Springer and Kirsch (1993) suggested that platypus and echidna separated more

recently than the rat and mouse. Many palaeontologists think that rats and mice diverged 8-12 MYA, but in contrast, molecular comparisons suggest a much older date, in the range from 20-35 MYA (Wilson *et al.* 1987). Many of the time estimates referred to in this section are based on the assumption of a correctly collaborated molecular clock. The existence and nature of such a clock, and the problems of its calibration, are discussed by Easteal *et al.* (1995).

1.2.4 Biochemical and haematological studies

The monotremes are frequently exposed, at least for short periods, to hypoxia and hypercapnia. Respiratory studies have shown that the echidna is physiologically well suited to burrowing. The oxygen consumption of echidnas is about half that expected from a similar sized placental mammal. Furthermore, the echidna seem able to withstand relatively high carbon dioxide concentrations (10-12% (v/v)) in the air (Bentley *et al.* 1967; Parer and Metcalfe, 1967a,b,c; Augee *et al.* 1971; Grigg *et al.* 1992; Bech *et al.* 1992).

Platypus and echidna red blood cells have the usual mammalian morphology in being non-nucleated, biconcave and disc-shaped. However, monotremes have relatively high levels of haemoglobin concentrations and their haemoglobin has a high oxygen carrying capacity (Parer and Metcalfe, 1967b; Parsons *et al.* 1971a,b; Whittington and Grant, 1983, 1984; Grigg *et al.* 1992; Bech *et al.* 1992). Parsons *et al.* (1971a,b) found that 90% of red blood cells in echidna contain alkali-resistant haemoglobin. Such haemoglobin is considered foetal in eutherian mammals but is found in high concentrations (up to 78%) in adult marsupials (Parer and Metcalfe, 1967b; Parsons *et al.* 1971a,b; Whittington and Grant, 1983, 1984; Grigg *et al.* 1992; Bech *et al.* 1992). The size and number of erythrocytes, together with the high haemoglobin level are likely to be adaptations to the hypoxic conditions experienced by monotremes during their time spent in burrowing and diving (Whittington and Grant, 1984; Grant, 1989a).

1.2.5 Chromosome studies

The DNA content of monotremes has been shown to lie within the normal range for eutherian mammals (Bick and Jackson, 1967a). Nevertheless, the study of monotreme chromosomes is difficult, for several reasons. Firstly, monotremes are very difficult to breed in captivity and are "protected" by law in the wild. However, a platypus cell line is available for *in vitro* chromosome study (Wrigley and Graves, 1988a). Secondly, the conditions for tissue and leucocyte culture necessary for obtaining good chromosome preparations, are very different to those of eutherian mammals (e.g. the monotreme cells must be cultured at 32°C). Thirdly, the genome of monotremes consists of many small size chromosomes which are virtually indistinguishable cytologically within or between species (Wrigley and Graves, 1988a,b). Fourthly, many of the "homologous chromosome" pairs are often heteromorphic, one chromosome being up to 25% longer than its partner due to a differential expansion or contraction of heterochromatic regions (Bick, 1992). Fifthly, satellited chromosomes are present in the platypus (chromosome 6S) and in the echidna (chromosomes 3S and 6S), and not only do the short arms of these satellited chromosomes contain different amounts of heterochromatin, but they can be differentially condensed in different tissues and in different cells of the same tissue of an individual at any one time (Wrigley and Graves, 1988a; Bick, 1992).

Monotreme chromosomes fall into two distinct size ranges: macrochromosomes and microchromosomes. The microchromosomes are morphologically similar to those of birds and reptiles but considerably larger than the typical microelements of birds. The macrochromosomes are regarded as "mammalian type" chromosomes (Bick and Jackson, 1967b; Bick and Sharman, 1975). It should be stressed that similar gross chromosomal morphology does not necessarily indicate genetic similarity.

The diploid chromosome number of the Australian echidna and New Guinea echidna is 64 in females and 63 in males. The New Guinea echidna chromosomes are almost identical in morphology (except for the presence or absence of a NOR-bearing satellite

chromosome 3) and G-band pattern to the Australian echidna chromosomes (Bick and Jackson, 1967a; Bick *et al.* 1973; Wrigley and Graves, 1988a). Bick and Sharman (1975) showed the diploid chromosome number of the platypus to be 2n=52 in both sexes.

The karyotypes of monotremes have several conserved features. The G-banding patterns show a considerable degree of conservation between species. Monotremes possess six large pairs of autosomes and many smaller autosomal pairs. On the basis of G-banding morphology, the six large pairs of autosomes are easily distinguishable from one another. However, many smaller autosomal pairs cannot be distinguished either within or between species. The three monotreme species also have a series of small unpaired chromosomes which participate together with the sex chromosomes in multivalent chain formation during male meiosis. Unpaired chromosomes a-d, in the platypus and a-f, in the echidnas, exist in both sexes (Murtagh, 1977; Wrigley and Graves, 1988a; Watson, 1990; Watson and Graves, 1992; Watson *et al.* 1992; Maccarone *et al.* 1992).

The chromosomal sex determination mechanism in monotremes is more complex than in most mammals where an XX female / XY male mechanism operates. The sex chromosomes in platypus are XX (female) and XY (male) which is a similar sex chromosome system to that of eutherians and marsupials, but the two echidna species show $X_1X_1X_2X_2$ (female) and X_1X_2Y (male) (Murtagh, 1977; Wrigley and Graves, 1988b; Watson *et al.* 1990, 1992; Watson and Graves, 1992). Like marsupials and eutherians, the Y chromosome in all three species of monotremes is found in males but not in females suggesting the Y chromosomes are involved in the major male sex-determination mechanisms. However, the function of the Y and its precise role in male determination is still debated (Watson *et al.* 1992). G-banding studies have revealed that the X chromosome has been highly conserved between the three extant monotremes. The platypus X, and X_1 of the two echidna species, are entirely homologous in their G-banding, which suggests that the X_1 from the two echidna X chromosomes (X_1 and X_2) is the original echidna X, and that the monotreme X chromosome has been largely conserved since the divergence of the platypus and

echidna lineage. The monotreme Y is much larger than its counterpart in eutherians and marsupials, and shares a region of cytological homology with the X chromosome which is unusual among mammals (Wrigley and Graves, 1988b; Watson, 1990; Watson and Graves, 1992; Watson *et al.* 1992; Bick, 1992).

Comparative studies of diverse chromosomal banding patterns and chromosomal mapping of homologous genes have provided information about the evolution of mammalian chromosomes (O'Brien and Graves, 1990). Wrigley and Graves (1984) have pointed out that a comparative study of gene arrangements in eutherians, marsupials and monotremes may provide an insight into the evolution of mammalian genome organisation and may even lead to an identification of the ancestral karyotype and gene map of mammals (Wrigley and Graves, 1984; Watson, 1990; Maccarone *et al.* 1992; Watson and Graves, 1992). Gene mapping has advanced rapidly, with continuing somatic cell genetic studies and *in situ* hybridisation, as well as the increasing use of DNA markers in linkage studies. Most physical gene mapping data from monotremes have been obtained using *in situ* hybridisation with heterologous probes.

Around 40 gene loci have been chromosomally mapped in the platypus, and six in the Australian echidna (Graves *et al.* 1993). No gene mapping has been undertaken in the New Guinea echidna. Four of the mapped loci in the Australian echidna are X linked, and the remaining two are autosomal. The four X-linked loci are in the same chromosomal position in the platypus and echidna (Graves and Watson, 1991). However, there are differences in the locations of the two autosomal loci in the different species. A ribosomal RNA gene, for example, has been localised to chromosome 6p in the platypus and to chromosomes 3p and 6p in the Australian echidna (Watson *et al.* 1991; Graves *et al.* 1993).

Graves *et al.* (1993) summarised various gene loci assignment in humans, marsupials and monotremes. The human β-like globin genes are closely syntenic to *CAT* (catalase) and *HRAS-1* (cellular homologue of the c-Ha-ras-1 oncogene) on the short arm of chromosome 11. These three loci are also syntenic and occur in the same order

along the chromosome 3q in the Tammar Wallaby (Marsupialia), *Macropus eugenii*. Of these three genes, only β-globin has been chromosomally mapped in the platypus - its location being 4q (Watson *et al.* 1991; Graves *et al.* 1993). The β-globin genes have not been mapped in the echidna. Therefore, mapping the β-globin genes on the echidna chromosomes may provide additional comparative data to investigate mammalian genome evolution. It may also enhance the opportunities to describe the chromosome changes and their relationships with speciation during mammalian evolution.

1.2.6 Monotreme globins

The amino acid sequences of myoglobin and of the α- and β-chains from the major adult-expressed haemoglobins have been determined in the platypus and echidna (Whittaker *et al.* 1972, 1973, 1978; Dodgson *et al.* 1974; Whittaker and Thompson, 1974, 1975; Fisher and Thompson, 1976; Castillo *et al.* 1978). No haemoglobin studies have been conducted in New Guinea echidna.

The amino acid sequences of platypus and echidna β-globins differ from eutherian adult-like β-globins at fewer sites (on average 24%) than do the marsupial adult-like β-globins (on average 28%). Conversely, the platypus and echidna α-globins differ from eutherian α-globins at more sites (on average 29%) than do marsupial α-globins (on average 26%). From this observation it was concluded that 'the amino acid sequences of the monotreme haemoglobins would not support any palaeontological estimate that gave a date for marsupial divergence from eutherians more recent than that for monotremes' (Whittaker *et al.* 1972; Whittaker and Thompson, 1974). However, this conclusion would be invalid if it was found that α- or β-globin genes did not evolve at a constant rate in the lineages compared. Furthermore, it is unlikely that the differences described are statistically significant.

As the platypus is semi-aquatic and the echidna is terrestrial, it would seem reasonable to anticipate some difference in haemoglobin characteristics such as oxygen dissociation curves, Bohr effects and the effect of 2,3-diphosphoglycerate. However,

the sequences of the α- and β-chains from the major haemoglobins of both animals show no change from the horse sequences or each other in the residues involved in the Bohr effect or in the binding site of 2,3-diphosphoglycerate, and also show high levels of cooperatively at various temperatures and pH values (Whittaker *et al.* 1972, 1973; Whittaker and Thompson, 1974; Hosken, 1978).

Using cellulose acetate electrophoresis, echidna red blood cells have been shown to contain two forms of haemoglobin, Hb I (major form, 80% of the total) and Hb II (minor form - 20% of the total) (Hosken and Boettcher, 1978), perhaps analogous to the Hb A and Hb A2 forms of haemoglobin found in human adults (Cooper *et al.* 1973; Hope *et al.* 1990). Oxygen affinity studies showed that Hb II had a markedly higher oxygen affinity and lower cooperatively than the Hb I. This higher affinity is thought to be related to a difference in the DPG interaction (Hosken and Boettcher, 1978; Griffiths, 1989). Comparisons between the Hb I and Hb II forms of echidna haemoglobin showed that the β-chains were identical, but the α-chains differed at 9 or 10 residues depending on the polymorphic allele involved (see below).

Using starch gel electrophoresis, polymorphic variants of echidna Hb I and Hb II were detected in a sample of 136 animals collected from different geographical locations in Australia (Cooper *et al.* 1973). Three common Hb I polymorphic variants designated Hb IA, Hb IB and Hb IC, and two Hb II polymorphic variants, Hb IIA and Hb IIB were observed. Hb IB was found mainly in eastern Australia, while Hb IA was found in central and western Australia. Polymorphic variants of Hb II were found only in Tasmania (Cooper *et al.* 1973; Thompson *et al.* 1973; Dodgson *et al.* 1974; Hosken and Boettcher, 1978).

The above findings are most simply interpreted by proposing the existence of at least two α-like globin loci in the echidna (Cooper *et al.* 1973; Thompson *et al.* 1973; Dodgson *et al.* 1974; Hosken and Boettcher, 1978; Hope *et al.* 1990). However, the molecular basis for the various forms of haemoglobin remains unresolved (Whittaker and Thompson, 1974, 1975; Whittaker *et al.* 1978; Hope *et al.* 1990).

Multiple haemoglobins have also been found in the platypus. Three different forms were detected by ion exchange chromatography (Whittaker and Thompson, 1974; Whittaker *et al.* 1978) and two by cellulose acetate electrophoresis (Cooper *et al.* 1973). However, the molecular basis for these variants also has not been investigated. Therefore, further research on the genetics and molecular biology of monotreme haemoglobins is required to determine the extent to which variant proteins result from inter- or intra-locus genetic variation.

Studies on monotreme globins take on an added interest and significance because the structure and function of these genes may well have evolved in concert with unique features of monotreme reproduction and lifestyle that impinge on respiration. These features include the delivery of oxygen to the embryo after birth by diffusion through the egg membrane; the requirement, at a very early stage of development, for lung function in a pouch where non-ambient levels of oxygen and carbon dioxide are likely to occur; and the oxygen stress placed on tissues during hibernation, burrowing and diving (Parsons *et al.* 1971a,b; Bentley *et al.* 1967; Whittington and Grant, 1983; Hope *et al.* 1990; Grigg *et al.* 1992; Bech *et al.* 1992). Comparative studies that include marsupial and monotreme haemoglobins are likely, therefore, to provide useful information on the relationships between haemoglobin structure, function and evolution. Little is known about the chromosome organisation of the haemoglobin genes in monotremes and there is a need for more research to determine the number, organisation, structure and regulation of these genes. Such research is important to understanding globin gene evolution, and will help to reconstruct the course of mammalian evolution.

1.3 MOLECULAR EVOLUTION AND PHYLOGENY

1.3.1 Molecular evolution and evolutionary clocks

Zuckerkandl and Pauling (1965) were the first to suggest that genes and their protein products might evolve at approximately constant rates over time, and that this rate constancy could be used to determine phylogenetic relationships among organisms (or genes) and for dating evolutionary events, particularly where there are insufficient data from fossils (Springer and Kirsch, 1989; Springer, 1990, 1995). Although a considerable body of evidence lends general support to the concept of a molecular clock, the idea remains controversial. Most of the debate on this topic is concerned with: i) methods of quantifying rates of evolution, ii) methods for calibrating the molecular clock, and iii) errors associated with molecular clock predictions of divergence times from measures of molecular similarity (Li *et al.* 1985; Wilson *et al.* 1987; Hope *et al.* 1990; Behe, 1990; Hillis *et al.* 1996).

There are two extreme views that attempt to explain both the levels of molecular polymorphism within species and the extent of molecular divergence between taxa. One is the neutral mutation theory of molecular evolution (Kimura, 1968). This theory is that most of the mutations that become fixed during evolution are "neutral" - i.e. they have little no effect on the function of the protein and the fitness of the organism. These changes become fixed in populations by random genetic drift (Wilson *et al.* 1977; Kimura, 1979, 1983). The other hypothesis is that mutational changes in protein sequence are either advantageous or deleterious and spread through populations by natural selection (Clarke, 1970; Richmond, 1970). Balancing selection is invoked to explain polymorphism while directional selection acting on advantageous substitutions explains their fixation (Perler *et al.* 1980; Felsenstein, 1988). Based on intra- and inter-specific comparisons in *Drosophila*, it has been suggested that the neutral theory can explain evolutionary changes at some but not all loci (Brookfield and Sharp, 1994).

DNA sequence comparisons and DNA/DNA hybridisation studies have revealed some examples where rates of molecular evolution are not constant in different taxa (Goodman *et al.* 1985; Wu and Li, 1985; Britten, 1986; Li and Tanimura, 1987; Li *et al.* 1987; Springer and Kirsch, 1989). For example, it has been claimed that rodent DNA has evolved faster than hominoid or bird DNA (Wu and Li, 1985; Goodman *et al.* 1985). Goodman and his colleagues (1985) have emphasised that rates of nucleotide substitution have slowed down during the evolution of higher primates and during hominoid descent. Using the "relative rate test", Wu and Li (1985) compared the coding regions of 11 genes from rodents and humans, using genes from either the goat or rabbit as outgroups. They concluded that there was a significantly faster rate of nucleotide substitution in the rodent lineage compared to in the human lineage and that the differences in substitution rate could be accounted for by the shorter generation times of rodents.

The cause of variation in rates of nucleotide and amino acid substitution is a central unresolved issue in molecular evolution (Kimura,1983; Li and Tanimura, 1987; Easteal, 1989, 1990; Easteal and Collet, 1994; Easteal *et al.* 1995). Various reasons for differences in substitution rates among mammalian lineages have been proposed, including: differences in metabolic rate, DNA repair efficiency, exposure to mutagens, nucleotide generation times, number of DNA replications in germ line cells, and organismal generation times (Wu and Li, 1985; Wilson *et al.* 1987; Sibley *et al.* 1988; Gaut *et al.* 1992; Hillis *et al.* 1996).

Easteal suggested that molecular evolutionary rates (including rates for β-globin genes) are generally uniform both among primates and between primates and other mammalian orders and there is stochastic uniformity of the molecular evolutionary rate among the eutherian orders (Easteal, 1985, 1989, 1990, 1991; Easteal and Collet, 1994; Easteal *et al.* 1995; Easteal and Herbert, 1997). He suggested that the previous studies, which indicated a rate slow down in the hominoid lineage (see above), were based on an incorrect interpretation of the poor fossil record. He concluded that more

extensive investigations were required to resolve the apparent inconsistencies in the interpretations by different workers (Easteal, 1991).

In test of the molecular clock hypothesis by sequence comparisons of orthologous protein coding genes, it is necessary to distinguish sites which evolve at different rates, or genes which are no longer under selective constraints. First, in comparisons of homologous DNA sequences from coding regions, it is always found that the DNA sequence divergence is greater than the amino acid divergence (Li *et al.* 1985). The excess divergence in the nucleotide sequences is due to synonymous or silent substitutions which are point mutations which do not alter the encoded amino acid. Second, comparisons of homologous DNA sequences show that introns diverge more rapidly than exons. This can be explained by the weaker selective constraints on introns than on coding regions (Czelusniak, *et al.* 1990; Mouchiroud and Bernardi, 1993). Thirdly, in sequence comparisons of orthologous genes and pseudogenes, it is always found that pseudogenes evolve more rapidly than their functional paralogues. Nucleotide substitutions, deletions and insertions accumulate in pseudogenes with few selective constraints, whereas substitutions in functional genes, unless they involve synonymous mutations, are constrained by natural selection (Li *et al.* 1981; Patterson, 1987).

Several methods have been proposed for estimating the numbers of synonymous and nonsynonymous substitutions (Miyata and Yasunaga 1980; Li *et al.* 1985; Nei and Gojobori, 1986; Li, 1993). Li *et al.* (1985) estimated the number of synonymous and nonsynonymous nucleotide substitutions among a large number of mammalian genes. They showed that the rates of nonsynonymous and synonymous substitutions are extremely variable among different genes. In general, the rates of synonymous substitutions are much greater than that of nonsynonymous substitutions. This variation in substitution rates is determined by two factors: i) the rate of mutation, and ii) the probability of fixation of mutation. Mutations that result in amino acid substitutions (nonsynonymous substitutions) have a higher chance of causing a deleterious effect on the function of the protein than synonymous changes. Consequently, the majority of nonsynonymous mutations will be eliminated from the

population by purifying selection. The result will be a reduction in the rate of substitution at nonsynonymous sites. In contrast, synonymous changes have a better chance of being neutral, and more of them will be fixed in a population but they may be constrained by such factors as tRNA availability (Ikemura and Ozeki, 1983; Ikemura, 1985) and the binding strength between codons and anticodons (Grosjean and Fiers, 1982; Li *et al.* 1985; Li and Graur, 1991).

1.3.2 Molecular phylogeny

Increasingly, DNA sequence data have been used for systematic and evolutionary studies. Once homologous protein or DNA sequences are identified in the species under study, they can be used to reconstruct the path of evolutionary descent of these species. This can be accomplished with computer algorithms which reconstruct evolutionary trees that account for the descent of sequences by the fewest number of base changes (Goodman *et al.* 1987b; Swofford and Olsen, 1990).

It is worth nothing that a gene tree is not necessarily the same as the phylogeny of the group of species from which the gene sequences were sampled (Nei, 1987; Pamilo and Nei, 1988). The chance that a gene tree and species tree differ is high when the gene used is polymorphic or when the species considered split relatively recently. In these cases, ideally, one must examine independently evolving genes to estimate the correct species tree (Saitou and Nei, 1986; Pamilo and Nei, 1988; Doyle, 1992). However, the study of the evolution of multigene families can provide very useful information (Miyamoto and Cracraft, 1991) on gene duplication, intragenic recombination, exon shuffling, horizontal transfer, and gene convergence. Such studies can also be used to test for heterogeneity in the rates of evolution amongst different branches of a phylogeny (Easteal, 1990), and to identify sequences subject to selection (Moritz and Hillis, 1996).

In re-constructing species phylogenies, it is necessary to distinguish between two type of homology, orthology (reflecting phylogeny of species) and paralogy (reflecting phylogeny of genes within those species) (Patterson, 1987; Goodman, *et al.* 1987b).

For deriving species trees, it is essential to use orthologous sequences, and it may be difficult, especially when using data from members of a multi-gene family, to distinguish orthologous from paralogous sequences. This distinction is made more difficult by the occurrence of events such as concerted evolution through gene conversion and unequal crossing over, because some portions of a pair of aligned sequences may have an orthologous relationship while others have a paralogous relationship. Therefore, concerted evolution is likely to confuse phylogenetic estimates of taxa or genes because sequences with many convergent residues will be depicted as sharing a more recent common ancestor than is in fact the case (Patterson, 1987; Goodman *et al.* 1987b; Moritz and Hillis, 1996). Goodman *et al.* (1987a) suggested that a way of overcoming these problem is to extend the alignment to a much larger number of sequence positions in order to increase the pool of variable positions and thereby lessens the chances of an unequal distribution of homoplasy among the contemporary species represented by the sequences.

There are many different tree-making methods for analysing molecular data (Nei, 1987; Felsenstein, 1988), and the methods, many of which remain controversial, have certain advantages and disadvantages, and have different relative efficiencies in recovering the "correct" phylogenetic tree. The relative efficiencies depend on various factors such as the shape of the true tree, transition/transversion ratios, and the numbers and rates of nucleotide substitutions (Miyamoto and Cracraft, 1991). One major problem in studying the relative efficiencies of different methods is that the true tree is usually unknown for any set of organisms or any set of DNA sequences, so that it is difficult to judge which tree is correct. Computer simulation can avoid some such problems (Miyamoto and Cracraft, 1991).

There are three main methods to find the optimal phylogenetic trees under the selected criterion applied to DNA and protein sequence data. First, with distance analysis, evolutionary distance is computed for all pairs of operational taxonomic units or DNA sequences. Sequence data is transformed into pairwise evolutionary distances based on an appropriate model of molecular evolution. The corrected distances are then taken to be estimates of the true evolutionary distance, which reflects the actual mean

number of changes per site that have occurred between a pair of sequences since their divergence from a common ancestor (Felsenstein, 1984, 1988; Saitou and Nei, 1987; Miyamoto and Cracraft, 1991; Swofford *et al.* 1996). Many of these methods (e.g. trees constructed using the UPGMA approach) produce a phenogram which could be interpretated as a phylogeny if the assumption of substitution rate constancy is made.

The second approach, the maximum likelihood method, evaluates a hypothesis about the evolutionary history of nucleotide substitutions. It is conjectured that a hypothesis with a higher probability of specifying the relationships between taxa under a particular model of sequence evolution is preferable to one with a lower probability (Felsenstein, 1981; Bishop and Friday, 1987; Saitou, 1988; Swofford *et al.* 1996). Models of sequence evolution used in maximum likelihood analyses can be manipulated to take account of differences in transversion and transition rates, rate variation at different sites and base composition.

The third approach is the method of maximum parsimony. In this method, the DNA sequences of ancestral species are inferred from those of extant species, considering a particular tree topology, and the minimum number of evolutionary changes that are required to explain all the observed differences among the sequences is computed. This number is obtained for all possible topologies, and the topology which shows the smallest number of evolutionary changes is chosen as the final tree (Cedergren *et al.* 1988; Miyamoto and Cracraft, 1991). The comparison of parsimony and likelihood methods show that parsimony methods seek solutions that minimise the amount of evolutionary changes required to explain the data, whereas likelihood methods attempt to estimate the actual amount of change according to an evolutionary model (Swofford *et al.* 1996).

Perhaps the best explanation for the principles underlying the maximum parsimony approach is provided by Goodman (1996):

> "Evolutionary theory predicts that in the descent of homologous sequences (either amino acid or nucleotide) from their common

> ancestors the sequence identities at homologous positions more frequently result from the absence of mutations at these positions in the descendant sequences sharing identities and, conversely, less frequently result from parallel or convergent back mutations. The maximum parsimony method finds that branching arrangement which in minimising change maximises perceived homologies amongst descendant sequences. Thus as long as the real homologies (identical sequences characters directly inherited in the real phylogeny from common ancestors) outnumber identities due to homoplastic (convergent, parallel, and back) mutations, the branching arrangement in the maximum parsimony tree should approximate that branching order that actually occurred during evolution, at least over most regions of the maximum parsimony tree".

Felsenstein (1985) proposed using the resampling test of bootstrapping to estimate level of support for internal branches in phylogenetic analysis. The bootstrap method creates pseudoreplicate data sets by randomly sampling the original character matrix with replacement to create new character matrices of the same size as the original (Felsenstein, 1985). The frequency of which a given branch is found in the minimal trees upon analysis of these pseudoreplicate data sets is recorded as the bootstrap proportion. The proportion can be used to estimate "confidence limits" of internal branches in the optimal phylogenetic tree and to interpret the probability that a phylogenetic estimate represents the true phylogeny (Helm-Bychowski and Wilson, 1986; Hillis and Bull, 1993; Hillis *et al.* 1996). It should be emphasised that bootstrap values are not confidence limits in the normally accepted statistical meaning of the word. They do, however, give some indication of the "robustness" of a phylogenetic tree to tolerating perturbations to the data used to construct the tree.

Using the various tree-making methods, an increase in the number of sequences considered may enhance the probability of obtaining the correct tree. However, there has been considerable controversy over whether different data sets should be analysed separately or combined in one analysis (Yang, 1996), and more research is needed to

solve this problem. The effects of different patterns of nucleotide substitution, such as varying rates among different sites, also needs further study. Statistical tests for topological differences between trees is an on-going area of research and debate (Miyamoto and Cracraft, 1991).

CHAPTER 2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Tissue samples

Australian	
Echidna tissue:	Blood was collected from an animal (Access No. 950528) at the Adelaide Zoo by Dr. D. Schultz, Chief Veterinary Officer. Liver tissue was collected from animal on Kangaroo Island by Dr. P. Rismiller, Pelican Lagoon Research and Wildlife Center.
New Guinea	
Echidna tissue:	Liver sample was obtained from Dr. S. Donnellan, Evolutionary Biology Unit, South Australian Museum.
Platypus tissue:	A liver sample was obtained from Dr. N. Gemmell, La Trobe University, and was stored in 70% ethanol. DNA was extracted from a line of cultured cells provided by Prof. J. A. M. Graves, La Trobe University.

2.1.2 Chemical reagents and other materials

Hybond-^{TM} N+	*Amersham Life Science*
Acrylamide and bis-acrylamide (electran grade), butan-1-ol, Gurr giemsa stain solution, silane A-174	*BDH*
Bio-Gel® P-60 resins (50~100 and 100~200 mesh)	*BioRad*
Biotinylated anti-streptavidin, ethidium bromide, FITC-streptavidin	*Boehringer Mannheim*
Blocking reagent, BCIP, NBT	*Bresatec Ltd*
Trypsin 1:250	*Difco*
Dulbecco's modified Eagle medium (DME), colcemid	*GIBCO BRL*
Phytohaemaglutinin A (PHA)	*Murex*
Repel-silane, sepharose ® CL-6B	*Pharmacia*
Agarose, IPTG, low melting Agarose, X-gal	*Progen*

Ampicillin, mineral oil (molecular biology reagent M-5904), BSA (fraction V), ficoll, phenylenediamide dihydrochloride (Sigma P1519), polyethylene glycol 8000, spermidine, triton X-100, tween 20	*Sigma Chemical Company*
Agar, tryptone and yeast extract	*Oxoid*

2.1.3 Enzymes

Agarase, DNAse I, Klenow DNA polymerase I, lysozyme, proteinase K, restriction endonucleases, RNAse A, T4 DNA ligase	*Boehringer Mannheim*
Taq polymerase	*Bresatec Ltd*
DNA polymerase I / DNase I mixture	*Gibco BRL*

2.1.4 Kits and special reagents

Sequenase™ version 2.0 DNA sequencing kit	*Amersham Life Science*
fmole ® DNA sequencing system	*Promega*
Wizard™ PCR preps DNA purification system	*Promega*

2.1.5 DNA

PUC 19	*Bresatec Ltd*
λ *Hind III* and *EcoRI* Spp-1 DNA size markers	*Progene*
Plasmid vectors, *PGM3zf*(+) and *PGM3zf*(-)	*Promega*

2.1.6 Nucleotides and radio-isotopes

α^{32}P-dATP	*Amersham Life Science*
Biotin-11-dUTP	*Boehringer Mannheim*
dNTP set	*Promega*

2.1.7 *E. coli* strains

E. coli KW251 (F-, supE44, galK2, melB1, hsdR2, mcrB1, mcrA-, argA81, TH10, recD1014)

E. coli JM109 (recA1, supE44, endA1, hsdR17, gyrA96, relA1, thiΔ (lac-proAB)

2.1.8 β-globin gene probes used in this project

Name of probe	Source of probe and reference	Insert sequence	Use
pSG2H (467 bp *Hind* III fragment)	ε-globin gene from *S. crassicaudata* Cooper and Hope (1993)	Contains the terminal (3') 12 bp from the first exon, the entire first intron (120 bp), the entire second exon (223 bp) and the first (5') 112 bp of the second intron	• screening an Australian echidna genomic library, EGL. 1.1 • Southern analysis of hybridising lambda clones, λEL. 1.1 and λEL.1 .2, obtained from EGL.1.1
pTAG-2 (856 bp *Eco*RI fragment)	Australian echidna β1-globin gene Robert Shroff (1990)	Contains 122 bp of the first intron, the entire second exon (223 bp), the entire second intron (465 bp) and the first (5') 54 bp of the third exon	• screening an Australian echidna genomic library, EGL. 1.2 • screening a platypus genomic library, PGL. 1.1 • Southern analysis of hybridising lambda clones, λEPL.1 (echidna), λPL. 1.1 and λPL. 1.2 (both platypus) • PCR amplification of β-globin sequences from New Guinea echidna genomic DNA

2.1.9 Buffers and solutions

Denaturing solution	1.5M NaCl, 0.5M NaOH
Neutralising buffer	1.5M NaCl, 0.5M Tris-HCl pH 7.2, 1mM EDTA

10x oligo labelling buffer	0.5M Tris-HCl pH 6.9, 1M $MgSO_4$, 1mM DTT, 0.6mM each of dTTP, dGTP and dCTP
Polynucleotide kinase buffer	50mM Tris-HCl pH 7.5, 10mM $MgCl_2$, 5mM DTT
1x Denhards solution	0.02% (w/v) Ficoll, 0.02% (w/v) BSA
L-broth	1.0% (w/v) Bacto-tryptone, 0.5% (w/v) Bacto-yeast extract, 1.0% (w/v) NaCl adjusted to pH 7.5 with NaOH
Phage buffer	20mM Tris-HCl pH 7.4, 100mM NaCl, 10mM $MgSO_4$
SM buffer	100mM NaCl, 10mM $MgSO_4$, 50mM Tris-HCl pH 7.5, 0.1% (w/v) gelatin
1x SSC	0.15M NaCl, 0.015M Na_3 citrate, 0.01M EDTA pH 8.0
1x SSPE	0.18M NaCl, 10mM NaH_2PO_4, 1mM EDTA pH 7.4
Stop buffer	0.1% (w/v) SDS, 10mM Tris-HCl pH 7.5, 1mM EDTA
1x TAE	0.4M Tris, 5mM NaAcetate, 1mM EDTA adjusted to pH 8.2 with acetic acid
1x TBE	89mM Tris, 89mM Boric Acid, 2mM EDTA, pH to 8.3 with HCl
1x TE	0.01M Tris-HCl pH 8.0, 5mM EDTA
10x ligation buffer	300mM Tris-HCl, pH 7.5, 100mM $MgCl_2$, 100mM DTT, 10mM ATP
0.1M IPTG stock solution	1.2g IPTG add H_2O to 50ml final volume. Filter-sterilise and store at 4°C
X-gal stock solution	50mg/ml stock in N,N'dimethyl formamide
LB agar	Add 15g agar to 1 litre of LB medium
TB top agar	1% (w/v) Bacto-Tryptone, 0.5% (w/v) NaCl, 0.8% (w/v) agar; autoclave to melt the agar and on cooling to 60°C add 1ml of 1M $MgSO_4$
LB/ampicillin/IPTG/X-gal plates	LB agar plates were supplemented with ampicillin (100µg/ml), IPTG (0.5mM) and X-gal (80µg/ml)
10x loading buffer for Agarose gels	0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene

	cyanol, 15% (w/v) Ficoll 400 in water
Cell culture medium	10% (w/v) foetal calf serum, 50µg/ml streptomycin sulfate, 60µg/ml benzyl penicillin, 100µg/ml glutamine in DME (Dulbecco's modification of Eagle's medium)

Reagents for chromosome studies and *in situ* hybridization

Buffer 1	15ml 1M Tris, 4.5ml 5M NaCl, 130.5ml dH_2O
Blocking solution	0.1g blocking reagent in 20ml Buffer 1
Buffer 2	5ml 1M Tris, pH 9.5, 1ml 5M NaCl, 2.5ml 1M $MgCl_2$, 41.5ml dH_2O
Staining solution	1:250, NBT: Buffer 2, 1:250, BCIP: Buffer 2
Sörenson phosphate buffer	Solution A: 9.08g KH_2PO_4 per litre, Solution B : 9.46g Na_2HPO_4 per litre. Add equal volumes of A and B to give pH 6.8
RNase treatment	100µg/ml RNase in 2x SSC
Probe solution	50% (w/v) formamide, 10% (w/v) dextran sulfate, 2x SSC, 250mg/ml salmon sperm DNA, 20ng/µl probe (Biotin- labelled) 0.125% (w/v) SDS in dH_2O
Washing solution I	2x SSC, 50% (v/v) formamide
Washing solution II	0.5% (w/v) SSC
Blocking solution	3% (w/v) BSA, 0.1% (w/v) Tween 20, 4x SSC
Nick translation mixture	1x nick translation buffer (0.5M Tris-HCl, pH 7.8, 0.05M $MgCl_2$, 0.5mg/ml nuclease-free bovine serum albumin), 0.5mM of each dCTP, dGTP, dATP in 10mM Tris-HCl pH 7.5, 0.65mM of dTTP, 0.4mM Biotin-11-dUTP, 1µl 100mM DTT
Detection buffer	1% (w/v) BSA, 0.1% (w/v) Tween 20, 4x SSC
Chromosome staining Solution (propidium iodide and or DAPI in antifade solution)	100mg p-phenylenediamide dihydrochloride in 10ml PBS, with 0.5M carbonate-bicarbonate buffer (0.42g $NaHCO_3$

	in 10ml adjusted to pH 9.0 with NaOH), add 90ml glycerol, filtered with 0.22µm membrane
Hypotonic solution	0.4% (w/v) KCl
Fixative solution	3:1 methanol / glacial acetic acid

2.1.10 Nucleotide sequences

Nucleotide sequences were obtained from the Gene Bank and EMBL data bases, which were accessed through the Australian Genome Information Service (ANGIS), Sydney University.

Species	β-globin genes	Locus ID.
Capra hircus (Goat)	β^F, β^A	CHHBGA, CHHBBAA
	β^C, ε^I	CHHBBB, CHEBGLI
	ε^{II}	CHEBGLII
	ε^V	CHHBBEVP
Didelphis virginiana (North American Opossum)	ε, β	OPOHBBE, OPOHBBB
Galago crassicaudatus (Primate)	ε, γ, δ and β	GCHBEGEB
Gallus gallus (Chicken)	ε, and β	GGHBBRE
Homo sapiens (Human)	ε, γ, δ and β	HSHBB
Mus musculus (Mouse)	ε^y, bh0, bh1, and β (minor and major)	MMBGCXD
Oryctolagus cuniculus (Rabbit)	ε, γ and β	OCBGLO01
Sminthopsis crassicaudata (Fat tailed Dunnart: Marsupialia)	ε	SCHBBEGN
	β	SCHBBHEMO

2.2 METHODS

2.2.1 Constructing monotreme genomic libraries

2.2.1.1 Preparation of high molecular weight genomic DNA

Liver tissue (0.1g) was homogenised in liquid nitrogen to make a fine powder. The powder was resuspended in 2ml of 10mM Tris-HCl pH 8.0, 1M EDTA pH 8.0, SDS solution was added to a final concentration of 1% (w/v) and RNase A was added to a final concentration of 20µg/ml. After incubation for 1 hour at 37°C, Proteinase K was added to a final concentration of 100µg/ml and the solution was incubated at 37°C overnight. The DNA was then extracted three times using phenol/chloroform and once with chloroform. Sodium acetate (pH 5.6) was added to a final concentration of 0.3M, followed by 2 volumes of ice cold 100% (v/v) ethanol and the mixture was left at -20°C for at least 1 hour. The DNA was spooled out, washed in 70% (v/v) ethanol, dried under vacuum and dissolved in distilled water (0.5~1µg/µl) for storage at -20°C (Biln and Stafford, 1976).

2.2.1.2 Partial digestion of genomic DNA with *Sau3A*

High MW DNA (100µg) was dissolved in 200µl of *Sau*3A enzyme digestion buffer and incubated for at least 30 min on ice. *Sau*3A was added to a final concentration of 0.04U/µg DNA and the solution was incubated at 37°C for 1 hour and then placed at 70°C for 15 minutes to inhibit further digestion.

2.2.1.3 Separation and recovery of partially digested DNA using a glycerol gradient

The 10~40% (v/v) glycerol gradient was made by adding 2.5ml each of 10% (v/v) and 40% (v/v) glycerol solution into a 5ml polyallomer tube. The gradient was left for 2 hours to equilibrate. Partially digested genomic DNA was loaded onto the glycerol

gradient and this gradient was centrifuged at 35,000 g for 17 hours at 10°C. After centrifugation, the base of the polyallomer tube was punctured with a 21 gauge needle and a volume of 100~200µl was collected as 10~20 fractions, each of 10µl, into 1.5ml Eppendorf tubes. Each fraction was mixed with 1µl of 10x loading buffer and electrophoresed through a 0.8% (w/v) agarose gel. The fractions which contained the 9~23 kb size range of DNA were selected and precipitated overnight with 3 volumes of cold ethanol at -70°C. The precipitated DNA was centrifuged at 13,000 rpm for 10 min at 4°C to pellet the DNA, the supernatant was discarded and the pellet was washed with 70% (v/v) ethanol and centrifuged for a further 5 min. The supernatant was removed and the pellet dried under vacuum. The pellet was then resuspended in 16µl of sterile water, 2µl of which was electrophoresed on a 0.8% (w/v) agarose gel to confirm the size range and concentration of DNA.

2.2.1.4 Library construction

The procedure followed was as specified in the *Promega* Technical Bulletin No. 005. A single colony of *E coli* KW 251 was cultured overnight at 37°C with vigorous shaking in 50ml of L-broth supplement with 0.2% (w/v) maltose and 10mM $MgSO_4$. Vector Lambda GEM-11 DNA (0.5µg) and 0.3~0.5µg of the 9~23 kb DNA were ligated in a final volume of 1µl of 1x ligation buffer, containing 1 Weiss Unit of T4 DNA ligase. Incubation was at room temperature for three hours. Phage packaging mix (50µl) was added to the entire ligation reaction and left at RT for two hours. The packaged phage suspension was diluted with 500µl of phage buffer and 25µl of chloroform and stored at 4°C. The packaged phage suspension was serially titred with phage buffer to make appropriate dilutions of the packaging extracts. Appropriately diluted recombinant phage (100µl) was allowed to adsorb to 100µl of *E. coli* KW251 at 37°C for 20 min and added TB top agarose at 45°C (5ml). This mixture was immediately poured onto LB plates. Plates were incubated overnight at 37°C.

2.2.1.5 Plaque lifting

The conditions for constructing plaques onto Hybond^{TM}-N+ membranes were as set out in the Protocols for Nucleic Acid Blotting and Hybridisation (*Amersham Life Science*). Hybond^{TM}-N+ membrane was placed on the surface of agar plates for one minute. The membranes and agar plates were marked with a needle for correct orientation. The membranes were removed and placed DNA side up for seven minutes on Whatman 3MM paper soaked with denaturing solution. The membranes were then transferred to 3MM paper (soaked with neutralising solution) for three minutes and the neutralising step was repeated once. Membranes were washed in 2x SSC and air-dried. DNA was fixed to the membrane by placing the membrane onto 3MM paper soaked in 0.4M NaOH for 20 min and washing in 2x SSC. The membranes were sealed in plastic bags and stored at 4°C.

2.2.1.6 Oligo-labelling of DNA restriction fragments

Oligo-labelling was performed according to the protocol of Hodgson and Fisk (1987). Probe DNA (200ng) and random hexamer primers (60ng/µl) were mixed and the volume made up to 12µl with H_2O. DNA was denatured by boiling for three minutes and then placed on ice. Oligo-labelling buffer (1x), 50 µCi of α^{32}P-dATP and Klenow enzyme (2U/µl) were mixed, centrifuged for 1~2 sec and the mixture was incubated at RT for two hours. Stop buffer was added to a final volume of 50µl to terminate the reaction.

2.2.1.7 Hybridisation of Hybond^{TM} N+ filters with radio-labelled DNA probes

Colony blotting of lambda phage, and Southern blotting, were done using the method described by Church and Gilbert (1984). Hybond^{TM}-N+ nylon membranes were pre-hybridised in 10~30ml of hybridisation solution for 1~3 hours in a hybridisation bottle at 65°C. Radio-labelled DNA probes were denatured at 100°C for five minutes and added to the hybridisation bottle. Hybridisation was carried out overnight at 65°C.

Filters were then rinsed in 2x SSC, 0.1% (w/v) SDS, washed twice in 1x SSC, 0.1% (w/v) SDS at RT for 10~20 min each and then 0.1x SSC, 0.1% (w/v) SDS at 37~45°C for another 15 min. Filters were then wrapped in plastic wrap and exposed to X-ray film, with and without an intensifying screen, at -70°C for the required exposure time. After autoradiography, filters were stripped of hybridising probe by boiling them for several minutes in 0.1x SSC, 0.1% (w/v) SDS. Filters were then either re-probed or stored at -20°C.

2.2.1.8 Column separation of unincorporated radionucleotides

A column was constructed using 150µl of 50~100 Biogel resin overlayed by 450µl of 100~200 Biogel resin in a 500µl Eppendorf tube. A hole was punched in the bottom of the tube with a 27 gauge needle. The column was placed into a 1.5ml of Eppendorf tube and centrifuged in a swing bucket centrifuge for 3 min at 1,000 rpm and washed through with 100µl of stop buffer. The solution containing the labelled probe was added to top of the column and the column was centrifuged at 1,000 rpm for 2 min. The probe which was collected in the outer 1.5ml Eppendorf tube was diluted by adding 50µl of stop buffer and then transferred again to the column, which was centrifuged at 1,000 rpm. Probe (100µl) was recovered from the outer tube. The specific activity of the probe was determined by performing a scintillation count on a 1µl aliquot in 1ml of scintillation fluid in a Beckman scintillation counter.

2.2.1.9 Preparation of lambda phage DNA

The method for isolating recombinant lambda phage DNA was as described in the Technical Bulletin No. 055 from *Promega*. A single recombinant phage plaque was picked from an agar plate by removing a plug of agar containing the plaque with a micropipette. The agar plug was gently expelled into a 1.5ml microcentrifuge tube containing 100µl of phage buffer. The tube was left overnight at 4°C to allow phage particles to diffuse into solution from the top agar. The lambda lysate was mixed with 300µl of an overnight culture of *E. coli* KW 251 cells, and incubated at 37°C for 20

min. The culture was used to inoculate 100ml of pre-warmed (37°C) LB supplemented with 10mM $MgSO_4$, and this was incubated at 37°C for 5~6 hours with vigorous shaking. Chloroform (500µl) was added to ensure all cells had lysed and the culture was incubated at 37°C for a further 15 min. The cell debris was pelleted by centrifugation at 5,000 rpm for 10 min and the supernatant was transferred to a sterile tube and stored at 4°C.

Recombinant phage DNA was isolated using a miniprep protocol from *Promega*. Pancreatic DNase I and RNase A, each at a final concentration of 1µg/ml, were added to the supernatant and the solution was incubated at 37°C for 30 min. NaCl and polyethylene glycol (PEG) 8000 were added to final concentrations of 1M and 10% (w/v), respectively and the mixture was left on ice for 1~3 hour. Phage particles were pelleted by ultracentrifugation at 15,000 rpm in a Beckman SW55i swing-out rotor for 20 min at 4°C. The phage particles were resuspended in 1ml of SM buffer by gently vortexing, and centrifuged at 15,000 rpm for 2 min to remove phage cell debris. Phage DNA was recovered after single extractions with phenol, phenol-chloroform (1:1) and chloroform-isoamyl alcohol (24:1). The phage DNA was precipitated by adding an equal volume of isopropanol and incubating at -70°C for 20 min. After centrifugation, DNA was washed with 1ml of 70% (v/v) ethanol, dried under vacuum, and resuspended in an appropriate volume of distilled water.

2.2.2 Characterisation of lambda clones and recombinant plasmids

2.2.2.1 DNA digestion with restriction endonuclease

Digestion reactions were carried out under conditions recommended by the manufacturers of restriction enzymes. In most cases, enzyme constituted no more than 1/10 of the total reaction volume. Reactions were stopped by the addition of an appropriate volume of 10x loading buffer.

2.2.2.2 Agarose gel electrophoresis

Electrophoresis was conducted on horizontal agarose gels 0.8~1.5% (w/v) prepared in 1x TAE buffer. Electrophoresis on large gels (15.5x20.3 cm) was carried out overnight at 25 V (20~30 mA). Electrophoresis of mini-gels used 80~100 V (60~120 mA) for 60~80 min. All gels were stained in 10µg/ml ethidium bromide for 5 to 10 min and photographed under UV (254 nm) light.

2.2.2.3 Southern transfer using the alkaline blot procedure

DNA fragments from agarose gels were transferred to Hybond^{TM}-N+ nylon membrane using the procedures recommended by the manufacturer (*Amersham*: Protocol for Nucleic Acid Blotting and Hybridisation), a modification of the procedure developed by Southern (1975). After electrophoresis, the ethidium bromide-stained gel was de-purinated in 0.25M HCl for 25 min, rinsed with distilled water and transferred to the capillary alkali blotting set-up which consisted of a tray filled with 0.4M NaOH and a platform with a wick made from the three sheets of Whatman 3MM paper soaked in 0.4M NaOH. The gel was placed on the wick and surrounded with Parafilm. The Hybond^{TM}-N + filter was placed on the gel making sure there were no air-bubbles between the gel and the filter. Three sheets of 3MM paper cut to the same size as the gel were placed on top of the membrane. The gel and apparatus were wrapped in plastic-wrap to prevent evaporation and left for at least four hours. After that, the filter was rinsed in 2x SSC, wrapped and stored at -20°C for hybridisation.

2.2.2.4 Purification of DNA fragments

DNA fragments to be used as probes and/or sub-cloned into plasmid vectors for further characterisation, were prepared as follows.

LMT Agarose gels

DNA samples were electrophoresed on horizontal low melting agarose gels (1~1.5% (w/v) at 60 V for 2~3 hours. The desired DNA fragments were excised, placed in a Eppendorf tube and incubated in a 60°C water bath for 10~20 min until the agar melted. DNA was purified using the Wizard™ PCR preps DNA purification kit *(Promega)*, using the vacuum manifold purification procedures from the manufacturers protocol. The aqueous agar was mixed with 1ml of resin and forced through a mini-column using an attached syringe. The column was washed with 2ml of 80% (v/v) isopropanol and dried by short spin at 13,000 rpm. The appropriate amount of distilled water was then added and the column left for 1 min at RT. The minicolumn was centrifuged for 20 seconds to elute the bound DNA fragment.

Freeze-squeeze

The "freeze-squeeze" technique for elution of DNA from agarose gel was modified according to Thuring *et al.* (1975). Agarose containing the desired band was excised from an agarose gel, placed in an Eppendorf tube, frozen overnight at -20°C and centrifuged at 13,000 rpm for 10 min. The supernatant was transferred to a new sterile tube and extracted with an equal volume of phenol/chloroform. DNA was precipitated with an equal volume of 5M ammonium acetate and 2 volumes of ethanol and pelleted by centrifugation. DNA was washed once in 70% (v/v) ethanol, vacuum dried and resuspended in H_2O.

2.2.3 Preparation of plasmid subclones

2.2.3.1 Ligation of DNA fragment into plasmid *PGM3zf*

Plasmid vector, *PGM3zf*(+) or *PGM3zf*(-) (*Promega*) and DNA to be cloned were each cleaved with the same restriction enzyme. The insert DNA was prepared by the methods described above. After restriction digestion, plasmid DNA was

dephosphorylated with 1U of alkaline phosphatase in 1x alkaline phosphatase buffer at 37°C for 30 min. The DNA was extracted once with phenol/chloroform and precipitated. Ligation were performed by mixing 0.1µg of the digested *PGM3zf*(+) DNA and 0.3µg of the cut insert DNA in a total reaction volume of 20µl 1X ligase buffer. T4 DNA ligase (1U) was added and the mixture was incubated at 14°C for 3~4 hours.

2.2.3.2 Preparation of competent *E. coli* JM109 cells

Competent cells were prepared using the calcium chloride method (Sambrook *et al.* 1989). L-broth (50ml) was inoculated with 2ml of an overnight culture of *E. coli* JM 109 cells and the culture incubated with vigorous shaking at 37°C until an O.D. of 0.3 to 0.5 (at 600 nm) was reached. The bacteria cells were pelleted by centrifugation and resuspended in 25ml of ice cold 0.1M $MgCl_2$ solution. The cells were then centrifuged at 4,000 rpm for 10 min at 4°C and resuspended in 12.5ml of ice-cold 0.1M $CaCl_2$. After incubation on ice for 20 min, the cells were again collected by centrifugation at 4,000 rpm for 10 min at 4°C and resuspended in 1.6ml of ice-cold 0.1M $CaCl_2$ and 15% (w/v) of glycerol. The cell suspension was aliquoted into 200µl lots in Eppendorf tubes and stored at -80°C.

2.2.3.3 Transformation of *E. coli* with plasmid DNA

Competent cells (200µl) were mixed with 0.1µg of the ligated plasmid vector and held on ice for 30 min with occasional mixing. The mixture was incubated at 42°C for 2 min and placed back on ice for 30 min. LB medium (1ml) was added to the transformed cells and the mixture was incubated at 37°C for 60 min. The transformed cells were plated onto 1% (w/v) LB agar plate supplemented with 25mg/ml ampicillin, 10mg/ml IPTG, and 40mg/ml X-gal. The plates were incubated at 37°C overnight.

2.2.3.4 Preparation of plasmid DNA

This protocol was a modification of the alkaline method of Sambrook *et al.* (1989). A single colony of the appropriate host bacterial strain, which contained the ligated plasmid, was inoculated into 10ml of LB media containing 100µg/ml ampicillin. The culture was incubated at 37°C overnight with vigorous shaking. A 1.5ml aliquot of this culture was centrifuged at 13,000 rpm in an Eppendorf tube. The supernatant was discarded and pelleted bacterial cells were resuspended in 200µl of GTE solution (50mM glucose, 25mM Tris-HCl pH 8.0 and 10mM EDTA pH 8.0) by vigorous vortexing. A freshly prepared solution in volume of 300µl (0.2M NaOH, 1% (w/v) SDS) was added and the mixture was incubated for 5 min at RT. 300µl of 3M sodium acetate (pH 4.6) was then added and the mixture was incubated for a further 5 min and then centrifuged at 13,000 rpm for 10 min. The supernatant was transferred to a fresh tube and extracted once with phenol/chloroform. DNA was precipitated with 1 volume of isopropanol and harvested by centrifugation at 13,000 rpm for 15 min at 4°C. The pellet was washed with 70% (v/v) ethanol, dried in a vacuum desiccator and resuspended in 20µl of distilled water.

2.2.3.5 Rapid boiling plasmid mini preparations

Bacterial cells from the overnight culture were transferred to a 1.5ml Eppendorf tube and centrifuged at 13,000 rpm for 1 min. The supernatant was flicked off and the pellet was resuspended with 200µl of STET buffer (0.1M NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA and 5% (v/v) Triton X-100). Freshly prepared lysozyme (10µl at a concentration of 25mg/ml) was added and the mixture was heated 100°C for 45 sec. Following centrifugation at 13,000 rpm for 15 min, the supernatant was transfer to clean Eppendorf tube and the same DNA harvesting steps as outlined in Section 2.2.3.4 were followed.

2.2.4 DNA sequencing

2.2.4.1 Purification and denaturation of plasmid DNA

The plasmid DNA for sequencing was prepared from either of the methods described in Sections 2.2.3.4 or 2.2.3.5. The DNA was denatured with alkali solution (1M NaOH and 1mM EDTA) and then purified using a Sepharose® CL-6B column. RNase A (2U) was added 3~5μg plasmid DNA in a volume of 20μl, and incubated at 37°C for 15 min. Alkali solution (5μl) was added and the solution incubated again at 37°C for 20 min. One drop of glass beads was added to cover the hole in the bottom of a 0.5ml tube which was made using a 26 gauge needle, and the tube was filled with Sepharose® CL-6B column which had been equilibrated in 1mM Tris-HCl (pH 8.0), 1mM EDTA, 0.2% (w/v) sodium azide. The 0.5ml microfuge tube was placed inside a 1.5ml Eppendorf tube and centrifuged three times at 1,500 rpm for 4 min to dry the column. The denatured DNA was added to the column which was centrifuged at 1,500 rpm for 4 min. The eluate contained the purified single stranded plasmid DNA used for sequencing.

2.2.4.2 DNA sequencing with the *Sequenase* kit

Sequencing the purified plasmid DNA followed the manufacture instructions (Sequenase™ version 2.0 DNA sequencing kit, *Amersham*). After the reaction had finished, the mixtures were denatured for 2 min at 90°C and 3μl of each sequencing DNA sample was loaded into the well of denatured 5% (w/v) polyacrylamide gel. Electrophoresis was carried out using a Model S2 sequencing gel electrophoresis system (*Bethesda Research Laboratories*) at 1,800 V (30~50 mA, 100 W) in 1x TBE. After electrophoresis, the gel was fixed in a solution of 20% (v/v) methanol, 10% (v/v) glacial acetic acid for 20 min. The gel was dried at 65°C for 2 hours and then exposed to X-ray film overnight at RT.

2.2.4.3 DNA sequencing with the *fmol* sequencing kit

The purified PCR products was directly sequenced with fmol DNA sequencing system (*Promega*). The sequencing protocol followed the manufacturer's description. The primer and the PCR condition for sequencing were the same as those used for the PCR reactions (Table 2.2.1).

2.2.5 Polymerase Chain Reaction

2.2.5.1 Synthesis and processing of oligonucleotide PCR primers

PCR primers were synthesised on 30mM columns using the *Beckman* Oligo 100 DNA synthesiser. The raw primer solutions recovered after synthesis were purified as follows: Butanol (10 volumes) was added to 120µl of the raw primer solution and the mixture was vortexed for 15 seconds and then centrifuged at 13,000 rpm for 5 min at RT. The supernatant was removed and the DNA pellet vacuum dried and resuspended in 120µl of dH_2O.

2.2.5.2 PCR primer design

The primers were designed to be 16~24 nucleotides in length, with 40~60% GC content and matching annealing temperatures. The primers for plasmid sequencing were designed on the basis of previously sequenced nucleotide sequences. All primer sequences are listed in Appendix 1.

PCR amplification of β-globin sequences from monotreme genomic DNA used three different sets of primers. The PCR primer sequences were designed on the basis of monotreme β-globin genes: i) B F1/Lee R primers from platypus β1-globin gene, ii) Lee F2/LeeR (platypus β2-globin gene), and iii) 707 F/737 R primers were designed on the basis of Australian echidna β1-globin gene and used to amplify the New Guinea echidna β1-globin gene (Table 2.2.1).

Table 2.2.1 PCR primer sequences and conditions for PCR amplification of monotreme β-globin genes.

Forward Primer	Reverse Primer	Use of Primers	PCR Conditions
i) B F1 5'-AAAGTGAACATCAATGAGTTG-3'	**Lee R** 5'-CCACAAGTACCACTGA-3'	Amplification of Platypus β1-globin gene	95 °C 0.7 min 55 °C 1 min 72 °C 2 min 34 CYCLES
ii) Lee F2 5'-CATGTGAACGTCAATGAACTC-3'	**Lee R** 5'-CCACAAGTACCACTGA-3'	Amplification of Platypus β2-globin gene	95 °C 0.7 min 55 °C 1 min 72 °C 2 min 34 CYCLES
iii) 707 F 5'-GGCCAATCAGCTCTGAG-3'	**737 R** 5'-GCCTACTCCCGCAGTAC-3'	Amplification of New Guinea Echidna β1-globin gene	95 °C 0.7 min 63 °C 1 min 72 °C 2 min 34 CYCLES

2.2.5.3 RCR conditions

PCR amplifications were performed using an automatic A.R.N. electronic temperature controller. The PCR reaction mix contained 2~5mM $MgCl_2$, 0.25mM of each dNTP (dATP, dCTP, dGTP, and dTTP), 1x *Taq* polymerase reaction buffer, 0.5μg of each primer, 0.5~2μg of template DNA, 1U of *Taq* polymerase (*Bresetec*), and distilled water to a volume of 50μl. A drop of mineral oil (*Sigma*) was added to the PCR reaction mix to prevent evaporation. Samples were heated for 3 min at 95°C and then cycled using the specified conditions. Negative controls contained all of the reagents mentioned above with the exception of template DNA.

2.2.5.4 Purification of PCR products

The PCR reaction products were purified using the Wizard™ PCR preps DNA purification kit (*Promega*). Conditions were as described by the manufacturer. The purification buffer (100μl) was added to PCR products and the same vacuum manifold purification procedures were followed as described in Section 2.2.2.4.

2.2.6 Chromosome preparation and *in situ* hybridisation

2.2.6.1 Echidna blood cell culture and chromosome preparation

Echidna blood was taken from the fore limbs of an animal (Access No. 950528) at the Adelaide Zoo. Blood samples (2~3ml) were transferred into heparinised tubes and subsequently used for blood cell culture. Blood (0.5~1ml) was added to cell culture flasks containing 10ml of Dulbecco's modification of Eagle's medium supplemented with 10% (v/v) foetal calf serum, 60µg/ml benzyl penicillin, 50µg/ml streptomycin sulphate, 100µg/ml glutamine and 10µg/ml phytohaemaglutinin A. The cells were cultured in a CO_2 incubator at 32°C for 3 days. After 3 days, colcemid (10ng/ml) was added to the culture and after an additional 2.5 hours incubation, cells were harvested by centrifugation and resuspended in 6ml of pre-warmed (37°C) hypotonic solution (0.4% (w/v) potassium chloride) and incubated for 30 min at 34°C. The cells were prefixed by resuspending in 1ml of freshly prepared ice cold fixative solution (methanol-glacial acetic acid, 3:1 (v/v)) and centrifuged at 1,200 rpm for 10 min. The cells were then washed three times with fixative solution and left in fixative overnight at -20°C. Chromosomes were spread on cleaned glass microscope slides by applying 2~3 drops of cell suspension to each slide from a height of 15~20cm. Slides were air dried and placed in slide boxes for storage at RT.

2.2.6.2 Preparation of probe DNA for fluorescent *in situ* hybridisation

The probe was biotin labelled using nick translation (Leitch *et al.* 1994). 1U of DNA polymerase I / DNase I were mixed with 0.5µg of probe DNA and the volume made up to 50µl with distilled water. The reaction mixture was incubated at 15°C for 2 hours and the reaction was stopped by addition of 2.4µl of 0.5M EDTA. To precipitate the labelled DNA, 75µl of 100% (v/v) ethanol and 5.5µl of 2M LiCl were added and incubated overnight at -20°C. The DNA was pelleted by centrifugation, washed with 70% (v/v) ethanol, dried in vacuum and resuspended in 20µl 1x TE.

Biotin labelled rDNA probe for use as a control was provided by Andreas Houben, Department of Genetics, University of Adelaide.

2.2.6.3 Estimation of labelling efficiency

After nick translation, the biotin labelled DNA was checked for incorporation of the labelled nucleotide using a spot test. Then 0.5µl each of biotin labelled probe and control biotin labelled DNA (*Bresatec*) were spotted onto Hybond^{TM}-N+ membrane. The membrane was incubated at 80°C for 1 hour and washed in Buffer 1 solution (0.1M Tris-HCl pH 7.5 and 0.15M NaCl) at RT for 5 min with shaking. The membrane was transferred to blocking solution and incubated for 30 min. The membrane was washed twice in Buffer 1 (5 min each) and incubated for 30 min with shaking in streptavidin-alkaline phosphatase (diluted 1:5000 in Buffer 1). The membrane was then washed in Buffer 2 solution (0.1M Tris-HCl pH 9.5, 0.1M NaCl and 0.05M $MgCl_2$) for 10 min and stained in staining solution (1:250, NBT, 1:250, BCIP in Buffer 2) for 10~20 min in the dark. After colour development, the filter was washed with distilled water and dried in air.

2.2.6.4 Chromosome pre-treatment

200µl of RNase A (0.1mg/ml in 2x SSC) was dropped onto the air-dried slide containing the chromosome spreads and incubated at 37°C for 90 min in a 2x SSC moisture chamber. Slides were washed four times (each 2 min) in 2x SSC at room temperature with shaking. Slides were dehydrated in 70%, 90% and 96% (v/v) ethanol at RT for 3 min each and air-dried.

2.2.6.5 Denaturation of chromosomal DNA and probe hybridisation

Slides were placed in 70% (v/v) deionised formamide in 2x SSC solution at 70°C for 3 min, and dehydrated in the ice-cold ethanol series 70%, 90%, 100% (v/v) for 3 min each with gently swirling. Slides were then dried in air. Biotin labelled probe was mixed with hybridisation mixture, 50% (v/v) deionised formamide, 10% (w/v) dextran

sulphate, 2x SSC, made up to 15µl with distilled water. The probe DNA (50ng) was denatured for 10 min at 80°C and placed immediately on ice. Denatured mixture (10µl) was added to each slide, covered with a coverslip and sealed with rubber cement. The slides were incubated overnight in a 2x SSC moisture chamber.

2.2.6.6 Post-hybridisation washing

After hybridisation, slides were first washed at 42°C in solution I (50% (v/v) deionised formamide and 2x SSC) and secondly at 60°C in solution II (0.5% (w/v) SSC) each three times for 5 min, to remove unspecific probe binding. 40µl of blocking solution (3% (w/v) BSA, 0.1% (v/v) Tween 20, 4x SSC) was added to each slide and incubating was carried out at 37°C for 30 min in a 2x SSC moisture chamber.

2.2.6.7 Detection and amplification of hybridised biotinylated probe

For detection of biotin-labelled probes, the biotin-FITC-streptavidin system was used. Slides were loaded with 20µl of 20µg FITC-streptavidin per 1ml of detection buffer (1% (w/v) BSA, 0.1% (v/v) Tween 20, 4x SSC) and incubated in the dark for 60 min at 37°C in a 2x SSC moisture chamber. Slides were washed in the dark in 4x SSC, 0.1% (v/v) Tween 20 at 42°C three times for 5 min each. Anti-streptavidin was used to increase the FITC signal. Slides were treated with 20µl of 20µg biotinylated anti-streptavidin in 1ml of detection buffer and incubated for 30 min at 42°C in a 2x SSC moisture chamber in the dark and washed in 4x SSC, 0.1% (v/v) Tween 20 at 42°C, three times for 5 min each. The detection step was then repeated as above, but for an incubation period of 30 min only.

2.2.6.8 Chromosome staining

Chromosomes were counterstained with propidium iodide (1µg/ml) and/or DAPI (1µg/ml) in anti-fade (glycerol-para-phenylene-diamine mixture: 1mg/ml p-phenylenediamine in 1mM sodium phosphate, 15mM NaCl, 50% (v/v) glycerol). The

chromosomes were covered with a coverslip at 4°C for 30 min. Hybridisation signals on the chromosomes were visualised using a fluorescence microscope with excitation filter system (UV filter for DAPI and blue filter for FITC signal) and epifluorescence signals were recorded on Fuji film 400 ASA and stored on the computer of photometrix program using the photometrix series 200 CCD camera system.

2.2.6.9 Chromosome G-banding after *in situ* hybridisation

After screening and recording the *in situ* hybridisation signal, chromosomes were trypsin G-banded. Slides were washed three times 5 min each with 4x SSC, 0.1% (v/v) Tween at RT and dried at 60°C for 2 hours. The slides was treated with 0.028% (w/v) trypsin solution in Sörensen buffer for 0.8~2 min, washed with 1x PBS and then stained with 10% (v/v) Giemsa staining solution for 1 hour.

2.2.7 DNA Sequence analysis

Searches for eukaryotic transcription elements were carried out using the program SIGSCAN (Prestridge, 1991). The compilation and comparison of nucleotide sequences was performed using various STADEN programs (Staden, 1980, 1984) and the GCG sequence analysis software package version 8 (The University of Wisconsin Genetics Computer Groups, Madison). Multiple sequence alignments were carried out using the CLUSTALW program (Higgins *et al.* 1988). BLASTN and FAST were used to search nucleotide sequence databases. All programs were accessed through ANGIS (Australian Genome Information Service), Sydney University.

2.2.8 Phylogenetic analysis

2.2.8.1 Relative rate test and synonymous/non-synonymous substitutions

Nucleotide substitution rates for β-globin genes were calculated using relative rate test (Wu and Li, 1985). Rates of non-synonymous substitution per site at non-degenerate

and two fold degenerate sites, and synonymous substitutions per site at two-fold and four-fold degenerate site, were estimated using the procedure of Li (1993).

2.2.8.2 Phylogenetic tree construction

Phylogenetic analyses were carried out using computer programs in the PHYLIP package (Felsenstein, 1993), the PAUP 3.1 (Swofford, 1993), the test version PAUP(*) and the MEGA program (Kimura *et al.* 1993).

Phylogenetic analysis based on coding region of β-globin genes was performed using maximum parsimony, maximum likelihood and distance methods. PAUP 3.1, a test version of PAUP(*) and PHYLIP (DNAPARS) were used for the maximum-parsimony analysis. The PHYLIP (DNAML) package (Felsenstein, 1993) and a test version of PAUP(*) were used to perform maximum-likelihood analyses. The program MEGA (Kimura *et al.* 1993) and the test version of PAUP(*) were used to generate distance trees. Bootstrap analyses, based on 500-1000 replications, were carried out to provide estimates of the confidence that can be place in various nodes of the maximum-parsimony and distance trees.

CHAPTER 3 RESULTS

3.1 CHARACTERISATION OF β-GLOBIN GENES IN MONOTREMES

3.1.1 Identification of globin clones in echidna genomic DNA libraries

Two echidna genomic DNA libraries were available for use. One (EGL.1.1) was made by R. A. Shroff (1990) from DNA obtained from Dr. D.L. Hayman, Department of Genetics, University of Adelaide. The source of the DNA was an animal from Kangaroo Island, South Australia. The other library (EGL.1.2) was made by C. L. Marchant (1994). Both genomic libraries were made using *Sau*3A partially digested, size- selected (9-15 kb) DNA ligated to the vector lambda phage *Gem*-11™ (*Promega*). Shroff isolated two positive clones named as λEL.1.1 and λEL.1.2 from EGL. 1.1 using a β-globin gene probe, pSG2H, from *S. crassicaudata* (Section 2.1.8). The two clones, which had 15 kb and 13.6 kb sized inserts respectively, showed identical restriction patterns suggesting that they were derived from the same β-globin region. Marchant (1994) isolated several positive lambda clones from EGL.1.2 but analyses of these clones showed that they did not contain β-globin sequences.

Re-screening of the EGL.1.2 genomic library was undertaken and one lambda clone, called λEPL.1 (Fig. 3.1.1) was isolated and characterised with six different restriction enzymes (Fig. 3.1.2 A) and Southern analysis was carried out using pTAG-2 as a probe (Fig. 3.1.2 B).

Information about the various β-globin gene probes used in this project is given in Section 2.1.8 (page 55).

3.1.2 Southern analysis and restriction mapping of the echidna clones λEL.1.1 and λEPL.1

Southern analysis (using the marsupial β-globin gene probe pSG2H) and restriction mapping of λEL.1.1 were conducted by Shroff (1990). The 4 kb *Bam*HI and 856 bp

*Eco*RI restriction fragments which hybridised to the probe were subcloned into the plasmid vector *PGM3zf*(-) (*Promega*) and were named pTAG-3 and pTAG-2, respectively. The 1.6 kb *Bam*HI (pTAG-4) and 0.5 kb *Bam*HI / *Xba*I (pTAG-5) fragments were also subcloned for further characterisation (Shroff, 1990) (Fig. 3.1.3). pTAG-3 was digested with *Sac*I and *Sma*I to give 0.35 kb *Bam*HI/*Sac*I, 0.85 kb *Sac*I/*Sac*I, 1.4 kb *Sac*I/*Sma*I and 1.3 kb *Sma*I/*Bam*HI fragments. These fragments were subcloned into *PGM3zf*(-) and the subclones used for sequencing. Figure 3.1.3 shows the restriction enzyme map of λEL.1.1 and the locations of each subcloned plasmid. pTAG-3 contains three exons and two introns of the echidna β-globin gene (Fig. 3.1.3).

λEPL.1, a positive clone from the EGL.1.2 echidna library, containing an insert of about 16 kb, was digested with five restriction enzymes that recognise specific 6 base sequences. These digestions were carried out with individual enzymes and with enzyme pairs. The digested DNA samples were electrophoresed on a 0.8% (w/v) agarose gel and transferred into a Hybond-N+ membrane for Southern analysis using pTAG-2 as a probe. The results are shown in (Fig. 3.1.2). The restriction digestion patterns of λEPL. 1 "overlap" with those of λEL. 1.1 (Fig. 3.1.4). However, λEPL.1 contains a longer 5' flanking region of the putative echidna adult β-globin gene than λEL.1.1 (Fig. 3.1.4).

3.1.3 Construction of a platypus genomic library and identification of globin clones

Two platypus genomic libraries were made, each from a different genomic DNA source. The first library, named as PGL.1.1, was made from DNA extracted from cultured cells obtained from the laboratory of Prof. J. A. M. Graves, La Trobe Univeristy. The second library, named as PGL.1.2, was made from DNA extracted from liver tissue provided by Dr. N. Gemmell, La Trobe University. In each case genomic DNA was partially digested with *Sau*3A and ultracentrifuged in a 10-40% (v/v) glycerol gradient solution. After centrifugation, the base of the centrifuge tube was punctured with a 21 gauge needle and 19 fractions (each 200 μl) were collected in 1.5 ml Eppendorf tubes.

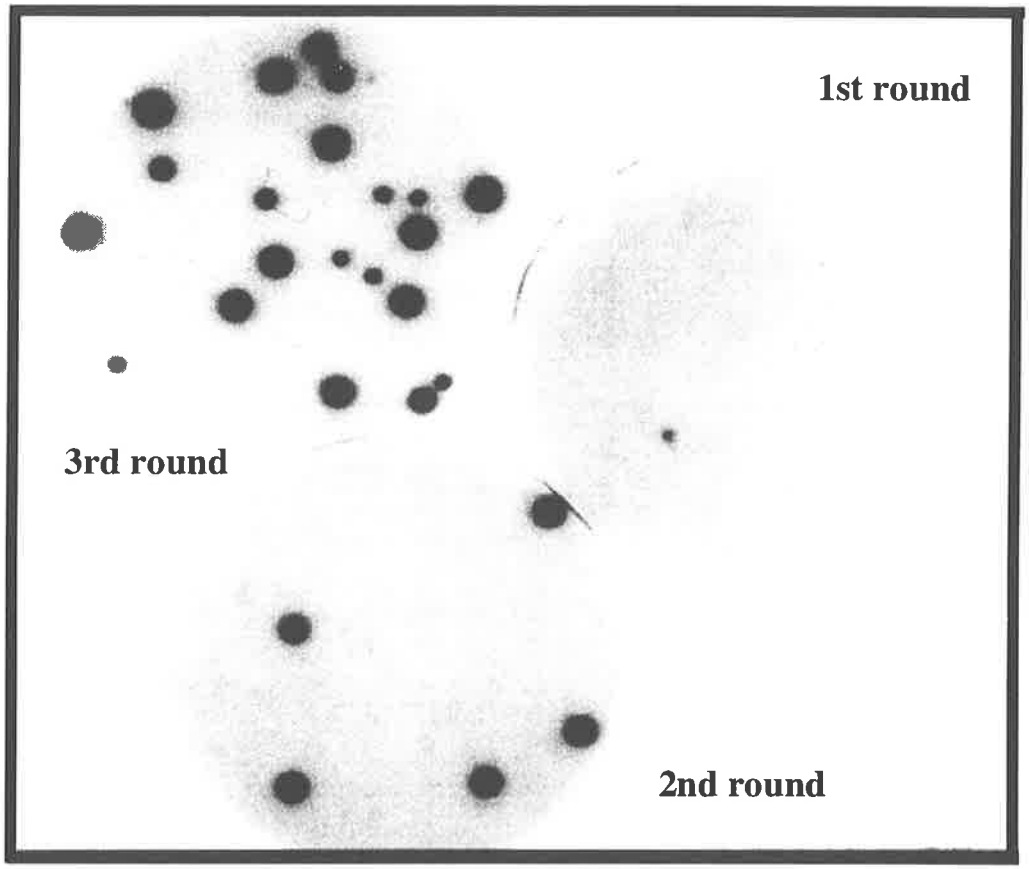


Figure 3.1.1 Autoradiographs showing hybridisation of the probe pTAG-2 to clone λEPL.1 in the echidna library EGL.1.2. Three rounds of consecutive low density plating achieved a pure culture of the lambda clone.

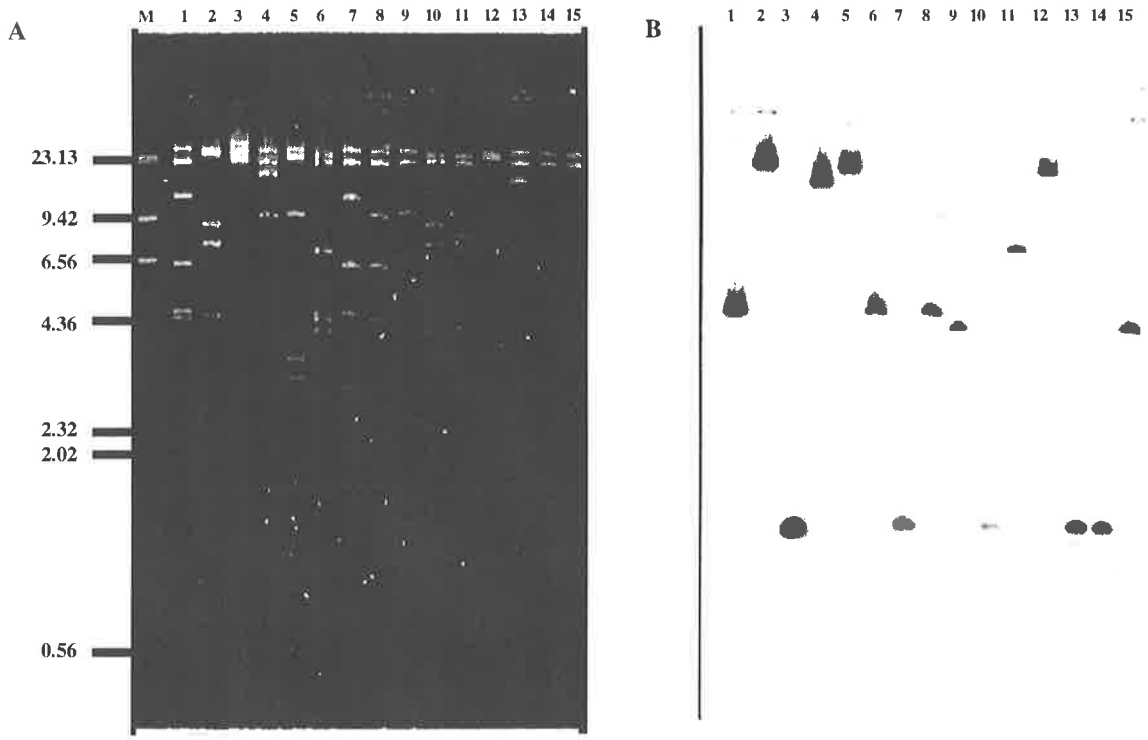


Figure 3.1.2 Restriction digestion and Southern analysis of λEPL. 1.

A: A agarose gel (0.8 % (w/v)) showing restriction fragments of λEPL.1. The tracks are: M- λ*Hind*III DNA size markers 1- *Bam*HI, 2- *Hind*III, 3- *Eco*RI, 4- *Sal*I, 5- *Xba*I, 6- *Bam*HI/*Hind*III, 7- *Bam*HI/*Eco*RI, 8- *Bam*HI/*Sal*I, 9- *Bam*HI/*Xba*I, 10- *Hind*III/*Eco*RI, 11- *Hind*III/*Sal*I, 12- *Hind*III/*Xba*I, 13- *Eco*RI/*Sal*I, 14- *Eco*RI/*Xba*I, 15- *Sal*I/*Xba*I.

B: Autoradiograph of the Southern filter taken from the gel shown in A. The probe was pTAG-2, the 856 bp *Eco*RI fragment of λEL.1.1. Lengths of hybridising fragments were estimated by comparing the autoradiograph to a photograph of the original agarose gel stained with EtBr.

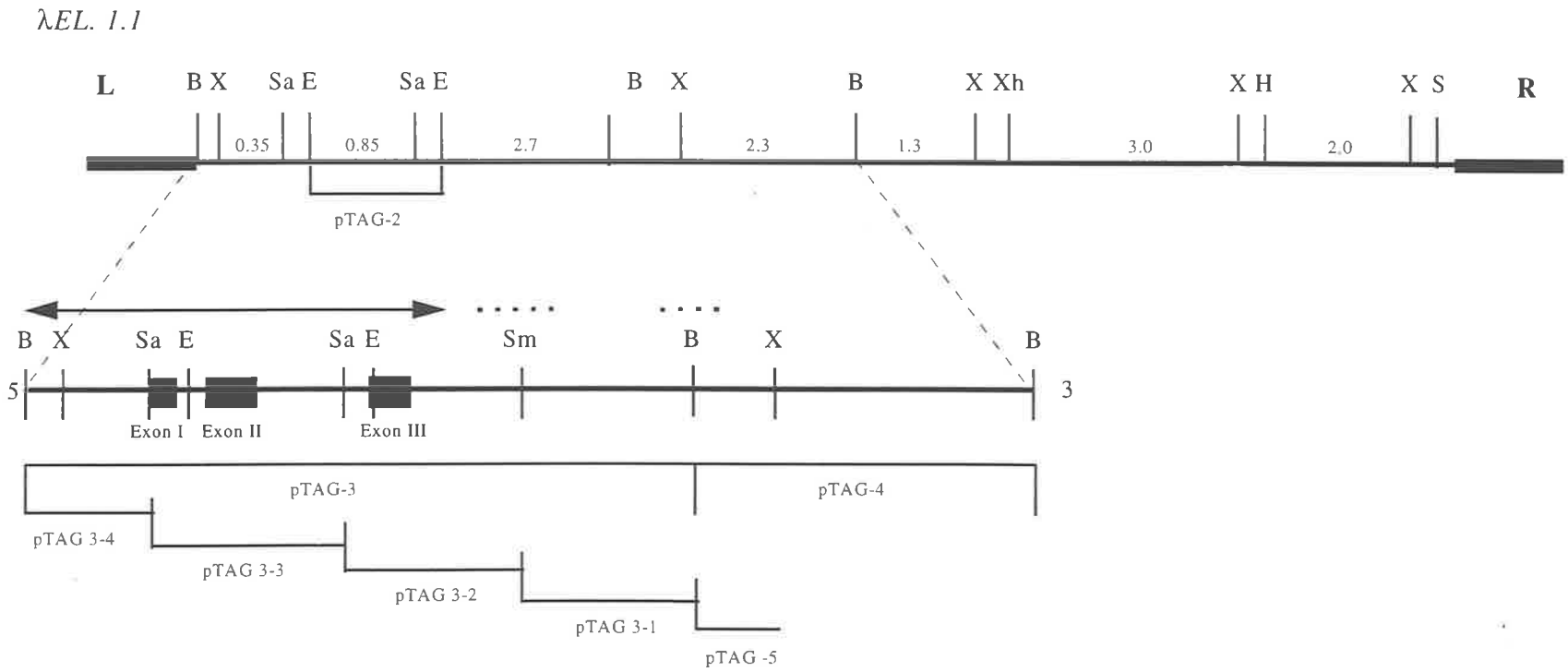


Figure 3.1.3 Restriction map of λEL.1.1 and the location of plasmid subclones used for DNA sequence determination. Approximate distances between restriction sites are given in kb. Exons are indicated by black boxes and the approximate extent of the sequenced region (2,308 bp) is indicated by a double arrow. The partially sequenced region (one strand only sequenced) is shown as a line of dots. L and R refer to the left and right arms of λ*GEM*-11. Restriction sites are: B = *Bam*HI, E = *Eco*RI, H = *Hind*III, Sa = *Sac*I, S = *Sal*I, Sm = *Sma*I, X = *Xba*I, Xh = *Xho*I.

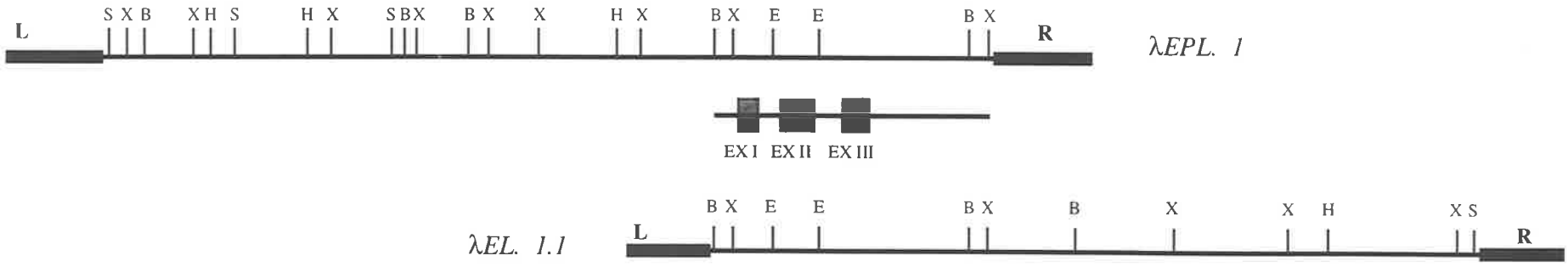


Figure 3.1.4 Restriction map of λEPL.1 showing the region of overlap between λEPL.1 and λEL.1.1. Exons are indicated by stippled boxes. L and R refer to the left and right arms of λ*GEM*-11. Restriction sites are: B = *Bam*HI, E = *Eco*RI, H = *Hind*III, S = *Sal*I, X = *Xba*I.

Each of the fractions was electrophoresed on 0.8% (w/v) agaorose gel (Fig. 3.1.5 A). The fraction containing DNA fragments in the size range 9-15 kb was selected (Fig. 3.1.5 B). The size-selected genomic DNA was ligated to the lambda phage vector *Gem*-11™.

The primary titre of the two libraries was $5x10^5$ (PGL.1.1) and $6x10^6$ (PGL.1.2) plaque forming units (pfu), and the average insert size was 15-16 kb. Two clones in the PGL.1.1 library showed significant hybridisation to the echidna β-globin gene probe. The probe, pTAG-2, used to detect these clones, was a 856 bp *Eco*RI fragment containing parts of intron I and II and all regions of exon II of the echidna β-globin gene (Section 2.1.8). These two positive clones are referred to as λPL.1.1 and λPL.1.2 (Fig. 3.1.6).

3.1.4 Southern analysis and restriction mapping of the platypus clones λPL.1.1 and λPL.1.2

The results of restriction digestion and Southern analyses of the clones λPL.1.1 and λPL.1.2 are shown in Fig. 3.1.7 and Fig. 3.1.9, respectively. The insert of λPL.1.1 was estimated to be 16 kb in size and the insert of λPL.1.2 to be 15 kb in size. Gels were Southern blotted onto Hybond-N+ membranes and the filters were hybridised to pTAG-2, the echidna β-globin gene probe (see Fig. 3.1.7 B for λPL.1.1 and Fig. 3.1.9 B for λPL.1.2). Based on the restriction and Southern analyses, restriction maps of λPL.1.1 (Fig. 3.1.8) and λPL.1.2 were derived (Fig. 3.1.10).

The 6.8 kb *Eco*RI fragment of λPL.1.1, which strongly hybridised to the echidna β-globin probe (Fig. 3.1.7 B lane 3) and the 3.8 kb *Eco*RI/*Sal*I fragment of λPL.1.1, were subcloned into the plasmid *PGM3zf*(+) vector for sequencing. The subclones were named pobeg-1 and pobeg-2, respectively (Fig. 3.1.8).

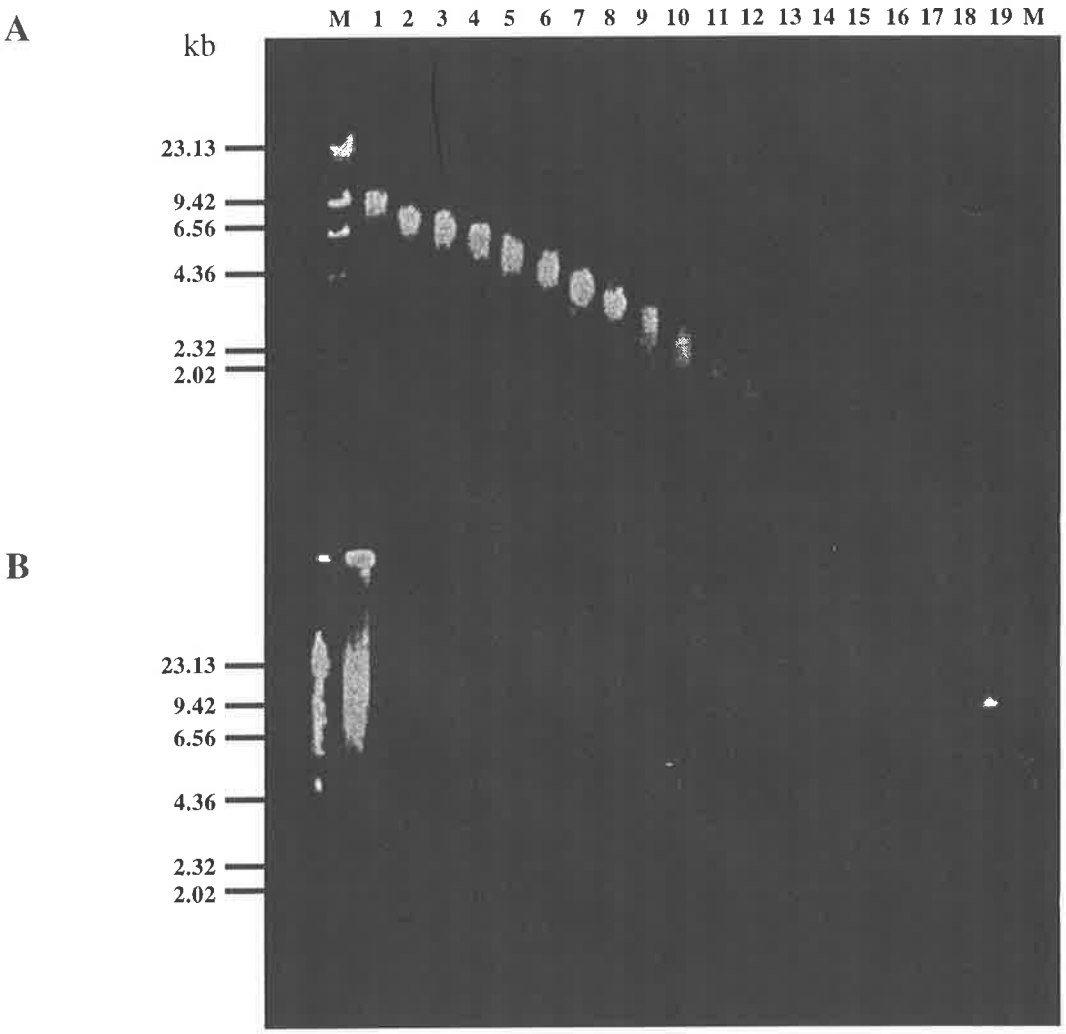


Figure 3.1.5 Partial digestion with *Sau*3A, and size separation on 0.8% (w/v) agarose gel, of platypus genomic DNA

A: Track M: λ *Hind*III DNA size markers; tracks 1-19: 2 µl aliquots of each of the 19 fractions collected from the glycerol gradient.

B: Track 1: pooled DNA from samples that contained DNA fragments between 9 and 23 kb.

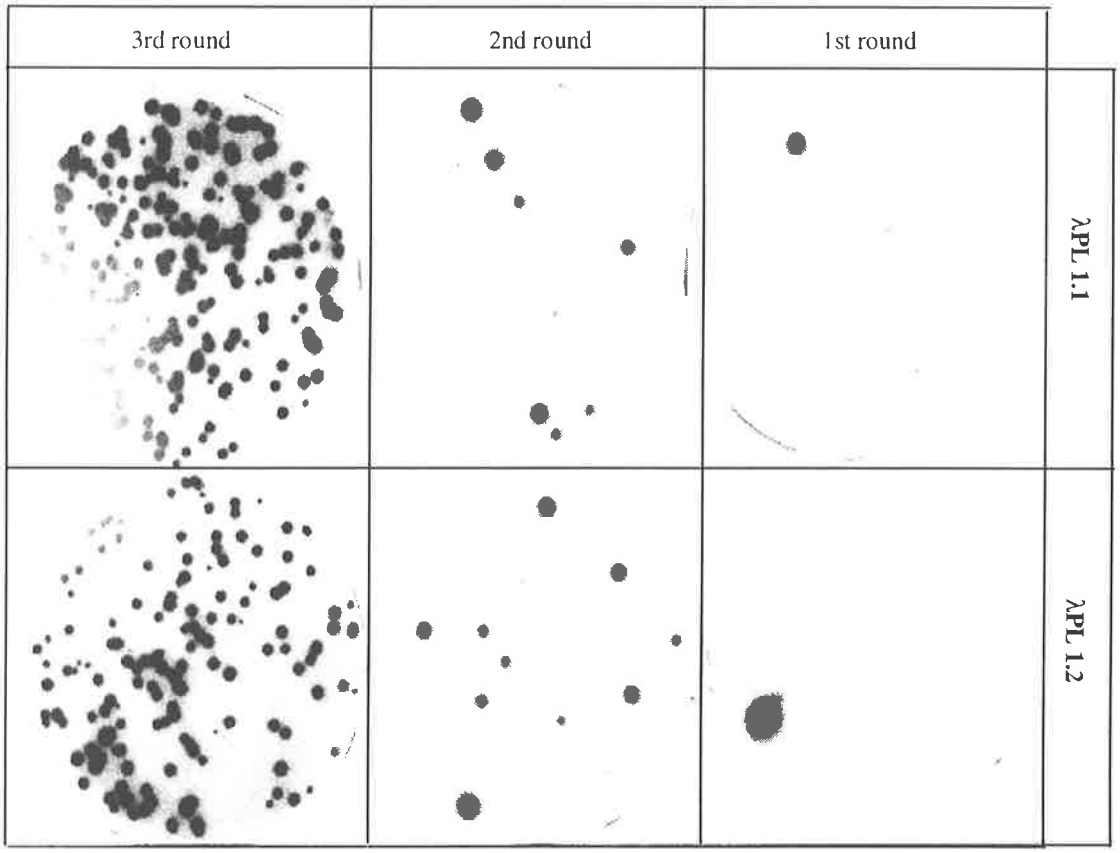


Figure 3.1.6 Autoradiographs of plaque lift filters from the platypus genomic DNA library, PGL.1.1. Two lambda clones, λPL.1.1 and λPL.1.2, show positive hybridisation to the *echid* β1-globin gene probe, pTAG-2. Three rounds of consecutive low density plating gave a pure culture of the lambda clones.

The subcloned pobeg-1 DNA was digested with *Xba*I to create smaller fragments suitable for analysis. This digestion gave two *Eco*RI/*Xba*I fragments 2.9 kb and 3.9 kb in length. These fragments were subcloned into *PGM3zf*(+) and the subclones named pobeg 1-1 and pobeg 1-2 (Fig. 3.1.8). The subclone pobeg-2 was digested with *Bam*HI and *Pst*I to generate 0.5 kb *Eco*RI/*Bam*HI, 1.3 kb *Bam*HI/*Pst*I and 2.0 kb *Pst*I/*Sal*I fragments. These fragments were subcloned into *PGM3zf*(+) and the subclones named pobeg 2-1, pobeg 2-2 and pobeg 2-3, respectively (Fig. 3.1.8). DNA sequencing, and sequence comparisons with β-like globin genes from other mammals, indicated that pobeg 1-1 contained exon II, exon III, Intron I and Intron II of a β-globin gene and pobeg 2-1 contained exon I (Fig. 3.1.8).

The pattern of restriction fragments of the platypus clone λPL.1.2 is very similar to the restriction patterns of the echidna clones, λEL.1.1 and λEPL.1 (Fig. 3.1.9). Southern analysis of restriction fragments of λPL.1.2 showed that a 0.85 kb *Eco*RI fragment and a 4 kb *Bam*HI fragment hybridised to the β-globin probe pTAG-2 (Fig. 3.1.9 B lane 1 and 3). The length of these two fragments appears to be the same as the length of fragments in the two echidna clones λEL.1.1 and λEPL.1 that hybridise to the marsupial β-globin probe pSG2H, and to the echidna probe pTAG-2 (Fig. 3.1.2 B lane 1 and 3).

The 4.3 kb *Xba*I fragments that hybridised to pTAG-2 (Fig. 3.1.9.B lane 6) was subcloned. This subclone (pobg-1) was digested with *Eco*RI to make three other subclones (pobg 1-1, pobg 1-2 and pobg 1-3) which contained 1.3 kb *Xba*I/*Eco*RI, 0.85 kb *Eco*RI/*Eco*RI and 2.4 kb *Eco*RI/*Xba*I fragments respectively. These three subclones were used for further characterisation (Fig. 3.1.10).

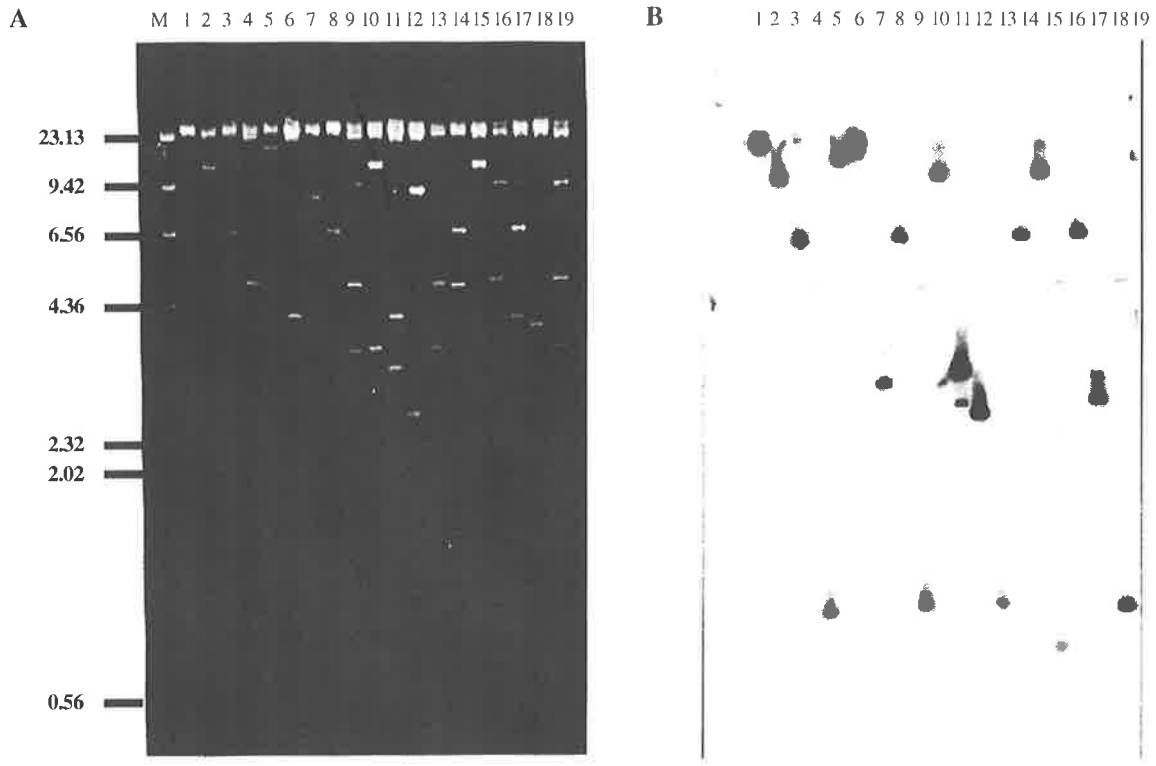


Figure 3.1.7 Restriction digestion and Southern analysis of λPL. 1.1.

A: A agarose gel (0.8 % (w/v)) showing restriction fragments of λPL. 1.1. The tracks are: M- λ*Hind*III DNA size markers 1- *Bam*HI, 2- *Hind*III, 3- *Eco*RI, 4- *Sac*I, 5- *Sal*I, 6- *Xba*I, 7- *Bam*HI/*Hind*III, 8- *Bam*HI/*Eco*RI, 9- *Bam*HI/*Sac*I, 10- *Bam*HI/*Sal*I, 11- *Bam*HI/*Xba*I, 12- *Hind*III/*EcoR*I, 13- *Hind*III/SacI, 14- *Hind*III/SalI, 15- *Hind*III/*Xba*I, 16- *Eco*RI/*Sac*I,17- *Eco*RI/*Sal*I, 18- *EcoRI*/*Xba*I, 19- *SacI*/*Sal*I, 20- *SalI*/*Xba*I.

B: Autoradiograph of the Southern filter taken from the gel shown in A. The probe used was pTAG-2. Sizes of hybridising fragments were estimated by comparing the autoradiograph to a photograph of the original agarose gel stained with EtBr.

λPL. 1.1

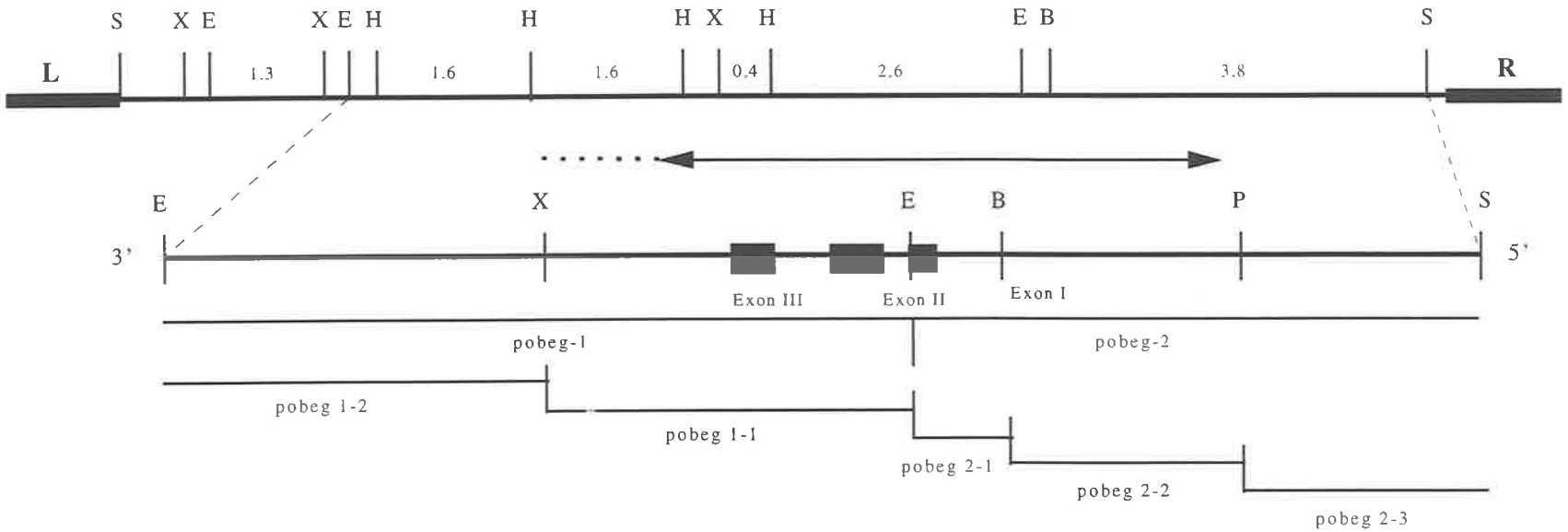


Figure 3.1.8 Restriction map of λPL.1.1 and location of the plasmid subclones used for sequence analysis. Approximate distances between restriction sites are given in kb. Exons are indicated by black boxes and the approximate extent of the region sequenced from both DNA strands (3.909 kb) is indicated by a double arrow. The region sequenced from a single strand only (1.141 kb) is shown as a line of dots. L and R refer to the left and right arms of λ*GEM*-11. Restriction sites are: B = *Bam*HI, E = *Eco*RI, H = *Hind*III, P = *Pst*I, S = *Sal*I, X = *Xba*I.

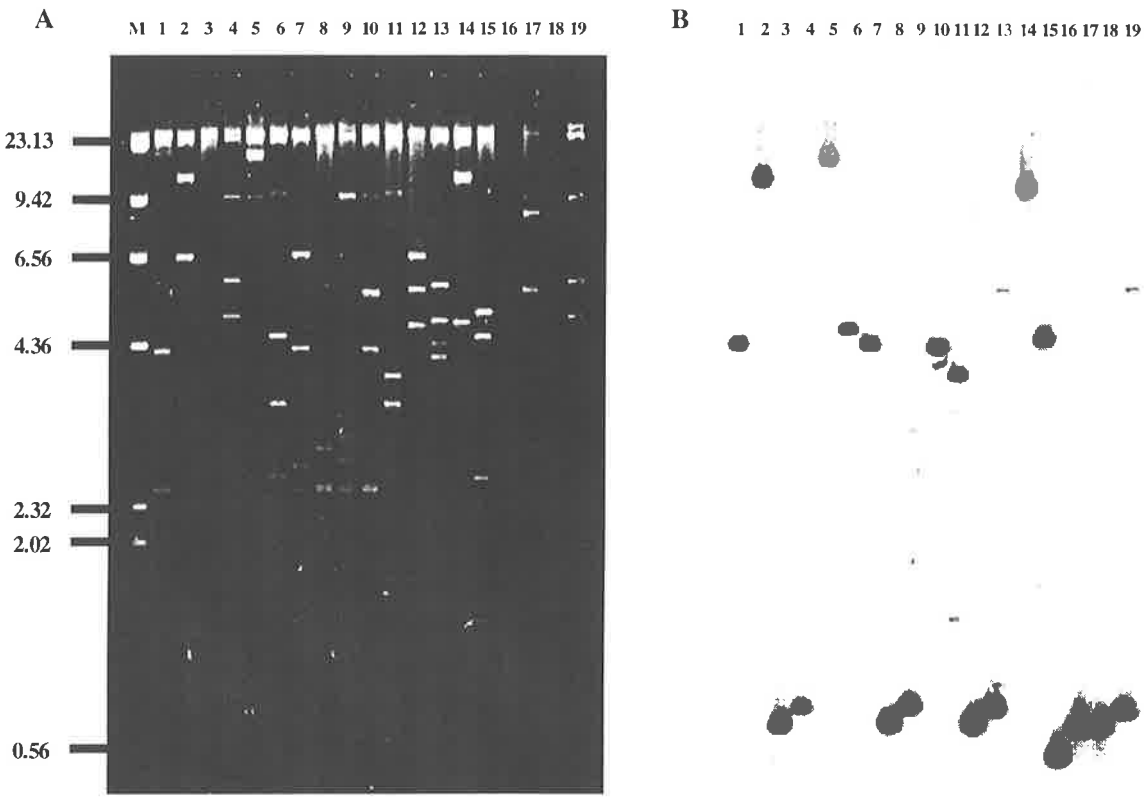


Figure 3.1.9 Restriction digestion and Southern analysis of λPL. 1.2.

A: An agarose gel (0.8 % (w/v)) showing restriction fragments of λPL. 1.2. The tracks are: M- λ*Hind*III DNA size markers, 1- *Bam*HI, 2- *Hind*III, 3- *Eco*RI, 4- *Sac*I, 5- *Sal*I, 6- *Xba*I, 7- *Bam*HI/*Hind*III, 8- *Bam*HI/*Eco*RI, 9- *Bam*HI/*Sac*I, 10- *Bam*HI/*Sal*I, 11- *Bam*HI/*Xba*I, 12- *Hind*III/*Eco*RI, 13- *Hind*III/*Sac*I, 14- *Hind*III/*Sal*I, 15- *Hind*III/*Xba*I, 16- *Eco*RI/*Sac*I, 17- *Eco*RI/*Sal*I, 18- *Eco*RI/*Xba*I, 19- *Sac*I/*Sal*I, 20- *Sal*I/*Xba*I.

B: Autoradiograph of the Southern filter taken from the gel shown in A, and hybridised to pTAG-2. Sizes of hybridising fragments were estimated by comparing the autoradiograph to a photograph of the original agarose gel stained with EtBr.

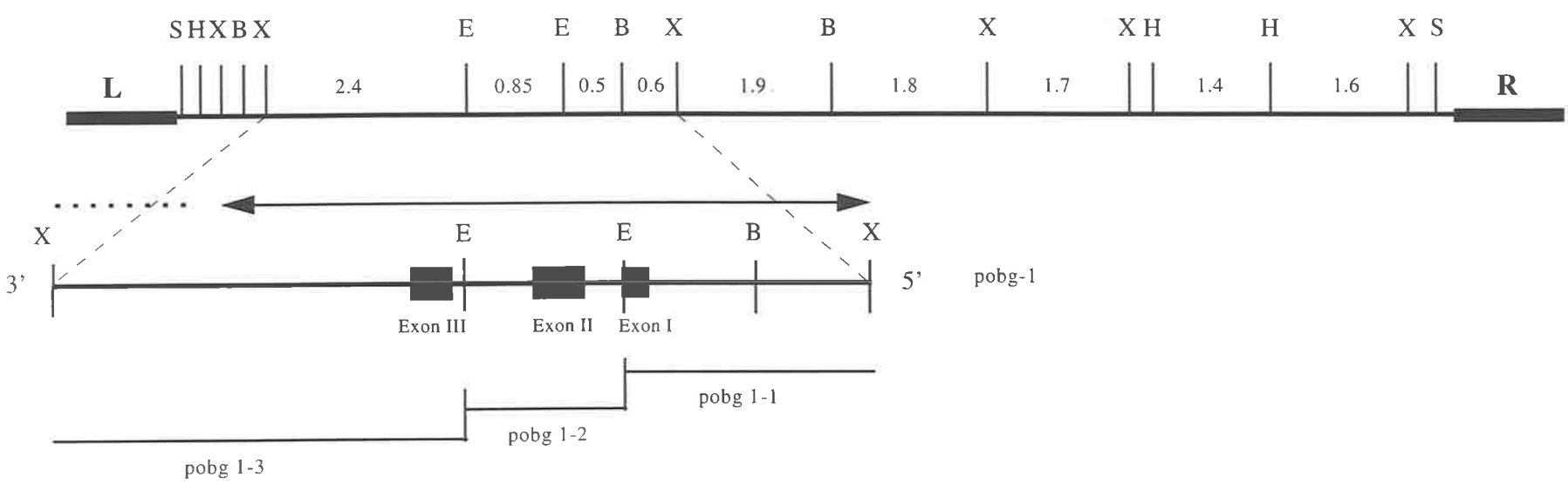


Figure 3.1.10 Restriction map of the λPL.1.2 insert and the position of fragments that were subcloned and used for sequencing of the *platy* β2-globin gene. Approximate distances between restriction sites are given in kb. Exons are indicated by black boxes and the approximate extent of the region sequenced from both DNA strands (3.177 kb) is indicated by a double arrow. The region sequenced from a single strand only is shown as a line of dots. L and R refer to the left and right arms of λ*GEM*-11. Restriction sites are: B = *Bam*HI, E = *Eco*RI, H = *Hind*III, S = *Sal*I, X = *Xba*I.

3.1.5 Sequencing of the putative β-globin gene from λEL.1.1

The sequencing strategy chosen was to use known restriction sites in pTAG-3 to generate subclones in *PGM3zf*(-) which were then sequenced. The four subcloned plasmids from pTAG-3, (pTAG 3-1, pTAG 3-2, pTAG 3-3 and pTAG 3-4), pTAG-5, and pTAG-4 were used as template DNA for sequencing (Fig. 3.1.3). Where necessary, for example to sequence across restriction sites at the ends of subclones, oligonucleotides, designed on the basis of previously ascertained sequences, were used as primers for further sequencing. Primers used for sequencing are detailed in Appendix 1. Identification and delineation of specific regions within the sequences was based on similarity to β-globin genes from other mammalian species. The 2,308 bp sequences from pTAG 3-3, 3-4 and parts of pTAG 3-2 are shown in Fig. 3.1.11 and the β-globin gene contained within this sequence was named *echid* β1. The nucleotide sequences from the partially sequenced regions from the subclones pTAG 3-1, pTAG 3-2, pTAG-4 and pTAG -5, are given in Appendix 2.

The complete nucleotide sequence of the putative β-globin gene from the echidna (*echid* β1) is presented in Fig. 3.1.11. This sequence, which is 2,308 bp in length, includes a 378 bp flanking region to the 5' side of the putative cap site, and a 730 bp region to the 3' side of the putative poly(A) signal site. Three exons are located in the positions characteristic to all vertebrate β-globin genes and encode a β-globin polypeptide of the expected length, 146 amino acids. The sequence of this protein are identical to those determined by Whittaker *et al.* (1972). The promoter elements identified in the 5' flanking region include "CACCC", "CCAAT" and "ATA" boxes located at -92, -78, and -33 from the putative cap site (CAG), respectively. Initiation (ATG) and termination (TGA) codons, open reading frames and consensus intron/exon splice junctions (AG/GT), as well as a putative poly (A) adenylation signal 92 bp downstream of the termination codon, were identified (Fig. 3.1.11). The lengths of the first and second introns are 146 bp and 465 bp, respectively. Two regions of trinucleotide (AAT) repeats were identified, one upstream (6 repeats) and the other downstream (16 repeats) of the gene (Fig. 3.1.11).

```
         10         20         30         40         50         60
GGATCCACAGCAGACTGTGCAACCATGGGGCTCTGGGAGGAGCAACATCTTGGGCTGGGA
AAGGAGGCCTGGGCCTGGTGACATGGAAAATAATAATAATAATAATGATGGCATTTATTA
AGTGCTTACAAAGGGGGCTAAAGGTCTGGACCTGGTGGGTGGGGAGTTGGGGGGGTGGTT
TCTCCAGCTCTCAGAGAGGCAGATTTGCATCCCTGCACCAGCCCTTTCAATACCTCTCCT
GCCTGCCAGGCCCGCATGACACAGGTCACGCCCCCAGTGTGGACCGCACCCCTGTCTGGG
CCAATCAGCTCTGAGGCTGAGAGTAGATGGAGCTCTGGGCAGGGGATAAAAAGCCAAGCT
AAACCTGGAGTTGATATCCAGTCTTTGCACATCTGAATCTTCAAGCAAACTTGCAAATCC

      INIValHisLeuSerGlySerGluLysThrAlaValThrAsnLeuTrpGlyHis
GACAAAATGGTGCATTTGAGTGGTTCTGAGAAGACTGCTGTCACCAACCTGTGGGGGCAT
ValAsnValAsnGluLeuGlyGlyGluAlaLeuGlyAr
GTGAACGTCAATGAACTCGGTGGCGAGGCCCTTGGCAGGTATGACCTCGAGGGCAGAATT
TTGAATTCCTGTGTTTGGGCCCACCTAGCTCTCAAAGGGCCTCCAGAGGGTAACCAGGTG
GAAACTCCAAGTTTCTCACTCTGTCTTTCTCTCTCTTTCCTTTGCCCTCCCTCTTACCCA

    gLeuLeuValValTyrProTrpThrGlnArgPhePheGluSerPheGlyAspLeuS
CCAGGCTGCTGGTCGTCTACCCCTGGACCCAGAGGTTCTTCGAATCCTTCGGTGACCTGT
erSerAlaAspAlaValMetGlyAsnAlaLysValLysAlaHisGlyAlaLysValLeuT
CCAGCGCCGATGCTGTGATGGGAAACGCCAAGGTCAAGGCCCATGGTGCCAAGGTGCTGA
hrSerPheGlyAspAlaLeuLysAsnLeuAspAsnLeuLysGlyThrPheAlaLysLeuS
CCTCCTTCGGCGATGCCCTGAAGAACCTCGACAACCTCAAGGGAACCTTCGCCAAGCTGA
erGluLeuHisCysAspLysLeuHisValAspProGluAsnPheAsn
GCGAACTGCACTGCGACAAGCTGCACGTGGACCCCGAGAATTTCAATGTGAGTACCCCAG
CCTTAGACTAGACCTGGCCATCTGACTGCGGGAGCTGGCATGGGAGCTGGAAGGAAAAGG
GGCTGAACGCCTGGGGAAGAAATGATGGGGTCCTGAGTGTGTGTTGGGCAAAGAGATAGC
CTGGGGCATCTGAAGGAAATGAGTCAGAATGAGAGCCTTTGAGGGAGAAAGGACCAGGGG
TTGCTACTGTGGGATACTTTACGAAGTAAGTGTCGGTCAGAGTGGGGGCTATCATTTGGG
GGTACAGAAGGAGAGGAGGGGACTGTGGTCAAGTTATGCATTAGGAGGCTTTAGGAGATA
GGAACGGGGTGTGGTAGGCATTTTGATATTGAGCCAAGGCCAGCGATGGTGCCTGGTAAG
AGCTCTGGGGTGTGAGAGTCACTGGGCTCAGACATTCTGCCTTGGATCAGGTCCCCACTA

                                        ArgLeuGlyAsnValLeuValValValL
TTCACCCTGCCCTGTGTCCATTTCTCCCACAGCGCCTGGGTAACGTGCTGGTCGTGGTCC
euAlaArgHisPheSerLysGluPheThrProGluAlaGlnAlaAlaTrpGlnLysLeuV
TGGCCCGTCACTTCAGCAAGGAATTCACCCCCGAGGCCCAGGCTGCCTGGCAGAAGCTGG
alSerGlyValSerHisAlaLeuAlaHisLysTyrHisTER
TGTCTGGTGTTTCCCACGCCCTGGCCCACAAGTACCACTGAGTCCTGAGCTGCCCCAGTG
ACTTCTGCGCCAGGTCCCATGACCTGCCCGTGGGAGTGCAGCACCACAGGCCCCTGAGAG
CCAGCCTGGCCCAATAAAGAGCTTTCAATCCAGTTTCTCCTCCGCTCTGCCTCTCTCTGT
GCGCCACTTTTTGGGGAAGGTCCCTACGGGATGGGGAGGGGGAGATGGGACCCTCTCCTC
CAGCTGCCTACTCCCGCAGTACTGTAGGGAGCTGCTCTAAATCTCCTCCATGGTAGCCTG
GGTGAGGAAGAGGGGCTGAACTGGTTGGGGTGAAGGAAGGGCTTTCTTGGAGGACTCTTG
GGGCATCTCCAGTTTGGCTGGGAAAAGGGAGGGCAAGGGAAGATGAGGGGGACGGTCTCT
GGATCTCAATGTAGCTAGAGGGAAGGGCCTGAAGGCACTCCCCAAGTGCCCTGGCCCCAG
ACTCCCTAGAATAGGGTCTAAATTAGGCTACCTGCTACTTGCCTCCAGAATAACCTCTGG
GGACAACCACCCCTTAGGCAGGTTTGGCATCTCCAAATAAGGAAAGCTACCAGGTGAGGG
GCCCACCAGCCCCATACACAATGTCCAGGACCCTGGCCCTTCCTTTTCTCTTCTCCCAAG
CCATCTCCCAGGGGCCGGGTAATAATAATAATAATAATAATAATAATAATAATAATAATA
ATAATAATAACAGCATTTATTAAATGCTTACTATGTGTAAAGCACTGTTCTAAGCACTGG
GGAGGTTACAAGGTGATCAGAGAAGCAACATGGTTCAATGGAAAGAGCACAGGCTTTGGA
GTCAGAGGTCATGGGTTCAAATCCCTAC
```

Figure 3.1.11 Complete nucleotide sequence and the corresponding amino acid sequence of the β-globin gene (*echid* β1) from the Australian echidna. The putative promoter elements in the 5' flanking regions of the gene "CACCC" and "CAAT", an RNA polymerase II binding site ("ATA"), a putative CAP site (CAG), translation initiation (ATG), termination (TGA) and poly (A) (AATAAA) signals in the 3' flanking regions are shown in bold. Trinucleotide repeats (AAT) are underlined. The conceptually translated amino acid sequence, shown using the standard three letter code, is 146 residues in length. The lengths of the 1st and 2nd introns of the gene are 146 bp and 465 bp, respectively. The length of the entire DNA sequence was 2,308 bp and the sequence was determined from both DNA strands.

3.1.6 Sequencing the platypus β-globin gene from λPL.1.1

Fig. 3.1.12 shows the sequence derived from λPL.1.1 using the same sequencing strategy as that adopted for the *echid* β1-globin gene. The sequences of the primers used are shown in Appendix 1. In total, 5,050 bp of sequence was determined from three plasmids (pobeg 1-1, pobeg 2-1 and pobeg 2-2) which were subclones of λPL.1.1 (Fig. 3.1.8). Most of this sequence (3.909 kb) was derived from both DNA strands, while 1.141 kb (containing most of the 3' flanking region) was sequenced from one of the DNA strands only (Fig. 3.1.8). The gene contained within λPL.1.1 was named *platy* β1 (Fig. 3.1.12). The contribution each subclone made to the complete sequence was: pobeg 2-2: 1.942 kb containing the 5' flanking region of the *platy* β1; pobeg 2-1: 0.491 kb containing exon I and parts of intron I; pobeg 1-1: 2.617 kb containing exons II and III, intron II, and most of the 3' flanking region (Fig. 3.1.8).

Three exons of the *platy* β1-globin are located in the characteristic positions of all vertebrate β-globin genes and these encode a β-polypeptide of the expected length of 146 amino acids. The translated codons give a sequence that is identical to the amino acid sequence of the major β-globin polypeptide determined by Whittaker and Thompson, (1975) from an adult platypus. The promoter elements which were identified in the 5' flanking region include "CACCC", "CCAAT" and "ATA" boxes located at exactly the same positions (-92, -78, and -33, respectively) as those of the *echid* β1-globin gene. Initiation and termination codons, open reading frames and consensus intron/exon splice junctions AG/GT, as well as a putative poly (A) adenylation signal 71 bp downstream of the termination codon, have been identified (Fig. 3.1.12). The lengths of the first and second introns are 153 bp and 440 bp, respectively. No repetitive sequences were detected.

Fig 3.1.12

```
         10         20         30         40         50         60
GAACTTTCGAGGTGCCAGACTGTGGGCAAATTCTGGGAAAGGTCCAGTCCCAGAAGGTTC
CCTGGAAGAGACAACTTTTCAATAGAATTTTTAGGAAGGGGTGTCCCAGAGTTTGACCTT
GAGTCCCGAAGAAAACCAATGATGAAAATGAAAAAGACATGAGCCAGTGGAACGAGTGCC
TGTAATTTGGAGAGGGGAAAGGGGAAAGGTGTTCTAGAGGAGGCTGAGTGCTAATAGGAT
AAGCGATCCTCTGGTGGAAATGAAGGAAACTGCCCGCCTGACACTGCCTCATCCTTCCTC
TGCACTCTCGAAACAGGCACCCTTCCAAGATGAAACATCCATTATCCTTGACTCCTCTCT
CTCACTCAACCCACACATTCAGTCCATCCCTAAATCCTTTCAGTTTCTCCTTCACAACAT
CGCTTAAATCTGCCCTTTCCACTCCATCCAAACTGCTACCATGTTAATAAAATCACTCAT
CCTATCCCGCCTGGATTTCTGCATCAGCCTCCTTGCAGACCACCCAACCTCCTGTCTCTT
CCCACTTCACGCTGCTGCCTGGATCATTTTTCTTCAAAAATGTTCAGGACATATCTTCCC
ACTCCTCAAAAAACTCCAGGGTTTGCCGATCCACCTACATATCAAGGAAAAACTCCTCAC
CATTGGTTTTAAAACACTCAATCACCTTTCTTCCTCCTCCCTCTCCTCAGTACTCTCCTA
CTACAACCCAGACCGCGCACTTCACTTCTCTAATGCTAACCTTCTCAATGTGCCTAGAAC
TTTGTCCTATCTTCCTGCCAACCCCTCGCCCCCCATCCTGTTTCTGGCCTGGAACACCCT
CCCTCTTGAATCTGACAATTATTCTCCCTCATTTCAAAGCCTTCTTGAAGGCACATCTCC
ATTAGGCCTTCCCTAAGCCCTCCTTTTTTCTTCTCCCACTTCCTTCTGTGTCGCCTTGAC
TTGCTCCCTTTGCTTTTTTCCCCTTCCCAGCCCCACAGCACTTATGTACATATCTGTAAT
TTATTTACTTATATTAATGTCTGTCTCCCTCCTCCAGACTGTAAGTTCATTGCAGGCAGG
GAATGTACCTGTTTATTGTTATATTCTACTCTCCCAAGCGCTGAATACAGTGTTCTGCAC
ATGGTAAACTCTCAATAAATACAAATAAATGAATGAATGAATCATAGGATAATGATAATA
ATAATGTTGGTATTTGTTAAGCGCTTACTATTGGCGGAGCACTGTTCTAAACGCTGGGGT
AGATAAAGGGTAATCAGTTTGTCCCACGTGAGGTTCCCAGTCTTAATCCCCATTTTACAG
ATGAGGGAACCGAGGCCCAGAGAAGTTAAGTAACTTGCCCATCGTCACACAGCTGACAAG
TGACAGAGCCGGGATTCGAACCCATGACCTCTGACTTCCAAACCCGGGCTCTTTCCACTG
AGCCACGCTGTTGCAGCCTCTGCTCTATCCCTGGGATTGCTGTGGTTTTTCCCAGGCTGC
CTAACCTAGGTGTAAGTAGCCTCCTAGAGTTGTGATGTTCAGACAGAGCCACAGAATCCA
TCCTCCTGCCTCTGAGCACAGCTCAGACCCAGCATGTCTCCCCACATCTTGAAAAAGCTG
GCTCAGTGGGGACTGAGCAGCCTCCCTGGGGAATGTGTCCTACAGCTCAACACAACCCAA
ATCCTTCCTGCTGTAGTCGCAGCCCCAATGCATGGTGGGAACCAAGGATGCTGTGGATTC
AATATTTCCACCCACGCCTGGGCACAAGTGGGACTCCTCAGGTTTCCCTTCTTTGGGCCA
GGCTGGACCAGTTGCTCCTCTCCCACCTAACACAAGGGGAGGTCAGGCTTTTAGTTCCTA
CCAGGGGTGCACATCAGACAAGTCAAAGAAAGAAACATGCTGTGGTCTGTGGAGGGAGAT
CTGGGGACAGAGCTCCAGCAATGGATCCATGGAAGACCACGTAACCATGGGCATCTGGGA
GGAGTGACATCTTGGGCTGGCAAAGGAGGACTGGGCCTGGGGATAGGGAGAAGGGGGATG
AAGGTCTGGACCCGGGTGGTGGGGGTGGGGGGGTATTTTCTCCAGCTCTCAGAGAGCAGA
TTTGCATCCCCGCACCAACCTCTCCTGCCTGCCAGCCCTGCATGATACAGGTCACGCCCT
CAGCGGGACCG CACCC CAATCTGGG CCAA TCATCTCTGAGGCCTAGGGTAGGTGGAGCTC
TGGGCAGGGG ATA AAAGGCCAAGCTAAACCTAGAGTCAGTACC CAG TCTTTGCACGTCTG
                                                  INI ValHisLeuSerGlyGlyGluLysSe
CATCTTCAAGCGAACTTGCAAATCCGACAAA ATGGTGCATCTGAGTGGTGGTGAGAAGAG
rAlaValThrAsnLeuTrpGlyLysValAsnIleAsnGluLeuGlyGlyGluAlaLeuGl
TGCCGTCACCAACCTGTGGGGGAAAGTGAACATCAATGAGTTGGGTGGCGAAGCCCTTGG
yAr
CAG GTACGACCTCGAAGGGAGAATTTCGAATTCCTGTATTTGGGCCCCACCTAGCTCTCA
AAGGGCCTCCAGAGCGTGACCAGGTAGAAACTCCAGGTCTCTCACTCTGTCACACTCTGT
                                                  gLeuLeuValValTyrProTrpTh
CTCTCTCTCTCCTTTGCCCTCGCTCTTACCCACCAG GCTGCTGGTCGTCTACCCCTGGAC
rGlnArgPhePheGluAlaPheGlyAspLeuSerSerAlaGlyAlaValMetGlyAsnPr
CCAGAGGTTCTTCGAAGCCTTTGGTGACCTGTCCTCCGCCGGCGCTGTGATGGGCAACCC
oLysValLysAlaHisGlyAlaLysValLeuThrSerPheGlyAspAlaLeuLysAsnLe
CAAGGTCAAGGCCCATGGTGCTAAGGTGCTGACCTCCTTCGGCGATGCCCTGAAGAACCT
uAspAspLeuLysGlyThrPheAlaLysLeuSerGluLeuHisCysAspLysLeuHisVa
CGACGACCTCAAGGGCACCTTTGCCAAGCTGAGCGAACTGCACTGCGACAAGCTGCACGT
lAspProGluAsnPheAsn
GGACCCTGAGAATTTCAAT GTGAGTCTCTGGCCCTGGGATAGACCTGCCCCTCTGACCCT
GCAGGAACTGGCATGGGAGGTGGAAGAAAAAGGTGCCGAAGCCCCTGGGGGAGAAAGGTT
GGGGGTCCTGAGTGTCTCATTCTCAGGGGTCAGAATGAGAGACTTTGAGGGAGAAAGGAG
CAGCAGTTGCTACTGAGGGATACTTTACCAAATGATTGTCAGTCAGAGTCCGGGTTATCA
TTTGGGGGTAAGGAAGGAGAGGAGGGGGCTGTGGTCAAGTTATGCCTTAAGAGGCTTTAG
GAGATAGAGATGGGTGTGGTAGGTGTTTTGATATTGAGCCAAGGCCATGGATAGTGCCTG
GCAAGAGCTCTGGGTTGTGACAGTCACCTGGGTTCAGACATTCTGACCTGGATCAGATCC
```

```
                                                    ArgLeuGlyAsnValLeuIle
CCAGTATTCACCCCTGCCCTGTGTCCATTTCTCCCACAGCGTCTGGGTAACGTGCTGATC
ValValLeuAlaArgHisPheSerLysAspPheSerProGluValGlnAlaAlaTrpGln
GTGGTCCTGGCCCGTCACTTCAGCAAGGACTTCAGCCCCGAGGTCCAGGCTGCCTGGCAG
LysLeuValSerGlyValAlaHisAlaLeuGlyHisLysTyrHisTER
AAGCTGGTGTCTGGTGTTGCCCACGCCCTGGGCCACAAGTACCACTGAGTCCTGAGCTGC
CCCGGTGGCTTCTGCATCAGGTCCCATGACCACACTCCCCTGAGGCCCTATTCTGGCCCA
ATAAAGAGCTTTCAGTCCAGTTTCTCCTCCGCTCTGCCTCTCTCTGTGCGCCTCTTTGTG
GGGAGGGTCCTATGGGGTGGGGAGAGGGAGATGGGACCCTCTCCTCCTGGCTATCTCCAG
CTGCCTACTTCCACAGGGCTGCAGGGAGCTGCTTTTAAATCCCCTCCATGGTGGCCTGGG
TGGGGAAGAGGGGCTGAACTGTTGGGGGCAAAGGAAGGGCTTTCTTGGAGGACTCTAAGG
GAGAGAAAGGGAGGATGGCGGGGACAGTCTATGGACCTCAATATAGCTAGAGCAAAGGGC
CTGAACTGATTACTCCCCAAGAGCCCTGACCCCAGAGTCCCCAGAATCTTGGCTAAGTTA
GGCTCCCTGCCACTTACCCCAAGAGCAGGAGAGTGGGCAACCTCTAGGGACAGCCACCCC
TTAGGCAGGTTCGACATTCCCAGGTGGAGAAAGCTACCAGATGAGGGGCCCACCAGTCCC
ATATCCAGTGCCCAGGACATTGGCTTTTCAAATTGGCTTTTGAAATTGACATTTTCTCCA
AAGCAACTCCCGGGTGCTCCCAAGCCCTTCCTCTTGCTGGTCCCACCATGCCTAACCCAG
CCATCGTAAGGCCTTCCTGCGGGCAGTGGTAACTGAGGGTGAATGCCCCATCATCACCTC
CCATTCATTATCTTAGTTCAACCACGGCCCCTAGGACAACCCCTCAACTCCAGATCCCAT
ACGCCTCAAGACAAATATCCATCAACCTCCACTCTTGTTTTTGCCTCCCTCAGACATCAC
ATCCTTGCTCCTGTGTCCCTTTTTTTATAGTATTTGTTTAGTGCTTACTATGTGTCAAGC
ACTGTTCTAAGCATTGTGGTAGATAGGCATAAACCAGGTCAGACATAGTTCCCTGTCCCA
TATGGCATTCATAGTCCAAGTAGGAGGAACAGCAGGTATTTCCTCTCCATTTTACAGATG
AGGAAAAGGAGGCACATAGAACTTGAGTGACATATCCAAAATCACACAGCAGTCACGTGG
CAGAGCTGGGTTTAGAACCTAAGTCATCGGACTTCCAGGCCCATGCTCTTTCCACTAGGC
TAAGGTGCTTCCCTCGTATTCCATGCCCCTGCCCTTCTCCCGACCTGCACTCCCACCTTT
AGCTGCATCCTCTGAAAACATTCAGACCATATTTCCCCTGCCTCAAAAAACTTGAGGGGT
TGCCCATCCACCTCCACATCAAACAAAAACTCCTCACCATGGCTTTAAAGCACTTAATCA
CCTTCCCCCCTCCTACCTCACCTCACTACTCTCCTACTACATCCCAGACCTCACACTTTG
TTCCTCTAATGCTAACCTTCTCACTGTACCTCGATCTCATCTACTCGCACCACTGACCAC
TTGCCCACATCCTACCTCTGGCCTGGAGCACCCTCCCTCCTCATATCATTGCTCTCCCGA
ACTTCAAAACCTTATTGAAGGCACATCCCCTCCACGAAGCTTTCCCTGACTAAGCCCCAC
CTCTCCTCTTCTCCCCTCCCTTCCTCGCCACTCTTACTTGCTCCTTCTTTCATCCCCATC
CCCACAGCACATTTGTATATGTCTGTAATTTATTTAATTTGTATGAATATCTCTCTCCCC
CTCCTAGAGTC
```

Figure 3.1.12 Complete nucleotide sequence and the corresponding amino acid sequence (shown using the standard three letter code) of the *platy* β1-globin gene. The putative "CACCC" and "CAAT" promoter elements in the 5' flanking region of the gene, an RNA polymerase II binding site ("ATA"), a putative CAP (CAG) site, translation initiation (ATG), termination signals (TGA) and poly (A) (AATAAA) in the 3' flanking region are shown in bold. The conceptually translated amino acid sequence is 146 residues in length. The lengths of the 1st and 2nd introns of the gene are 153 bp and 440 bp respectively. Most of the DNA sequence (5,050 bp) shown was determined from both DNA strands, but 1.141 kb of 3' flanking sequence was determined from one strand only.

3.1.7 Sequencing the platypus β-globin gene from λPL.1.2

Regions of the platypus β-globin gene contained in the three subclones, pobg 1-1, pobg 1-2 and parts of pobg 1-3 (3.177 kb) (Fig. 3.1.10) were sequenced from both DNA strands. The gene was named *platy* β2 (Fig. 3.1.13). The partial sequences of pobeg 1-3 are shown in Appendix 3 and are marked in Fig. 3.1.10 by a dotted line. The promoter elements in the 5' flanking region of the *platy* β2 such as the "CACCC", "CCAAT" and "ATA" boxes located at -93, -75, and -33, respectively, are located at positions very similar to those in the *echid* β1 and *platy* β1 globin genes. Initiation and termination codons, open reading frames and consensus intron/exon splice junctions AG/GT, as well as a putative poly (A) adenylation signal 99 bp downstream of the termination codon, have been identified (Fig. 3.1.13). The lengths of the first and second introns are 147 bp and 460 bp, respectively. Two repetitive sequences were apparent, a trinucleotide repeat "AAT", and a tetranucleotide repeat "TGAA". The trinucleotide repeats were located in the same position as similar repeats in the *echid* β1-globin gene: 6 repeats in the 5' flanking region and 18 repeats in the 3' flanking region. The 11 tetranucleotide repeats were found in the 5' flanking region of the *platy* β2-globin gene (Fig. 3.1.13). The "TGAA" repeats were also identified around 2 kb downstream of the *echid* β1-globin gene (Appendix 2).

The restriction map of λPL.1.2 is very similar to that of the echidna clone λEL.1.1 (Fig. 3.1.10). Furthermore, a comparison of the DNA sequences of the *platy* β2 and *echid* β1 shows a high level of sequences identity in flanking regions (95%), two introns (97%) and coding regions (98%) (Table 3.1.2 and Table 3.1.3). The coding regions show only one nucleotide difference. The translated amino acid sequence was exactly the same for both genes (Fig. 3.1.14).

10 20 30 40 50 60
AGTTTAAACTCATTGCAAGCAGGGAATTTACCTGTTTATTGTTATATTGTATCTACTCCC
AAGCGCTGAGTACAGTACTCTTGCACGTGGTAAGCTCTTAATAGATATGAATGAATGAAT
GAATGAATGAATGAATGAATGAATGAATGAATCATAGGCTGTTGCAGCTTCCTTCAATCT
ATCCCTGGATTGCTGTGTCTTTTCCCAGGCTGCCTAACCTAAGTGCAGGTAGCCCCCTAG
AGTTGTAATGGTCAGACAGGGCCACGAAATCCATCCCCCTGCCTTTTAGCATGTCTCCCC
TCATGTTGAAAAAGTTGGCTCAGCGGGGACTGAGCAGCCTCCCTGTGGAATGTGTCCTAC
AGCTCATCACAGCCCAAATCCTCCCTGCTGTAGTTGCAGCCCCAGTGCATGGTGGGACCA
GGGAGGCTGTGGATTCAATTTTCCCACCTACACCTGGGCACAGGTGGGACCCTTCAGCTT
CCCTTTCTCTGGTACAGGCTGGCAGGTGCTCCTCTCCCACCCAATGGCTTCGGGGCTCAG
GTGAGGTCTAGGGGCTACTGCAGGGCCCACCAGAAGCCAAGCACAAGGTGAGGTCAGGCT
TTTAGTTCTAACCAGGGTGCAGATCAGACAAGCCGAACAAAGAAACGTTTTGGCCTATGG
AGGGGGATCTGGGGACAGAGCTCCAGCAGTGGATCCACAGCAGACTGTGCAACCATGGGG
CTTTGGGAGGAGCAACATCTTGGGCTGGGAAAGGAGCCTGGGCCTGGTGACATGGAAAAT
AATAATAATAATAATGATGGCATTTATTAAGTGCCTACAAAGGGGGCCAAAGGTCTGGAC
CTGGTGGGTGGGGAGTTGGGGGGGTGGTTTCCCCAGCTCTCAGAGAGGAGATTTGCATCC
CTGACACCAGACCCTTTCAATACCTCGTCCTGCCTGCCAGCCCCGCATGACACCAGGTCA
CAGACCCCCAGTGTGGACCG**CACCCC**TGTCTGGG**CCAA**TCAGCTCTGAGGCTGAGAGTAG
ATGGAGCTGCTGGGCAGGGG**ATA**AAAAGCCAAGCTAAACCTGGAGTTGATATC**CAG**TCTT
INIValHisLeuSerGlyS
TGCACATCTGAATCTTCAAGCAAACTTGCAAATCCGACAAA**ATGGTGCATTTGAGTGGTT**
erGluLysThrAlaValThrAsnLeuTrpGlyHisValAsnValAsnGluLeuGlyGlyG
CTGAGAAGACTGCTGTCACCAACCTGTGGGGGCATGTGAACGTCAATGAACTCGGTGGCG
luAlaLeuGlyAr
AGGCCCTTGGCAGGTATGACCTCGAGGGGAGAATTTTGAATTCCTGTGTTTGGGCCGACC
TAGCTCTAAAAGGGGCCTCCAGAGGGTAACCAGGTGGAAACTCCAAGTTTCTCACTCTGT
gLeuLeuValValTyrProT
CTTTCTCTCTCTTTCCTTTGCCCTCCCTCTTACCCACCAG**GCTGCTGGTCGTCTACCCCT**
rpThrGlnArgPhePheGluSerPheGlyAspLeuSerSerAlaAspAlaValMetGlyA
GGACCCAGAGGTTCTTCGAATCCTTCGGTGACCTGTCCAGCGCCGATGCTGTGATGGGAA
snAlaLysValLysAlaHisGlyAlaLysValLeuThrSerPheGlyAspAlaLeuLysA
ACGCCAAGGTCAAGGCCCATGGTGCCAAGGTGCTGACCTCCTTCGGCGATGCCCTGAAGA
snLeuAspAsnLeuLysGlyThrPheAlaLysLeuSerGluLeuHisCysAspLysLeuH
ACCTCGACAACCTCAAGGGCACCTTCGCCAAGCTGAGCGAACTGCACTGCGACAAGCTGC
isValAspProGluAsnPheAsn
ACGTGGACCCCGAGAATTTCAATGTGAGTCCCCAGCCTTAGACTAGACCTGGCCATCTGA
CTGCGGGAGCTGGCATGGGAGCTGGAAGGAAAGCTGAAGCCCCTGGGGAAGAAATGATGG
GGTCCTGAGTGTGTGTTGGGCAAAGAGATAGCCTGGGGCATCTGAAGGAAATGAGTCAGA
ATGAGAGCCTTTGAGGGAGAAAGGACCAGGGGTTGCTACTGTGGGATACTTTACGAAGAA
GTGTCGGTCAGAGTGGGGGCTATCATTTGGGGGTACAGAAGGAGAGGAGGGGACTGTGGT
CAAGTTATGCATTAGGAGGCTTTAGGAGATAGGAACGGGGTGTGGTAGGCATTTTGACAT
TGAGCCAAGGCCAGCGATGGTGCCTGGTAAGAGCTCTGGGGTGTGAGAGTCACTGGGCTC
AGACATTCTGCCTTGGATCAGGTCCCCACTATTCACCCTGCCCTGTGTCCATTTCTCCCA
ArgLeuGlyAsnValLeuValValValLeuAlaArgHisPheSerLysGluPheThr
CAG**CGCCTGGGTAACGTGCTGGTCGTGGTCCTGGCCCGTCACTTCAGCAAGGAATTCACC**
ProGluAlaGlnAlaAlaTrpGlnLysLeuValSerGlyValSerHisAlaLeuAlaHis
CCCGAGGCCCAGGCTGCCTGGCAGAAGCTGGTGTCTGGTGTTTCCCACGCCCTGGCCCAC
LysTyrHis**TER**
AAGTACCACTGAGTCCTGAGCTGCCCCAGTGACTTCTGCGCCAGGTCCCATGACCTGGCC
CGTGGGAGTGCAGCACCACAGGCCCCTGAGAGCCAGCCTGGCCC**AATAAA**GAGCTTTCAA
TCCAGTTTCTCCTCCGCTCTGCCTCTCTCTGTGCGCCACTTTTTGGGGAAGGTCCTACGG
ATGGGGAGGGGGAGATGGGACCCCTCTCCTCCAGCTAGCCTACTCCCGCAGTGCTGTAGG
GAGCTGCTCTAAATTCCTCCATGGCTAGCCTGGGTTGGGGCAAGAGGGCTGCAACTGGTT
GGGGATAAGGAAGGCTTTCTTGGAGGACTCTTGGGGCATCTCCAGTTTGGCTGGGAAAAG
GGAGGGCAAGGGAAGATGAGGGGGACGGTCTCTGGATCTCAATGTAGCTAGAGGGAAGGG
CCTGAAGGCACTCCCCAAGTGCCTGGCCCCAGACTCCCTAGAATAGGGTCTAAATTAGGC
TACCTGCTACTTGCCTCCAGAATAACCTCTGGGGACAACCACCCCTTAGGCAGGTTTGGC
ATCTCCAAATAAGGAAAGCTACCAGGTGAGGGGCCCACCAGCCATACACAATGCCCAGGA
CCCTGGCCTTCTTTTCTCTTCTCCAAGGCGGTAATAATAATAATAATAATAATAATAATA
ATAATAATAATAATAATAATAATAATAACAGCATTTATTAAATGCTTACTATGTGTAAAG
CACTGTTCTAAGCACTGGGGAGGTTACAAGGTGATCAGAGAAGCAACATGGTTCAATGGA
AAGAGCACAGGCTTTGGAGTCAGAGGTCATGGGTTCAAATCCTACTCTGCCAACTGTCAG
CTGTGTGACTATGGACAAGTCACTTAACTTCTCTGTGCCTTGGTTCCCTCATCTGTAAAA
TGGGGATTAAGACTGACCCCCACGTGGACAATCTGATCACCTTGTATCCCCCCAGCGCTT
AGAACAGTGCTTTACACATAGTAAGGCTTAACAAATACCATCATTATTATTATGAGG

3.1.8 Is platy β2 an authentic gene in the platypus ?

There are two questions to answer about the *platy* β2-globin gene: i) why is *platy* β2 so similar to the *echid* β1-globin gene? ii) If the *platy* β2-globin gene exists in the platypus, is the gene expressed ? No evidence was obtained that *platy* β2-globin is expressed. However, experiments were carried out to provide evidence for the existence of an echidna like β-globin gene, *platy* β2, in the platypus. These experiments made use of PCR, restriction digestion and direct sequencing of PCR products, sequence and location comparisons of boundary regions of the tetranucleotide "TGAA" repeat, and restriction map comparisons between λPL1.2, λEL.1.1 and λEPL.1.

3.1.8.1 PCR amplification of the putative *platy* β2 gene from platypus genomic DNA

Primer sets were designed to specifically amplify each β-globin gene from platypus genomic DNA: the set B F1/Lee R for *platy* β1 and the set Lee F2/Lee R for *platy* β2 (The primers sets use the same reverse primer). The PCR conditions and the sequences of the primers are show in Table 2.2.1. The Lee F2 and B F1 primer sequences are located in the same position of exon I in the two genes, and the reverse primer sequence, which includes the termination codon, is located at the end of exon III of *platy* β1 and *platy* β2 (Fig. 3.1.14). To confirm the specificity of each primer set, three lambda DNA clones, λPL.1.1 (*platy* β1), λPL.1.2 (*platy* β2) and λEL.1.1 (*echid* β1) were used as control DNA samples for PCR (Fig. 3.1.15 lane 1, 2, 4, 8, 9, 11). The PCR amplification using the Lee F2/Lee R primer set amplified PCR products of the predicted sizes from lambda DNA λEL.1.1 (Fig. 3.1.15 lane 4) and λPL.1.2 (Fig. 3.1.15 lane 2). No amplification was obtained from λPL.1.1 (Fig. 3.1.15 lane 1). In contrast, the PCR reactions using the B F1/Lee R primer set gave a PCR product with λPL.1.1 DNA (Fig. 3.1.15 lane 8), but not with λEL.1.1 (Fig. 3.1.15 lane 11) or λPL.1.2 (Fig. 3.1.15 lane 9).

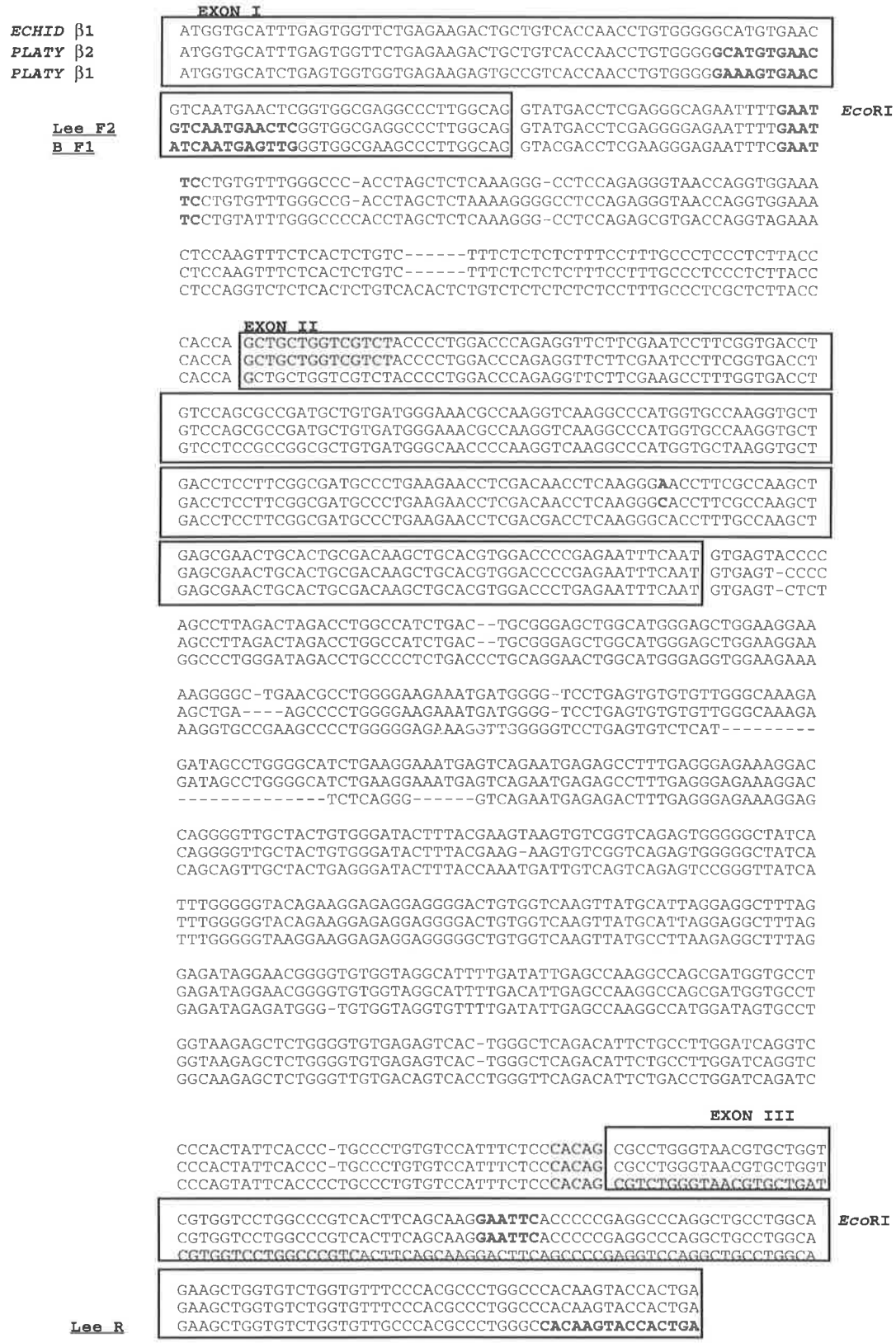


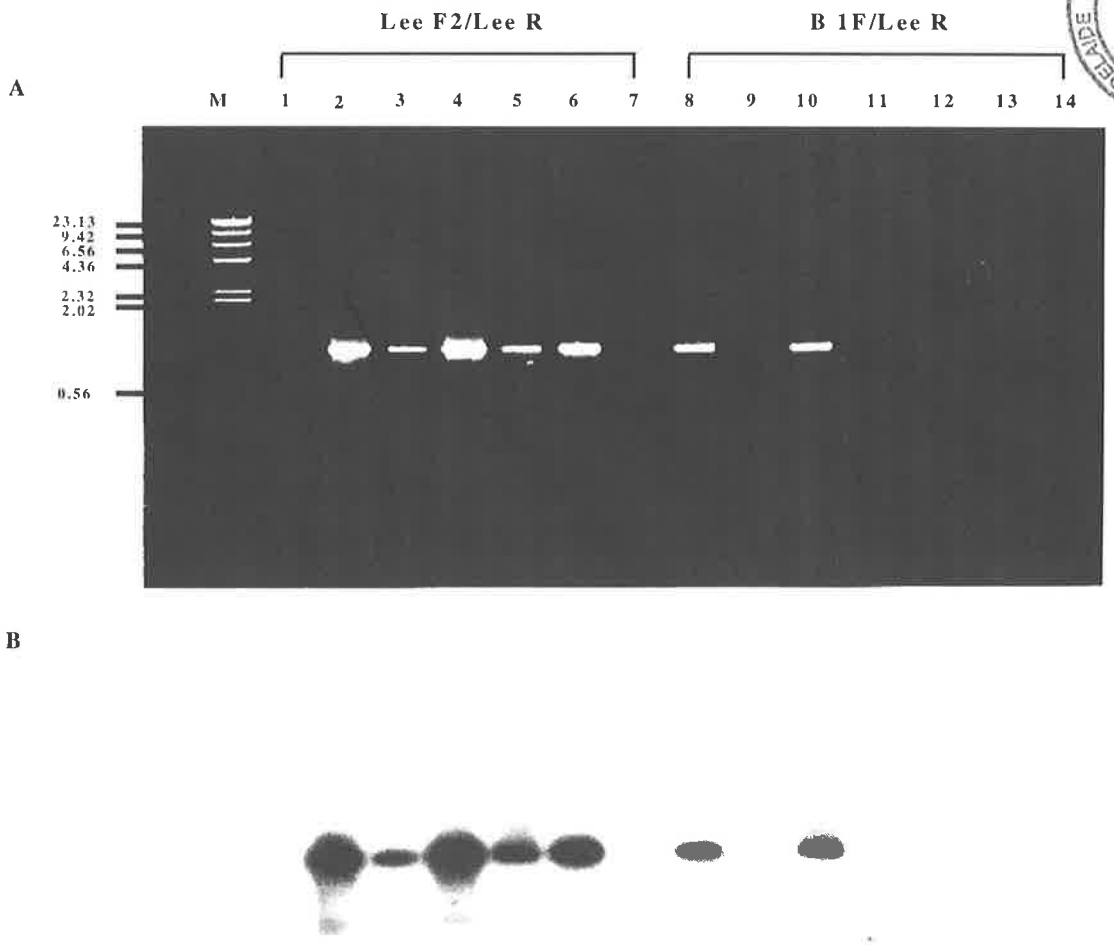
Figure 3.1.14 DNA sequence comparisons of *echid* β1-, *platy* β2- and *platy* β1-globin gene sequences. Exons are shown in boxes and the PCR primer sequences (and different nucleotide sites between *echid* β1 and *platy* β2) and *Eco*RI sites are shown in bold letters.

These results indicate that the Lee F2/Lee R primers specifically bind to *platy* β2 and *echid* β1, and the B F1/Lee R specifically bind to the *platy* β1. PCR amplification of monotreme genomic DNA using the two different primer sets showed that Lee F2/Lee R primers amplified a target sequence in platypus, echidna and New Guinea echidna genomic DNAs (Fig. 3.1.15 lane 3, 5 and 6, respectively). However, the B F1/Lee R primer set showed PCR products only when used with platypus DNA (Fig. 3.1.15 lane 10). No amplification was obtained when the two echidna genomic DNA samples were used (Fig. 3.1.15 lane 12 and 13). This result suggests that there are no sequences in the two echidna genomic DNAs that bind to the primer B F1.

The platypus genomic DNA used for PCR amplification, and that used for constructing the platypus genomic library PGL.1.1 from which the two lambda clones, λPL.1.1 and λPL.1.2 were isolated, were obtained from different animals. The PCR reactions with each primer set were carried out using identical conditions (Table 2.2.1). The sizes of the PCR products obtained using each primer set were expected to be slightly different (Table 3.1.1) but not sufficiently different to enable distinction to be made on an agarose gel.

The expected PCR product sizes were calculated from the *echid* β1, *platy* β1 and *platy* β2-globin gene sequences (Table 3.1.1). The PCR products were electrophoresed on 1% (w/v) agarose gel (Fig. 3.1.15 A) and Southern transferred to Hybond-N+ for hybridisation with the *echid* β1-globin gene probe, pTAG-2 (Fig. 3.1.15 B). It can be seen that the PCR products bind to the globin gene probe suggesting the amplified PCR products contain the β-globin gene sequences or at least the exon II region.

LIBRARY UNIVERSITY OF ADELAIDE



3.1.15 PCR amplification of *platy* β1 and *platy* β2 genes from monotreme genomic DNA with Lee F2/Lee R and B F1/Lee R primers.

A: 1% (w/v) agarose gel of PCR products from lambda clones and monotreme genomic DNA. Lanes 1-7 are the amplified PCR products with Lee F2/Lee R primers and the lanes 8-14 are the amplified PCR products with B F1/Lee R primers.
The tracks are: M- λ*Hind* III size DNA markers.

1 and 8	λPL. 1.1 DNA	2 and 9	λPL. 1.2 DNA
3 and 10	platypus genomic DNA	4 and 11	λEL. 1.1 DNA
5 and 12	Australian echidna genomic DNA	6 and 13	New Guinea echidna genomic DNA
7 and 14	negative controls		

B: Autoradiograph of a Southern blot of the gel shown in A. The probe used was pTAG-2.

Table 3.1.1 The predicted sizes of PCR products with Lee F2/Lee R and B F1/Lee R PCR primers and *Eco*RI digestion of the PCR products.

	Echid β1	*NGechid* β1	*Platy* β1	*Platy* β2
Lee F2/Lee R	1003 bp	1004 bp	-	1001
B1 F/Lee R	-	-	987	-
*Eco*RI digestion	67/856/80	65/856/81	67/920	67/853/81

3.1.8.2 *Eco*RI restriction digestion of PCR products

The *echid* β1- and *platy* β2-globin genes each contain two *Eco*RI sites which flank a 0.85 kb fragment that includes the 3' part of intron I, all of exon II, all of intron II and a 5' region of exon III (Fig. 3.1.14). However, the *platy* β1-globin gene has only a single *Eco*RI site (Fig. 3.1.14). Therefore, *Eco*RI digestion of the PCR products amplified with the two primer sets, Lee F2/Lee R and B F1/Lee R, is expected to produce fragments of different length (Table 3.1.1). Based on the *echid* β1, *NGechid* β1, *platy* β1 and *platy* β2 globin gene sequences, it is possible to the predict the *Eco*RI restriction fragment sizes of different PCR products, such as the PCR products, from Lee F2/Lee R (853-856 bp and two small 65-81 bp fragments) and from B F1/Lee R (987 bp and 67 bp fragments) (Table 3.1.1).

Fig. 3.1.16 A shows the results of electrophoresis of the *Eco*RI digested PCR products shown in Fig. 3.1.15. The Lee F2/Lee R PCR product from platypus genomic DNA gives an *Eco*RI fragment of 0.856 kb (Fig. 3.1.16 A lane 7) which is the same size as the *Eco*RI fragments obtained from the PCR products of λEL.1.1, λPL.1.2, Australian echidna and New Guinea echidna genomic DNAs (Fig. 3.1.16 A lane 3, 4, 5 and 6). The *Eco*RI fragment of the B F1/Lee R PCR products obtained from the λPL.1.1 and platypus genomic DNA is slightly bigger than 856 bp, i.e. about 987 bp (Fig. 3.1.16 A lane 9 and 10). This result supports the conclusion that the Lee F2/Lee R PCR product from platypus genomic DNA contains derived from the *platy* β2-globin gene.

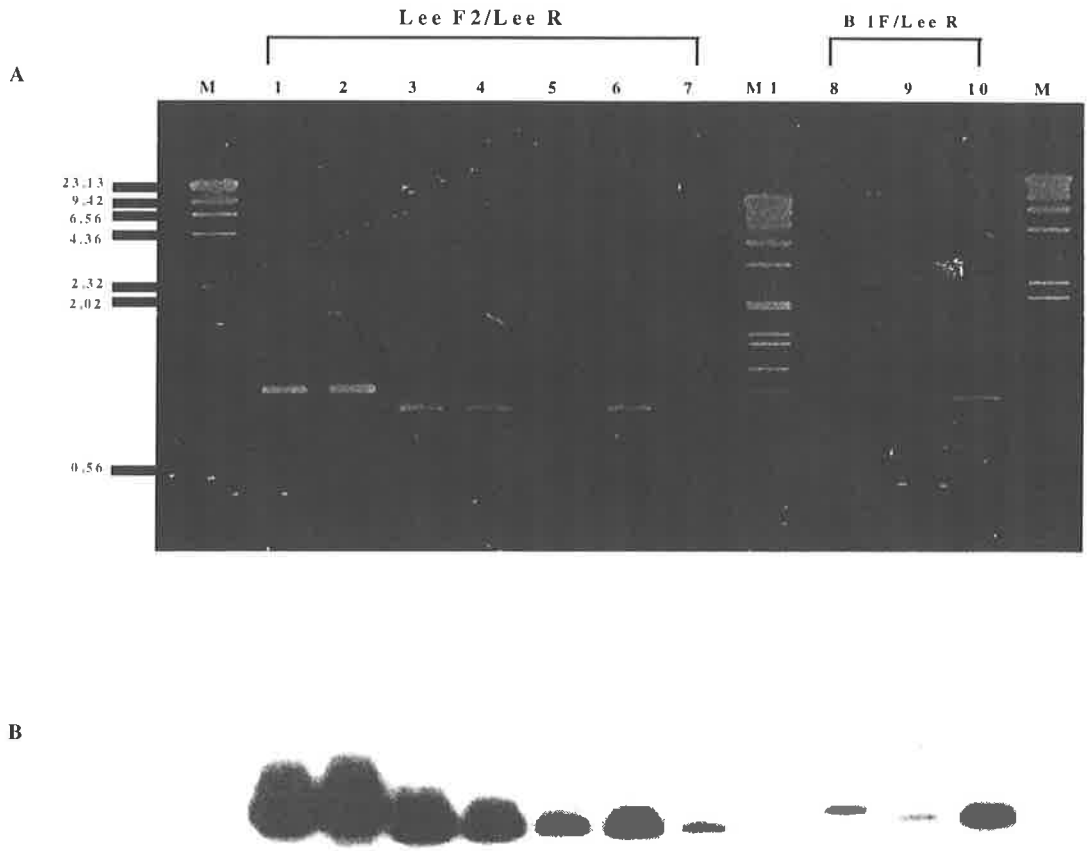


Figure 3.1.16 *Eco*RI digestions of PCR products amplified from lambda clones and monotreme genomic DNAs using the primer sets, Lee F2/Lee R and B F1/Lee R.

A: 1% (w/v) agarose gel of undigested and digested PCR products. The lanes 1-7 contains Lee F2/Lee R PCR products and the lanes 8-10 contain BF1/Lee R PCR products. The tracks are :

M- λ*Hind*III DNA size markers
M1- *Eco*RI Spp1 DNA size makers
1- uncut PCR product from λEL. 1.1
2- uncut PCR product from λPL. 1.2
3- cut PCR product from λEL. 1.1
4- cut PCR product from λPL. 1.2
5- cut PCR product from Australian echidna genomic DNA
6- cut PCR product from New Guniea echidna genomic DNA
7- cut PCR product from platypus genomic DNA
8- uncut PCR product from λPL 1.1
9- cut PCR product from λPL. 1.1
10- cut PCR produc from platypus genomic DNA

B: An autoradiograph of a Southern blot of the gel shown in A, hybridised with pTAG-2.

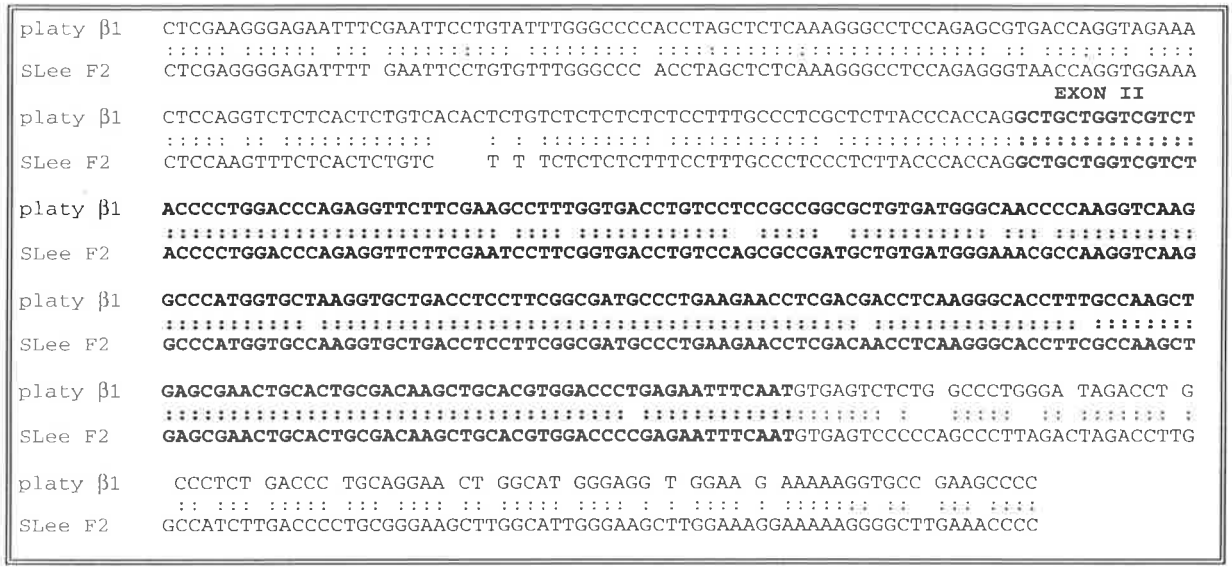
3.1.8.3 Direct sequencing of the Lee F2/Lee R PCR product

The Lee F2/Lee R PCR product from platypus genomic DNA was partially sequenced by automatic sequencing. The sequences from the Lee F2/Lee R PCR product which were named SLee F and SLee R (Appendix 4) were compared with *platy* β1 and *platy* β2 sequences (Fig. 3.1.17 A and B). Sequence alignments show that the partial PCR sequences have a higher level of sequence identity to *platy* β2-globin gene sequence (Fig. 3.1.17 B) than *platy* β1-globin gene sequence (Fig. 3.1.17 A). The sequence alignments reveal some mismatches and gaps between the partial PCR sequences and *platy* β2 sequences. This may be due to PCR amplification errors or technical errors accompanying the automatic sequencing. This result provides additional support for the hypothesis that the PCR product from Lee F2/Lee R is derived from part of the *platy* β2-globin gene.

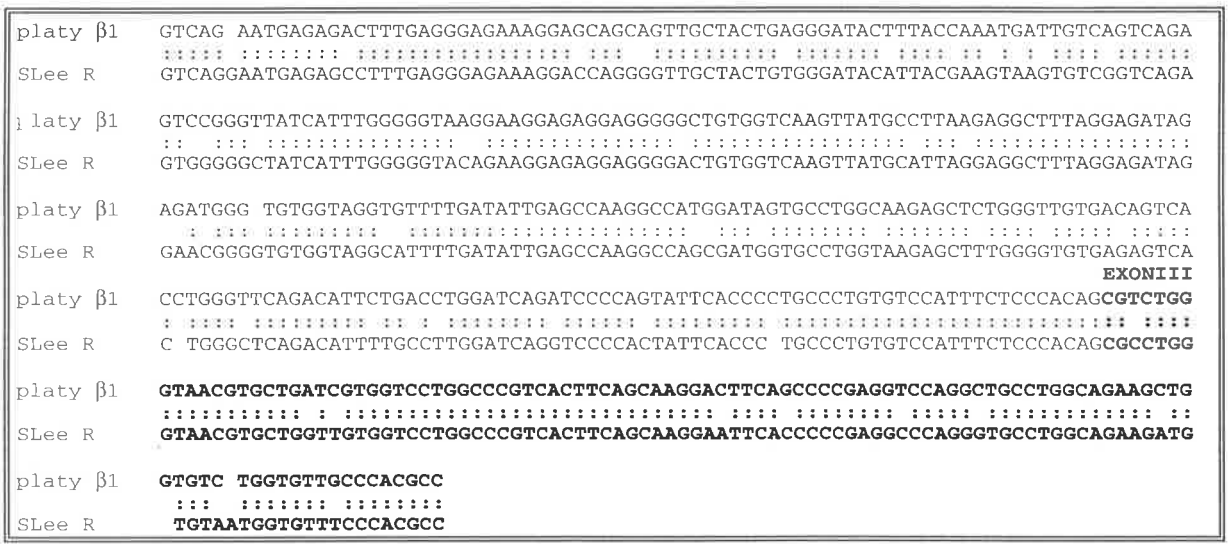
3.1.8.4 Alignments of "TGAA" repetitive boundary sequences

To provide further evidence that the *platy* β2-globin gene actually exists in the platypus, sequence comparisons were carried out between the partial sequences from one subclone of λEL.1.1 (pTAG-5) and *platy* β2. pTAG-5 contains a 500 bp *Bam*HI/*Xba*I insert which is located 2-3 kb downstream from the termination codon of *echid* β1 (Fig. 3.1.4). Interestingly, pTAG-5 and *platy* β2 contain 11 "TGAA" repetitive elements (Appendix 2 and Fig. 3.1.13, respectively). The "TGAA" repetitive sequences from the *platy* β2-globin gene are located in the 5' flanking region around 1 kb from the start codon (Fig. 3.1.13). Fig. 3.1.18 B shows the alignment of a 178 bp sequence containing the "TGAA" repeats in *platy* β2 and a partial sequence of pTAG-5 and a high level of sequence identity can be observed (Fig. 3.1.18 B). Although the compared sequences are very short, the high percentage of sequence identity needs to be explained. Furthermore, the "TGAA" repeats are located in different regions with respect to the globin genes: the repetitive element in λEL.1.1 is located 2-3 kb downstream of the *echid* β1-globin gene, while the repetitive element in λPL.1.2 is located in the 5' flanking region of the *platy* β2-globin gene (Fig. 3.1.18 A).

A1: *Platy* β1 and SLee F2



A2: *Platy* β1 and SLee R



B1. *Platy* β2 and SLee F2 sequences

```
platy β2    TATGACCTCGAGGGGAGAATTTTGAATTCCTGTGTTTGGGCCGACCTAGCTCTAAAAGGGGCCTCCAGAGGGTAACCAGG
            :::::: :::::::::::: ::::::::::::::::::::::::: :::::::::: :::::: ::::::::::::::::::
SLee F2     TATGAC TCGAGGGGAGA TTTTGAATTCCTGTGTTTGGGCCCACCTAGCTCTCAAAGGG CCTCCAGAGGGTAACCAGG
                                                                                    EXON II
platy β2    TGGAAACTCCAAGTTTCTCACTCTGTCTTTCTCTCTCTTTCCTTTGCCCTCCCTCTTACCCACCAGGCTGCTGGTCGTCT
            :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SLee F2     TGGAAACTCCAAGTTTCTCACTCTGTCTTTCTCTCTCTTTCCTTTGCCCTCCCTCTTACCCACCAGGCTGCTGGTCGTCT

platy β2    ACCCCTGGACCCAGAGGTTCTTCGAATCCTTCGGTGACCTGTCCAGCGCCGATGCTGTGATGGGAAACGCCAAGGTCAAG
            :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SLee F2     ACCCCTGGACCCAGAGGTTCTTCGAATCCTTCGGTGACCTGTCCAGCGCCGATGCTGTGATGGGAAACGCCAAGGTCAAG

platy β2    GCCCATGGTGCCAAGGTGCTGACCTCCTTCGGCGATGCCCTGAAGAACCTCGACAACCTCAAGGGCACCTTCGCCAAGCT
            :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SLee F2     GCCCATGGTGCCAAGGTGCTGACCTCCTTCGGCGATGCCCTGAAGAACCTCGACAACCTCAAGGGCACCTTCGCCAAGCT

platy β2    GAGCGAACTGCACTGCGACAAGCTGCACGTGGACCCCGAGAATTTCAATGTGAGTCCCC AGCC TTAGACTAGACCT G
            :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::: :::: :::::::::::: :
SLee F2     GAGCGAACTGCACTGCGACAAGCTGCACGTGGACCCCGAGAATTTCAATGTGAGTCCCCCAGCCCTTAGACTAGACCTTG

platy β2    GCCATCT GAC   TGCGGGA GCT GGCAT GGGA GCT GGAA GGAAA
            ::::::: :::   ::::::: ::: ::::: :::: ::: :::: :::::
SLee F2     GCCATCTTGACCCCTGCGGGAAGCTTGGCATTGGGAAGCTTGGAAAGGAAA
```

B2. *Platy* β2 and SLee R

```
platy β2    CCCCTGGGGAAGAAATGATGGGGTCCTG AGTGTGTGTTGGGCAAAGAGATAGCCTGGGGCATCT GAAGGAAATGAGTC
            ::::::::::::::::::::::::: ::::: :::::::::::::::::::::::::::::::: : :: ::::::::::::
SLee R      CCCCTGGGGAAGAAATGATGGGTTCCTGGAGTGTGTGTTGGGCAAAGAGATAGCCTGGGGCATTTTGAGGGAAATGAGTC

platy β2    AG AATGAGAGCCTTTGAGGGAGAAAGGACCAGGGGTTGCTACTGTGGGATACTTTACGAAG AAGTGTCGGTCAGAGTG
            :: ::::::::::::::::::::::::::::::::::::::::::::::::::::::::: :::::::: ::::::::::::::::
SLee R      AGGAATGAGAGCCTTTGAGGGAGAAAGGACCAGGGGTTGCTACTGTGGGATACATTACGAAGTAAGTGTCGGTCAGAGTG

platy β2    GGGGCTATCATTTGGGGGTACAGAAGGAGAGGAGGGGACTGTGGTCAAGTTATGCATTAGGAGGCTTTAGGAGATAGGAA
            ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SLee R      GGGGCTATCATTTGGGGGTACAGAAGGAGAGGAGGGGACTGTGGTCAAGTTATGCATTAGGAGGCTTTAGGAGATAGGAA

platy β2    CGGGGTGTGGTAGGCATTTTGACATTGAGCCAAGGCCAGCGATGGTGCCTGGTAAGAGCTCTGGGGTGTGAGAGTCACTG
            ::::::::::::::::::::::: :::::::::::::::::::::::::::::::::::: ::::::::::::::::::
SLee R      CGGGGTGTGGTAGGCATTTTGATATTGAGCCAAGGCCAGCGATGGTGCCTGGTAAGAGCTTTGGGGTGTGAGAGTCACTG
                                                                                    EXON III
platy β2    GGCTCAGACATTCTGCCTTGGATCAGGTCCCCACTATTCACCCTGCCCTGTGTCCATTTCTCCCACAGCGCCTGGGTAAC
            :::::::::::: :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
SLee R      GGCTCAGACATTTTGCCTTGGATCAGGTCCCCACTATTCACCCTGCCCTGTGTCCATTTCTCCCACAGCGCCTGGGTAAC

platy β2    GTGCTGGTCGTGGTCCTGGCCCGTCACTTCAGCAAGGAATTCACCCCCGAGGCCCAGGCTGCCTGGCAGAAGCTGGTGTC
            :::::::: :::::::::::::::::::::::::::::::::::::::::::::::::::: :::::::::::::: :: :::
SLee R      GTGCTGGTTGTGGTCCTGGCCCGTCACTTCAGCAAGGAATTCACCCCCGAGGCCCAGGGTGCCTGGCAGAAGATG TGTA

platy β2     TGGTGTTTCCCACGCC
             ::::::::::::::::
SLee R      ATGGTGTTTCCCACGCC
```

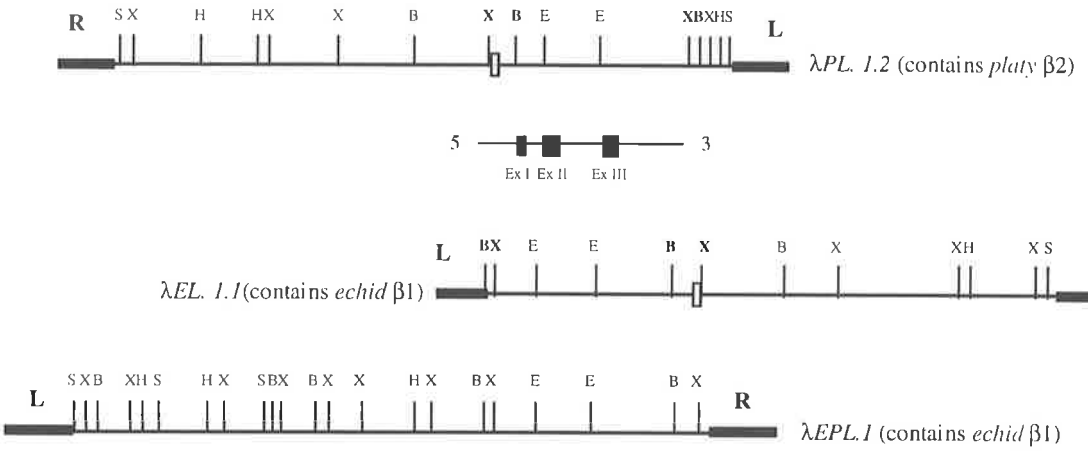
Figure 3.1.17 Alignments of the partial DNA sequences (SLee F and SLee R) obtained from the Lee F2/Lee R PCR product and *platy* β1 and *platy* β2 sequences, respectively.

A: Sequence alignments of the SLee F and the *platy* β1 (A1), and the SLee R and the *platy* β1 (A2).

B: Sequence alignments of the SLee F and the *platy* β2 (B1), and the SLee R and the *platy* β2 (B2).

The matched bases are shown as (:) and the exon regions are marked in bold letters.

A:



B :

```
Platy β2   5′ TTTAAACTCATTGCAAGCAGGGAATTTACCTGTTTATTGTTATATTGTATCTACTCCCAA
pTAG-5     5′ TTTAAACTCATTGCAAGCAGGGAATTTACCTGTTTATTGTTATATTGTA CT CTCCCAA
              ********************************************** ** *******

platy β2       GCGCTGAGTACAGTACTCTTGCACGTGGTAAGCTCTTAATAGATATGAATGAATGAATGA
pTAG-5         GCGCTGAGTACAGTACTCTTGCACGTGGTAAGCTCTTAATAAATATGAATGAATGAATGA
               ************************************** ******************

platy β2       ATGAATGAATGAATGAATGAATGAATGAATCATAGGCTGTTGCAGC TTCCTTCAATC  3′
pTAG-5         ATGAATGAATGAATGAATGAATGAATGAATGA AGGCTGTTGCAGCCTTCCTTCAATC  3′
               ****************************** * ************* ***********
```

Figure 3.1.18

A: Comparison of the restriction maps of λEL.1.1, λEPL.1 and λPL.1.2. The location of **"TGAA"** repetitive regions are indicated by hatched boxes. The *echid* β1 and *platy* β2 globin genes are drawn to approximate scale and their three exons are indicated by black boxes. L and R refer to the left and right arms of λ*GEM*-11. Restriction sites are: B = *Bam*HI, E = *Eco*RI, H = *Hindl*II, S = *SalI*, X = *XbaI*.

B: Alignment of **"TGAA"** repeats and their boundary sequences in the *platy* β2 and partial sequences from a plasmid subclone pTAG-5 of λEL.1.1. The conserved bases are represented by "*" and **"TGAA"** repeats are shown in bold letters.

The locations of the *Bam*HI and *Xba*I restriction enzyme sites that flank the "TGAA" repeat regions in λPL.1.2 and λEL.1.1 are reversed in their orientation (Fig. 3.1.18 A). A comparison between the restriction maps of λPL.1.2 and λEPL.1 (which is a lambda clone that overlaps with λEL.1.1 and contains an extended 5' flanking region from *echid* β1) shows that the restriction sites occur with a similar arrangement around coding regions of the β-globin genes, but differ significantly in the upstream regions (Fig. 3.1.18 A).

All these data suggest that the *platy* β2-globin gene does exist in the platypus and that the coding and non-coding regions are very similar in sequence to the *echid* β1-globin gene, but the regions outside the β-globin gene such as the 5' and 3' flanking regions are very different. This topic will be discussed further in Section 4.2.

3.1.9 PCR amplification and identification of a β-globin gene from the New Guinea echidna

As shown in Fig. 3.1.15 (lane 6), a putative β-globin gene was PCR amplified from New Guinea echidna genomic DNA using the Lee F2/Lee R primers. The PCR product was not expected to contain all regions of the β-globin gene (see Fig. 3.1.14) as the Lee F2 primer sequence is located within exon I of *echid* β1 (Fig. 3.1.14). It was necessary to use a different primer set (707 F/737 R) to amplify the entire β-globin gene. PCR amplification using the 707 F/737 R primer set amplified two fragments: one of 1.5 kb and the other 0.58 kb in length (Fig. 3.1.19 A lane 1 and 2). The PCR conditions and primer sequences are shown in Table 2.2.1. Southern analysis using pTAG-2 (an 856 bp *echid* β1-globin probe) was used to identify the band containing the β-globin gene among these products (Fig. 3.1.19 B). Only the 1.5 kb fragment hybridised to the globin gene probe (Fig. 3.1.19 B lane 1 and 2).

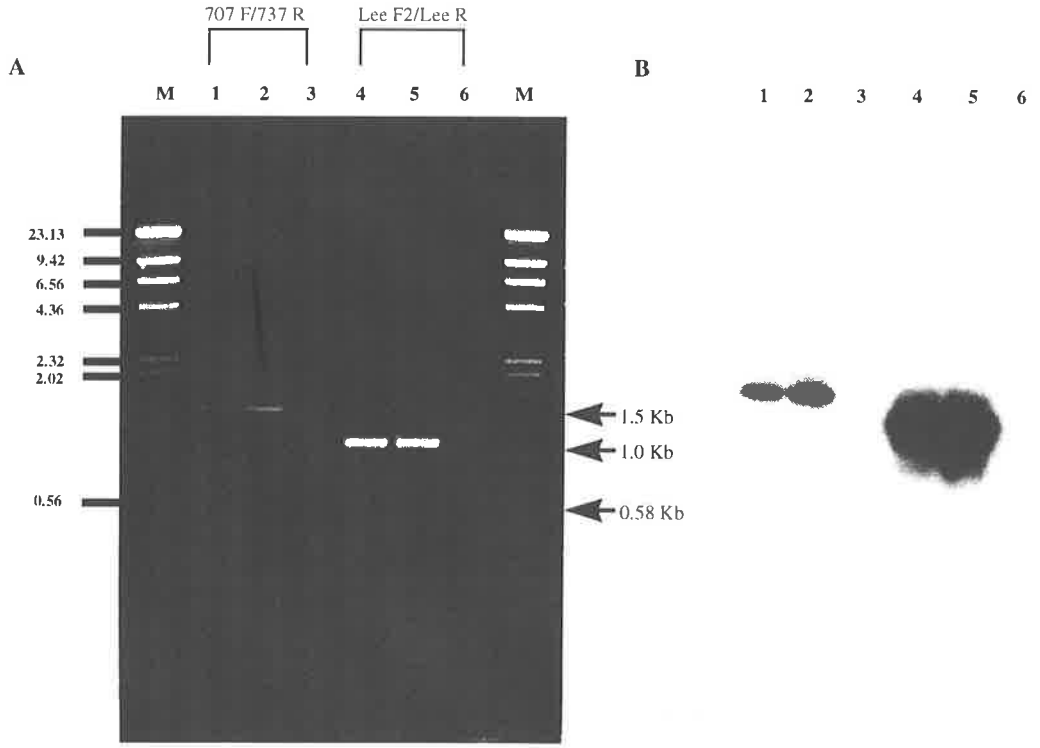


Figure 3.1.19 PCR amplification of a putative β-globin gene from the New Guniea echidna.

A: 1% (w/v) agarose gel showing the PCR products from New Guinea echidna genomic DNA. The tracks are:

M: λ*Hind*III DNA markers,

Lane 1 and 2: PCR products amplified with 707 F/737 R primers

Lane 4 and 5: PCR products amplified with Lee F2/Lee R primers

Lane 3 and 6: negative control of 707 F/737 R and Lee F2/Lee R, respectively.

B: Autoradiograph of a Southern blot of the gel shown in "A". The probe used was pTAG-2.

To confirm that this PCR product contained a β-globin gene, the fragment was sequenced using a direct sequencing system, and independently, the Lee F2/Lee R PCR product was also sequenced. A sequence comparison of the two PCR products showed them to contain an identical region and suggested that the 1.5 kb fragment contains the entire β-globin gene. Therefore, the 707 F/737 R PCR product was sequenced and the β-globin gene contained in this sequence was named *NGechid* β1 (New Guinea echidna β1-globin gene) (Fig. 3.1.20). Insufficient 5' sequence was obtained to cover the 5' promoter elements. However, initiation and termination codons, open reading frames and consensus intron/exon splice junctions AG/GT, as well as a putative poly (A) adenylation signal 91 bp downstream of the termination codon, have been identified (Fig. 3.1.20). The lengths of the first and second introns are 147 bp and 460 bp, respectively. The three exons encode a β-polypeptide of the expected length of 146 amino acid residues. A comparison of the amino acid sequences shows that NGechid β1-globin has 6 and 15 amino acid residue differences to echid β1- and platy β1-globin polypeptides, respectively (Fig. 3.1.21). There are no amino acid differences at α1-β1, α1-β2, and haem contact sites between NGechid β1-globin and either echid β1- or platy β1-globin. However, at DPG binding sites, NGechid β1-globin differs from echid β1 at one amino acid residue (51) and from platy β1 at three amino acid residues (51, 122 and 125) (Fig. 3.1.21).

```
         10         20         30         40         50         60
                                                    INIValHisLeu
CAGTCTTTGCACATCTGCATCTTCAAGCAAACTTGCAAATCCGACAAAATGGTGCATTTG
SerGlySerGluLysThrAlaValThrThrLeuTrpGlyHisValAsnValAsnGluLeu
AGTGGTTCTGAGAAGACTGCTGTCACCACCCTGTGGGGGCATGTGAACGTCAATGAACTC
GlyGlyGluAlaLeuGlyAr
GGTGGCGAGGCCCTTGGCAGGTATGACCTCGAGGGGAGAATTTTGAATTCCTGCATTTGG
GCCCACCTAGCTCTCAAAGGGCCTCCAGAGGGTAACCAGGTGGAAACTCCAAGTTTCTCA
                                                    gLeuLeuValValT
CTCTGTCTTTCTCTCTCTTTCCTTTGCCCTCGCTCTTACCCACCAGGCTGCTGGTCGTCT
yrProTrpThrGlnArgPhePheGluSerPheGlyAspLeuSerSerAlaAsnAlaValM
ACCCCTGGACCCAGAGGTTCTTCGAATCCTTTGGTGACCTGTCCAGCGCCAATGCTGTGA
etGlyAsnProLysValLysAlaHisGlyAlaLysValLeuThrSerPheGlyAspAlaL
TGGGAAACCCCAAGGTCAAGGCCCATGGTGCCAAGGTGCTGACCTCCTTCGGCGATGCCC
euLysAsnLeuAspAspLeuLysGlyThrPheAlaLysLeuSerGluLeuHisCysAspL
TGAAGAACCTCGACGACCTCAAGGGCACCTTCGCCAAGCTGAGCGAACTGCACTGCGACA
ysLeuHisValAspProGluAsnPheAsn
AGCTGCACGTGGACCCCGAGAATTTCAATGTGAGTCCCCAGCCTTAGGGTAGACCTGGCC
ATCTGACTTCCGGAGCTGGCATTGGAGCTGGAAGGAAAAGGCTGAAGCCCCCTGGGGGAG
AAATGATGGGGTTCTGAGTGTGTGTTGGGCAAAGAGATAGCCTGGGGCATCTGAAGGAAA
TGAGTCAGAATGAGAGCCTTTGAGGGAGAAAGGACCAGGGGTTGCTACCGTGGGATACTT
TACGAAGTAAGTGTCACTCAGAGTGGGGGCTATCATTTGGGGGTAGAGAAGGAGAGGAGG
GGACTGTGGTCAAGTTATGCATTAGGAGGCTTTAGGAGATAGGGATGGGGTGTGGTAGGC
ATTTTGATATTGAGCCAAGGCCAGCGATGGTGCCTGGCAAGAGCTCTGGGGTGTGAGAGT
CACTGGGCTCAGACATTCTGCCTTGGATCAGGTCCCCACTATTTACCCTGCCCTGTGTCC
                ArgLeuGlyAsnValLeuValValValLeuAlaArgHisPheSerLy
ATTTCTCCCACAGCGCCTGGGTAACGTGCTGGTCGTGGTCCTGGCCCGTCACTTTAGCAA
sGluPheThrProGluAlaGlnAlaAlaTrpGlnLysLeuValSerAspValSerHisAl
GGAATTCACCCCCGAGGCCCAGGCTGCCTGGCAGAAGCTGGTGTCTGATGTTTCCCACGC
aLeuAlaAspLysTyrHisTER
CCTGGCCGACAAGTACCACTGAGTCCTGAGCTGCCCCAGTGACTTCTGCGCCAGGTCCCA
TGACCTGCCCGTGGGAGTGCAGCACCACAGGCCCCTGAGAGCCAGCCCGGCCCAATAAAG
AGCTTTTAGTCCAGTTTCTCCTCTGCTCTGCCTCTCTCTGTGCG
```

Figure 3.1.20 Complete nucleotide sequence and corresponding amino acid sequence of the New Guinea echidna β1-globin gene. A putative CAP site (CAG), translation initiation signal (ATG), termination signal (TGA) and poly (A) (AATAAA) signal in the 3' flanking region are shown in bold. The conceptually translated amino acid sequence gives a full length polypeptide (146 amino acids). The lengths of the 1st and 2nd introns of the gene are 146 bp and 464 bp, respectively. The length of the entire DNA sequence is 1,244 bp. The sequence was determined from both DNA strands.

```
NGECHID β1:  VHLSGSEKTAVTTLWGHVNVNELGGEALG RLLVVYPWTQRFFESF
ECHID β1:    VHLSGSEKTAVTNLWGHVNVVNELGGEALG RLLVVYPWTQRFFESF
PLATY β1:    VHLSGGEKSAVTNLWGKVNINELGGEALG RLLVVYPWTQRFFEAF

             GDLSSANAVMGNPKVKAHGAKVLTSFGDALKNLDDLKGTFAKLSEL
             GDLSSADAVMGNAKVKAHGAKVLTSFGDALKNLDNLKGTFAKLSEL
             GDLSSAGAVMGNPKVKAHGAKVLTSFGDALKNLDDLKGTFAKLSEL

             HCDKLHVDPENFN RLGNVLVVVLARHFSKEFTPEAQAAWQKLVSD
             HCDKLHVDPENFN RLGNVLVVVLARHFSKEFTPEAQAAWQKLVSG
             HCDKLHVDPENFN RLGNVLIVVLARHFSKDFSPEVQAAWQKLVSG

             VSHALADKYH
             VSHALAHKYH     6/146
             VAHALGHKYH    15/146
```

Figure 3.1.21 Alignment of the 146 amino acid residues of monotreme β-globin putative gene products. Residues are abbreviated by the standard one letter code. Differences in the sequences are shown in bold letters. NGechid β1 differs from echid β1 at 6 sites, and from platy β1 at 15 sites. Note: The putative platypus gene a *platy* β2 translates into a globin chain that is identical to that from *echid* β1.

3.1.10 Comparisons involving the four monotreme β-globin genes

3.1.10.1 DNA sequence comparisons

Fig. 3.1.22 shows the alignment of the four monotreme β-globin DNA sequences identified in the three monotreme species examined in this thesis. The exons (92-99%) and introns (83-97%) show a high level of sequence identity among the four genes (Table 3.1.2).

Table 3.1.2 Sequence identity (%) in the three exons (above) and two introns (below) of the four monotreme β-globin genes.

	Echid β1	*NGechid* β1	*Platy* β1	*Platy* β2
Echid β1		97	92	99
NGechid β1	95		92	98
Platy β1	83	83		93
Platy β2	97	95	83	

A sequence alignment of the 5' and 3' flanking regions of three of the four genes is shown in Appendix 5 (The *NGechid* β1-globin gene was excluded because of insufficient sequence data from these regions). The levels of sequence identity are shown in Table 3.1.3. It can be seen that the *echid* β1 sequences have higher identity to *platy* β2 than to *platy* β1 in both flanking regions.

Table 3.1.3 Sequence identity (%) in the 5' (above) and 3' (below) flanking regions of three monotreme β-globin genes.

	Echid β1	*Platy* β1	*Platy* β2
Echid β1		69	96
Platy β1	66		81
Platy β2	93	61	

Fig. 3.1.22

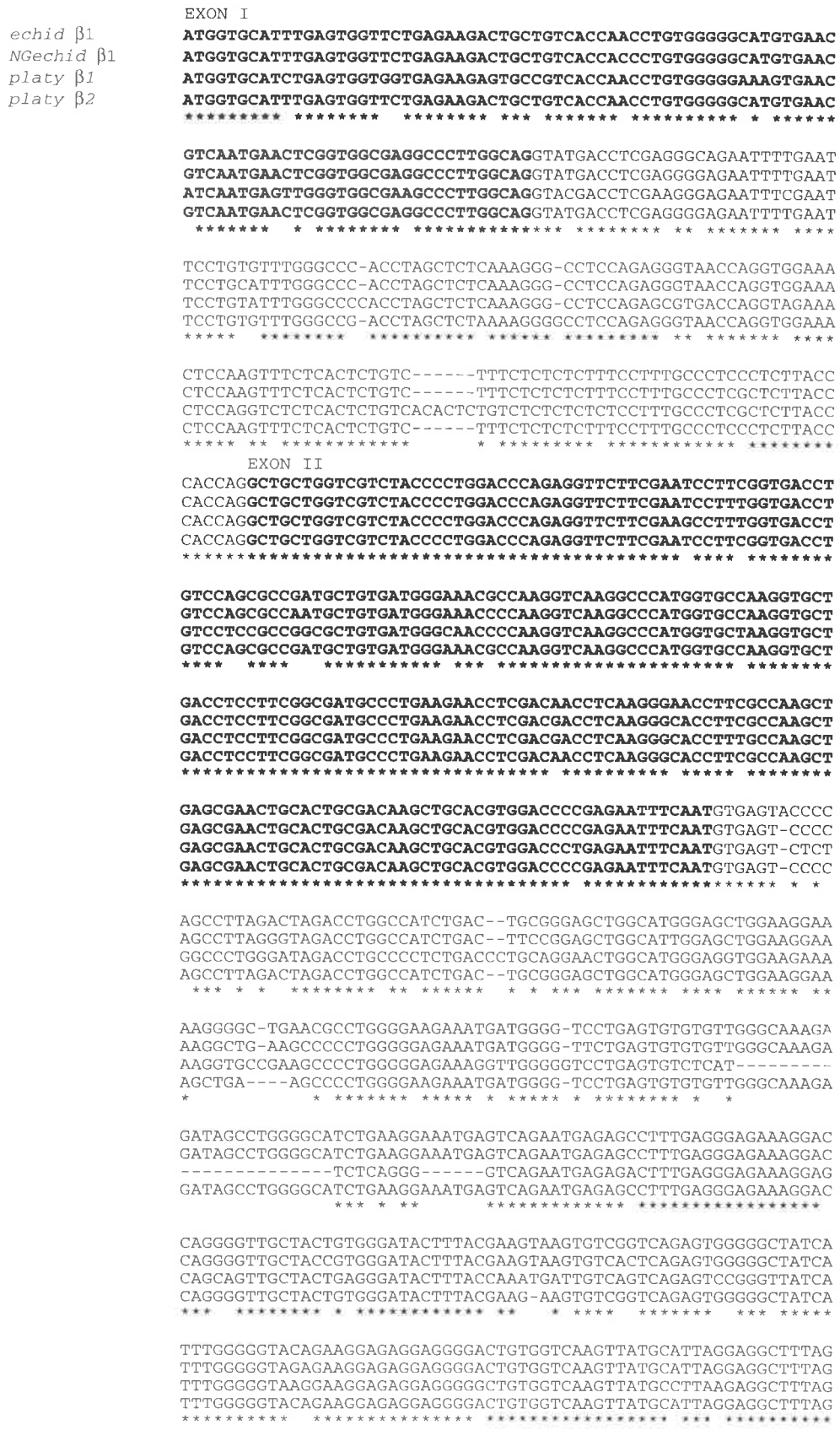




Figure 3.1.22 Multiple sequence alignment of four monotreme β-globin genes using the CLUSTALW computer program. The three exons are shown in bold letters, the conserved bases are represented by "*". Gap opening and extension penalties were "--" and "--".

3.1.10.2 Intron lengths and sequences

The intron lengths in each of the four monotreme β-globin genes are shown in Table 3.1.4. Intron I ranges in length from 146-153 bp, and intron II ranges in length from 440 to 465 bp. Data on β-globin intron lengths from a range of other species were extracted from GenBank (Appendix 6) and plotted as a frequency distribution (Fig. 3.1.23). Intron I lengths of the monotreme β-globin genes are larger than those of most other β-like globin genes, and intron II lengths are smaller than all other β-like globin genes. The intron lengths of marsupial β-like globin genes (ε-globin and β-globin from each of two species, *Didelphis virginiana* and *Sminthopsis crassicaudata*) are smaller (Intron I) and larger (Intron II) than the corresponding introns of the monotreme β-globins. The marsupial Intron II lengths are the largest described for any species (Appendix 6 and Fig. 3.1.23).

Table 3.1.4 Intron lengths of β-globin genes in monotremes.

	Echid β1	*NGechid* β1	*Platy* β1	*Platy* β2
Intron I	146	146	153	147
Intron II	465	464	440	460

To search for sequence identity between corresponding introns of different monotreme β-globin genes, and between the introns of monotreme β-globin genes and those of other mammalian β-globin (embryonic-like and adult-like) genes, pair-wise sequence comparisons were carried out using the GCG sequence analysis software packages COMPARISON (with a stringency setting of 12) and DOTPLOT (Appendix 7).

Intron sequences, being comparatively free from the effects of natural selection (in relation to coding sequences) evolve at a relatively rapid and uniform rate (Czelusniak *et al.* 1990; Mouchiround and Benardi, 1993). Under such conditions, significant sequence identity between pairs of homologous intron sequences is often difficult to detect, and the visual approach provided by the "dot plot" method is useful.

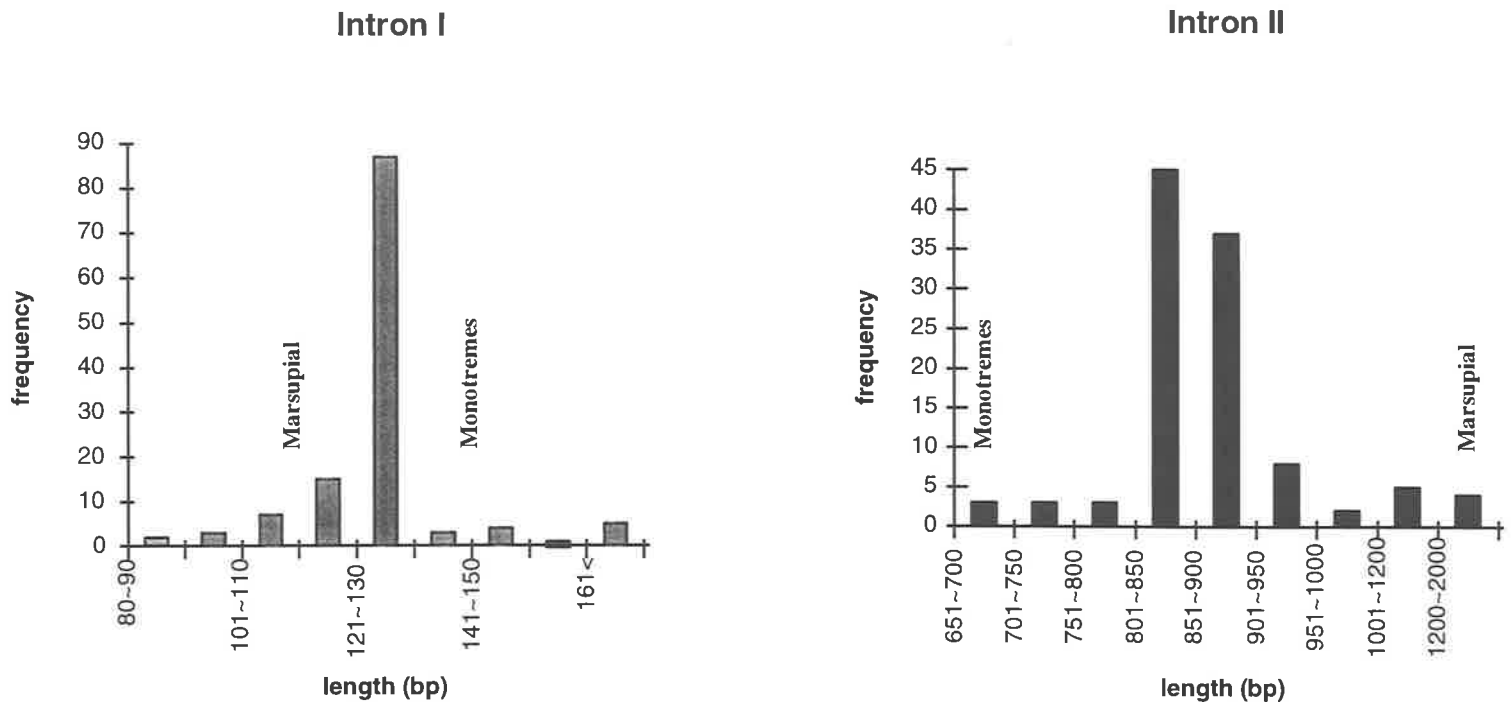


Figure 3.1.23 Length distributions of β-like globin gene introns. The genes and the intron lengths used in the analysis are listed in Appendix 6. The approximate lengths of the marsupial and monotreme introns are indicated.

To examine the rate at which sequence identity dissipates over time, orthologous globin intron I sequences from pairs of species that separated from one another at approximately known times, were compared (Appendix 8 - adapted from Wheeler, 1997). It can be seen that significant sequence identity remains after 80 million years, and that by 135 million years some sequence identity is still apparent. However, after about 200 million years, little if any significant sequence identity can be detected.

It can be seen (Appendix 7) that no significant sequence identity exists between corresponding introns in the monotremes and other mammalian β-globin genes. However, a high level of sequence identity exists amongst the monotreme β-globin intron sequences.

3.1.10.3 5' promoter regions

The location and sequences of 5' promoter motifs of the monotreme β-globin genes are highly conserved between the two monotreme species. In relation to the putative CAP sites, "CACCC" is located at -92 (*echid* β1 and *platy* β1) and -93 (*platy* β2), "CCAAT" is located at -78 (*echid* β1 and *platy* β1) and -75 (*platy* β2), and "ATA" is located at -33 in all three genes. In the 5' promoter regions of β-globin genes (embryonic-like and adult-like) from other mammalian species, a number of presumptive synapomorphies have been interpreted as evidence for orthology. The embedded sequences of the "ATA", "CCAAT" and "CACCC" boxes and the distances between them are well conserved among the orthologous β-globin genes (embryonic-like and adult-like specificities). There is also evidence that adult-like β-globins have two "CACCC" boxes, while the embryonic-like β-globins have a single "CACCC" box (Koop and Goodman, 1988; Cooper and Hope, 1993).

Nucleotide sequence alignments were carried out to examine the 5' promoter region of the monotreme β-globin genes for regions of sequence identity with each other, and with homologous regions from other mammalian β-globin genes. The other mammalian promoter sequences used were obtained from Koop and Goodman (1988), Cooper and Hope (1993) and Cooper *et al.* (1996).

Figure 3.1.24 Comparison of 5' promoter elements in monotreme β-globin genes and various marsupial and eutherian β-like globin genes. Blocks of sequences are shown, each containing a conserved sequence motif (in bold) that has been implicated in globin gene regulation (from Koop and Goodman, 1988). Dots represent regions of sequences that are not shown. The numbers in the columns on the right indicate the distances, upstream from the ATA box, of the various sequence motifs. Opossum, Dunnart, and Rabbit β-globin genes have additional CACCC boxes (not shown) at -93, -93, and -81 respectively.

Human β:	GACCT**CACCC**TGTGG.AGCCA**CACCC**TAGG.TTGG**CCAAT**CTAC.GGGC**ATA**AAAGTC	-74, -59, -45
Rabbit β:	GACCT**CACCC**TGCAG.AGCCA**CACCC**TGGT.TTGG**CCAAT**CTAC.GGGC**ATA**AAAGGC	-74, -59, -45
Goat β^A:	AGCCT**CACCC**TGTGG.AACCA**CAACT**TGGC.CGAG**CCAAT**CTGC.GGGC**ATA**AAAGGA	-74, -59, -45
Mouse β1:	AGCCT**GATTC**CGTAG.AGCCA**CACCC**TGGT.AGGG**CCAAT**CTGC.GAGC**ATA**TAAGGT	-75, -60, -46
Opossum β:	GGCCT**CATTC**CAGGT.AGCCA**CACCC**CAAC.TTAG**CCAAT**AGAC.AGGG**ATA**AAAGGC	-76, -60, -46
Dunnart β:	GGCTT**CATTC**CAGAT.AGCCG**CACCC**CAAC.AGAA**CCAAT**GGAC.AGGA**ATA**AAAGGC	-76, -60, -46
Echid β1:	TCACG**CCCCC**AGTGT.GACCG**CACCC**CTGT.TGGG**CCAAT**CAGC.GGGG**ATA**AAAAGC	-75, -59, -45
Platy β2:	ACAGA**CCCCC**AGTGT.GACCG**CACCC**CTGT.TGGG**CCAAT**CAGC.GGGG**ATA**AAAAGC	-76, -60, -46
Platy β1:	TCACG**CCCTC**AGCGG.GACCG**CACCC**CAAT.TGGG**CCAAT**CATC.GGGG**ATA**AAAGGC	-74, -59, -45
Dunnart ε:	AGCAC**CACCC**CCAAA. .CTGA**CCAAT**AGCT.GGAA**ATA**AAACAG	-72, -53
Opossum ε:	GACTC**CACCC**CACAG. .TTGA**CCAAT**AGCT.GCAA**ATA**AAAGGC	-87, -57
Mouse ε^Y:	GGCTC**CACCC**ATGAG. .CTGA**CCAAT**GGCT.AAGA**ATA**AAAGGC	-83, -55
Goat ε^I:	GACTC**CACCC**CTGAG. .TTGA**CCAAT**GACT.AAGA**ATA**AAAGGC	-84, -55
Rabbit ε:	GGCTC**CACCC**ATGAA. .TTGA**CCAAT**GACT.GGGA**ATA**AAAGGC	-84, -55
Human ε:	GACTC**CACCC**CTGAG. .TTGA**CCAAT**GACT.AAGA**ATA**AAAGGC	-83, -55

The alignments are shown in Figure 3.1.24. It can be seen that the monotreme β-globin promoter elements are more similar in number and location to the marsupial and eutherian β-globin gene promoter regions than to those of the marsupial and eutherian ε-globin genes (Fig. 3.1.24). The first "CACCC" box is not well conserved in sequence among the orthologous β-globin genes.

3.1.10.4 Amino acid sequence comparisons

Pairwise comparisons were carried out to search for amino acid sequence identity amongst the β-globin chains encoded by the three monotreme genes and β-like globin chains from various other mammals and birds (Table 3.1.5). It had previously been suggested that the monotreme β-globins differ from eutherian β-globins at fewer sites than do the marsupial β-globins (Whittaker and Thompson, 1975).

The average number of amino acid differences between pair of genes is shown below (taken from table 3.1.5).

	Monotreme β	Marsupial β	Eutherian β	Marsupial ε	Eutherian ε
Monotreme β	-	41	35.9	35.3	35.7
Marsupial β		-	44.6	40.8	41.1
Eutherian β			-	37.4	37.4
Marsupial ε				-	27
Eutherian ε					-

As shown above, the monotreme β-globins have about the same average number of amino acid residue differences to eutherian β-globins, eutherian ε-globins, and marsupial ε-globins. However, the monotreme β-globins differ most from the marsupial β-globins. The marsupial ε-globins reveal a lower number of amino acid residue differences to the eutherian ε-globins than the eutherian β-globins. The marsupial β-globins have a higher number of differences to eutherian β-like globins (ε- and β-globins) than to monotreme β-globins.

Table 3.1.5 β-like globin polypeptide sequence comparisons. The numbers of sites that contain different amino acid residues between pairs of globin chains are shown. Species and gene abbreviations are described in Sections 2.1.10 and 3.3.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. ECHID β1		6	14	42	43	37	35	31	39	31	36	46	37	31	34	34	39	34	31
2. NGECHID β1			15	42	43	41	37	34	41	32	39	46	39	32	37	35	41	36	33
3. PLATY β1				42	43	39	36	34	40	33	40	45	39	34	34	37	43	34	34
4. CHICKEN β					48	42	47	45	49	48	51	21	43	30	40	37	41	40	36
5. OPOSSUM β						33	46	46	49	45	48	48	48	41	44	45	47	45	42
6. DUNNART β							41	42	44	41	44	41	40	34	36	39	41	36	36
7. GALAGO β								16	29	20	37	49	39	32	38	36	37	37	33
8. HUMAN β									29	14	28	50	39	32	37	36	39	36	33
9. MOUSE β1										28	40	51	43	36	37	40	37	39	32
10. RABBIT β											29	47	37	31	34	37	36	39	30
11. GOAT β^{A}												54	46	39	41	44	43	44	38
12. CHICKEN ε													39	36	39	38	42	41	37
13. OPOSSUM ε														29	29	32	32	30	28
14. DUNNART ε															24	23	29	24	19
15. GALAGO ε																16	26	18	11
16. HUMAN ε																	27	21	15
17. MOUSE ε^{y}																		29	21
18. RABBIT ε																			21
19. GOAT ε^{II}																			

3.1.10.5 G+C content

Exon DNA sequences were extracted from 31 β-like globin genes (species and locus identifications from GenBank and EMBO are listed in Appendix 9 and Section 2.1.10, respectively). The average values of G+C and A+T contents were calculated from these sequences and compared with the average values of G+C and A+T contents of the four monotreme β-globin exon sequences.

Fig. 3.1.25 shows the distribution of the average values of G+C and A+T contents in each β-globin gene. The monotreme β-globin genes (Fig. 3.1.25, 1-4 on the abscissa) show a higher G+C content than all the other mammalian β-globin genes. The chicken β-globin genes (ε-globin and β-globin) (genes 5 and 17 on the abscissa) also have a high G+C content. The mammalian adult β-globin genes generally show a higher G+C content than embryonic β-globin genes (Fig. 3.1.25).

3.1.11 Southern analyses of monotreme genomic DNA

Southern blots of *Eco*RI digested genomic DNA from monotremes revealed two fragments (lengths 10 kb and 8.7 kb in the Australian echidna and 10 kb and 6.8 kb in the platypus) that showed specific hybridisation to the probe pSG2H (containing the entire second exon of ε-globin gene from *S. crassicaudata*- Section 2.1.8) under conditions of low stringency (Fig. 3.1.26 A) (Shroff, 1990). The λEL.1.1 restriction map (Fig. 3.1.4) shows two *Eco*RI restriction sites that are expected to give rise to a 856 bp *Eco*RI fragment containing exon II of *echid* β1. However, a fragment of this length is not seen in Fig 3.1.26 A. The most likely explanation for this apparent inconsistency is that the 856 bp DNA fragment was electrophoresed off the gel and, therefore, was not be detected. Further Southern analyses were carried out on *Eco*RI digested genomic DNA from the Australian echidna and platypus using the probe pTAG-2 (Fig. 3.1.26 B).

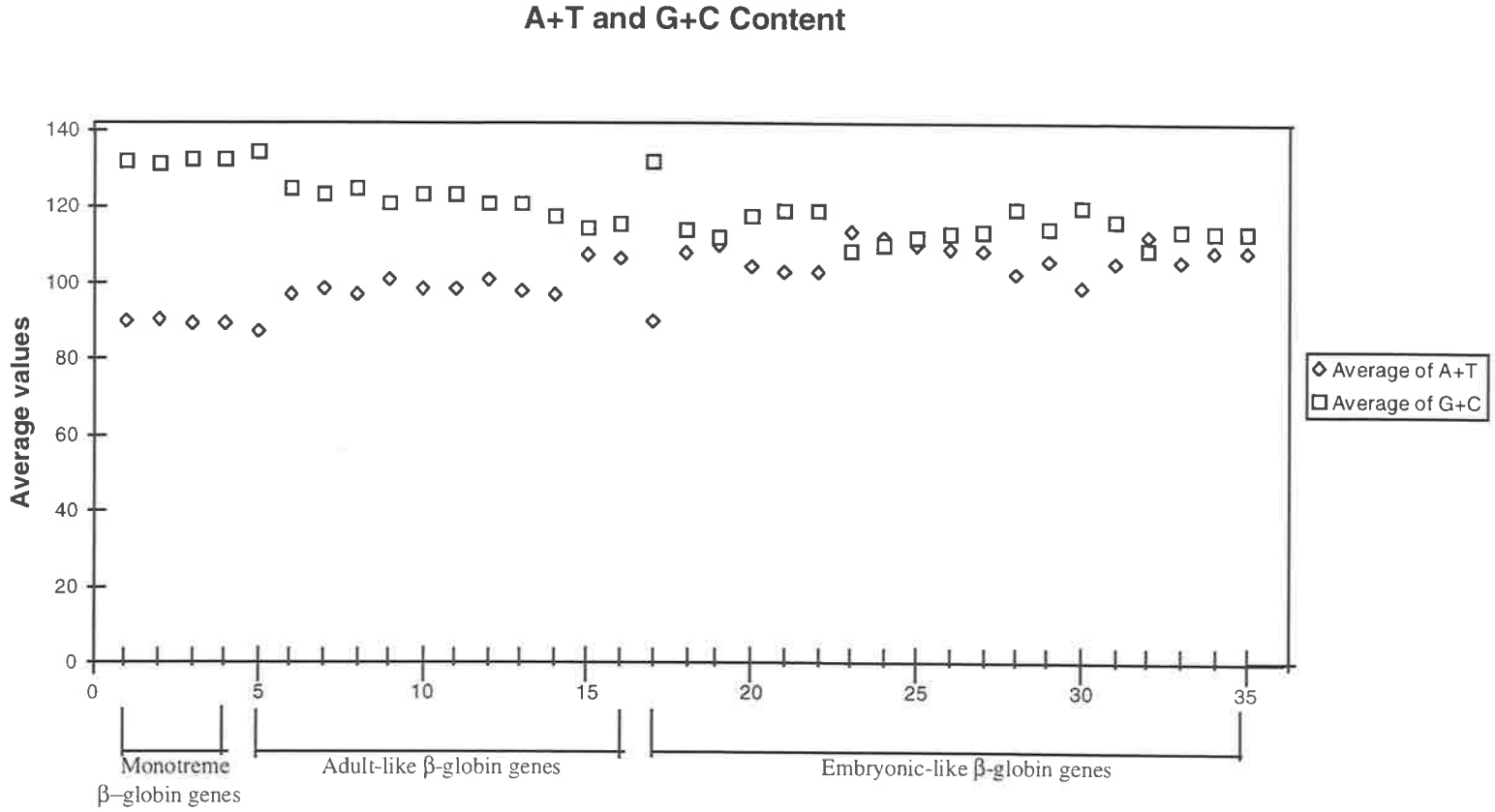


Figure 3.1.25 A comparison of A+T and G+C contents of β-like globin genes from mammals and birds. The β-like globin genes included and numbered on the abscissa are: 1 - Australian echidna β1, 2 - New Guinea echidna β1, 3 - platypus β1, 4 - platypus β2, 5 - chicken βh, 6 - galago (primate) β, 7 - galago δ, 8 - human β, 9 - human δ, 10 - mouse β1, 11 - mouse β2, 12 - rabbit β, 13 - goat β^A, 14 - goat β^C, 15 - opossum β, 16 - dunnart β, 17 - chicken ε2, 18 - galago ε, 19 - galago γ, 20 - human ε, 21 - human γ^A, 22 - human γ^G, 23 - mouse h0, 24 - mouse h1, 25 - mouse ε^y, 26 - rabbit ε, 27 - rabbit γ, 28 - goat ε^I, 29 - goat ε^{II}, 30 - goat β^F, 31 - goat ε^V, 32 - human ψ, 33 - rabbit ψ, 34 - opossum ε, 35 - dunnart ε.

Each of the digested genomic DNA samples revealed two large bands (10 kb and 8.7 kb for Australian echidna and 10 kb and 6.8 kb for platypus) which are the same band lengths as those shown in Figure 3.1.26 A. On these occasion, a DNA fragment of the expected length of 0.85 kb was detected in both species. However, the quality of Southern (Fig. 3.1.26 B) was poor, especially, the 0.85 kb *Eco*RI fragments in both monotreme DNA samples which were barely visible. Southern analyses of the two lambda clones which were isolated from monotreme genomic DNA libraries, λEL.1.1 from an echidna genomic DNA library (Shroff, 1990) and λPL.1.2 from a platypus genomic DNA library (Fig. 3.1.9 lane 3) show that the 0.85 kb *Eco*RI fragment hybridised to the β-globin gene probes (pSG2H and pTAG-2). Therefore, the 0.85 kb *Eco*RI fragment detected by Southern analysis of monotreme genomic DNA is likely to be the equivalent to the 0.85 kb fragments detected in the lambda clones, λEL.1.1 and λPL.1.2. Southern analyses of a platypus lambda clone λPL.1.1 show that an *Eco*RI fragment (6.8 kb) hybridised to the pTAG-2 β-globin probe (Fig. 3.1.7 lane 3). The 6.8 kb *EcoR*I fragment detected by Southern analysis of platypus genomic DNA is likely to be equivalent to the 6.8 kb fragment containing *platy* β1 detected in the lambda clone, λPL.1.1 (Fig. 3.1.26 A and B). This result suggests that each of the three hybridising bands detected on Southerns may correspond to a different β-like globin gene.

In the platypus, genomic Southern analysis shows fragments of the expected lengths (6.8 kb and 0.85 kb) based on the nucleotide sequences of *platy* β1 and *platy* β2. Therefore, the 10 kb fragment may represent another β-like globin gene. In echidna, only one β-globin gene was characterised in each species (*echid* β1 and *NGechid* β1), but Southern analysis showed the presence of three bands (10 kb, 8.7 kb and 0.85 kb), one of which (0.85 kb *Eco*RI fragment) is equivalent in length to the 0.85 kb *Eco*RI fragment found in *echid* β1. The other two bands may represent additional β-like globin genes in this species.

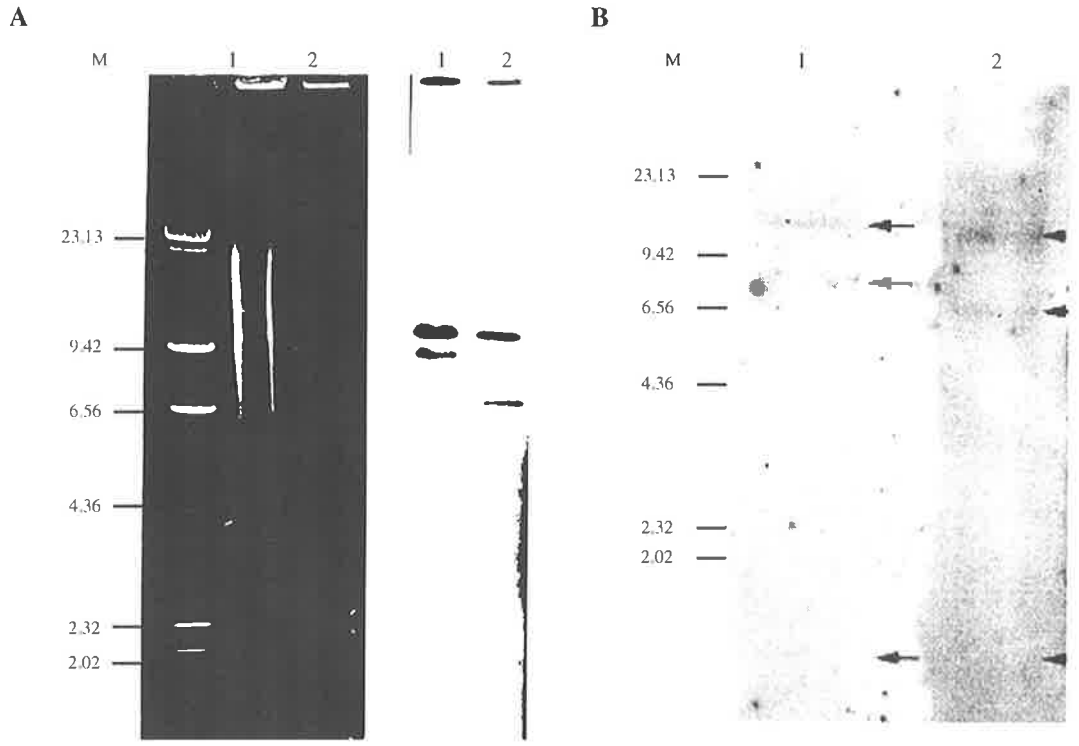


Figure 3.1.26 Southern analysis of monotreme genomic DNA using β-globin gene probes.

Lane M: λ*Hind*III DNA size marker

Lane 1: *Eco*RI digested Australian echidna DNA

Lane 2: *Eco*RI digested platypus DNA

A. On the left, an EtBr stained gel and on the right a Southern blot of the gel probed with pSG2H (Photographs kindly supplied by Robert Shroff, - Shroff (1990)).

B. Southern blot (from a different gel) probed with pTAG-2.

Arrows indicate hybridised fragments. An 0.85 kb fragment was faintly visible on the original autoradiograph, but, can not be seen in this reproduction.

3.2 LOCALISATION OF A β-GLOBIN GENE ON ECHIDNA CHROMOSOMES USING *IN SITU* HYBRIDISATION

3.2.1 G-banded karyotype of the Australian echidna

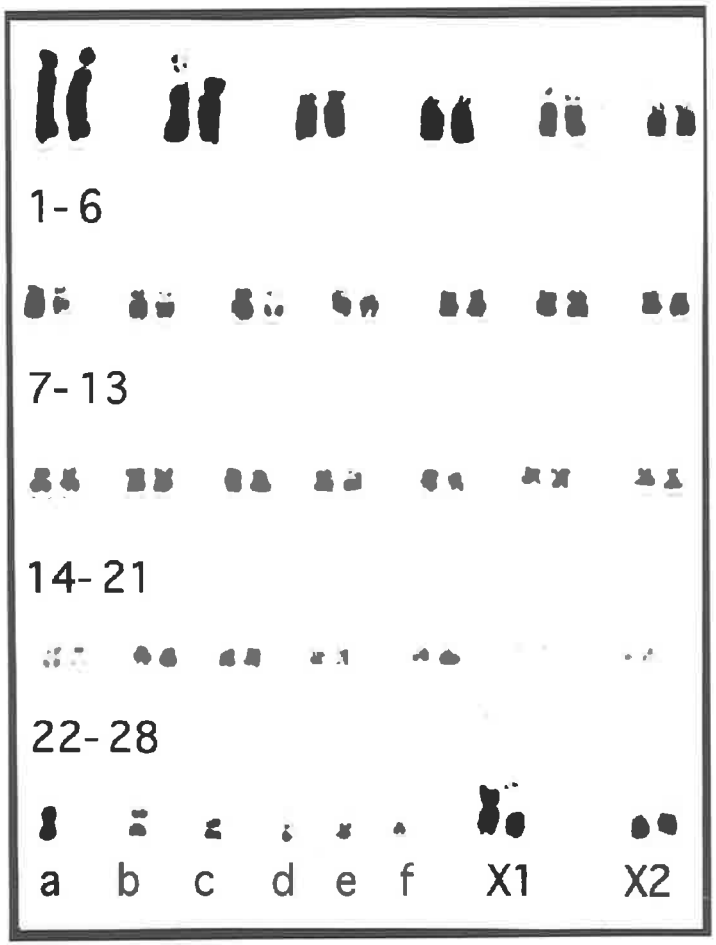
Australian echidna chromosomes were prepared from cultured blood cells, from an animal of unknown sex obtained from the Adelaide Zoo, using the protocol described in Section 2.2.6.1. Chromosomes were treated with trypsin and stained with Giemsa to cause G-banding. The sex of an echidna is difficult to determine morphologically. The G-banding karyotypes showed the sampled echidna to be a female with the karyotype 2n = 64 X1X1/X2X2 (Fig. 3.2.1). The G-banding chromosomes were identified according to the recommendations of Wrigley and Graves (1988a). However, many of the small sized chromosomes could not to be identified even with G-banding (Fig. 3.2.1). The diploid echidna genome consists of six large homologous pairs of chromosomes (Chromosome No. 1-6), many small homologous pairs of chromosomes grouping as 7-13, 14-21 and 22-28, and six (a-f) non-homologous chromosomes (Fig. 3.2.1).

3.2.2 Fluorescent *in situ* hybridisation (FISH) of Australian echidna chromosomes using a β-globin probe

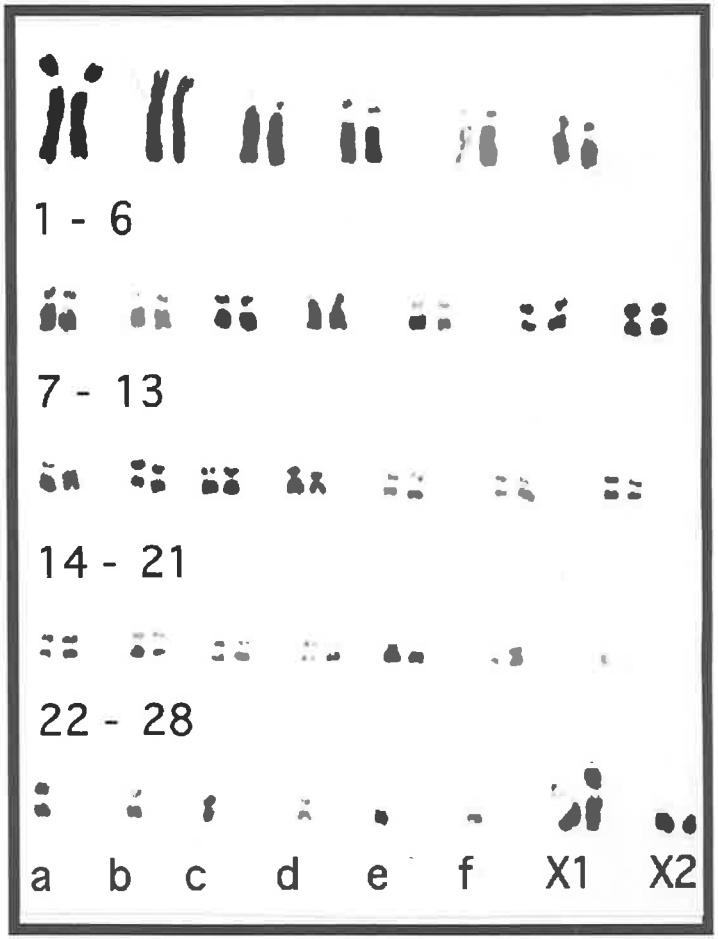
The location of *in situ* β-globin signals was analysed in relation to the G-banded karyotype shown in Figure 3.2.1. To optimise FISH conditions, *in situ* hybridisation was done first using a ribosomal RNA gene probe (VER 17), provided by Dr. Andreas Houben, Department of Genetics. VER 17 is a plasmid clone containing parts of the 18S, all of the 5.8S, and most of the 25S rDNA sequences, as well as the internal transcribed spacers, from *Vicia faba* (broad bean) (Yakura and Tanifuji, 1983). This clone can be used as a nucleolar organiser (NOR)-specific probe. Ribosomal RNA genes are highly conserved in sequences over many species and occur in high copy numbers. Therefore, the *in situ* signal should be strong. Sites of hybridisation using the ribosomal RNA probe were found on the NOR-region of chromosome 3p and $6p^{H}$ (Fig. 3.2.2 A). This confirms a previous

localisation of echidna 12S rDNA sequences, using *in situ* hybridisation of radio-labelled ribosomal RNA probe (Graves *et al.*, 1993).

The Lee F2/Lee R PCR product from λEL.1.1 (Fig. 3.1.15 lane 4), which contains *echid* β1, was used as a FISH gene probe for mapping β-globin gene sequences to the Australian echidna chromosomes. The biotin labelled probe was hybridised to denatured chromosomes and the hybridisation signals were detected using a streptavidin-FITC-anti-streptavidin-FITC system. Figures 3.2.2 (B and C) shows the epi-fluorescence signals recorded on Fuji film (400 ASA). A consistent signal was seen on one of the small chromosomes belonging to the 14-21 group in several cells (Fig. 3.2.2 B and C).



A



B

Figure 3.2.1 G-banded chromosomes of a female Australian echidna. A and B are are different cells from the same animal.

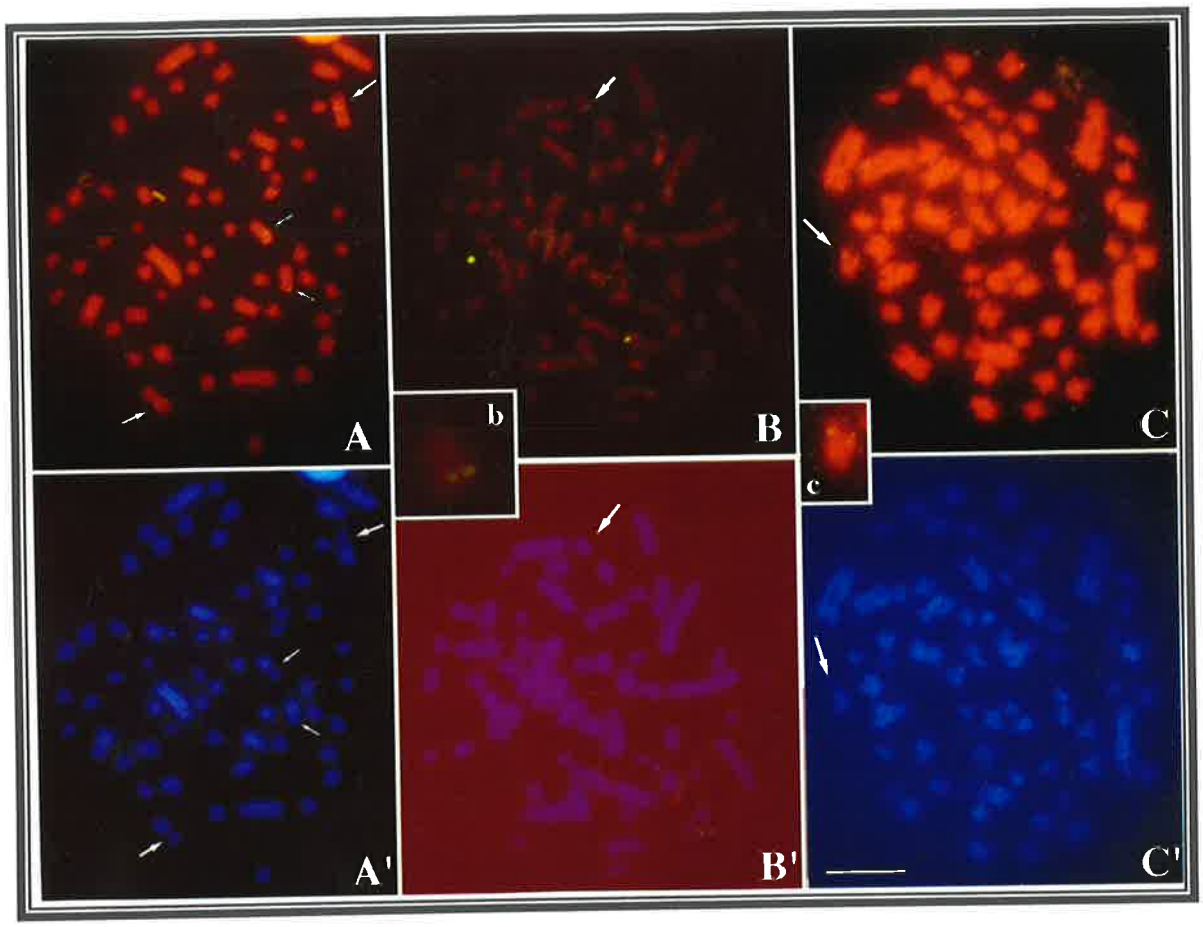


Figure 3.2.2 *In situ* hybridised cells from an Australian echidna.

A and A': Echidna chromosomes hybridised with the plant 12S ribosomal DNA probe VER17. **A:** FITC signals; **A'**: DAPI image of the same cell.

B, B', C and C': Echidna chromosomes hybridised with a biotin labelled β-globin gene probe (PCR product obtained λEL.1.1 using the primers Lee F2/Lee R) **B** and **C:** FITC signals; **B'** and **C':** DAPI images of the same cells.

b and **c:** magnified images of the chromosomes that showed specific hybridisation

Arrows indicate the *in situ* signals and a bar indicates the approximate size of chromosome 1, around 8 μm.

3.3 PHYLOGENETIC ANALYSIS USING β-GLOBIN GENE SEQUENCES

Phylogenetic analyses were carried out using a total of 30 β-like globin genes. Only coding regions (438 nucleotides) of each gene were used in the analyses, and initiation and termination codons were excluded (Goat βF, βA and βC globins are several codons shorter than the other sequences). The 30 β-globin genes comprised: the three monotreme genes (*echid* β1, *NGechid* β1 and *platy* β1) - one from each species; four marsupial genes, ε-globin and β-globin from each of two species; two genes from the chicken (used as outgroups), and 21 eutherian genes (including ε, γ, η δ and β) from species representing four major orders. A list of the genes used in the phylogenetic analyses, and their abbreviations, is given below. Locus identifications are indicated in Section 2.1.10, and DNA sequences are given in Appendix 10.

Species	Abbreviation used on the phylogenetic trees.
Capra hircus (Goat)	GOAT ε^{I}
	GOAT ε^{II}
	GOAT β^{F}
	GOAT β^{A}
	GOAT β^{C}
Didelphis virginiana (North American Opossum)	OPOSSUM ε
	OPOSSUM β
Galago crassicaudatus (Primate)	GALAGO ε
	GALAGO γ
	GALAGO δ
	GALAGO β
Gallus gallus (Chicken)	CHICK ε
	CHICK β^{H}
Homo sapiens (Human)	HUMAN ε
	HUMAN γ^{A}
	HUMAN δ
	HUMAN β
Mus musculus (Mouse)	MOUSE ε^{y}
	MOUSE h0
	MOUSE h1
	MOUSE β1
	MOUSE β2
Ornithorhynchus anatinus (Platypus)	PLATYPUS β1
	PLATYPUS β2
Oryctolagus cuniculus (Rabbit)	RABBIT ε
	RABBIT γ
	RABBIT β
Sminthopsis crassicaudata (Fat-tailed dunnart- Marsupialia)	DUNNART ε
	DUNNART β
Tachyglossus aculeatus (Australian Echidna)	ECHIDNA β1
Zaglossus bruijni (New Guinea Echidna)	NEW GUINEA ECHIDNA β1

3.3.1 Maximum parsimony analysis

The coding sequences of the 30 β-globin genes were analysed using the maximum parsimony algorithms: DNAPARS from the PHYLIP package (Felsentein, 1993), and heuristic search programs from PAUP 3.1.1 (Swofford, 1993) and from the test version of PAUP (*). Trees with the same general topology were obtained using the 2 different algorithms. A heuristic maximum-parsimony tree using the test version of PAUP is shown in Figure 3.3.1. The tree shows a primary dichotomy between the i) the monotreme β-like globins, and ii) the β-like globins from other mammals (marsupials and eutherians). The embryonic-like and adult-like globin genes of eutherians are each in separate monophyletic groups supported by 87% of bootstrap pseudoreplicaties. The marsupial ε-globin genes and eutherian embryonic-like β-globin genes also form a monophyletic group with bootstrap support <50%. The marsupial and eutherian β-globin genes are not monophyletic. Many nodes in the tree are not well supported by the bootstrap analysis, especially the dichotomy between monotreme β-globin genes and the other mammalian β-like globin genes. However, bootstrap support was high for monophyletic groups the monotreme β-globin genes (100%), the eutherian adult-like β-globin genes (87%), and the marsupial embryonic-like β-globin genes (93%) (Fig. 3.3.1). The monotremes comprise two distinct groups corresponding to the two families *Ornithorhynchidae* and *Tachyglossidae*, with monophyly of the echidna species supported by 99% of bootstrap pseudoreplicates (Fig. 3.3.1).

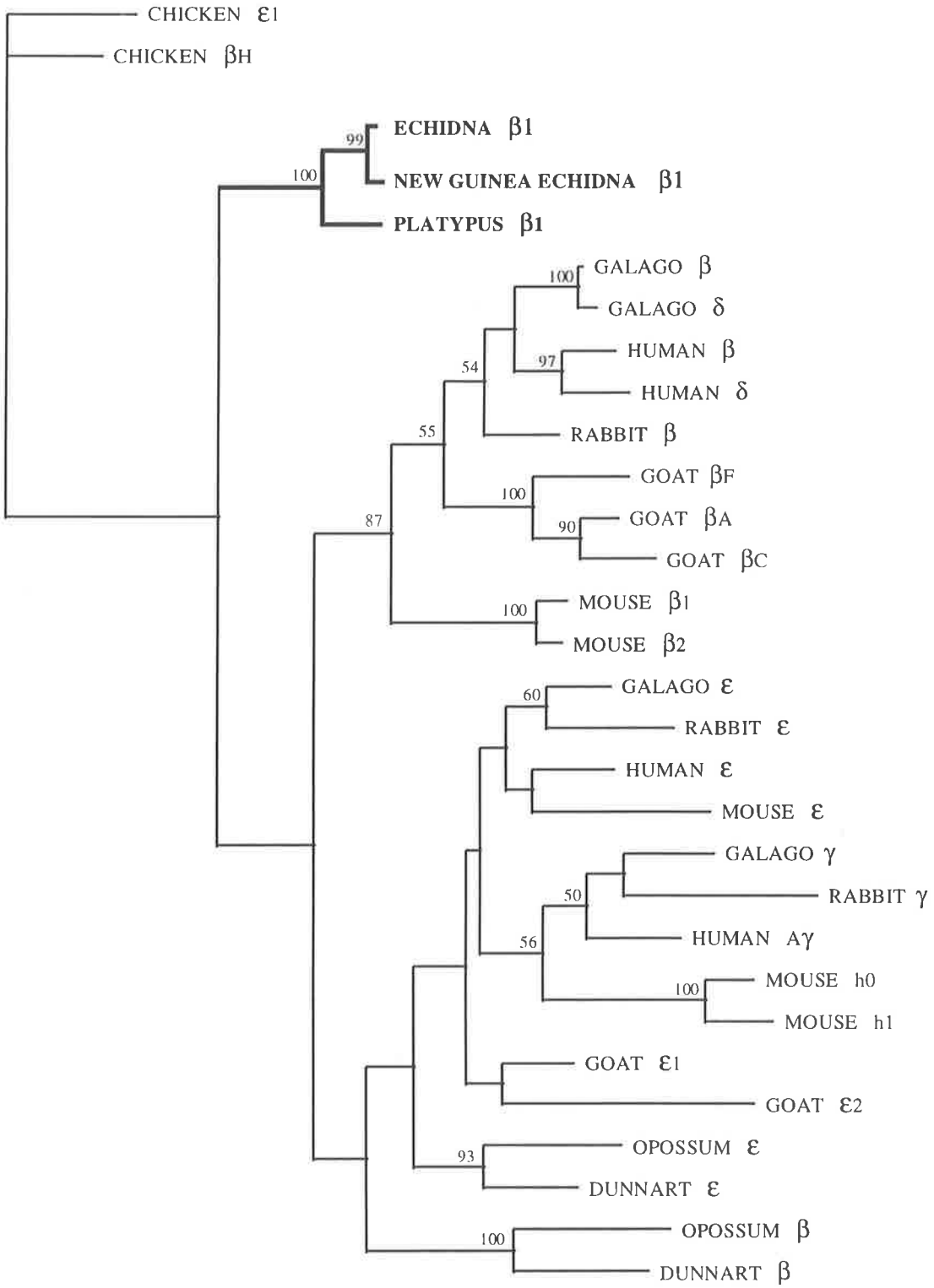


Figure 3.3.1 Maximum parsimony tree (length = 849) derived from 30 β-like globin genes using the heuristic algorithm in the test version of PAUP (*). Bootstrap values (%), based on 1,000 pseudo-replicates of the original data set, are shown on nodes of the tree. Branches lengths are drawn to approximate scale.

3.3.2 Distance tree

Distance trees were constructed using the test version of PAUP (*) and MEGA (Kimura *et al.* 1993). Distances were calculated by two different methods: the Kimura-2-parameter method and the maximum likelihood method using the HKY 85 model (Hasegawa, *et al.* 1985). Trees (Figure 3.3.2 and 3.3.3) were derived using the neighbor-joining approach in PAUP (*). Bootstrap values were based on 1,000 pseudoreplicates. The trees are almost identical and show the same topology as the maximum parsimony tree (Fig. 3.3.1). The distance trees show a dichotomy between monotreme and other mammalian β-like globin genes. The monotreme β-globin genes formed a monophyletic group to the exclusion of all other mammalian β-like globin genes (supported by bootstrap values of 73% and 69% - Fig. 3.3.2 and 3.3.3). The bootstrap values give strong support for monophyletic groupings of the monotreme β-globin genes (100%) and the eutherian adult-like β-globin genes (97% - 99%). The Australian echidna and New Guinea echidna form a monophyletic group supported by a 100% bootstrap value. The eutherian adult-like β-globin genes were separated into two main clades, one for mouse β-globin genes and the other for (((human, galago) rabbit) goat) β-globin genes (Fig. 3.3.2 and 3.3.3). The clade (((human, galago) rabbit) goat) is separated into two clades, ((human, galago), rabbit) and goat adult-like β-globin genes, supported by bootstrap values of 92-93%. The marsupial embryonic-like β-globin genes form a diphyletic group with eutherian embryonic-like β-globin genes supported by bootstrap values of 74% (Fig. 3.3.2) and 77% (Fig. 3.3.3). The eutherian embryonic-like β-globin genes divide into two clades, ε-globin genes and γ-globin genes (58-60 % bootstrap values). The marsupial β-globin genes form a based clade with the mammalian β-like globin genes (excluding the monotreme β-globin genes) supported by bootstrap values of 69%-73%.

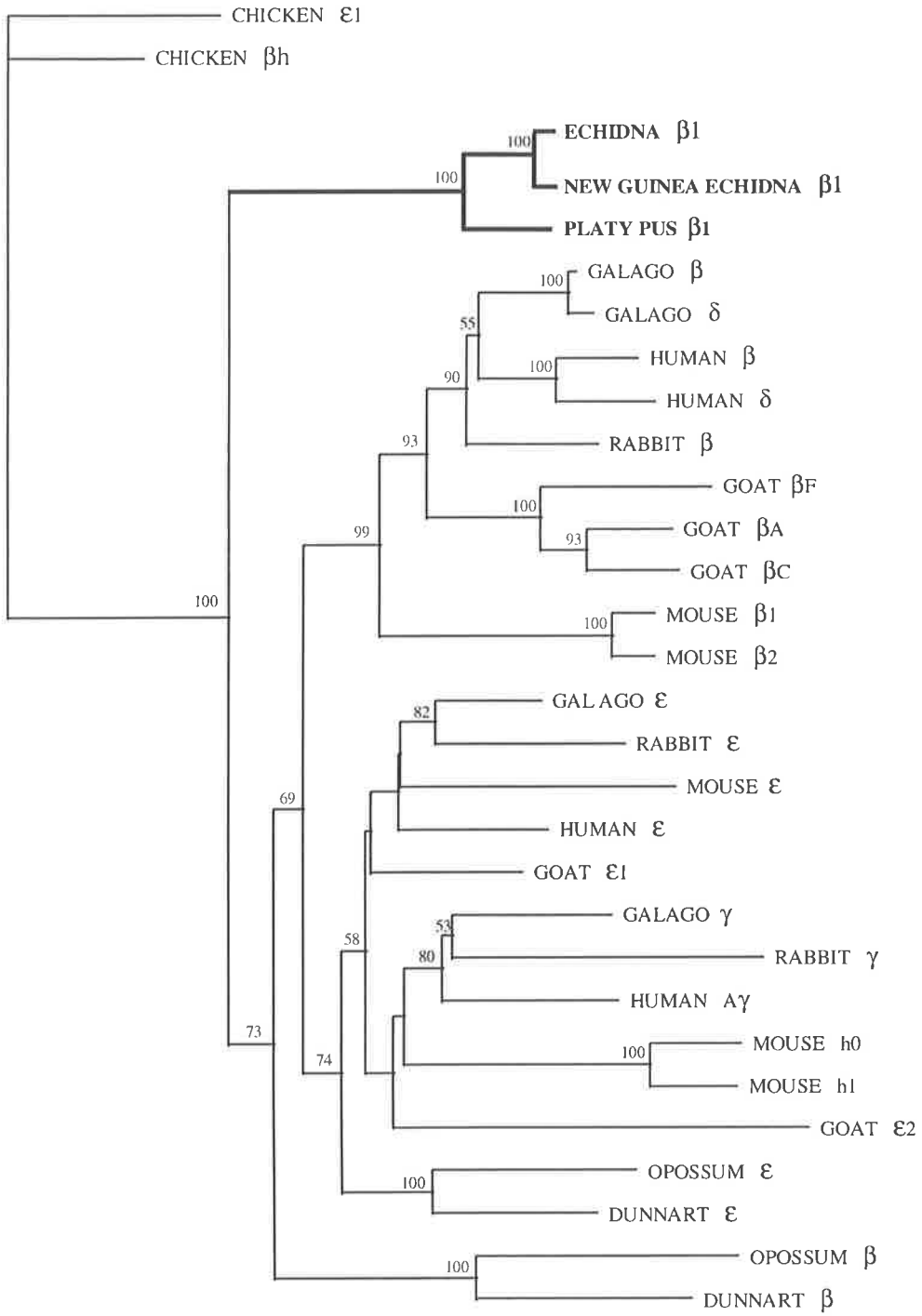


Figure 3.3.2 Cladogram representing the Kimura-2-parameter \ Neighbor-Joining tree of the 30 β-like globin genes. The tree was constructed using PAUP (*). Bootstrap values (%) based on 1,000 pseudo-replicates of the original data set, are shown. Branch lengths are drawn to approximate scale.

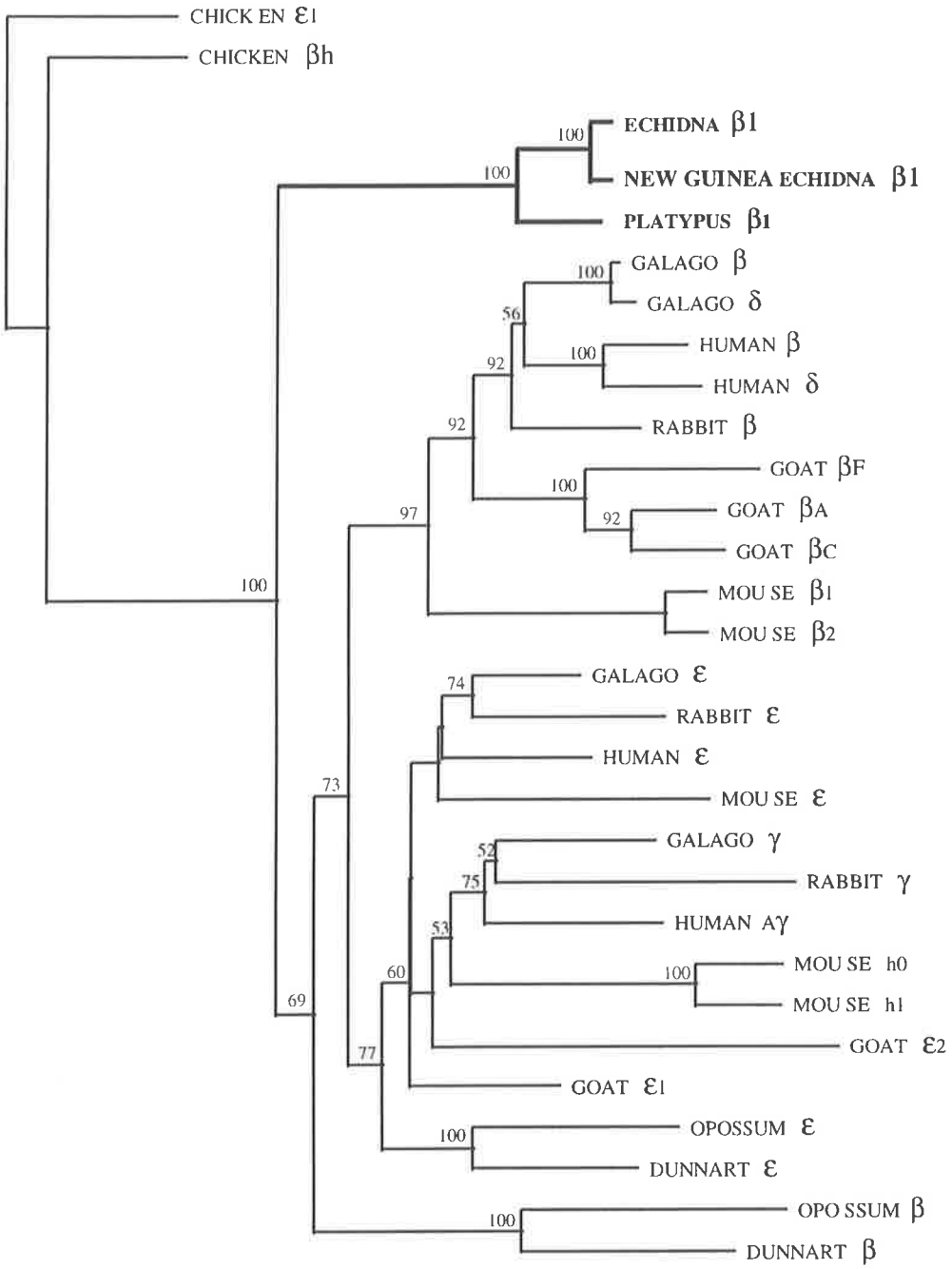


Figure 3.3.3 Neighbor-Joining tree constructed using PAUP (*). Distances were estimated by the maximum likelihood method. Bootstrap values (%) based on 1,000 pseudo-replicates of the original data set, are shown. Branch lengths are drawn to approximate scale.

3.3.3 Maximum likelihood analysis

Maximum likelihood analyses were carried out using two different data sets, one consisting of sequences from 30 β-like globin genes and the other containing a smaller set consisting of sequences from 18 genes. In the smaller data set, only sequences from embryonic (ε) and adult (β) globin gene were included. Different evolutionary models were assumed for the maximum likelihood analyses. These models made different assumptions about the pattern of nucleotide substitution, the variation of substitution rates over sites, and the transition-transversion ratio. The gamma distribution model assumes that nucleotide substitution rates are randomly distributed over different sites. The HKY 85 model (Hasagawa *et al.* 1985) takes account of inequalities in the frequency of bases and differences in transition vs. transversion rates (Yang, 1993, 1996). The transition-transversion ratio in the two models and the α-parameter for the gamma distribution were estimated by maximum likelihood from the smaller data set to reduce computational time (see above). The overall transition - transversion ratio and the α value for the gamma distribution model were estimated as 1.165 and 0.61, respectively. These values were then applied to an analysis of the 30 β-like globin genes. This transition - transversion ratio was used throughout subsequent analyses. Bootstrap analyses could not be performed due to computational time limits.

Figure 3.3.4 shows the maximum-likelihood phylogenetic tree derived using the HKY 85 model. The basic tree topology is identical to that of the maximum-parsimony and distance trees. The monotreme and other mammalian β-like globin genes form a major dichotomy above the tree root. The eutherian adult-like β-globin genes comprise a monophyletic group. The marsupial β-globin genes group with the marsupial ε-globin genes and the clade so formed joins the root of the eutherian early expressed globin genes (Fig. 3.3.4).

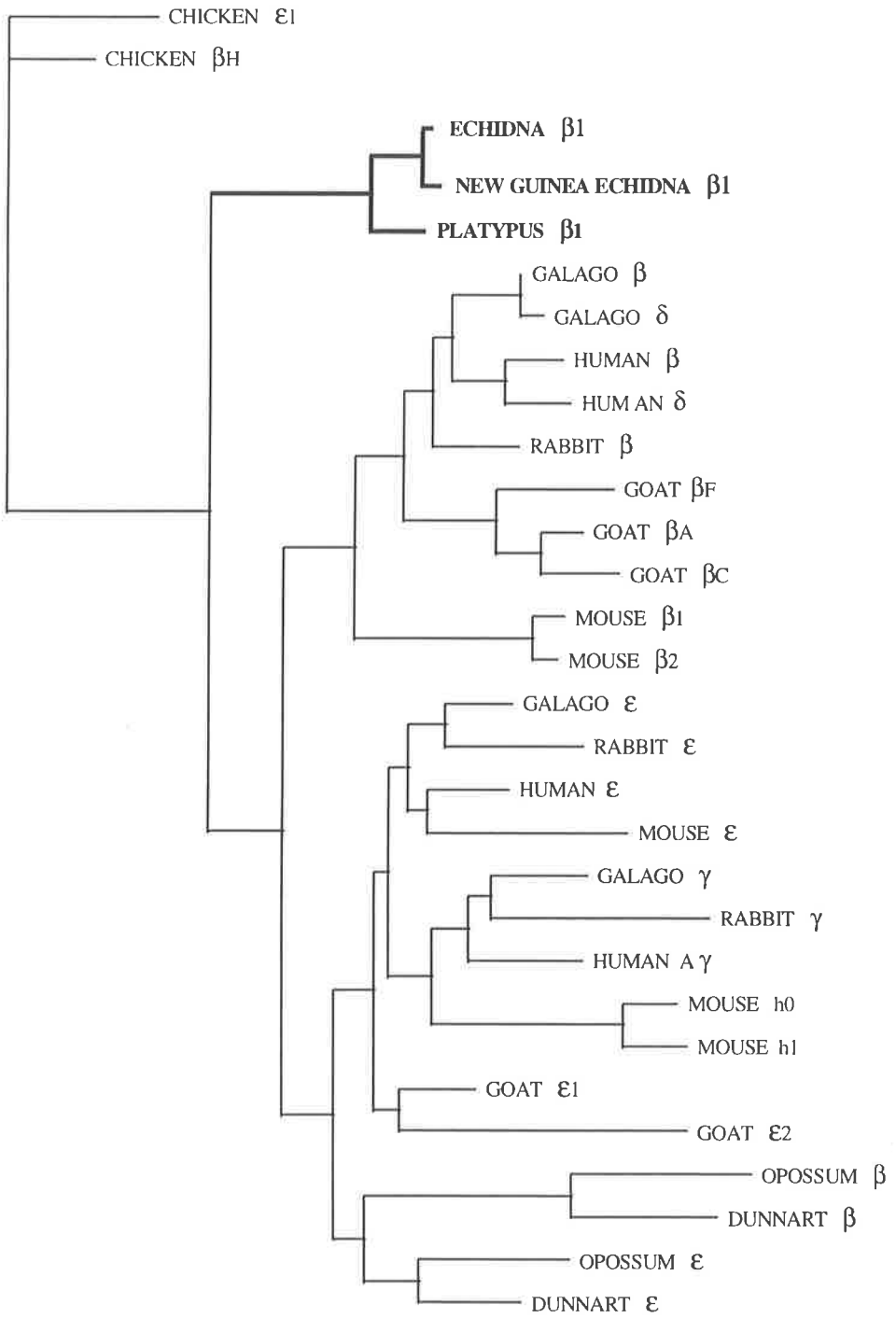


Figure 3.3.4 Maximum likelihood tree constructed from the 30 monotreme, marsupial and eutherian β-like globin nucleotide sequences using PAUP (*) with the HKY 85 model, and an estimated transition / transversion rate of 1.165.

The maximum-likelihood tree derived using the HKY 85 + gamma distribution model is shown in Figure 3.3.5. Interestingly, the tree topology is very different from the other trees. The marsupial ε-globin genes form a group that joins the root of all other globin genes. The remaining mammalian globin genes comprise two distinct groups: i) a group comprising the monotreme β-globin genes, the marsupial β-globin genes, and the eutherian adult-like β-globin genes, ii) a group comprising the eutherian embryonic-like globin genes (Fig. 3.3.5). The marsupial β-globin genes group with monotreme β-globin genes to the exclusion of the eutherian adult-like β-globin genes (Fig. 3.3.5).

3.3.4 Transitions and transversions

The frequencies of transitions and transversions between pairs of β-like globin genes (embryonic-like and adult-like) were plotted against genetic distance (Jukes-Cantor) to graphically examine whether these types of substitution were accumulating in a linear fashion or whether saturation had occurred. The Juke-Cantor distance values were obtained on the assumption that all the four possible nucleotide substitutions at a site occur with equal probability (Jukes, 1983). The proportion of transitions and transversions and the Jukes-Cantor distances were obtained using MEGA. From a data set of 28 genes, excluding chicken ε- and β-globin genes from the 30 β-like globin genes used for the phylogenetic analyses (see previously).

Results of the analysis (Fig. 3.3.6) show that transversion substitutions have accumulated in an approximately linear fashion but transitions show some evidence for saturation with respect to Jukes-Cantor distances.

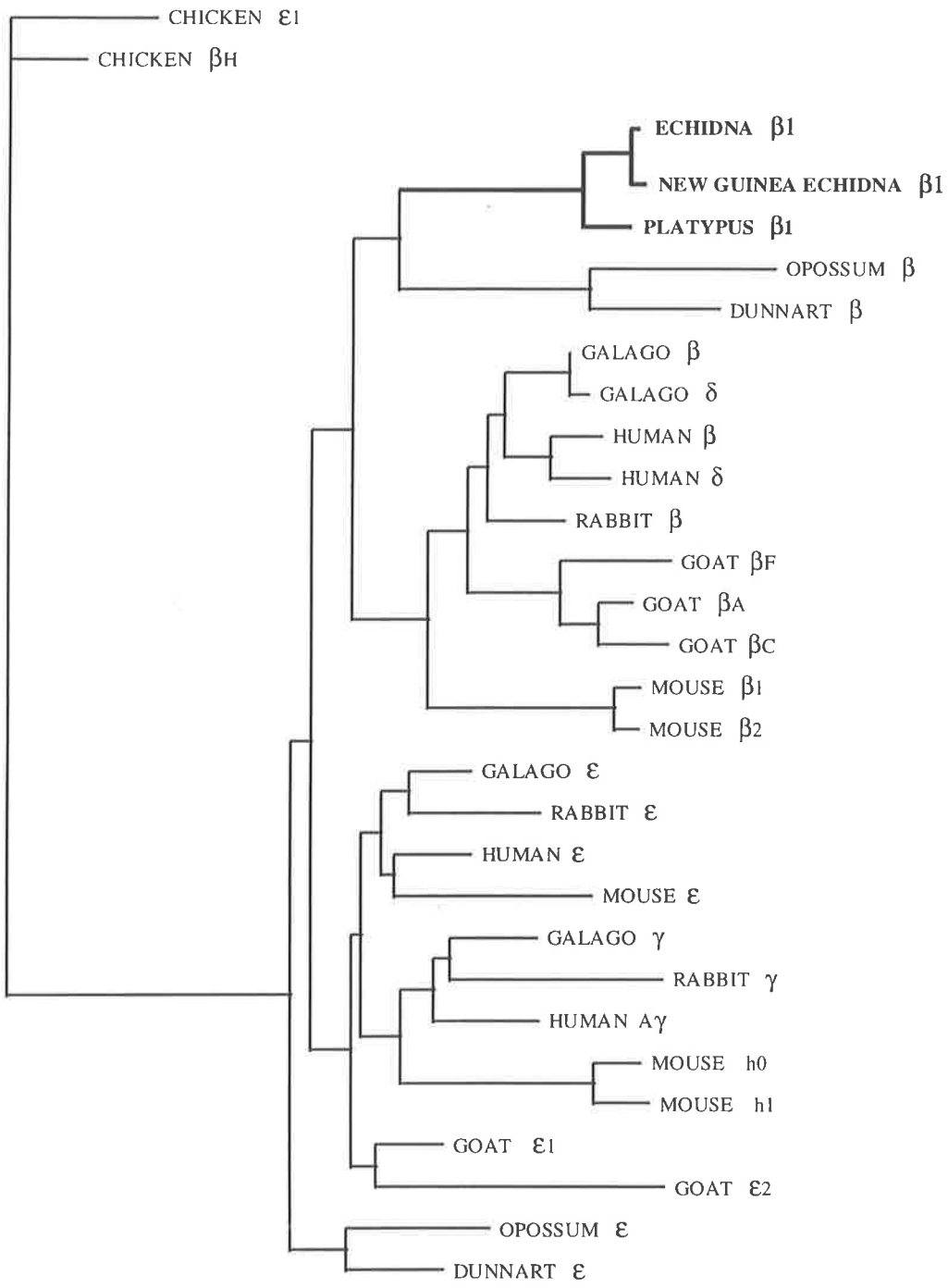
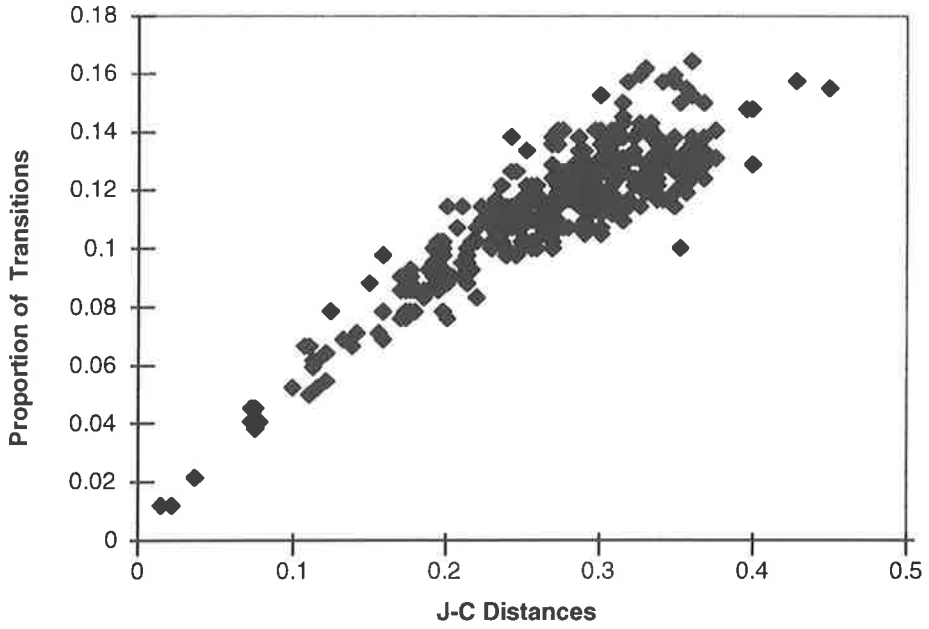


Figure 3.3.5 Maximum likelihood tree, constructed from 30 monotreme, marsupial and eutherian β-like globin nucleotide sequences, using PAUP (*) and the HKY 85 and gamma distribution models.

A.



B.

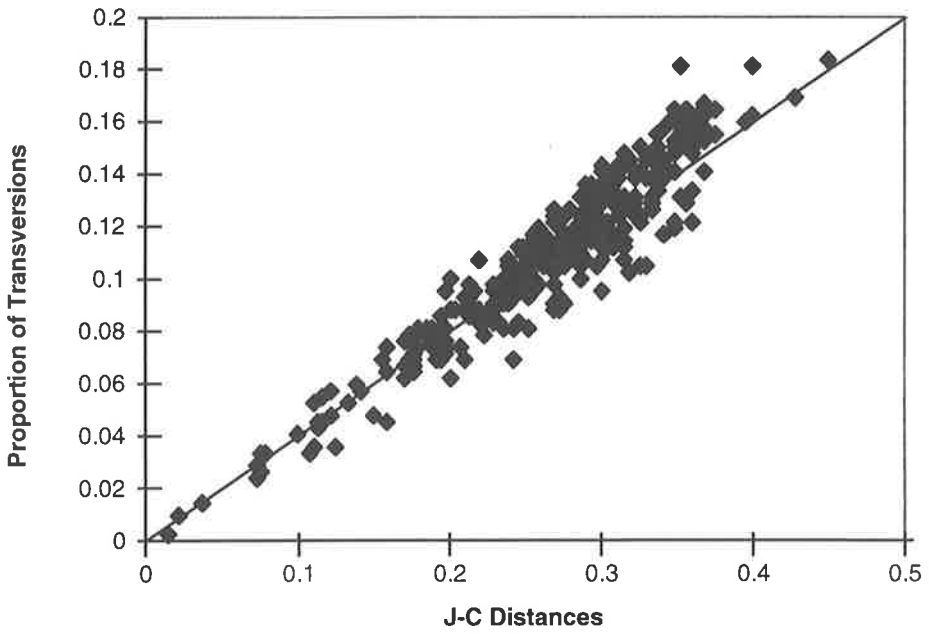


Figure 3.3.6 A plot of the proportion of sites with transitional (A) and transversional (B) nucleotide substitutions against the Jukes-Cantor genetic distances. Each point on the graph represents the values of the two variables from a comparison of the coding region of two aligned β-like globin genes.

3.3.5 Relative rates tests

The relative rate test of Wu and Li (1985) was applied to test for constancy of nucleotide substitution rates between different β-like globin gene lineages.

In such a test, one compares the minimum number of nucleotide substitutions that have occurred in the lineage joining a test sequence with a reference sequence (which must be an outgroup) to the minimum number of nucleotide substitutions that have occurred in the lineage connecting a second test sequence to the same reference sequence. Such an approach enables the number of substitutions, that have occurred down the separate branches of the test sequences since they separated from one another, to be determined. The molecular clock hypothesis predicts that, on average, these values should be the same.

In carrying out the relative rates tests, chicken β-globin can be confidently used as an outgroup when the test groups are mammalian β-like globin sequences. However, there is uncertainty about the order of branching events that separate the monotreme, marsupial and eutherian globin genes so for some comparisons, marsupial, eutherian and monotreme β-globins were used as an outgroup.

Rates were estimated for non-synonymous substitutions at non-degenerate and 2-fold degenerate sites, and synonymous substitutions at 2-fold and 4-fold degenerate sites. The sequences used for the tests were from chicken (chick β), eutherians (mouse ε and β, goat β, rabbit β, and human ε and β), marsupials (dunnart ε and β) and opossum (opos ε and β)) and monotremes (Australian echidna (*echid* β1), New Guinea echidna (*Ngech* β1) and platypus (*platy* β1)). In all, 71 pairs of β-like globin sequences were analysed (coding regions only).

The results of these tests, shown in Table 3.3.1, can be summarised as follows:

1. The marsupial adult β-globin genes appear to have higher substitution rates than the eutherian β-globin genes. Using the chicken as an outgroup, 7 of the 12 comparisons at non-synonymous sites were positive (i.e. the rates for marsupial sequences were higher

than those for eutherian sequences) and all 12 of the comparisons at synonymous sites were positive. When echidna β1-globin was used as outgroup, 6 of the 8 comparisons at non-synonymous sites were positive and all 8 of the comparisons at synonymous sites were positive. Overall, there appears to be no significant difference in the substitution rates between marsupial and eutherian ε-globins. A comparison between the mouse and dunnart ε-globin genes (using *echid* β1-globin as an outgroup) showed that the mouse gene evolved faster than the dunnart gene at non-degenerate sites (significant at the 5% level).

2. Comparisons between marsupial ε- and β-globin genes showed that overall, the marsupial β-globin genes had higher substitution rates than the marsupial ε-globin genes. Comparisons between the dunnart ε-globin and dunnart β-globin genes (using echidna β1 and chicken β as outgroups), showed that dunnart β-globin evolved faster than dunnart ε-globin at non-degenerate sites for non-synonymous substitutions and 2-fold degenerate sites for synonymous substitution (significance levels: 5% and 1% respectively).

3. In general, the eutherians β-globin genes had slightly higher substitution rate than the eutherian ε-globin genes. Mouse β-globin evolved faster than mouse ε-globin at non-degenerate sites for non-synonymous substitutions with significance at the 5% levels.

4. In general, the opossum β-like globin genes (ε- and β-globin) had higher substitution rates than the dunnart β-like globin genes. The mouse β-globin gene had higher substitution rates than the human β-globin gene at non-synonymous sites (1% significance level).

5. Comparisons of the monotreme β-globin genes and marsupial β-globin genes show that the marsupial β-globin genes evolved more rapidly than the monotreme β-globin genes (with 5% significant level of differences). The eutherian β-globin genes evolved slightly faster than the monotreme β-globin genes with no significant levels of difference.

6. The monotreme β-globin genes evolved faster than marsupial ε-globin genes and eutherian ε-globin genes at non-synonymous sites (5% significant level) but at synonymous sites the monotreme β-globin genes evolved more slowly.

7. For comparisons among the three monotreme β-globin genes, various marsupial, eutherian and chicken β-globin genes were used as outgroups. Comparisons between the *echid* β1 and *platy* β1 and between the *NGechid* β1 and *platy* β1 showed no significant differences of substitution rates. The *NGechid* β1 gene evolved slightly faster than *echid* β1 but the difference values were close to zero.

3.3.6 Synonymous and non-synonymous comparisons

Table 3.3.2 shows the synonymous (S) and non-synonymous (NS) values obtained from pair wise comparisons involving chicken, marsupial and eutherian embryonic-like and adult-like β-globin genes and monotreme β-globin genes, using the method of Li (1993). When the Australian echidna β-globin gene was compared with each of the other monotreme β-globin genes (*NGechid* β1, *platy* β1, and *platy* β2), it showed the lowest values (S / NS) with *platy* β2 (1% / 0%), followed by *NGechid* β1 (3% / 2%) and *platy* β1(12% / 6%). Comparisons of each of the monotreme β-globin genes with marsupial adult β-globin genes revealed that the monotreme genes had fewer substitutions compared to the dunnart (81-87% / 18-19%) than compared to the opossum (91-98% / 19-21%) at synonymous and non-synonymous sites, respectively. Comparisons of eutherian and monotreme β-globin genes showed that the primate β-globins (galago β) (56-61%) and rabbit β-globin genes (14-15%) had the lowest values at synonymous and non-synonymous sites, respectively. In general, the monotreme β-globins show a closer resemblance in sequence to eutherian ("adult") β-globin genes than they do to marsupial ("adult") β-globin genes at non-synonymous and synonymous sites (Table 3.2.3). The substitution values for other mammalian (eutherian and marsupial) ε-globin genes and monotreme β-globin genes are slightly higher than between other mammalian β-globin genes and monotreme β-globin genes both at synonymous and non-synonymous sites. Comparisons between monotreme β-globin genes and eutherian ε-globin genes revealed lower substitution values than those between monotreme β-globin and marsupial ε-globin

genes at synonymous and non-synonymous sites, with the exception of comparisons with the mouse ε-globin gene (95-103% / 18-19%). The divergence values between monotreme β-globin genes and marsupial ε-globin genes (opossum ε and dunnart ε) are slightly higher at synonymous sites but lower at non-synonymous sites than those between monotreme β-globin genes and marsupial β-globin genes.

3.3.7 Estimation of the separation time of the three monotreme species

As shown in Section 3.3.5, the results of relative tests showed no significant differences in the evolutionary rates of the three monotreme β-globin genes. Therefore, it is possible to estimated the time of separation of the three monotreme species, if it is assumed that the rates of nucleotide substitution in these genes are approximately constant over time.

The separation time of the three monotreme species was calculated using nucleotide substitutions in β-globin genes according to the method of Li and Graur (1991). The unknown divergence time between species A and B (T_2) is estimated by $T_2 = 2K_{AB}T_1/K_{AC}+K_{BC}$. The divergence time T1 is assumed to be the same as the time of divergence of a reference species "C" which is usually inferred from palaeontological data. K_{ij} (K_{AB}, K_{AC}, and K_{BC}) is the number of nucleotide substitution per site between species i and j. The K_{ij} values used were those given by Li (1993). The marsupial (S. *crassicaudata*) is used as the reference species "C". The nucleotide substitution values (K_{ij}) (at the non-degenerate sites) were obtained from different combinations of species, as listed below:

Species Compared	K_{AB}	K_{AC}	K_{BC}
Australian Echidna (A)/Platypus (B)	0.061±0.015	0.195±0.029	0.193±0.023
Australian Echidna (A)/New Guinea Echidna (B)	0.022±0.009	0.195±0.029	0.204±0.030
Platypus (A)/New Guinea Echidna (B)	0.069±0.016	0.193±0.023	0.204±0.030

If the separation of marsupial from monotreme is assumed to have occurred 150 MYA, then the divergence times of Australian echidna / platypus, Australian echidna / New Guinea echidna and platypus / New Guinea echidna can be estimated as 41.1-51.8, 11.5-20.3 and 46.2-56.7 MYA, respectively.

Table 3.3.1 Relative rates tests of bird, marsupial, eutherian and monotreme β-like globin genes*.

Species/Genes Compared			No. of Nonsynonymous Substitution Per Site								No. of Synonymous Substitution Per Site							
			Nondegenerate Sites				2-fold degenerate Sites				2-fold Degenerate Sites				4-fold Degenerate Sites			
1	2	3	K12	K13	K23	K13-K23	B12	B13	B23	B13-B23	A12	A13	A23	A13-A23	K12	K13	K23	K13-K23
Dunn β	Mouse β	Echid β1	0.219	0.191	0.179	0.012	0.121	0.091	0.141	-0.050	0.451	0.497	0.374	0.123	0.826	0.780	0.597	0.184
	No. of sites			281.0		(0.032)		82.0		(0.046)		82.0		(0.16)		75.0		(0.19)
Dunn β	Human β	Echid β1	0.208	0.191	0.155	0.036	0.099	0.091	0.047	0.044	0.481	0.497	0.395	0.103	0.899	0.780	0.657	0.123
	No. of sites			280.0		(0.031)		81.3		(0.038)		81.3		(0.16)		76.7		(0.19)
Dunn β	Goat β	Echid β1	0.223	0.195	0.176	0.019	0.093	0.092	0.106	-0.014	0.434	0.509	0.427	0.082	0.809	0.800	0.638	0.162
	No. of sites			276.3		(0.032)		82.3		(0.039)		82.3		(0.16)		73.3		(0.19)
Dunn β	Rabbit β	Echid β1	0.190	0.191	0.138	0.053	0.137	0.091	0.084	0.007	0.415	0.497	0.304	0.194	0.746	0.780	0.760	0.021
	No. of sites			279.7		(0.030)		83.2		(0.049)		83.2		(0.15)		75.3		(0.20)
Dunn β	Mouse β	Chick β	0.219	0.239	0.270	-0.031	0.121	0.065	0.118	-0.053	0.451	0.527	0.329	0.198	0.826	0.973	0.890	0.083
	No. of sites			282.3		(0.034)		83.7		(0.044)		83.7		(0.157)		72.0		(0.252)
Dunn β	Human β	Chick β	0.208	0.239	0.241	-0.002	0.099	0.065	0.063	0.002	0.481	0.527	0.294	0.234	0.899	0.973	0.711	0.261
	No. of sites			281.3		(0.033)		83.0		(0.037)		83.0		(0.157)		73.8		(0.231)
Dunn β	Rabbit β	Chick β	0.190	0.239	0.252	-0.013	0.137	0.065	0.062	0.003	0.415	0.527	0.327	0.200	0.746	0.973	0.745	0.227
	No. of sites			281.0		(0.032)		84.7		(0.043)		84.7		(0.152)		72.3		(0.224)
Opos β	Human β	Chick β	0.221	0.271	0.241	0.030	0.109	0.080	0.063	0.017	0.449	0.437	0.294	0.144	0.907	0.919	0.711	0.208
	No. of sites			282.0		(0.035)		83.3		(0.040)		83.3		(0.134)		72.8		(0.224)
Opos β	Mouse β	Chick β	0.239	0.271	0.270	0.001	0.156	0.080	0.118	-0.038	0.314	0.437	0.329	0.108	0.895	0.919	0.890	0.029
	No. of sites			283.0		(0.036)		84.0		(0.049)		84.0		(0.125)		71.0		(0.254)
Opos β	Rabbit β	Chick β	0.212	0.271	0.252	0.019	0.120	0.080	0.062	0.018	0.484	0.437	0.327	0.110	0.842	0.919	0.745	0.174
	No. of sites			281.7		(0.034)		85.0		(0.041)		85.0		(0.139)		71.3		(0.225)
Dunn ε	Human ε	Chick β	0.103	0.177	0.214	-0.037	0.034	0.093	0.089	0.004	0.289	0.321	0.334	-0.014	0.565	0.711	0.734	-0.023
	No. of sites			282.7		(0.023)		84.3		(0.024)		84.3		(0.104)		71.0		(0.181)
Dunn ε	Mouse ε	Chick β	0.129	0.177	0.207	-0.03	0.066	0.093	0.088	0.005	0.323	0.321	0.358	-0.037	0.773	0.711	0.794	-0.083
	No. of sites			281.7		(0.025)		86.0		(0.032)		86.0		(0.110)		70.3		(0.206)
Opos ε	Human ε	Chick β	0.130	0.211	0.214	-0.003	0.062	0.091	0.089	0.002	0.254	0.338	0.334	0.004	0.620	0.825	0.734	0.091
	No. of sites			283.0		(0.026)		83.7		(0.032)		83.7		(0.102)		71.7		(0.205)
Opos ε	Mouse ε	Chick β	0.142	0.211	0.207	0.004	0.066	0.091	0.088	0.003	0.373	0.338	0.358	-0.020	0.829	0.825	0.794	0.032
	No. of sites			282.0		(0.027)		85.3		(0.032)		85.3		(0.118)		70.5		(0.228)
Dunn ε	Human ε	Echid β1	0.103	0.136	0.161	-0.025	0.034	0.109	0.105	0.004	0.289	0.436	0.517	-0.081	0.565	1.049	0.703	0.346
	No. of sites			281.3		(0.022)		82.7		(0.025)		82.7		(0.14)	74.0			(0.25)
Dunn ε	Mouse ε	Echid β1	0.129	0.136	0.179	**-0.044***	0.066	0.109	0.131	-0.022	0.323	0.436	0.457	-0.020	0.773	1.049	1.035	0.014
	No. of sites			280.3		(0.024)		84.3		(0.034)		84.3		(0.14)		73.3		(0.29)

Species/Genes Compared			No. of Nonsynonymous Substitution Per Site								No. of Synonymous Substitution Per Site							
			Nondegenerate Sites				2-fold degenerate Sites				2-fold Degenerate Sites				4-fold Degenerate Sites			
1	2	3	K12	K13	K23	K13-K23	B12	B13	B23	B13-B23	A12	A13	A23	A13-A23	K12	K13	K23	K13-23
Dunn ε	Dunn β	Echid β1	0.155	0.136	0.191	**-0.056***	0.067	0.109	0.091	0.018	0.675	0.436	0.497	**-0.61****	0.849	1.049	0.780	0.268
	No. of sites			281.3		(0.027)		82.7		(0.034)		82.7		(0.18)		74.0		(0.28)
Opos ε	Opos β	Echid β1	0.204	0.168	0.198	-0.030	0.133	0.135	0.122	0.013	0.361	0.527	0.659	-0.132	1.105	0.896	0.939	-0.043
	No. of sites			282.3		(0.031)		82.3		(0.049)		82.3		(0.207)		73.3		(0.292)
Dunn ε	Dunn β	Chick β	0.155	0.177	0.239	**-0.061***	0.067	0.093	0.065	0.028	0.675	0.321	0.527	-0.207	0.849	0.711	0.973	-0.262
	No. of sites			282.7		(0.028)		84.3		(0.032)		84.3		(0.170)		71.0		(0.238)
Opos ε	Opos β	Chick β	0.204	0.211	0.271	-0.059	0.133	0.091	0.080	0.011	0.361	0.338	0.437	-0.099	1.105	0.825	0.919	-0.094
	No. of sites			283.7		(0.033)		84.0		(0.044)		84.0		(0.130)		70.3		(0.261)
Human ε	Human β	Echid β1	0.165	0.161	0.155	0.006	0.083	0.105	0.047	0.056	0.311	0.517	0.395	0.122	0.512	0.703	0.657	0.046
	No. of sites			280.0		(0.027)		81.3		(0.036)		81.3		(0.15)		76.7		(0.16)
Mouse ε	Mouse β	Echid β1	0.166	0.180	0.179	0.001	0.169	0.131	0.141	-0.009	0.439	0.457	0.374	0.082	0.779	1.035	0.597	0.438
	No. of sites			280.0		(0.028)		83.7		(0.054)		83.7		(0.15)		74.3		(0.24)
Human ε	Human β	Chick β	0.165	0.214	0.241	-0.027	0.083	0.089	0.063	0.026	0.311	0.334	0.294	0.041	0.512	0.734	0.711	0.023
	No. of sites			281.3		(0.029)		83.0		(0.035)		83.0		(0.105)		72.0		(0.172)
Mouse ε	Mouse β	Chick β	0.166	0.207	0.270	**-0.062***	0.169	0.088	0.118	-0.030	0.439	0.358	0.329	0.028	0.779	0.794	0.890	-0.096
	No. of sites			281.3		(0.030)		85.3		(0.051)		85.3		(0.125)		71.3		(0.232)
Mouse β	Human β	Echid β1	0.138	0.179	0.155	0.024	0.047	0.141	0.047	**0.094****	0.256	0.374	0.395	-0.021	0.395	0.597	0.657	-0.061
	No. of sites			278.7		(0.025)		81.0		(0.031)		81.0		(0.11)		78.3		(0.13)
Mouse β	Human β	Chick β1	0.138	0.274	0.245	0.029	0.047	0.120	0.064	0.056	0.256	0.337	0.299	0.037	0.395	0.916	0.728	0.188
	No. of sites			276.0		(0.028)		81.7		(0.029)		81.7		(0.100)		73.3		(0.184)
Mouse β	Human β	Dunn β1	0.138	0.219	0.208	0.010	0.047	0.121	0.099	0.022	0.256	0.451	0.481	-0.03	0.395	0.826	0.899	-0.072
	No. of sites			280.3		(0.027)		81.7		(0.029)		81.7		(0.14)		76.0		(0.170)
Dunn β	Opos β	Echid β1	0.145	0.191	0.198	-0.007	0.033	0.091	0.122	-0.031	0.386	0.497	0.659	-0.162	0.278	0.780	0.939	-0.159
	No. of sites			283.0		(0.027)		82.7		(0.025)		82.7		(0.203)		72.3		(0.176)
Dunn β	Opos β	Chick β1	0.145	0.243	0.275	-0.032	0.033	0.066	0.081	-0.015	0.386	0.503	0.447	0.056	0.278	1.002	0.947	0.055
	No. of sites			280.3		(0.029)		83.0		(0.023)		83.3		(0.152)		68.3		(0.179)
Opos ε	Dunn ε	Chick β	0.111	0.211	0.177	0.034	0.068	0.091	0.093	-0.002	0.165	0.338	0.321	0.017	0.274	0.825	0.711	0.114
	No. of sites			282.0		(0.024)		84.0		(0.033)		84.0		(0.086)		72.0		(0.152)
Opos ε	Dunn ε	Echid β1	0.111	0.168	0.136	0.033	0.068	0.135	0.109	0.026	0.165	0.527	0.436	0.090	0.274	0.896	1.049	-0.153
	No. of sites			280.7		(0.023)		82.3		(0.036)		82.3		(0.124)		75.0		(0.205)
Human ε	Mouse ε	Chick β	0.091	0.214	0.207	0.007	0.056	0.089	0.088	0.001	0.316	0.334	0.358	-0.023	0.421	0.734	0.794	-0.060
	No. of sites			282.7		(0.022)		85.7		(0.030)		85.7		(0.111)		69.7		(0.168)

Species/Genes Compared			No. of Nonsynonymous Substitution Per Site								No. of Synonymous Substitution Per Site							
			Nondegenerate Sites				2-fold degenerate Sites				2-fold Degenerate Sites				4-fold Degenerate Sites			
1	2	3	K12	K13	K23	K13-K23	B12	B13	B23	B13-B23	A12	A13	A23	A13-A23	K12	K13	K23	K13-23
Echid β1	Dunn β	Mouse β	0.191	0.179	0.219	-0.040	0.091	0.141	0.121	0.020	0.497	0.374	0.451	-0.077	0.780	0.597	0.826	-0.230
	No. of sites			281.0		(0.031)		82.0		(0.042)		82.0		(0.154)		75.0		(0.189)
Echid β1	Dunn β	Human β	0.191	0.155	0.208	-0.053	0.091	0.047	0.099	-0.052	0.497	0.395	0.481	-0.087	0.780	0.657	0.899	-0.241
	No. of sites			280.0		(0.030)		81.3		(0.038)		81.3		(0.159)		76.7		(0.203)
Echid β1	Opos β	Mouse β	0.198	0.179	0.239	-0.060	0.122	0.141	0.156	-0.015	0.659	0.374	0.314	0.060	0.939	0.597	0.895	-0.298
	No. of sites			281.7		(0.032)		82.3		(0.049)		82.3		(0.145)		74.0		(0.22)
Echid β1	Opos β	Human β	0.198	0.155	0.221	**-0.065***	0.122	0.047	0.109	-0.062	0.659	0.395	0.449	-0.054	0.939	0.657	0.907	-0.249
	No. of sites			280.7		(0.031)		81.7		(0.043)		81.7		(0.164)		75.7		(0.214)
Echid β1	Dunn β	Chick β	0.191	0.230	0.239	-0.009	0.091	0.099	0.065	0.035	0.497	0.238	0.527	-0.289	0.780	0.604	0.973	-0.369
	No. of sites			282.0		(0.032)		83.3		(0.038)		83.3		(0.16)		72.7		(0.21)
Platy β1	Dunn β	Chick β	0.193	0.240	0.243	-0.003	0.069	0.102	0.066	0.037	0.477	0.184	0.503	**-0.319***	0.814	0.576	1.002	-0.426
	No. of sites			277.7		(0.032)		83.3		(0.033)		83.3		(0.146)		71.0		(0.226)
Echid β1	Opos β	Chick β	0.198	0.230	0.271	-0.041	0.122	0.099	0.080	0.019	0.659	0.238	0.437	-0.200	0.939	0.604	0.919	0.316
	No. of sites			282.7		(0.033)		83.7		(0.043)		83.7		(0.143)		71.7		(0.216)
Platy β1	Opos β	Chick β	0.195	0.240	0.275	-0.035	0.111	0.102	0.081	0.021	0.638	0.184	0.447	-0.262	0.783	0.576	0.947	-0.371
	No. of sites			278.3		(0.033)		83.7		(0.042)		83.7		(0.141)		70.0		(0.215)
Echid β1	Human β	Chick β	0.155	0.230	0.241	-0.011	0.047	0.099	0.063	0.036	0.395	0.238	0.294	-0.056	0.657	0.604	0.711	-0.107
	No. of sites			279.7		(0.029)		82.3		(0.028)		82.3		(0.10)		76.0		(0.17)
Echid β1	Mouse β	Chick β	0.179	0.230	0.270	-0.040	0.141	0.099	0.118	-0.019	0.374	0.238	0.329	-0.091	0.597	0.604	0.890	-0.286
	No. of sites			280.7		(0.031)		83.0		(0.048)		83.0		(0.11)		74.3		(0.20)
Platy β1	Human β	Chick β	0.170	0.240	0.245	-0.005	0.061	0.102	0.064	0.038	0.370	0.184	0.299	-0.115	0.606	0.576	0.728	-0.152
	No. of sites			275.3		(0.031)		83.0		(0.032)		83.0		(0.098)		74.3		(0.164)
Platy β1	Mouse β	Chick β	0.194	0.240	0.274	-0.034	0.125	0.102	0.120	-0.018	0.350	0.184	0336	-0.151	0.664	0.576	0.916	-0.340
	No. of sites			276.3		(0.033)		83.0		(0.046)		83.0		(0.104)		72.7		(0.210)
Echid β1	Opos ε	Chick β	0.168	0.230	0.211	0.018	0.135	0.099	0.091	0.008	0.527	0.238	0.338	-0.100	0.896	0.604	0.825	-0.221
	No. of sites			281.3		(0.030)		83.0		(0.046)		83.0		(0.119)		73.7		(0.209)
Ngech β1	Opos ε	Chick β	0.173	0.229	0.211	0.018	0.135	0.111	0.091	0.020	0.435	0.230	0.338	-0.108	0.942	0.596	0.825	-0.230
	No. of sites			281.3		(0.030)		83.0		(0.046)		83.0		(0.113)		73.7		(0.210)
Platy β1	Opos ε	Chick β	0.160	0.236	0.211	0.025	0.141	0.101	0.091	0.010	0.477	0.201	0.338	-0.138	0.933	0.564	0.825	-0.261
	No. of sites			281.0		(0.029)		84.0		(0.046)		84.0		(0.113)		73.0		(0.210)

Species/Genes Compared			No. of Nonsynonymous Substitution Per Site								No. of Synonymous Substitution Per Site							
			Nondegenerate Sites				2-fold degenerate Sites				2-fold Degenerate Sites				4-fold Degenerate Sites			
1	2	3	K12	K13	K23	K13-K23	B12	B13	B23	B13-B23	A12	A13	A23	A13-A23	K12	K13	K23	K13-23
Echid β1	Dunn ε	Chick β	0.136	0.230	0.177	**0.052***	0.109	0.099	0.093	0.006	0.436	0.238	0.321	-0.083	1.049	0.604	0.711	-0.107
	No. of sites			281.0		(0.026)		83.7		(0.042)		83.7		(0.110)		73.3		(0.198)
Ngech β1	Dunn ε	Chick β	0.140	0.229	0.177	**0.052***	0.109	0.111	0.093	0.018	0.436	0.230	0.321	-0.091	1.071	0.596	0.711	-0.115
	No. of sites			281.0		(0.027)		83.7		(0.042)		83.7		(0.110)		73.3		(0.199)
Platy β1	Dunn ε	Chick β	0.132	0.236	0.177	**0.059***	0.119	0.101	0.093	0.008	0.445	0.201	0.321	-0.120	1.048	0.564	0.711	-0.147
	No. of sites			280.7		(0.026)		84.7		(0.043)		84.7		(0.108)		72.7		(0.196)
Echid β1	Human ε	Chick β	0.161	0.230	0.214	0.016	0.105	0.099	0.089	0.010	0.517	0.238	0.334	-0.097	0.703	0.604	0.734	-0.130
	No. of sites			282.0		(0.029)		83.3		(0.041)		83.3		(0.118)		72.7		(0.178)
Ngech β1	Human ε	Chick β	0.165	0.229	0.214	0.015	0.105	0.111	0.089	0.022	0.446	0.230	0.334	-0.105	0.671	0.596	0.734	-0.138
	No. of sites			282.0		(0.029)		83.3		(0.041)		83.3		(0.113)		72.7		(0.175)
Platy β1	Human ε	Chick β	0.157	0.236	0.214	0.022	0.119	0.101	0.089	0.012	0.380	0.201	0.334	-0.134	0.650	0.564	0.734	-0.170
	No. of sites			281.7		(0.029)		84.3		(0.043)		84.3		(0.105)		72.0		(0.172)
Echid β1	Mouse ε	Chick β	0.180	0.230	0.207	0.022	0.131	0.099	0.088	0.011	0.457	0.238	0.358	-0.120	1.035	0.604	0.794	-0.190
	No. of sites			281.0		(0.030)		85.0		(0.044)		85.0		(0.116)		72.0		(0.205)
Ngech β1	Mouse ε	Chick β	0.189	0.229	0.207	0.022	0.136	0.111	0.088	0.023	0.459	0.230	0.358	-0.128	1.033	0.596	0.794	-0.198
	No. of sites			281.0		(0.031)		85.0		(0.046)		85.0		(0.116)		72.0		(0.204)
Platy β1	Mouse ε	Chick β	0.176	0.236	0.207	0.029	0.139	0.101	0.088	0.013	0.416	0.201	0.358	-0.157	0.976	0.564	0.794	-0.230
	No. of sites			280.7		(0.030)		86.0		(0.045)		86.0		(0.110)		71.3		(0.200)
Echid β1	Ngech β1	Platy β	0.022	0.060	0.068	-0.008	0.000	0.025	0.025	0.000	0.025	0.088	0.088	0.000	0.013	0.105	0.090	0.015
	No. of sites			278.7		(0.009)		82.0		(0.00)		82.0		(0.021)		77.3		(0.014)
Echid β1	Ngech β1	Chick β	0.022	0.234	0.233	0.0002	0.000	0.101	0.113	**-0.012****	0.026	0.221	0.213	0.008	0.013	0.617	0.608	0.008
	No. of sites			276.3		(0.011)		81.7		(0.0016)		81.7		(0.029)		74.0		(0.026)
Echid β1	Ngech β1	Opos β	0.022	0.202	0.215	-0.014	0.000	0.124	0.129	**-0.004****	0.025	0.627	0.542	0.086	0.031	0.968	0.976	-0.0065
	No. of sites			277.3		(0.011)		81.0		(0.0006)		81.0		(0.068)		73.7		(0.041)
Echid β1	Ngech β1	Dunn β	0.022	0.195	0.204	-0.009	0.000	0.092	0.092	0.000	0.025	0.509	0.439	0.0699	0.013	0.800	0.807	-0.007
	No. of sites			276.7		(0.011)		80.7		(0.00)		80.7		(0.052)		74.7		(0.032)
Echid β1	Ngech β1	Mouse β	0.022	0.182	0.201	-0.019	0.000	0.143	0.147	**-0.005****	0.026	0.352	0.326	0.0266	0.013	0.609	0.573	0.0357
	No. of sites			275.3		(0.011)		80.3		(0.001)		80.3		(0.038)		76.3		(0.024)
Echid β1	Ngech β1	Human β	0.022	0.158	0.167	-0.009	0.000	0.048	0.048	0.000	0.025	0.374	0.321	0.053	0.013	0.671	0.642	0.029
	No. of sites			274.3		(0.011)		79.7		(0.00)		79.7		(0.038)		78.0		(0.025)
Ngech β1	Platy β1	Chick β	0.069	0.233	0.240	-0.007	0.025	0.113	0.102	0.010	0.089	0.213	0.184	0.028	0.092	0.609	0.576	0.0324
	No. of sites			276.0		(0.020)		82.7		(0.021)		82.7		(0.051)		73.3		(0.063)
Ngech β1	Platy β1	Opos β	0.069	0.215	0.195	0.021	0.025	0.129	0.111	0.018	0.089	0.542	0.638	-0.096	0.092	0.975	0.783	0.192
	No. of sites			277.0		(0.019)		82.0		(0.022)		82.0		(0.118)		73.0		(0.144)

Species/Genes Compared			No. of Nonsynonymous Substitution Per Site								No. of Synonymous Substitution Per Site							
			Nondegenerate Sites				2-fold degenerate Sites				2-fold Degenerate Sites				4-fold Degenerate Sites			
1	2	3	K12	K13	K23	K13-K23	B12	B13	B23	B13-B23	A12	A13	A23	A13-A23	K12	K13	K23	K13-23
Ngech β1	Platy β1	Dunn β	0.069	0.204	0.193	0.0112	0.025	0.093	0.069	0.023	0.089	0.439	0.477	-0.038	0.092	0.807	0.814	-0.0065
	No. of sites			276.3		(0.019)		81.7		(0.021)		81.7		(0.085)		74.0		(0.092)
Ngech β1	Platy β1	Human β	0.069	0.167	0.170	-0.003	0.025	0.048	0.061	-0.014	0.089	0.321	0.370	-0.049	0.092	0.642	0.606	0.0362
	No. of sites			274.0		(0.0185)		80.7		(0.0195)		80.7		(0.068)		77.3		(0.073)
Echid β1	Platy β1	Chick β	0.061	0.234	0.240	-0.006	0.025	0.101	0.102	-0.002	0.089	0.221	0.184	0.036	0.107	0.617	0.576	0.041
	No. of sites			275.8		(0.019)		82.7		(0.021)		82.7		(0.05)		73.3		(0.069)
Echid β1	Platy β1	Opos β	0.061	0.202	0.195	0.0071	0.025	0.124	0.111	0.013	0.089	0.627	0.637	-0.011	0.107	0.968	0.783	0.186
	No. of sites			277.0		(0.018)		82.0		(0.022)		82.0		(0.123)		71.0		(0.137
Echid β1	Platy β1	Dunn β	0.061	0.195	0.193	0.0018	0.025	0.092	0.069	0.0023	0.089	0.509	0.477	0.032	0.107	0.800	0.814	-0.014
	No. of sites			276.3		(0.018)		81.7		(0.021)		81.7		(0.09)		74.0		(0.094)
Echid β1	Platy β1	Mouse β	0.061	0.182	0.194	-0.0124	0.025	0.143	0.125	0.018	0.089	0.352	0.350	0.0023	0.107	0.609	0.664	-0.055
	No. of sites			275.0		(0.018)		81.3		(0.023)		81.3		(0.071)		75.7		(0.081)
Echid β	Platy β	Human β	0.061	0.158	0.170	-0.012	0.025	0.048	0.061	-0.014	0.089	0.374	0.370	0.004	0.107	0.671	0.606	0.065
	No. of sites			274.0		(0.017)		80.7		(0.020)		80.7		(0.071)		77.3		(0.079)

※ Table columns 1, 2 and 3 show test sequence 1, test sequence 2 and the reference sequence (outgroup) respectively. A_{ij}, B_{ij} and K_{ij} are, respectively, the number of transitional substitutions, the number of transversional substitutions, and the total number of substitutions between species i and species j. The numbers in the parentheses are the standard errors. If the values of (K13-K23), (B13-B23), or (A13-A23) are negative, this is indicated with "-" and implies a lower substitution value for the branch connecting test sequence 1 to the reference sequence, than test sequence 2 to the reference sequence. ** Significant at 1 % level, * Significant at 5 % level. Relative rate tests were carried out according to Wu and Li (1985) and the nucleotide substitution values were obtained from the method of Li (1993).

Table 3.3.2 Synonymous (above) and non-synonymous (below) divergence values (%, ± s.e.) for pairs of coding regions from β-like globin genes, estimated using the method of Li (1993). Comparisons are based on 146 codons, the start and stop codons being excluded.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. ECHID β1		**3**	**12**	**1**	64	87	98	64	67	69	61	68	96	91	103	67	80	84	79
		±2	**±3**	**±1**	±13	±15	±17	±11	±11	±12	±10	±11	±17	±15	±18	±11	±14	±14	±13
2. NGECHID β1	**2**		**10**	**1**	63	81	92	60	61	64	56	63	94	86	99	62	77	79	70
	±1		**±3**	**±1**	±13	±14	±16	±10	±10	±11	±10	±10	±17	±15	±17	±10	±13	±13	±12
3. PLATY β1	**6**	**6**		**10**	58	88	91	65	62	78	57	63	97	89	95	73	82	81	66
	±1	**±1**		**±3**	±12	±15	±16	±11	±11	±14	±10	±11	±16	±15	±17	±12	±14	±13	±11
4. PLATY β2	**0**	**2**	**6**		65	86	98	62	64	67	59	65	94	88	100	65	79	83	74
	±0	**±1**	**±1**		±11	±14	±17	±11	±11	±11	±10	±11	±17	±15	±17	±11	±13	±14	±13
5. CHICKEN β	23	23	23	23		106	97	83	70	71	62	69	67	76	79	68	86	63	70
	±3	±3	±3	±3		±20	±18	±14	±12	±12	±11	±12	±11	±13	±14	±12	±15	±11	±12
6. DUNNART β	19	19	18	19	22		41	88	90	83	81	96	99	112	111	87	85	86	83
	±3	±3	±3	±3	±3		±8	±15	±17	±15	±14	±17	±17	±20	±21	±15	±14	±15	±15
7. OPOSSUM β	19	21	19	19	25	13		83	84	95	88	99	81	101	113	86	85	83	84
	±3	±3	±3	±3	±3	±2		±15	±16	±18	±16	±18	±15	±19	±22	±15	±15	±14	±15
8. MOUSE β	19	21	20	19	26	21	24		48	47	38	42	82	83	80	66	84	63	58
	±3	±3	±3	±2	±3	±3	±3		±9	±8	±7	±8	±14	±15	±14	±11	±15	±11	±10
9. GOAT β	18	16	18	18	27	21	23	17		28	22	28	81	67	78	50	59	64	50
	±3	±2	±3	±3	±3	±3	±3	±3		±6	±5	±6	±13	±11	±14	±9	±10	±11	±9
10. RABBIT β	14	14	15	14	24	19	21	14	11		22	26	66	57	70	50	75	57	49
	±2	±2	±2	±2	±3	±3	±3	±2	±2		±5	±5	±11	±10	±12	±9	±13	±10	±9
11. GALAGO β	15	16	15	15	22	18	20	11	14	5		22	68	62	64	50	71	55	41
	±2	±2	±2	±2	±3	±3	±3	±2	±2	±1		±5	±11	±10	±11	±9	±12	±10	±8
12. HUMAN β	14	15	16	14	23	20	21	13	12	5	6		81	68	82	63	74	67	52
	±2	±2	±2	±2	±3	±3	±3	±2	±2	±1	±1		±13	±12	±14	±11	±13	±11	±9
13. DUNNART ε	14	14	14	14	17	15	17	17	19	13	12	14		27	80	49	64	59	58
	±2	±2	±2	±2	±3	±2	±2	±3	±3	±2	±2	±2		±6	±15	±9	±12	±10	±10
14. OPOSSUM ε	18	18	17	17	20	17	20	21	22	15	14	16	11		80	49	56	55	57
	±3	±3	±3	±3	±3	±3	±3	±3	±3	±2	±2	±2	±2		±14	±9	±10	±10	±10
15. MOUSE ε	18	19	19	18	19	18	21	18	19	16	14	17	13	13		63	64	50	48
	±3	±3	±3	±3	±3	±3	±3	±3	±3	±2	±2	±3	±2	±2		±11	±12	±9	±9
16. GOAT ε	16	17	15	16	18	15	18	14	17	13	12	15	8	11	8		42	35	34
	±2	±3	±2	±2	±3	±2	±3	±2	±3	±2	±2	±2	±2	±2	±2		±8	±7	±7
17. RABBIT ε	16	17	15	16	21	16	19	19	21	17	15	16	10	12	11	8		33	42
	±2	±3	±2	±2	±3	±2	±3	±3	±3	±3	±2	±2	±2	±2	±2	±2		±7	±8
18. GALAGO ε	16	17	15	16	18	17	18	16	19	15	14	16	10	11	9	4	6		29
	±2	±2	±2	±2	±3	±2	±3	±2	±2	±2	±2	±2	±2	±2	±2	±1	±1		±6
19. HUMAN ε	16	17	17	16	21	17	19	17	19	16	14	16	10	13	9	5	8	6	
	±2	±3	±3	±2	±3	±3	±3	±3	±3	±2	±2	±2	±2	±2	±2	±1	±2	±1	

CHAPTER 4 DISCUSSION

4.1 Structure of monotreme β-globin genes

In this project, β-like globin genes were isolated and characterised from each of the three monotreme species. *Echid* β1 was isolated from an Australian echidna genomic DNA library, using a marsupial ε-globin (pSG2H) gene probe. *Platy* β1 and *platy* β2 were isolated from a platypus genomic library using *echid* β1 (pTAG-2) as a probe, and *NGechid* β1 was successively amplified using PCR from New Guinea echidna DNA using primers designed on the basis of the *echid* β1 sequence.

The following structural features of the monotreme β-globin genes suggest that these genes are functional :

- Each gene contains three exons characteristic of all vertebrate β-globin genes and the combined exons contain a single open reading frame.

- The four genes conceptually translate to β-polypeptides of the expected length of 146 amino acids.

- The introns interrupt the coding region at the same positions as those found in all vertebrate β-globin genes (i.e. intron 1 splits codon 30 and intron 2 separates codons 104 and 105) and the intron sequences contain donor and acceptor splice signals that conform to the GT-AG rule.

- In three of the genes (*echid* β1, *platy* β1 and *platy* β2), the promoter elements identified in the 5' flanking region include "CACCC", "CCAAT" and the RNA polymerase II binding site (ATA) located at -92 ~ -93, -75 ~ -78, and -33, respectively from the putative cap site (CAG). These elements are characteristic of functional globin genes.

- Initiation (ATG) and termination (TGA) codons and a putative poly (A) adenylational signal 71~99 bp downstream of the termination codon are present in all four genes.

The following additional features of the monotreme β-like globin genes provide evidence that they are functionally expressed in post-natal stages:

- The conceptually translated coding region of *echid* β1 gives an amino acid sequence that is identical to that determined by Whittaker *et al.* (1972) from an analysis of β-globin chains chromatographically fractionated from the major haemoglobin component (Hb I) of adult echidna blood. Translation of *platy* β1 gives an amino acid sequence that is identical to the major adult β-globin sequences determined by Whittaker and Thompson (1975) from an adult platypus. A comparison of the amino acid sequence conceptually translated from *NGechid* β1 and the other monotreme β-globin chains show a high level of sequence identity - 96% with the echid β1-chain and 90% with the platy β1-chain.

- Two features of the promoter regions that can be used to distinguish the β- and ε-globin genes of eutherians and marsupials are: i) the ε- and β-globin genes have different spacings of their promoter motifs (5'- "CACCC" - "CAATT" - "ATA" - 3'), (ii) adult-like β-globins have two "CACCC" boxes, while the embryonic-like β-globins have a single "CACCC" box (Koop and Goodman, 1988; Cooper and Hope, 1993). The β-globin promoter elements in *echid* β1, *platy* β1 and *platy* β2 are more similar in number and location to the marsupial and eutherian β-globin promoter regions than to those of the marsupial and eutherian ε-globin genes (see Figure 3.1.24).

- The four monotreme β-globins have leucine at position 3, a characteristic residue of the adult-like β-globin genes of eutherian and marsupials. Koop and Goodman (1988) noted the replacement of leucine at position 3 of putative proto β-globin with phenylalanine in the putative proto ε-globin. They suggested that such a substitution would be expected to reduce DPG binding and increase oxygen

affinity of ε-globin over that of maternal β-globin, thereby favouring oxygen transport from mother to embryo. Phenylalanine is found at position 3 in marsupial ε and eutherian ε, γ, and η chains, whereas leucine occupies this site in marsupial and eutherian adult β-globins. All these observations suggest that the characteristic promoter motifs of the mammalian adult-like β-globin genes, and the presence of a leucine residue at position 3 in the β-globin, evolved prior to the divergence of the monotreme, marsupial and eutherian lineages.

4.1.1 Features of the monotreme β-globin gene introns and flanking sequences

The four monotreme β-globin genes have the longest first introns (146-153 bp) and the shortest second introns (440-465 bp) of any described functional adult mammalian β-globin gene. Interestingly, intron length comparisons between the marsupial β-like globin genes (include embryonic and adult-expressed genes) and monotreme β-globin genes show that the length of intron I of marsupial globin genes is very short and the length of intron II is very long in contrast to the monotreme β-globin gene introns (see Fig. 3.1.23). Comparisons of intron lengths of functional β-like globin genes from various groups including birds, marsupial and eutherians (Appendix 6) show that on average, intron I of bird β-globin genes are the shortest in length (101 bp) followed by marsupials (114 bp), eutherians (129 bp) and monotremes (148 bp). Marsupials have the longest intron II (1,576 bp) followed by eutherians (831 bp), birds (721 bp) and monotremes (457 bp). In general, intron sizes within each group of mammals (monotremes, marsupial and eutherians) are well conserved. The causes and significance of variation in intron length are unknown.

Pair-wise comparisons of intron sequences showed that no significant sequence identity exists between corresponding introns in the monotremes and other mammalian β-like globin genes (Appendix 7, A and B). A comparison of the dot plots shown in Appendix 7 (A) with those shown in Appendix 8 indicates that the intron I sequences of monotremes and other mammals (marsupials and eutherians) are likely to have last shared a common ancestor greater than 150 MYA.

The dot plot comparisons amongst different monotreme introns (Appendix 7, C and D) indicate a high level of sequence identity, as would be expected of the introns of recently separated orthologous genes. Comparisons between the dot plots shown in Appendix 7 (C) and those shown in Appendix 8 indicate that the monotreme globin genes are likely to have been derived from a common ancestral gene 60 MYA or less.

Tetra-nucleotide microsatellites consisting of a "TGAA" repeats (11 repeats) are found in the *echid* β1 and *platy* β2 genes. The repeats are located at approximately 2-3 kb downstream from the termination codon of the *echid* β1 and in the 5' flanking regions around 1 kb from the start codon of the *platy* β2-globin gene. Two regions of "AAT" trinucleotide repeats were also detected in the flanking regions of monotreme β-globin genes. These repetitive sequences may provide useful PCR based genetic markers for population studies in monotremes.

4.1.2 G+C content of the monotreme β-globin genes

As shown in the Figure 3.1.24, the monotreme β-globin genes have high G+C (GC) content compared to other mammalian β-like globin genes. Their level of GC content is more similar to that of chicken β-like globin genes.

Analyses of DNA sequence data have revealed a positive correlation between the GC levels of genes, and the GC level of the larger DNA regions in which they are embedded (Bernardi *et al.* 1985; Ikemura, 1985; Aota and Ikemura, 1986; Ikemura and Aota, 1988; Wolfe and Sharp, 1993). Each gene cluster is contained within a very long chromosomal segment (much larger than 200 kb) of homogeneous base composition called an isochore (Bernardi, 1989). The "genome hypothesis" of Grantham *et al.* (1980) suggests that there is some intrinsic genome-wide pressure favouring the adoption of a particular (G+C)/(A+T) ratio in particular species. Forsdyke (1995) has suggested that fine-tuning of the (G+C)/(A+T) ratio could have been a key component of the post-zygotic isolation process leading to speciation.

The mammalian β-like globin genes are generally low in GC content, and are embedded in a low GC region (the AT rich isochore). On the other hand, the α-like globin genes are GC rich and are embedded in the GC rich isochore fraction (Li and Graur, 1991). In chickens, both α- and β-globin genes are GC rich, and both are embedded in GC rich regions (Reitman *et al* 1993).

The similar levels of GC in monotreme and avian β-like globin genes suggests that the ancestral mammalian β-like globin genes had a high GC level, and that after the separation of the mammalian lineages, the GC level of marsupial and eutherian β-like globin gene lineages was reduced. Given that a high GC content is likely to be the plesiomorphic condition, the low GC content of marsupials and eutherians is probably a synapomorphy suggesting that marsupials and eutherians share a more recent common ancestor.

4.2 The *Platy* β2-globin gene

The *platy* β2-globin gene, isolated from a platypus genomic DNA library, showed an extremely high level of nucleotide sequence identity with the *echid* β1-globin gene in coding (99%) and intron (97%) regions. These two genes, *platy* β2- and *echid* β1-globin, showed the lowest divergence values at synonymous and non-synonymous sites of all pairs of monotreme β-globins (Table 3.3.2). The *platy* β2 is more similar in sequence to *echid* β1 than it is to *platy* β1 and than *echid* β1 is to *NGechid* β1.

These observations are surprising and unexpected for the following reason. Li *et al.* (1985) suggested that the average rate of non-synonymous substitutions in mammalian β-globin genes is 0.87×10^{-9} per year. Based on the assumption that the rate of nucleotide substitution in β-like globin genes are approximately constant, the divergence values at non-synonymous sites between *platy* β2 and *echid* β1 would lead to an estimate of a very recent separation time of these two genes. Estimates of the time that the platypus and echidna lineages and the New Guinea echidna and Australian echidna lineages separated are 30-60 MYA and 1-2 MYA respectively,

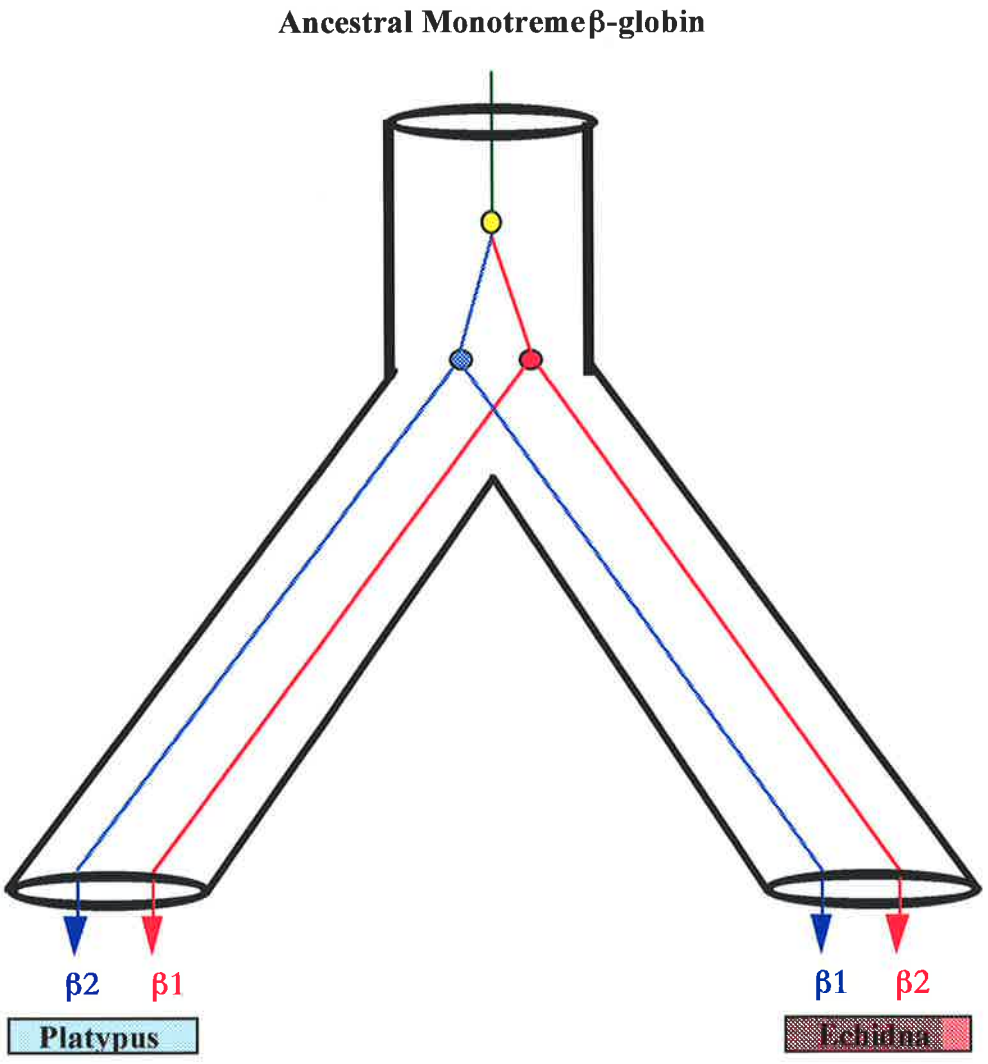
based on fossil and molecular data (DNA-DNA hybridisation and sequence analyses involving mitochondrial 12SRNA and protamine p1 genes) (Griffiths *et al.* 1991; Westerman and Edwards, 1992; Springer and Kirsch, 1993; Retief *et al.* 1993; Gemmell and Westerman, 1994). Even the lower end of these time estimates is well above the estimates deduced from the DNA sequence comparisons, of the time of separation of *platy* β2 and *echid* β1.

The difficulty in interpreting the similarity between the *platy* β2 and *echid* β1 genes in an evolutionary context (see above) suggests that alternative explanations for these observations are required. Amongst the possibilities are:

- A gene duplication event, that occurred before the separation of the platypus and echidna lineages, gave rise to two paralogous β-like globin lineages in monotremes (Fig. 4.1). The subsequent speciation of platypus and echidna resulted in the presence in each group of a pair of paralogous genes, represented in the platypus by *platy* β1 and *platy* β2, and in the echidna by *echid* β1 and a putative, but as yet unidentified, orthologue of *platy* β1. After the speciation event, the orthologous genes evolved independently in the two monotreme lineages. The orthologues *echid* β1 and *platy* β2 evolved very slowly (perhaps 20x more slowly than "normal" for β-globin genes), possibly due to some unknown selective constraints, and the *platy* β1 and putative *echid* β2 evolved relatively far more quickly. For this to be the correct explanation, it would be necessary to propose that both the coding regions and the introns evolved relatively slowly in the *echid* β1 and *platy* β2 branches. This would seem to be very unlikely.

- Horizontal gene transfer, possibly mediated by a virus, could possibly have introduced a β-globin gene from the echidna to the platypus. Busslinger *et al.* (1982) have suggested that horizontal gene transfer between species could explain the temporal pattern of evolution in the three different histone gene clusters in sea urchin species. The vertebrate genome contains a number of sequences that are homologous to retroviruses and there are several examples of such sequences being transferred between vertebrate species.

Figure 4.1 An hypothetical phylogenetic tree of monotreme β-globin genes that could account for the existence of *platy* β2 in the platypus. Gene lineages (thin lines) have been fitted within a species phylogeny (thick lines). A green line depicts the ancestral monotreme β-globin gene. The circle filled with yellow represent a gene duplication, producing paralogous genes, that occurred before the separation of the echidna and platypus lineages. A blue line depicts one of the globin gene lineages following a duplication event, and the red line depicts the other globin lineage. Note: *echid* β2-globin has not been detected. The tree is not drawn to a time scale.



For example, type-C virogenes with very similar sequences have been detected in the cellular DNA of all Old World Monkeys and cats (Li and Graur, 1991). The theory of horizontal gene transfer is a salutary challenge to currently held views of a totally coherent evolution and orderly transmission of genes in eukaryotes. It would seem unlikely that horizontal transmission between the monotremes could account for the similarity of *platy* β2 and *echid* β1. Both genes appear to be intact and include intron and flanking sequences; this would be unexpected if viral vectors were responsible. However, nothing is known about the parasites in monotremes, and the possibility of horizontal transmission can not be completely disregarded.

- DNA contamination could have lead to the mistaken conclusion that the *platy* β2 was derived from the platypus, when in fact it was of echidna origin. The platypus genomic DNA library may have been constructed from DNA contaminated with echidna DNA. However, several lines of evidence suggest this is not the case: i) DNA used to make the platypus library (PGL.1.1) was prepared at La Trobe University (Department of Genetics and Human Variation) from a platypus cell line. Echidna cell lines were not being grown in this laboratory and therefore cross contamination between cell lines is unlikely to have occurred. ii) PCR was used to determine whether the *platy* β2 gene existed in an independent sample of platypus genomic DNA extracted from liver. A PCR product was amplified using *platy* β2 specific primers, and directly sequenced. The sequence showed a very high level of identity to *platy* β2 (Fig. 3.1.17 B), suggesting that contamination of the DNA used to make the platypus genomic DNA library is unlikely. iii) The restriction maps of the two lambda clones λPL.1.1 (containing *platy* β2) and λEL.1.1 (containing *echid* β1) are different (Fig. 3.1.18). If the platypus library was contaminated with previously isolated *echid* β1, then the restriction maps should be the same.

The existence of the *platy* β2 globin gene in the platypus is very difficult to explain in an evolutionary context and further research is required using PCR and genomic Southern analyses, with a larger number of platypus and echidna DNA samples. While the PCR technique is unlikely to amplify *Eco*RI DNA fragments as large as those (6 - 10 kb) shown in Figure 3.1.26, it could be used to amplify the smaller (0.85 kb) *Eco*RI

fragment. It proved difficult, possibly for technical reasons, to detect hybridising fragments using Southern analyses (Fig. 3.1.26) although these experiments were repeated several times. Southern analysis could be extended by the use of additional restriction enzymes. Further research is required to confirm the species of origin of *platy* β2-globin gene. If such research definitely rules out contamination as a cause of the findings, then one would be forced to invoke an evolutionary model such as that depicted in Fig. 4.1 support for this model would be provided by the discovery of the putative *echid* β2 gene in the echidna.

4.3 Existence of ε-like globin genes in monotremes

Despite repeated attempts, an ε-like globin gene could not be isolated from the echidna or platypus, either by screening genomic libraries with marsupial (*S. crassicaudata*) ε-globin probes or by using PCR strategies (with specific primers designed from bird, marsupial and eutherian ε-globin genes). Southern analysis of *Eco*RI restricted monotreme genomic DNA (Fig. 3.1.26 B) using as a probe an 856 bp *Eco*RI fragment containing exon 2 and part of exon 3 of the e*chid* β1-globin gene (see Section 2.1.8), showed hybridisation to the expected 856 bp fragment as well as two additional larger fragments (8.7 kb and 10 kb in echidna and 6.8 kb and 10 kb in platypus) (Fig. 3.1.26 B). The two hybridising fragments in the platypus (6.8 kb and 0.85 kb) could be accounted for by the *platy* β1 and *platy* β2 genes, respectively. The third fragment (10 kb) may contain a β-like globin gene expressed early in platypus development, and this gene could be an orthologue of the ε-globins of marsupials and eutherians. This putative gene is yet to be isolated and identified. However, these observations provide evidence for the existence of two and possibly three β-like globin genes in monotreme genomes.

Recently, Holland and Gooley (1997) reported the presence of a third β-like globin gene in an Australian marsupial, *M. eugenii* (Tammar Wallaby). They found four distinct haemoglobin types using isoelectric focusing of neonatal Tammar blood proteins. Purification and partial amino acid sequencing of the β-like globin chains

present in these early expressed globins revealed two different chains. One of these appeared to be orthologous to *S. crassicaudata* and *D. virginiana* ε-globin. However, the other (named ω-globin) contained a number of unexpected features.

Comparisons of the ω-globin amino acid sequence to β-like chains from a range of vertebrate species showed that ω-globin most closely resembled avian β-globin chains. Phylogenetic analyses also showed the partial ω-globin amino acid sequence was monophyletic with avian β-like globin sequences to the exclusion of mammalian β-like globin sequences (Hope and Goodman, personal communication).

A model to account for these observations has been proposed by Holland, Hope, Cooper, Gooley, Wheeler and Goodman (unpublished). It is proposed that a gene duplication event occurred in a β-like globin gene before the separation of birds and mammals. One of the duplicate genes so formed was the progenitor of the ω-globin gene and the β-like globin genes found in birds, while the other was the progenitor of all the other mammalian β-like globin genes so far described. A prediction of this model is that a ω-globin orthologue would exist in monotremes, and such a gene may be expressed early in monotreme development. In this context it would be particularly interesting to use ω-globin specific probes to search for this gene (or pseudogene) in monotremes.

One way of identifying early expressed globin genes in monotremes would be carry out iso-electric focusing on globin proteins derived from the blood of developing embryos and neonates. If globin proteins were found that were restricted to early developmental stages, their polypeptide chains could be sequenced, and PCR primers could be designed to enable the corresponding genes to be isolated and characterised. However, such an approach would be very difficult given the practical problems of obtaining material from monotreme embryos and neonates. Another approach would be to use Northern analysis with ε-specific probes from birds, marsupials and eutherians. Again, appropriate material for RNA extraction would be difficult to

obtain, and probes may not cross hybridise specifically to monotreme ε-like globin sequences.

4.4 Chromosomal location of β-globin genes in monotremes

In situ hybridisation on Australian echidna chromosomes with a ribosomal RNA gene probe confirmed the previous localisation of this gene to 3p, $6P^{H}$ (Graves *et al.* 1993) (Fig. 3.2.2 A). The β-globin gene has been localised to platypus chromosome $4q^{H}$ using *in situ* hybridisation with radio-labelled human and mouse β-globin cDNA probes (Watson *et al.* 1991). In the present project, the *echid* β1-globin gene was localised to one of the small echidna chromosomes (group 14-22) (Fig. 3.2.2 B and C).

Around 40, and six gene loci have been localised on the platypus and Australian echidna chromosomes, respectively (Graves *et al.* 1993). The six gene loci from echidna were localised to the X-chromosome (four) and autosomes (two). The four X-linked gene loci in the echidna were also localised to the platypus X chromosome (Graves and Watson, 1991). However, there are differences in the locations of the two autosomal loci between the two monotreme species. The ribosomal RNA gene, for example, was localised to chromosome 6p in the platypus and 3p and 6p in the Australian echidna (Watson *et al.* 1991; Graves *et al.* 1993).

Considering that the platypus and echidna have different karyotypes (see Section 1.2.5) and that some differences exist in chromosomal sizes and G-banding pattern amongst the six large chromosomes (chromosome 1-6) in both species, the different chromosomal location of the β-globin loci in these two species is not entirely unexpected. It is possible that the ancestral genome of platypus and echidna possessed a large chromosome homologous to platypus chromosome 4 and that one or more fission events broke up this chromosome in the echidna lineage, one of the products being the smaller chromosome on which the β-globin locus is located. In contrast, chromosome 4 in the platypus may have arisen through fusion events from smaller chromosomes including the chromosome in the 14-22 group presently found in the

echidna. It is impossible to distinguish between these scenarios given the fragmentary data available and given the low resolution of G-banding of monotreme chromosomes. In order to understand the karyotypic evolution in monotremes, there is a need for more research including detailed G-banding comparisons between the echidna and platypus chromosomes, and the localisation of more genes in monotremes. It would be interesting, for example, to localise in the platypus and echidna, gene loci that are known in eutherian mammals to be associated with the β-globin locus as part of a syntenic group, e.g. *CAT* and *HRAS-1*.

4.5 Rates of evolution of β-globin genes in mammals

One of the main issues in molecular evolution is whether rates of substitution are constant or not in the gene lineages of different taxa (Goodman *et al.* 1985; Wu and Li, 1985; Easteal, 1985, 1989, 1990, 1991; Britten, 1986; Li and Tanimura, 1987; Li *et al.* 1987; Springer and Kirsch, 1989; Easteal and Collet, 1994; Easteal and Herbert, 1997). If constancy of substitution rates can be shown to exist for a set of data, then it may be possible to calibrate a molecular clock and draw conclusion about the timing of evolutionary events such as gene duplications and taxa separations.

Goodman and his colleagues (1985) have emphasised that rates of nucleotide substitution have slowed down during the evolution of higher primates and during hominoid descent. Using the "relative rate test" Wu and Li (1985) compared the coding regions of 11 genes from rodents and humans, using genes from either the goat or rabbit as outgroups. They concluded that there was a significantly faster rate of nucleotide substitution in the rodent lineage than in the human lineage and that the differences in substitution rate could be accounted for by the shorter generation times of rodents. However, Easteal (1989, 1991) conducted a series of relative rate tests of β-like globin genes using sequences from 17 eutherian mammals and suggested that molecular evolutionary rates are generally uniform both among primates and between primates and other mammalian orders and there is stochastic uniformity of the molecular evolutionary rate among the eutherian orders. He concluded that more extensive investigations were required to resolve the apparent inconsistencies in the

interpretations by different workers (Easteal, 1985, 1989, 1990, 1991; Easteal and Collet, 1994; Easteal and Herbert, 1997).

In the present study, relative rates tests were carried on 71 pairs of β-like globin gene sequences from mammals (Table 3.3.1). Several results are highlighted:

- significant differences were observed in the substitution rates along paired phylogenetic branches. However, given the larger number of tests carried out, it would be expected that some of these significant differences would have occurred by chance.
- when different outgroups were used to test substitution rates along the same pair of phylogenetic branches, some consistency in the results was observed, as judged by pooling positive and negative values. However, only the chicken β-globin can be used with confidence as an outgroup for comparisons involving monotreme, marsupial and eutherian β-like globin genes, so the results obtained using other outgroups must be treated with caution.
- Substitution rates in mouse β-globin genes were significantly higher than those of human β-globin genes at non-synonymous sites (1% significant level) supporting previous results by Wu and Li (1985).
- There was a consistent trend in the results indicating that marsupial β-globin genes evolve at a higher rate than eutherian β-globin genes at synonymous and non-synonymous sites.
- Mammalian β-globin genes evolve more rapidly than mammalian ε-globin genes. Of particular interest was the finding that *S. crassicaudata* β-globin evolved significantly faster than *S. crassicaudata* ε-globin at non-degenerate and degenerate sites (see Table 3.3.1). This finding supports the hypothesis of Koop and Goodman (1988) that the rates of substitution in ε-globin genes are slower than in β-globin genes.
- Evidence is presented that the monotreme β-globin genes evolved at a slower rate than the marsupial and eutherian β-globin genes (Table 3.3.1). Similar findings have been made by Gemmell and Westerman (1994) based on a study of the mitochondrial 12S ribosomal RNA. Among the mammalian adult-like β-globin

genes, the monotreme β-globin genes appear to have evolved the slowest followed by eutherian β-globin genes and marsupial β-globin genes.

Based on the relative rate tests reported here, there was some evidence for differences in rates of evolution of different β-like globin gene lineages between different mammalian groups. However, taken overall, no strong and consistent trends were apparent.

4.6 Evolutionary relationships of the monotreme β-globin genes

It has been argued that the four monotreme β-globin genes (*echid* β1, *NGechid* β1, *platy* β1 and *platy* β2) are likely to be adult expressed globin genes. Based on the following observations, there is evidence that the monotreme β-globin genes originated from a single progenitor β-globin gene, present before the monotremes diverged.

First, a comparison between *platy* β1- and *echid* β1-globin genes in their putative 5' promoter motifs reveals that these regions are highly conserved in location and sequence. Secondly, a comparison of the nucleotide sequences of the four monotreme β-globin genes show that these genes have a high level of sequence identity in coding and non-coding regions. In particular, the intron sequences show significant sequence identity (Appendix 7). The percent nucleotide identity in the coding regions ranges from 92% to 99%, and in the introns from 83% to 97% (Table 3.1.2). The divergence values among the four monotreme β-globin coding sequences at synonymous (1-12%) and non-synonymous sites (0-6%) are low. Thirdly, the various phylogenetic trees show that the monotreme β-globin genes form a monophyletic group to the exclusion of all other mammalian β-like globins, with 100% bootstrap support.

These results suggest that the four genes have evolved from one common ancestral β-like globin gene present in a progenitor of the monotremes.

4.7 Time of separation of the monotreme species

Relative rate tests carried out among the monotreme β-globin genes reveal no significant difference in evolutionary rates (Section 3.3.5). Therefore, with some degree of confidence, one can assume the operation of a molecular clock. Based on such an assumption, it is estimated that the echidna and platypus lineages separated about 41-57 MYA and the Australian echidna and New Guinea echidna separated about 11-20 MYA. The first of these estimates is in conformity with the generally accepted divergence time estimates of 28-54 MYA for amino acid sequence comparisons of α-globins, β-globins and myoglobins (Whittaker *et al.* 1978; Richardson, 1987) and 62-78 MYA for DNA-DNA hybridisation data (Westerman and Edwards, 1991, 1992). Based on analyses of protamine p1 and mitochondrial 12S rDNA sequences, the divergence time of the platypus and echidna lineages was estimated to be more recent, i.e. 18-27 MYA and 25-30 MYA respectively (Retief *et al.* 1993; Gemmell and Westerman, 1994) (Table 4.1).

The estimate for the time of separation of the two echidna species is far older than the estimates of 1-2 MYA obtained from DNA-DNA hybridisation and mitochondrial 12S rDNA sequence comparisons (Westerman and Edwards, 1992; Gemmell and Westerman, 1994) (Table 4.1). However, the date estimated from this study is consistent with the finding of *Tachyglossus* and *Zaglossus* fossils, dated at 5 MYA, from both Australia and New Guinea (Flannery, 1990). This fossils imply the separation of these two species occurred before the Pleistocene. Other palaeontologists, however, suggest more recent separation of the two species, i.e. about 1-2 MYA (Griffiths *et al.* 1991).

It has been suggested that the DNA-DNA hybridisation technique is most effective for resolving phylogenies of taxa that separated 5-50 MYA (Hillis *et al.* 1996). Therefore, the estimates derived by Westerman and Edwards (1992) may not be optimal. Based on an analysis of mitochondrial 12S rDNA sequences, Gemmell and Westerman (1994) estimated the separation dates for the platypus and echidna and for

the two echidna lineages. Their original estimate for the platypus-echidna separation was 14-15 MYA, but they later adjusted the estimates to 25-30 MYA, suggesting that the original lower values may be underestimates because they were based only on the comparison of sequences from the conserved 3' portion of the 12S rDNA (Gemmell and Westerman, 1994). However, the authors did not adjust their estimates for the separation dates of the two echidna species (1-2 MYA) and it is possible that these values are also underestimates.

Table 4.1 The estimated separation dates (MYA) among monotreme species.

Taxa compared			Basis of Comparisons	Reference
Australian Echidna / Platypus	Australian Echidna / New Guinea Echidna	New Guinea Echidna / Platypus		
52 38 28	-	-	Hb α Hb β Myoglobin	Whittaker *et al.* (1978)
54	-	-	α+β+myoglobin	Richardson, (1987)
62-78	~1	-	DNA-DNA hybridisation	Westerman & Edwards, (1992)
18-27	-	-	protamine p1	Retief *et al.* (1993)
	1-2		fossil skulls	Griffiths *et al.* (1991)
25-30	1-2	-	mitochondrial 12S rDNA	Gemmell & Westerman, (1994)
41.1-51.8	**11.5-20.3**	**46.2-56.7**	**β-globin nucleotide sequences**	**Present study**

4.8 Evolution of the β-globin gene family in mammals

To investigate the evolution of the β-like globin genes in mammals, phylogenetic analyses of 30 adult and embryonic β-like globin genes from monotremes, marsupial and eutherians were carried out. As expected, the eutherian β-like globin genes were

separated into two main clades, one containing the embryonic-like and the other the adult-like genes. With one exception (see the maximum likelihood trees, Fig. 3.3.5) the marsupial ε-globin clade (comprising opossum and dunnart ε-globins) joined the root of the clade comprising the eutherian ε-like globin genes. This is the arrangement expected from the evolutionary model proposed by Koop and Goodman (1988) and supported by data from Cooper and Hope (1993) and Cooper *et al.* (1996). These studies showed that the first duplication of an ancestral β-like globin gene occurred 150-200 MYA before the separation of the marsupial and eutherian lineages. This duplication gave rise to the production of two major β-globin gene lineages, one of which became developmentally delayed in its expression (proto β-globin), and the other became expressed in embryonic tissues only (proto ε-globin) (Efstratiadis *et al.* 1980; Goodman *et al.* 1987a; Koop and Goodman 1988; Cooper and Hope, 1993; Cooper *et al.* 1996). About 100 MYA, a series of gene duplications began in a stem species of eutherians. The proto ε duplicated two times producing the progenitors of ε-, γ- and η-globin genes and the proto β duplicated once producing the progenitor of δ- and β-globin genes. The five genes were linked in the order: 5'-ε-γ-η-δ-β-3' (Goodman *et al.* 1984; Hardison, 1984, 1991; Harris *et al.* 1984).

In the majority of phylogenetic analyses the eutherian embryonic-like and adult-like β-globin genes were in separate monophyletic groups. The eutherian embryonic-like genes were resolved into three monophyletic groups: i) ε-like globin genes, ii) γ-like globin genes, iii) η-like globin genes (goat ε-globin genes) as expected. In general, phylogenetic tree patterns in this study support the hypothesis that the five-member ancestral β-globin cluster (5'-ε-γ-η-δ-β-3') existed prior to the radiation of the four eutherian orders, Rodents, Lagomorphs, Primates and Artiodactyls. If the above hypothesis is correct, the primate δ-globin gene is expected to group with the human δ-globin gene and the primate β-globin gene is expected to group with human β-globin gene. However, in humans and primates the β-globin genes group together with their respective δ-globin genes. It has been suggested that the δ-globin gene was involved in multiple gene conversions with the β-globin gene during the evolution of primates (Jeffreys *et al.* 1982; Martin *et al.* 1983; Hardison, 1991). Multiple gene conversions

are likely to confuse the phylogenetic affinities of these genes by introducing convergent residues which support alternative phylogenetic arrangements (Patterson, 1987; Goodman *et al.* 1987b). Therefore, it is likely that the anomalous grouping of the δ- and β-globins in primates is a result of gene conversion.

Of the phylogenetic trees produced in the present study, the marsupial β-globin genes did not form a sister clade to the eutherian β-globin genes. An extensive series of relative rates tests showed that there were no significant overall differences in the rates of nucleotide substitution of marsupial and eutherian ε-globin genes. However, the rates of molecular evolution of marsupial β-globin genes appeared to be higher than those of the eutherian β-globin genes, especially at synonymous sites where the rate was consistently higher in all comparisons (although not significantly so). This high rate of evolution at synonymous sites may have resulted in saturation or convergence at 3rd codon sites and this could, conceivably, lead to homoplasies in the data set and the anomalous position of the marsupial β-globin clade. It is also likely that the ε/β duplication occurred just prior to the divergence of marsupial and eutherian lineages and there are insufficient information sites to fully resolve this region of the phylogenetic tree.

4.8.1 Evolutionary relationships among monotreme, marsupial and eutherian β-like globin genes

In examining the evolutionary relationships between the monotreme, marsupial and eutherian β-like globins, the following observations need to be taken into account.

1. Comparisons of the 5' promoter regions show that in their sequence organisation, the monotreme β-globin genes most closely resemble eutherian and marsupial adult β-globin genes, rather than the embryonically and foetally expressed genes.

2. The four monotreme β-globins have leucine at position 3 which is a characteristic of the adult-like β-globin genes.

3. Most of the phylogenetic analyses (except one maximum likelihood tree- Fig. 3.3.5) show that the monotreme β-globin gene clade is ancestral to all other mammalian β-like globin genes (eutherian and marsupial ε- and β-like globin genes).

4. The monotreme β-globin genes have a higher G+C level than eutherian and marsupial β-like globin genes but similar levels to avian β-like globin genes.

5. Outgroup sequences used for rooting phylogenetic trees must be ancestral to all other taxa in the tree (which themselves are assumed to form a monophyletic group) while at the same time being as closely related as possible to the ingroup. With an ingroup comprising monotremes, marsupials and eutherians, representative of avians and/or reptilians must be used as outgroups. Given the absence of published DNA sequences for reptilian β-globin genes, chicken β-globin DNA sequences were used as outgroups. Conclusions drawn from phylogenetic analyses must take into account the very distant relationship of the chicken to the mammals. (Until their evolutionary relationships become more clearly resolved, monotremes can not be used with certainty as outgroups for marsupial/eutherian ingroups).

Two important and interrelated questions can be posed which related to the phylogenetic analysis of β-like globin genes from monotremes, marsupials and eutherians:

1. What are the evolutionary relationships of monotremes, marsupials and eutherians?
2. Did the gene duplication that gave rise to progenitors of the ε-like and β-like globins presently found in marsupials and eutherians, occur before or after the separation of the monotremes from other mammals (or the separation of the (monotreme, marsupial) stem from the eutherians)? [This duplication will be referred to as the β/ε duplication]

With respect to the first question, there are three possible phylogenies: ((Monotremes), (Marsupials, Eutherians)), ((Eutherians), (Monotremes, Marsupials)) and ((Marsupials), (Monotremes, Eutherians)).

Therefore, four models can be proposed that could account for the combined species and β-like globin gene phylogenies:

Model 1 encompasses the generally accepted arrangement of monotremes, marsupials and eutherians with monotremes forming an ancestral lineage to a "Therian" lineage consisting of eutherians and marsupials, and places the occurrence of the β/ε gene duplication after the monotreme split but before the marsupial and eutherian split.

Model 2 involves the same species tree as Model 1, but places the occurrence of the β/ε gene duplication before the monotreme split.

Model 3 encompasses the ((Eutherian), (Marsupial, Monotreme)) arrangement of taxa with the β/ε globin gene duplication occurring before the separation of stem eutherians from the stem (Marsupial, Monotreme).

Model 4 encompasses the ((Marsupial), (Eutherian, Monotreme)) arrangement of taxa with the β/ε globin gene duplication occurring before the separation of stem mammals (Marsupial, Monotreme, Eutherians).

The β-globin gene phylogenies (constrained within the appropriate species phylogenies) predicted by each model are given in Fig. 4.2.

The maximum parsimony trees, distance trees and maximum likelihood tree obtained, as exemplified by the trees shown in Fig. 3.3.1-3.3.4, most closely resemble the gene tree predicted by Model 1. The monotreme β-globin genes formed an ancestral monophyletic lineage to exclusion of a monophyletic group containing all other mammalian β-like globin genes. However, the low bootstrap values (69-73%) of this

ancestral node, the distantly related outgroup used, and the anomalous location of the marsupial β-globin clade, suggest that support for Model 1 must be treated with caution. One of the maximum-likelihood trees (Fig. 3.3.5) obtained using the HKY 85 and gamma distribution model, which accounted for differences in the transition and transversion rates and variable rates of evolution at different sites, gave a tree that supported Model 3 with the monotreme β-globin genes in monophyly with the marsupial adult β-globin genes. However, no bootstrap analysis representing the frequency of that particular monophyletic group was carried out due to the excessive computing time required for maximum likelihood analyses and the anomalous position of the marsupial ε-globin clade suggests that support for Model 3 in this case, should also be treated with caution.

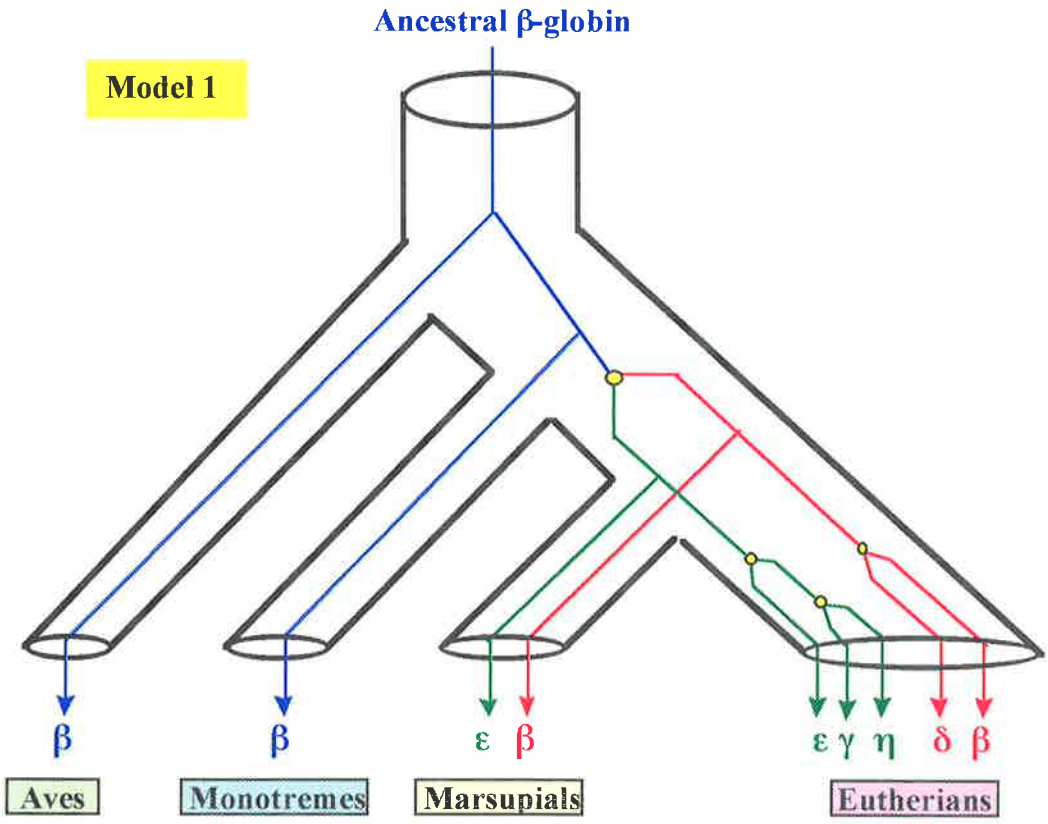
The monotreme β-globin genes were found to have a high GC content compared to eutherian and marsupial β-like globin genes, but have the same GC content as avian β-like globin genes. This observation also gives some support to the hypothesis that monotreme β-globin genes form a lineage that is ancestral to the marsupial and eutherian β-like globin genes.

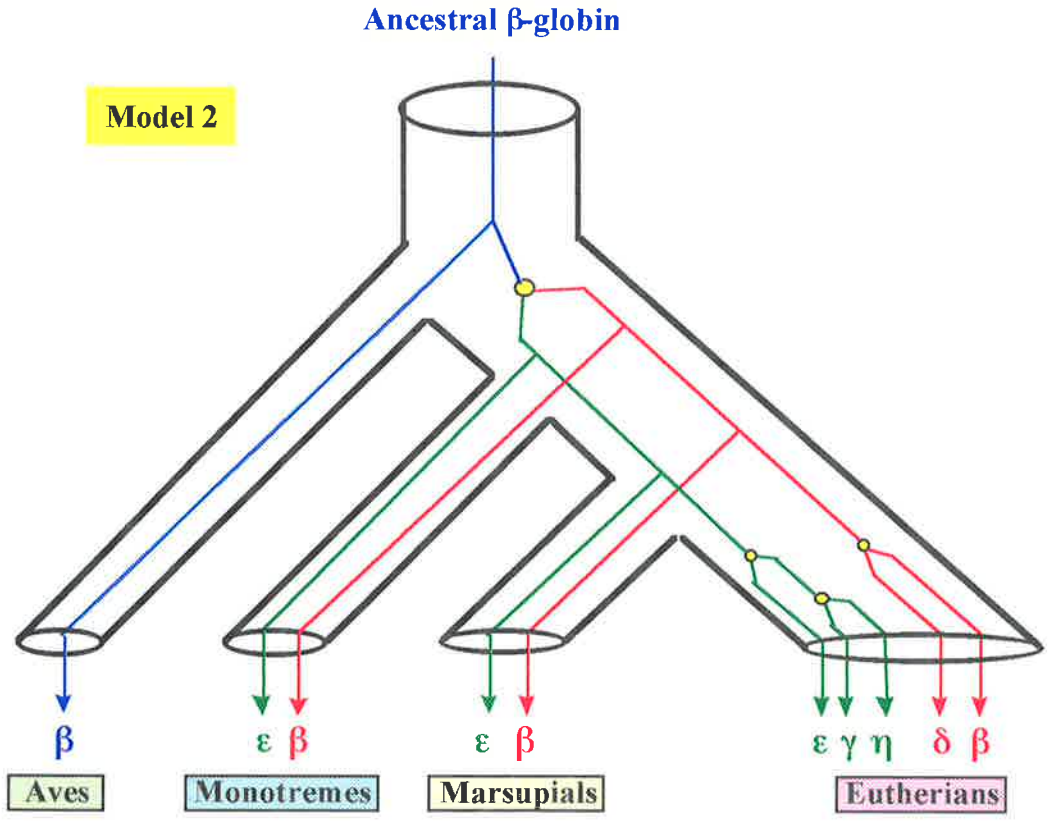
Taken overall, and bearing in mind the reservations raised, the results of this thesis provide support for Model 1, that proposes monotremes form an ancestral lineage outside a monophyletic group comprising marsupials and eutherians, and that the gene duplication that produced proto ε-globin and proto β-globin occurred after the divergence of the monotreme lineages from the common ancestor of marsupials and eutherians, but before the separation of the marsupial and eutherian lineages.

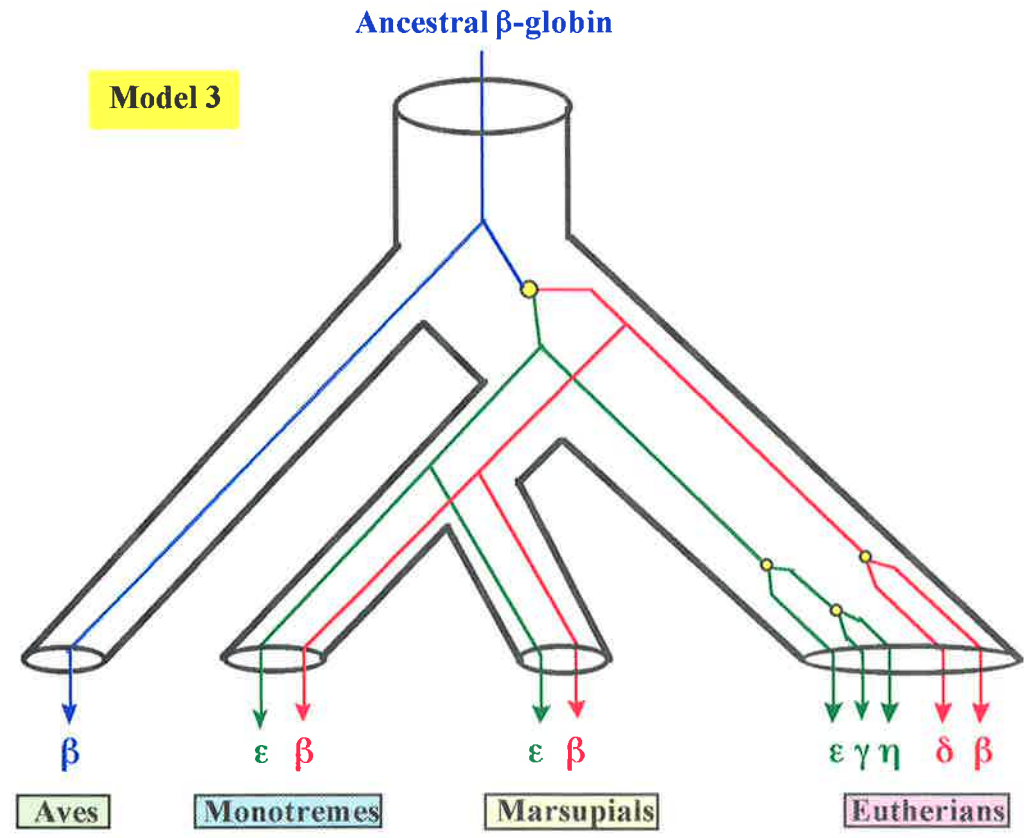
However, a key question that remains unanswered is: "Is there a β-like globin gene in monotremes that is directly descended from (i.e. is orthologous with) a putative ε-like globin gene progenitor that resulted from the β/ε duplication event?" The presence of such a gene would argue strongly against Model 1.

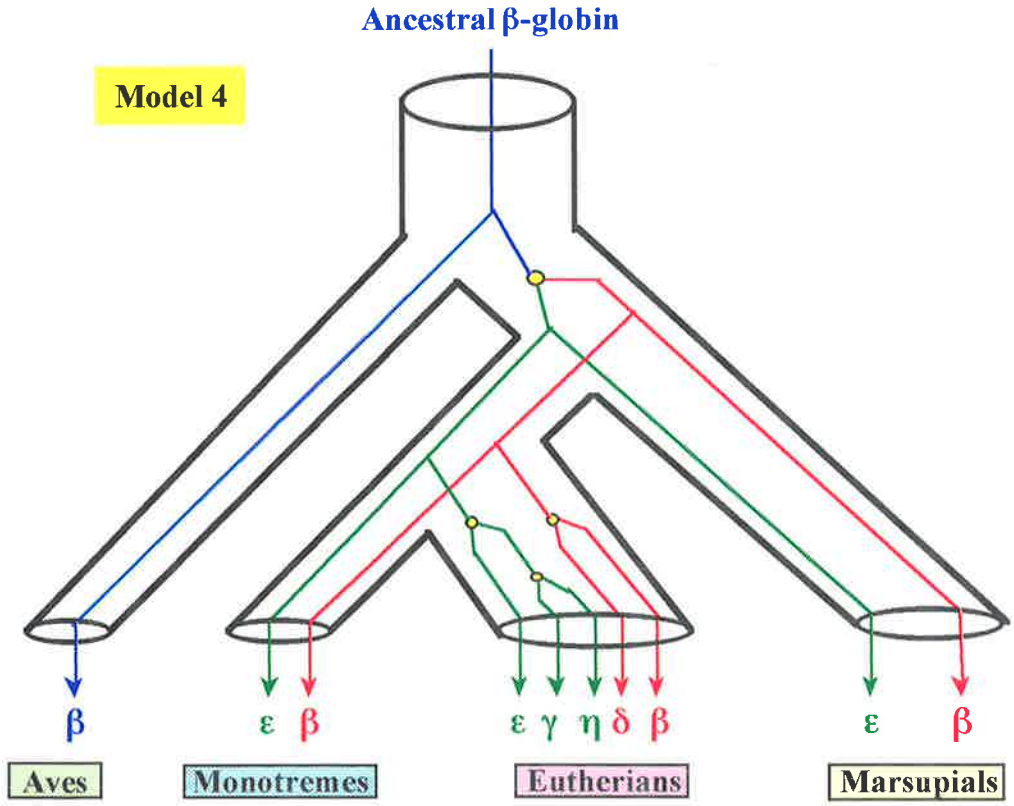
Although the evidence in favour of Model 1 is limited at this stage, at least the four models proposed in this thesis, provide a framework for interpreting further studies on β-globin gene evolution, and marsupial and monotreme species evolution.

Figure 4.2 Phylogenetic trees illustrating four possible evolutionary models. Gene lineages (thin colored lines) have been fitted within species lineages (thick lines) using the approach proposed by Goodman *et al.*(1975). Blue lines depict the ancestral β-globin gene, red lines depict proto β-globin and its descendants, green lines depict proto ε-globin and its descendants. Small yellow circles represent gene duplications producing paralogous genes. The trees are not drawn to a time scale. Within the avian branch, the various β-like globin genes and the duplications that gave rise to them are not indicated. Within the eutherian lineage, paralogous genes ε-, γ-, and η-globin of the ε-lineage and the paralogous genes β- and δ-globin of the β-lineage are shown. Only two β-like globin genes (ε-globin and β-globin) are shown for marsupials (Koop and Goodman, 1988; Cooper and Hope, 1993; Cooper *et al.* 1996). The recently discovered ω-globin gene of the marsupial *Macropus eugenii* (Holland and Gooley, 1997; Wheeler, Cooper, Hope, Holland, Gooley and Goodman (unpublished) is not shown because its phylogenetic relationships are still under investigation.


Ancestral β-globin
Model 1
β
β
ε β
ε γ η δ β
Aves
Monotremes
Marsupials
Eutherians


Ancestral β-globin
Model 2
β
ε β
ε β
ε γ η δ β
Aves
Monotremes
Marsupials
Eutherians


Ancestral β-globin
Model 3
β
ε β
ε β
ε γ η δ β
Aves
Monotremes
Marsupials
Eutherians


Ancestral β-globin
Model 4
β
ε β
ε γ η δ β
ε β
Aves
Monotremes
Eutherians
Marsupials

Evolutionary relationships of monotremes, marsupials and eutherians

The fragmentary fossil record and complete loss of adult teeth in extant monotreme species have been major problems to understanding the precise taxonomic relationships of monotremes to the other two extant mammalian groups, marsupials and eutherians (Pascual *et al.* 1992a,b; Archer *et al.* 1992). DNA-DNA hybridisation, and protamine p1 nucleotide sequence data have suggested a monotreme - (eutherian/marsupial) dichotomy (Westerman and Edwards, 1992; Retief *et al.* 1993). An analysis of mitochondrial 12S sequence data from platypus and echidna also support this dichotomy (Springer and Kirsch, 1993; Gemmell and Westerman, 1994). Recently, however, based on phylogenetic analyses of the protein-coding genes of the entire mitochondrial genome of platypus and representative eutherians and marsupials, Janke *et al.* (1996, 1997) suggested that monotremes and marsupials are sister groups within a clade "Marsupionta".

The phylogenetic analyses presented in this study support a monotreme-therian (eutherian and marsupial) dichotomy. However, it must be re-emphasised that further data are required to resolve this question. It would be useful, in this context, to carry out a phylogenetic analysis based on all the relevant molecular sequence data currently available. There is debate amongst molecular evolutionists about the appropriateness (and method) of combining sequence data from different regions of the genome, and carrying out phylogenetic analyses (e.g. using the maximum parsimony approach) on the combined data set (Yang, 1996). With the inevitable accumulation of DNA sequences from representative species of the three mammalian groups, and with progress towards resolving the use of combined data, it should become possible to re-examine, in a more "powerful" way, the relationships of marsupials, monotremes and eutherians.

CONCLUDING REMARKS

Conclusions from this project can be summarised as follows:

1. Four β-globin genes were isolated and sequenced from 3 monotremes species. The overall structure of these monotreme β-globin genes suggests that they are functionally expressed. The encoded amino acid sequences of two monotreme β-globin genes (*echid* β1 and *platy* β1) are identical to these determined by Whittaker *et al.* (1972) and Whittaker and Thompson (1975), respectively from haemoglobin isolated from adult animals, suggesting that these genes are "adult-like" β-globin genes (i.e. expressed in post-natal stages). Comparisons of the 5' promoter regions show that in their sequence organisation, the monotreme β-globin genes most closely resemble marsupial and eutherian adult β-globin genes, rather than the embryonic β-like globin genes.

2. On average the monotreme β-globin genes have the longest intron I and shortest intron II of any described functional adult β-globin gene from birds and mammals. The coding sequences of monotreme β-globin genes have a higher G+C content than marsupial and eutherian β-like globin genes, but are similar in G+C content to chicken β-like globin genes.

3. The presence of the *platy* β2 gene supports the existence of more than one β-like globin gene in monotremes. However, this gene was very similar in sequence to *echid* β1, both in its coding and non-coding regions, and further research is required to confirm the species of origin of *platy* β2-globin. The possibility that the mis-identication of this gene resulted from DNA contamination in the genomic library from which it was derived, can not be completely ruled out, although evidence is presented against such a possibility.

4. The genomic Southern analysis of platypus and echidna genomic DNA suggests that monotreme species may have more than one β-like globin gene.

5. Using *in situ* hybridisation, the β-globin gene locus has been localised to one of the small chromosomes (group 14-22) of the Australian echidna. This result differs from that obtained in the platypus, where the β-globin locus has been localised to $4q^{H}$ (Watson *et al.* 1991).

6. A series of tests on the relative rates of evolution of the coding regions of monotreme, marsupial and eutherian β-like globin genes provide some evidence that rates are not constant, and that the rate of substitution in ε-globin genes is slower than that of β-globin genes. The monotreme β-globin genes appear to have evolved the slowest followed by eutherian β-globin genes and marsupial β-globin genes among the mammalian adult-like β-globin genes.

7. The low nucleotide divergence values at synonymous and non-synonymous sites, the high percentage of sequence similarity at coding and non-coding sites among the three monotreme β-globin genes (*echid* β1, *NGechid* β1 and *platy* β1) and the monophyletic grouping of these genes in phylogenetic analyses with 100% bootstrap support, suggests that these genes evolved from one common ancestral monotreme β-like globin gene and they are orthologously related to each other.

8. Based on comparisons of β-globin gene substitution rates, the estimated time of divergence for the echidna and platypus lineages is 41-57 MYA. This date accords with the generally accepted divergence time (30-60 MYA) of the two monotreme families. However, the divergence time of Australian echidna and New Guinea echidna is estimated at 11-20 MYA which is far older than estimates obtained from DNA-DNA hybridisation and mitochondrial 12S rDNA sequence comparisons (1-2 MYA).

9. The phylogenetic analysis of β-like globin genes from birds, monotremes, marsupials and eutherians provides support to the following model of evolution. The monotreme lineage separated from a lineage that subsequently separated to give the marsupial and eutherian lineages, and a gene duplication that gave rise to the progenitors of the ε-like and β-like globins presently found in marsupials and eutherians, occurred after the separation of the monotremes from the (marsupial/ eutherian) branch, but before the marsupial/eutherian separation.

BIBILIOGRAPHY

Air, G. M., and E. O. P. Thompson. 1969. Studies on marsupial proteins II. amino acid sequences of the β-globin of haemoglobin from the grey kangaroo, *Macropus giganteus*. *Aust. J. Biol. Sci. 22*: 1437-1454.

Air, G. M., E. O. P. Thompson, B. Richardson, and G. Sharman. 1971. Amino acid sequences of kangaroo myoglobin and haemoglobin and the date of marsupial-eutherian divergence. *Nature 229*: 391-394.

Alan, F. S., P. Heath, S. Trusko, S. H. Boyer, W. Prass, M. Goodman, J. Czelusniak, L.-Y. E. Chang, and J. J. Slightom. 1984. The sequence of the gorilla foetal globin genes: evidence for multiple gene conversion in human evolution. *Mol. Biol. Evol. 1*: 371-389.

Antoine, M., and J. Niessing. 1984. Intron-less globin gene in the insect *Chironomus thummi thummi*. *Nature 301*: 795-798.

Aota, S. I., and T. Ikemura. 1986. Diversity in G+C content at the third position of codons in vertebrate genes and its cause. *Nucleic Acids Res. 14*: 6345-6355.

Archer, M., T. F. Flannery, A. Ritchie, and R. E. Molnar. 1985. First Mesozoic mammal from Australia-an early Cretaceous monotremes. *Nature 318*: 363-366.

Archer, M., F. A. JR.Jenkins, S. J. Hand, P. Murray, and H. Godthelp. 1992. Description of the skull and non-vestigial dentition of a Miocene platypus (*Obdurodon dicksoni*. Sp.) from Riversleigh, Australia, and the problem of monotreme origins. Pages 15-26 in M. Augee, ed. *Platypus and Echidnas*, 1st ed. The Royal Zoological Society of New South Wales, New South Wales, Australia.

Archer, M., P. Murray, S. Hand, and H. Godthelp. 1993. Reconsideration of monotremes relationships based on the skull and dentition of the Miocene *Obdurodon dicksoni*. Pages 75-94 in F. S. Szalay, M. J. Novacek, and M. C. Mckenna, eds. *Mammal phylogeny*. New York, Springer-Verlag.

Augee, M. 1984. Quills and Bills : Pages 567-570 in M. Archer, and G. Clayton, eds. *Vertebrate zoogeography and evolution in Australia*. Perth, Hesperian Press.

Augee, M., and B. A. Gooden. 1993. *Echidnas of Australia and New Guinea*. New South Wales Uni., NSW Australia.

Augee, M. L., R. W. Elsner, B. A. Gooden, and P. R. Wilson. 1971. Respiration and cardiac responses of a burrowing animal, the echidna. *Resp. Physiol. 11*: 327-334.

Bailey, W. J., J. L. Slightom, and M. Goodman. 1992. Rejection of the "flying primate" hypothesis by phylogenetic evidence from the ε-globin gene. *Science 256*: 86-89.

Barrie, P. A., A. J. Jeffreys, and A. F. Scott. 1981. Evolution of the β-globin gene cluster in man and the primates. *J. Mol. Biol. 149*: 319-336.

Bashford, D., C. Chothia, and A. M. Lesk. 1987. Determinants of a protein fold unique features of the globin amino acid sequences. *J. Mol. Biol. 196*: 199-216.

Bech, C., S. Nicol, and N. A. Andersen. 1992. Ventilation in the echidna, *Tachyglossus aculeatus*. Pages 134-139 in Augee, ed. *Platypus and echidnas*. New South Wales, Royal Zoological Society.

Behe, M. J. 1990. Histone deletion mutants challenge the molecular clock hypothesis. *TIBS 15*: 374-376.

Bentley, P. J., C. F. Herreid II, and S.-N. Knut. 1967. Respiration of a monotreme, the echidna, *Tachyglossus aculeatus*. *Am. J. Physiol. 212*: 957-961.

Benton, M. J. 1990. Phylogeny of the major tetrapod groups: Morphological data and divergence dates. *J. Mol. Evol. 30*: 409-424.

Bernardi, G. 1989. The isochore organisation of the human genome. *Ann. Rev. Genet. 23*: 637-661.

Bernardi, G. B., B. Olofsson, J. Filipski, M. Zerial, J. Salinas, G. Cuny, M. Meunier-Rotival, and F. Rodier. 1985. The mosaic genome of warm-blooded vertebrates. *Science 229*: 953-958.

Bick, Y. A. E. 1992. The meiotic chain of chromosomes of monotremata. Pages 64-68 in M. L. Augee, ed. *Platypus and echidnas*, 1st ed. NSW, Sydney, The Royal Zoological Society.

Bick, Y. A. E., and W. D. Jackson. 1967a. DNA content of monotremes. *Nature 215*: 192-193.

Bick, Y. A. E., and W. D. Jackson. 1967b. A mammalian X-O sex-chromosome system in the monotreme *Tachyglossus aculeatus* determined from leucocyte cultures and testicular preparation. *The American Naturalist 101*: 79-86.

Bick, Y. A. E., and G. B. Sharman. 1975. The chromosomes of the platypus (*Ornithorhynchus*: Monotremata). *Cytobios 14*: 17-28.

Bick, Y. A. E., C. Murtagh, and G. B. Sharman. 1973. The chromosomes of an egg-laying mammal *Tachyglossus aculeatus* (the echidna). *Cytobios 7*: 233-243.

Biln, N., and D. W. Stafford. 1976. A general methods for the isolation of high molecular weigh DNA from eukaryotes. *Nucleic Acids Res. 3*: 2303-2308.

Bishop, M. F., and A. E. Friday. 1987. Tetrapod relationships: the molecular evidence. Pages 123-139 in C. Patterson, ed. *Molecules and morphology in evolution: conflict or compromise?* NY, USA, Cambridge University.

Blachetot, A., V. V. Wilson, and A. J. Jeffreys. 1983. The seal myoglobin gene: an unusually long globin gene. *Nature 301*: 732-734.

Blake, C. C. F. 1978. Do genes-in-pieces imply proteins-in pieces? *Nature 273*: 267.

Blake, C. C. F. 1981. Exons and the structure, function and evolution of haemoglobin. *Nature 291*: 616.

Blaxter, M. L. 1993. Nemoglobins: divergent nematode globins. *Parasitol. Today 9*: 353-360.

Bogusz, D., C. A. Appleby, J. Landsmann, E. S. Dennis, M. J. Trinick, G. Brown, J. S. Lee, N. Brisson, and D. Verma. 1984. The evolution of a plant globin gene family. *J. Mol. Evol. 21*: 19-32.

Bogusz, D., C. A. Appleby, J. Landsmann, E. S. Dennis, M. J. Trinick, and W. J. Peacock. 1988. Functioning haemoglobin genes in non-nodulating plants. *Nature 331*: 178-180.

Bresnick, E. H., and G. Felsenfeld. 1994. Dual promoter activation by the human β-globin locus control region. *Proc. Natl. Acad. Sci. USA 91*: 1314-1317.

Brisson, N., and D. P. S. Verma. 1982. Soybean leghemoglobin gene family: normal, pseudo, and truncated genes. *Proc. Natl. Acad. Sci. USA 79*: 4055-4059.

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

Brookfield, J. F. Y., and P. M. Sharp. 1994. Neutralism and selectionism face up to DNA data. *TIG 10*: 109-111.

Brown, G., J. S. Lee, N. Brisson, and D. Verma. 1984. The evolution of a plant globin gene family. *J. Mol. Evol. 21*: 19-32.

Busslinger, M., S. Rusconi, and M. L. Birnstiel. 1982. An unusual evolutionary behaviour of a sea urchin histone gene cluster. *EMBO 1*: 27-33.

Butler, P. J., and D. R. Jones. 1982. The comparative physiology of diving in vertebrates. *Adv. Comp. Physiol. Biochem. 8*: 179-364.

Castillo, O., L. T. Jones, and H. Lehmann. 1978. The myoglobin of an echidna (*TACGYGLOSSUS ACULEATUS ACULEATUS*). *Biochimica et Biophysica. Acta 533*: 289-292.

Cedergren, R., M. W. Gray, Y. Abel, and D. Sankoff. 1988. The evolutionary relationships among known life forms. *J. Mol. Evol. 28*: 98-112.

Chang, J. C., R. Poon, K. H. Neumann, and Y. W. Kan. 1978. The nucleotide sequences of the 5' untranslated region of human γ-globin mRNA. *Nucleic. Acids Res. 5*: 1147-1162.

Chang, L.-Y. D., and J. L. Slightom. 1984. Isolation and nucleotide sequences analysis of the β-type globin pseudogene from human, gorilla and chimpanzee. *J. Mol. Biol. 180*: 767-784.

Choi, O.-R. B., and J. D. Engel. 1988. Developmental regulation of β-globin gene switching. *Cell 55*: 17-26.

Church, G. M., and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA 81*: 1991-1995.

Clarke, B. C. 1970. Darwinian evolution of proteins. *Science 168*: 1009-1011.

Clearly, M. L., E. A. Schon, and J. B. Lingrel. 1981. Two related pseudogenes are the result of a gene duplication in the goat β-globin locus. *Cell 26*: 181-190.

Clemens, W. 1989. Diagnosis of the class mammalia. Pages 401-406 *In: Fauna of Australia*, vol. Vol IB. Mammalia.

Cifelli, R. L. 1993. Early Cretaceous mammal from North America and the evolution of marsupial dental characters. *Proc. Natl. Acad. Sci. USA 90*: 9413-9416.

Collins, F. S., and S. M. Weissman. 1984. The molecular genetics of human haemoglobin. *Nucleic Acids Res 31*: 315-421.

Collis, P., M. Antoniou, and F. Grosveld. 1990. Definition of the minimal requirements within the human β-globin gene and the dominant control region for high level expression. *EMBO 9*: 233-240.

Cooper, D. W., J. L. Vandeberg, M. E. Griffiths, and E. H. M. Ealey. 1973. Haemoglobin polymorphisms in the echidna, *Tachyglossus aculeatus*. *Aust. J. Biol. Sci. 26*: 605-612.

Cooper, S. J., and R. M. Hope. 1993. Evolution and expression of a β-like globin gene of the Australian marsupial *Sminthopsis crassicaudata*. *Proc. Natl. Acad. Sci. USA 90*: 11777-11781.

Cooper, S. J. B., R. Murphy, G. Dolman, D. Hussey, and R. M. Hope. 1996. A molecular and evolutionary study of the β-globin gene family of the Australian marsupial *Sminthopsis crassicaudata*. *J. Mol. Biol. 17*: 1012-1022.

Craik, C. S., S. R. Buchman, and S. Beychok. 1981. Oxygen binding properties a product of the central exon of the β-globin gene. *Nature 291*: 87-90.

Crossley, M., and S. H. Orkin. 1993. Regulation of the β-globin locus. *Current Biol. 3*: 232-237.

Czelusniak, J., M. Goodman, D. Hewett-Emmett, M. L. Weiss, P. J. Venta, and R. E. Tashian. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature 298*: 297-300.

Czelusniak, J., M. Goodman, B. F. Koop, D. E. Tagle, J. Shoshani, G. Braunitzer, B. K. Kleinschmidt, W. W. DeJong, and G. Matsuda. 1990. Perspective from amino acid and nucleotide sequences on cladistic relationships among higher taxa of eutheria. Pages 545-572 in H. H. Genoways, ed. *Current Mammalogy*, 2nd ed. Plenum Press, New York.

Dayhoff, M. D. 1976. *Atlas of protein sequence and structure*, vol. 5. Washington, D.C, National Biomedical Research Foundation.

DeBoer, E., M. Antoniou, V. Mignotte, L. Wall, and F. Grosveld. 1988. The human β-globin promoter; nuclear protein factors and erythroid specific induction of transcription. *EMBO 7*: 4203-4212.

Disegni, G., G. Carrara, G. R. Tocchini-Valentini, C. C. Shoulders, and F. E. Baralle. 1981. Selective in vitro transcription of one the two *Alu* family repeats present in the 5' flanking region of the human ε-globin gene. *Nucleic Acids Res. 9*: 6709-6722.

Dixon, B., and B. Pohajdak. 1992. Did the ancestral globin gene of plants and animals contain only two introns ? *TIBS 17*: 486-488.

Dixon, B., B. Walker, W. Kimmins, and B. Pohajdak. 1991. Isolation and sequencing of a cDNA for an unusual haemoglobin from the parasitic nematode *Psudoterranova desipiens*. *Proc. Natl. Acad. Sci. USA 88*: 5655-5659.

Dixon, B., B. Walker, W. Kimmins, and B. Pohajdak. 1992. A nematode haemoglobin gene contains an intron previously though to be unique to plants. *J. Mol. Evol. 35*: 131-136.

Dodgson, S. J., W. K. Fisher, and E. O. P. Thompson. 1974. Studies on montremes proteins. IV Amino acid sequences of haemoglobin-IA of the echidna; a comparison of major haemoglobins from two geographical groups of echidnas. *Aust. J. Biol. Sci. 27*: 111-115.

Dolan, M., B. J. Sugarman, and J. D. Engel. 1981. Chromosomal arrangement of the chicken β-globin genes. *Cell 24*: 669-677.

Doolitle, W. F. 1987. What introns have to tell us: Hierarchy in genome evolution. *Cold Spring Harbore Symposia on Quantitative Biology LII*: 907-913.

Doyle, J. J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot 17*: 144-163.

Driscoll, M. C., C. S. Dobkin, and B. P. Alter. 1989. γκβ-thalassemia due to a *de novo* mutation deleting the 5' β-globin gene activation-region hypersensitive sites. *Proc. Natl. Acad. Sci. USA 86*: 7470-7474.

Easteal, S. 1985. Generation time and the rate of molecular evolution. *Mol. Biol. Evol. 2*: 450-453.

Easteal, S. 1989. The pattern of mammalian evolution and the relative rate of molecular evolution. *Genetics 124*: 165-173.

Easteal, S. 1990. The pattern of mammalian evolution and the relative rate of molecular evolution. *Genetics 124*: 165-173.

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

Easteal, S., and C. Collet. 1994. Consistent variation in amino-acid substitution rate, despite uniformity of mutation rate: Protein evolution in mammals is not neutral. *J. Mol. Biol. 11*: 643-647.

Easteal, S., C. Collet and D. Betty. 1995. *The mammalian molecular clock,* vol. 1. R.G Landes Company, Austin, Texas, U.S.A.

Easteal, S., and G. Herbert. 1997. Molecular evidence from the nuclear genome for the time frame of human evolution. *J. Mol. Evol. 44*: s121-s132.

Eaton, W. A. 1980. The relationship between coding sequences and function in haemoglobin. *Nature 284*: 183-185.

Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Sligtom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proundfoot. 1980. The structure and evolution of the human globin gene family. *Cell 21*: 653-668.

Ellis, J., K. C. Tan-Un, D. Michalovich, N. Yannoutsos, S. Philipsen, and F. Grosveld. 1996. A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human β-globin locus control region. *EMBO 15*: 562-568.

Engel, J. D. 1993. Developmental regulation of human β-globin gene transcription: a switch of loyalties. *TIG 9*: 304-309.

Enver, T., N. Raich, A. J. Ebens, T. Papayannopoulou, F. Costantini, and G. Stamatoyannopoulos. 1990. Developmental regulation of human foetal-to-adult globin gene switching in transgenic mice. *Nature 344*: 309-313.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol. 17*: 368-376.

Felsenstein, J. 1984. Distance methods for inferring phylogenies: a justification. *Evolution 38*: 16-24.

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

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Ann. Rev. Genet. 22*: 521-565.

Felsenstein, J. 1993. PHYLIP: Phylogeny Inference Package, Version 3.5. Computer program distributed by University of Washington, Seattle, USA

Fisher, W. K., and E. O. P. Thompson. 1976. Studies on monotremes proteins. VII. amino acid sequences of myoglobin from the platypus, *Ornithorhynchus anatinus*. *Aus. J. Biol. Sci. 29*: 57-72.

Flannery, T. 1990. *Mammals of New Guinea*, 1st ed. Robert Brown & Associates Pty.Ltd, Carina, Australia.

Forsdyke, D. R. 1995. A stem-loop "kissing" model for the initiation of recombination and the origin of introns. *Mol. Biol. Evol. 12*: 949-958.

Fritsch, E., R. M. Lawn, and T. Maniatis. 1980. Molecular cloning and characterisation of the human β-like globin gene cluster. *Cell 19*: 959-972.

Fuhrman, S. A., P. L. Deininger, P. LaPorte, T. Friedmann, and E. P. Geiduschek. 1981. Analysis of transcription of the human *Alu* family ubiquitous repeating element by eukaryotic RNA polymerase III. *Nucleic Acids Res. 9*: 6439-6456.

Garner, K. J., and J. B. Lingrel. 1988. Structural organisation of the β-globin locus of β-haplotype sheep. *Mol. Biol. Evol. 5*: 134-140.

Gaut, B. S., S. V. Muse, W. D. Clark, and M. T. Clegg. 1992. Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. *J. Mol. Evol. 35*: 292-303.

Gemmell, N. J., and M. Westerman. 1994. Phylogenetic relationships within the class Mammalia: a study using mitochondrial 12s RNA sequences. *J. Mammal. Evol. 2*: 3-23.

Gemmell, N. J., A. Janke, P. S. Western, J. M. Watson, S. Paabo, and J. A. M. Graves. 1994. Cloning and characterisation of platypus mitochondrial genome. *J. Mol. Evol. 39*: 200-205.

Gerstein, M., E. L. L. Sonnhammer, and C. Chothia. 1994. Volume changes in protein evolution. *J. Mol. Biol. 236*: 1067-1078.

Ghazal, P., A. J. Clark, and J. O. Bishop. 1985. Evolutionary amplification of a pseudogene. *Proc. Natl. Acad. Sci. USA 82*: 4182-4185.

Gilber, W. 1978. Why genes in pieces ? *Nature 271*: 501.

Gilbert, W. 1987. The exon theory of genes. *Cold Spring Harbor Symposia on Quantitative Biology LII*: 901-905.

Gö, M. 1981. Correlation of DNA exon region with protein structural units in haemoglobin. *Nature 291*: 9092.

Gö, M., and M. Nosaka. 1987. Protein architecture and the origin of introns. *Cold Spring Harbor Symposia on Quantitative Biology LII*: 915-924.

Goodman, M. 1989. Emerging alliance of phylogenetic systematics and molecular biology: a new age of exploration. Pages 43-61 in B. Fernholm, K. Bremer, and H. Jornvall, eds. *The hierarchy of life*. USA, Elsever Science.

Goodman, M.. 1996. Epilogue: a personal account of the origins of a new paradigm. *Molecular Phylogenetics and Evolution 5*: 269-285.

Goodman, M., and G. Moore. 1975. Darwinian evolution in the genealogy of haemoglobin. *Nature 253*: 603-608.

Goodman, M., B. F. Koop, J. Czelusniak, M. L. Weiss, and J. L. Slightom. 1984. The η-globin gene. Its long evolutionary history in the β-globin gene family of mammals. *J. Mol. Biol. 180*: 803-823.

Goodman, M., J. Czelusniak, and J. E. Beeber. 1985. Phylogeny of primates and other eutherian orders: a cladistic analysis using amino acid and nucleotide sequence data. *Cladistics 1*: 171-185.

Goodman, M., J. Czelusniak, B. F. Koop, D. A. Tagle, and J. L. Slightom. 1987a. Globins : a case study in molecular phylogeny. *Cold Spring Harbor Symposia on Quantitative Biology LII*: 875-890.

Goodman, M., M. M. Miyamoto, and J. Czelusniak. 1987b. Pattern and process in vertebrate phylogeny revealed by coevolution of molecules and morphologies. Pages 141-175 in C. Patterson, ed. *Molecular and morphology in evolution conflict or compromise ?*, 1st ed. New York, USA, Cambridge Uni.

Goodman, M., J. Pedwaydon, J. Czelusniak, T. Suzuki, T. Gotoh, L. Moens, F. Shishikura, D. Walz, and S. Vinogradov. 1988. An evolutionary tree for invertebrate globin sequences. *J. Mol. Evol. 27*: 236-249.

Grant, T. R. 1989a. *Ornithorhynchidae*. Pages 436-450 in *In: Fauna of Australia*, vol. Vol IB Mammalia.

Grant, T. 1989b. *The platypus: A unique mammal*. New South Wales University Press, Sydney.

Grantham, R., G. Gautier, M. Gouy, R. Mercier, and A. Pare. 1980. Codon catalogue usage and the genome hypothesis. *Nucleic. Acids Res. 8*: 49-62.

Graves, J. A. M., D. W. Cooper, L. M. McKenzie, and R. M. Hope. 1993. Genetic maps of marsupial and monotreme mammals. Pages 4.282-4.290 in J. Stephen, and O'Brien, eds. *Genetic maps: locus maps of complex genomes*, 6th ed. Plainview, New York, Cold Spring Harbour.

Graves, J. M. L., and J. M. Watson. 1991. Mammalian sex chromosomes: evolution of organisation and function. *Chromosoma 101*: 63-68.

Gregory, W. K. 1974. Branching evolution in egg-laying and marsupial mammals. in *Evolution emerging*.

Griffiths, M. G. 1968. Affinities of the prototheria. Pages 216-229 in M. Griffiths, ed. *Echidnas*.

Griffiths, M. 1978. *The biology of monotremes*, vol. 1. Academic Press, N.Y.

Griffiths, M. 1989. *Tachyglossidae*. Pages 407-435 in D. W. Walton, and B. J. Richardson, eds. *In: Fauna of Australia*, vol. Vol IB. Australian Government Publishing Service, Australia.

Griffiths, M., R. T. Wells, and D. J. Barrie. 1991. Observation on the skulls of fossil and extant echidnas (Monotremata: *Tachyglossidae*). *Aust. Mammal. 14*: 87-101.

Grigg, G. C., M. L. Augee, and L. A. Beard. 1992. Thermal relations of free-living echidnas during activity and in hibernation in a cold climate. Pages 160-173 in Augee, ed. *Platypus and echidnas*. New South Wales, Royal Zoological Society.

Grosjean, H., and W. Fiers. 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene 18*: 199-209.

Gumucio, D. L., D. A. Shelton, W. Zhu, D. Millinoff, T. Gray, J. H. Bock, J. L. Slightom, and M. Goodman. 1996. Evolutionary strategies for the elucidation of *cis* and *trans* factors that regulate the developmental switching programs of the β-like globin genes. *Mol. Phyl. and Evol. 5*: 18-32.

Haigwood, N. L., C. L. Jahn, C. A. Hutchison, and M. H. Edgell. 1981. Locations of three repetitive sequence families found in BALB/c adult β-globin clones. *Nucleic Acids Res. 9*: 1133-1150.

Hardies, S., M. H. Edgell, and C. A. Hutchison III. 1984. Evolution of the mammalian β-globin gene cluster. *J. Biol. Chem. 259*: 3748-3756.

Hardison, R. C. 1984. Comparison of the β-like globin gene families of rabbits and human indicates that the gene cluster 5'-ε-γ-δ-β-3' predates the mammalian radiation. *Mol. Biol. Evol. 1*: 390-410.

Hardison, R. C. 1991. Evolution of globin gene families. Pages 272-289 in R. K. Selander, A. G. Clark, and T. S. Whittan, eds. *Evolution at molecular level*, 1st ed. Sinauer Associates Incorporated, Massachusetts.

Hardison, R. C., and R. E. Gelinas. 1986. Assignment of orthologous relationships among mammalian α-globin genes by examining flanking regions reveals a rate of evolution. *J. Mol. Biol. 3*: 243-261.

Hardison, R. C., and J. B. Margot. 1984. Rabbit globin pseudogene ψβ2 is a hybrid of δ- and β-globin gene sequences. *Mol. Biol. Evol. 1*: 302-316.

Hardison, R., J. Xu, J. Jackson, J. Mansberger, O. Selifonova, B. Grotch, J. Biesecker, H. Petryowska, and W. Miller. 1993. Comparative analysis of the locus control region of the rabbit β-like gene cluster: HS3 increases transient expression of an embryonic ε-globin gene. *Nucleic Acids Res. 21*: 1265-1272.

Harris, S., P. A. Barrie, M. L. Weiss, and A. J. Jeffreys. 1984. The primate ψβ1 gene: an ancient β-globin pseudogene. *J. Mol. Biol. 180*: 785-801.

Harris, S., J. R. Thackeray, A. J. Jeffreys, and M. L. Weiss. 1986. Lemur β-globin genes. *Mol. Biol. Evol. 3*: 465-484.

Hartzog, G. A., and R. M. Myers. 1993. Discrimination among potential activators of the β-globin CACCC elements by correlation of binding and transcription properties. *Mol. and Cell. Biol. 13*: 44-56.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol. 22*: 160-174.

Hawkins, J. D. 1988. A survey on intron and exon lengths. *Nucleic. Acids Res. 16*: 9893-9908.

Helm-Bychowski, K. M., and A. C. Wilson. 1986. Rates of nuclear DNA evolution in pheasant-like birds: evidence from restriction maps. *Proc. Natl. Acad. Sci. USA 83*: 688-692.

Higgins, D. G., A. J. Bleasby, and A. Fuchs. 1988. CLUSTAL V: improve software for multiple sequence alignment. *CABIOS 8*: 189-191.

Hill, A., S. Hardies, M. Phillips, M. Davis, C. Hutchison III, and M. Edgell. 1984. Two early embryonic β-globin sequences. *J. Biol. Chem. 259*: 3739-3747.

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol. 42*: 182-192.

Hillis, D. M., B. K. Mable, and C. Moritz. 1996. Application of molecular systematics: the state of the field and a look to the future. Pages 515-543 in D. M. Hillis, C. Moritz, and B. K. Mable, eds. *Molecular systematics*, 2nd ed. Sunderland, Massachusetts, USA, Sinauer Asso. Inc.

Hodgson, C. P., and R. Z. Fisk. 1987. Hybridisation probe size control: optimised oligolabelling. *Nucleic Acids Res. 15*: 6295.

Holland, R. A. B., and A. A. Gooley. 1997. Characterization of the embryonic globin chains of the marsupial Tammar wallaby, *Macropus eugenii. Eur. J. Biochem. In Process.*

Honig, G. R., and J. G. Adams III. 1986. *Human hemoglobin genetics*. New York, Springer-Verlag.

Hood, L., J. H. Campbell, and S. C. R. Elgin. 1975. The organisation, expression and evolution of antibody genes and other multigene families. *Ann. Rev. Genet. 9*: 305-353.

Hope, R. M., S. Cooper, and B. Wainwright. 1990. Globin macromolecular sequences in marsupials and monotremes. Pages 147-171 in J. A. M. Graves, R. M. Hope, and D. W. Cooper, eds. *Mammals from pouches and eggs: Genetics, breeding and evolution of marsupials and monotremes*, 1st ed. CSIRO, Australia.

Hopson, J. A. 1970. Classification of non-therian mammals. *J. Mammal. 51*: 1-9.

Hosbach, H. A., T. Wyler, and R. Weber. 1983. The *Xenopus laevis* globin gene family: Chromosomal arrangement and gene structure. *Cell 32*: 45-53.

Hosken, R. W. 1978. Oxygen binding of monotreme haemoglobin I. Oxygen binding parameters of the haemoglobins of platypus, *Ornithorhynchus anatinus*, and echidna, *Tachyglossus aculeatus*. *Aust. J. Biol. Sci. 31*: 327-336.

Hosken, R. W., and B. Boettcher. 1978. Oxygen Binding of monotreme haemoglobins. II. fixation of functional non-equivalent haemoglobins in the echidna population. *Aust. J. Biol. Sci. 31*: 337-344.

Hughes, R. L. 1993. Monotremes development with particular reference to the extra-embryonic membranes. *J. Exp. Zoo 266*: 480-494.

Hughes, S. H., E. Stubblefield, F. Payvar, J. D. Engel, J. B. Dodgson, D. P. Spector, B. Cordell, R. T. Schimke, and H. E. Varmus. 1979. Gene localisation by chromosome fractionation: globin genes are on at least two chromosomes and three estrogen-inducible genes are on three chromosomes. *Proc. Natl. Acad. Sci. USA 76*: 1348-1352.

Hyldig-Nielsen, J. J., E. O. Jensen, K. Paludan, O. Wiborg, R. Garrett, P. Jorgensen, and K. A. Marcker. 1982. The primary structure of two leghemoglobin genes from soybean. *Nucleic Acids Res. 10*: 689-701.

Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol. 2*: 13-34.

Ikemura, T., and H. Ozeki. 1983. Codon usage and transfer RNA content: organism-specific codon-choice patterns in reference to iso-accepter content. *Cold Spring Harbor Symposia on Quantitative Biology 42*: 1087-1097.

Ikemura, T., and S. I. Aota. 1988. Global variation in G+C content along vertebrate genome DNA: possible correlation with chromosome band structures. *J. Mol. Biol. 203*: 1-13.

Iwaasa, H., T. Takagi, and K. Shikama. 1991. Amino acid sequences of haemoglobin from yeast (*Candida norvegensis*). *Zool. Sci. 8*: 1134.

Jahn, C. L., C. Hutchison III, S. J. Philips, S. Weaver, N. L. Haigwood, C. F. Voliva, and M. H. Edgell. 1980. DNA sequences organisation of the β-globin complex in the BALB/c mouse. *Cell 21*: 159-168.

Janke, A., N. J. Gemmell, G. Feldmaier-Fuchs, A. V. Haeseler, and S. Paabo. 1996. The mitochondrial genome of a monotremes- the platypus (*Ornithorhynchus anatinus*). *J. Mol. Evol. 42*: 153-159.

Janke, A., X. Xu, and U. Arnason. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. *Proc. Natl. Acad. Sci. USA 94*: 1276-1281.

Jeffreys, A. J. 1979. DNA sequence variants in the γ^G-, γ^A-, δ- and β-globin genes of man. *Cell 18*: 1-10.

Jeffreys, A. J. 1982. Evolution of globin genes. Pages 157-176 in G. A. Dover, and R. B. Flavell, eds. *Genome evolution*, vol. Systematic association special 20. Academic Press.

Jeffreys, A. J., V. Wilson, D. Wood, and J. P. Simon. 1980. Linkage of adult α- and β-globin genes in *X. laevis* and gene duplication by tetraploidization. *Cell 21*: 555-564.

Jeffreys, A. J., P. Barrie, S. Harris, D. H. Fawcett, Z. J. Nugent, and A. Boyd. 1982. Isolation and sequences analysis of a hybrid δ-globin pseudogene from the brown lemur. *J. Mol. Biol. 156*: 487-503.

Jeffreys, A. J., S. Harris, P. Barrie, D. Wood, A. Blanchetot, and S. M. Adams. 1983. Evolution of gene families: the globin genes. Pages 175-195 in D. S. Bendall, ed. *Evolution from molecules to men*, 1st ed. Darwin College, Cambridge, NY, USA.

Jellie, A., W. P. Tate, and C. N. A. Trotman. 1996. Evolution history of intron in a multidomain globin gene. *J. Mol. Evol. 42*: 641-647.

Jenkins, F. A. 1990. Monotremes and the biology of Mesozoic mammals. *Neth. Zool. 4*: 5-31.

Jensen, E. O., K. Paludan, J. J. Hyldig-Nielsen, P. Jorgensen, and K. A. Marcker. 1981. The structure of a chromosomal leghaemoglobin gene from soybean. *Nature 291*: 677-679.

Jhiang, S., J. Garey, and A. Riggs. 1988. Exon-intron organisation in genes of the earthworm and vertebrate globins. *Science 240*: 334-336.

Jimenez, G., K. B. Gale, and T. Enver. 1992. The mouse β-globin locus control region: hypersensitive sites 3 and 4. *Nucleic Acids Res. 20*: 5797-5803.

Jukes, T. H. 1983. Evolution of the amino acid code. Pages 191-232 in M. Neil, and R. K. Koehn, eds. *Evolution of genes and proteins*, 1st ed. USA, ISBN.

Kao, W.-Y., P. M. Trewitt, and G. Bergtrom. 1994. Intron-containing globin genes in the insect *Chironomus thummi*. *J. Mol. Evol. 38*: 241-249.

Karlsson, S., and A. W. Nienhuis. 1985. Developmental regulation of human globin genes. *Ann. Rev. Biochem. 54*: 1071-1108.

Kemp, T. S. 1983. The relationships of mammals. *Zoo. J. Linn. Soci. 77*: 353-384.

Kemp, T. S. 1988. A note on the Mesozoic mammals, and the origin of therians. Pages 22-29 in M. J. Benton, ed. *The phylogeny and classification of the tetrapods*, vol. 2 mammals.

Kermack, D. M., and K. A. Kermack. 1984. *The evolution of Mammalian characters.* NSW.

Kielan-Jaworowska, Z., A. W. Crompton, and F. A. Jenkins. 1987. The origin of egg-laying mammals. *Nature 326*: 871-873.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature 217*: 624-626.

Kimura, M. 1979. The neutral theory of molecular evolution. *Scientific American 241*: 94-104.

Kimura, M. 1983. *The neutral theory of molecular evolution.* London, Cambridge Univ.

Kimura, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolution Genetic Analysis, Version 1.01. Computer program distributed by the Pennsylvania State University, PA, USA.

Klinger, H. P. 1987. *Cytogenetics and cell genetics; human gene mapping 9*, 9th ed. S. Karger Medical and Scientific publication, Switzerland.

Kloek, A. P., D. R. Sherman, and D. E. Goldberg. 1993. Novel gene structure and evolutionary context of *Caenorhabditis elegans* globin. *Gene 129*: 215-221.

Knochel, W., E. Korge, A. Basner, and W. Meyerhof. 1986. Globin evolution in the genus *Xenopus*: comparative analysis of cDNA coding for adult globin polypeptides of *X. borealis* and *X. tropicalis. J. Mol. Evol. 23*: 211-223.

Konkel, D. A., J. V. J. R. Maizel, and P. Leder. 1979. The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. *Cell 18*: 865-873.

Koop, B. F., and M. Goodman. 1988. Evolutionary and developmental aspects of two haemoglobin β-chain genes (εM and βM) of opossum. *Proc. Natl. Acad. Sci. USA 85*: 3893-3897.

Koop, B. F., M. Goodman, P. Xu, K. Chan, and J. L. Slightom. 1986. Primate η-globin DNA sequences and man's place among the great apes. *Nature 319*: 234-238.

Kotkow, K. J., and S. H. Orkin. 1995. Dependence of globin gene expression in mouse erythroleukemia cells on the NF-E2 heterodimer. *Mol. and Cell. Biol. 15*: 4640-4647.

Lacy, E., and T. Maniatis. 1980. The nucleotide sequence of a rabbit β-globin pseudogene. *Cell 21*: 545-553.

Lacy, E., R. C. Hardison, D. Quon, and T. Maniatis. 1979. The linkage arrangement of four rabbit β-globin genes. *Cell 18*: 1273-1283.

Leder, A., S. Swan, F. Ruddle, P. D'Eustachio, and P. Leder. 1981. Dispersion of α-like globin genes of the mouse to three different chromosomes. *Nature 293*: 196-198.

Lee, S. J., G. G. Brown, and D. P. Verma. 1983. Chromosomal arrangement of leghemoglobin genes in soybean. *Nucleic Acids Res. 11*: 5541-5553.

Leitch, A. R., T. Schwarzacher, D. Jackson, and I. J. Leitch. 1994. *In situ hybridisation.* Oxford, United Kingdom, BIOS Scientific Publishers.

Lewin, R. 1984. First mRNA splicing intermediate. *Science 225*: 153-156.

Li, Q., B. Zhou, P. Power, T. Enver, and G. Stamatoyannopoulos. 1991. Primary structure of the goat β-globin locus control region. *Genomics 9*: 488-499.

Li, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol. 36*: 96-99.

Li, W.-S., and D. Graur. 1991. *Fundamentals of molecular evolution.* Sinauer Associates, INC, Massachusetts, USA.

Li, W.-H., and M. Tanimura. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature 326*: 93-96.

Li, W., T. Gojobori, and M. Nei. 1981. Psudogenes as a paradigm of neutral evolution. *Nature 292*: 237-239.

Li, W.-H., C.-I. Wu, and C.-C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *J. Mol. Biol. 2*: 150-174.

Li, W.-H., M. Tanimura, and P. Sharp. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol. 25*: 330-342.

Lingrel, J. B., T. M. Townes, S. G. Shapiro, S. M. Wernke, P. A. Liberator, and A. G. Menon. 1985. Structural organisation of the α- and β-globin loci of the goat. Pages 67-79 in G. Stamatoyannopoulos, and A. W. Nienhuis, eds. *Experimental approaches for the study of haemoglobin switching*. New York, Liss.

Little, P. F. R. 1982. Globin pseudogenes. *Cell 28*: 683-684.

Liu, X., and J. E. Mertz. 1995. HnRNP L bind a *cis*-acting RNA sequence element that enable intron-independent gene expression. *Genes & Development 9*: 1766-1780.

Loomis, W. F., and M. E. Gilpin. 1986. Multigene families and vestigial sequence. *Proc. Natl. Acad. Sci. USA 83*: 2143-2147.

Maccarone, P., J. M. Watson, D. Francis, L. Selwood, I. Kola, and J. A. M. Graves. 1992. The evolution of human chromosome 21: evidence from in situ hybridisation in marsupial and a monotreme. *Genomics 13*: 1119-1124.

Magram, J., K. Chada, and F. Costantini. 1985. Developmental regulation of a cloned adult β-globin in transgenic mice. *Nature 315*: 338-340.

Malcom, S., P. Barton, and M. A. Ferguson-Smith. 1981. The chromosomal distribution of repetitive DNA sequences within the human β-globin gene cluster. *Hum. Genet. 57*: 388-393.

Maniatis, S. T., E. F. Fritsch, J. Lauer, and R. M. Lawn. 1980. The molecular genetics of human haemoglobins. *Ann. Rev. Genet. 14*: 145-178.

Marchant, C. L. 1994. Isolation of the embryonic-like β-globin sequences in the Australian Echidna, *Tachyglossus aculeatus*. Honour Thesis, University of Adelaide, Australia.

Martin, S. L., K. A. Vincent, and A. C. Wilson. 1983. Rise and fall of the δ globin gene. *J. Mol. Biol. 164*: 513-528.

Miyamoto, M. M., and J. Cracraft. 1991. *Phylogenetic analysis of DNA sequences*. New York, Oxford.

Miyata, T., and T. Yasunaga. 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J. Mol. Evol. 16*: 23-36.

Moon, A. M., and T. J. Ley. 1991. Functional properties of the β-globin locus control region in K562 erythroleukemia cells. *Blood 77*: 2272-2284.

Moritz, C., and D. M. Hillis. 1996. Molecular systematics: context and controversies. Pages 1-27 in D. M. Hillis, C. Moritz, and B. K. Mable, eds. *Molecular systematics*, 2nd ed. Sunderland, Massachusetts, USA, Sinauer Asso. Inc.

Mouchiroud, D., and G. Bernardi. 1993. Composition properties of coding sequences and mammalian phylogeny. *J. Mol. Evol. 37*: 109-116.

Murray, P. F. 1984. Furry egg-layers: The monotreme radiation. Pages 571-583 in M. Archer, and G. Clayton, eds. *Vertebrate zoogeography and evolution in Australia*. Perth, Hesperian Press.

Murtagh, C. E. 1977. A unique cytogenetic system in monotremes. *Chromosoma 65*: 37-57.

Nassif, N. T., W. K. Glenn, and A. G. Mackinlay. 1994. The organisation of the β-globin gene of the bivalve mollusc *Anadara trapezia* and its evolutionary relationship to other invertebrate and vertebrate globin genes. *J. Mol. Evol. 39*: 47-55.

Nei, M. 1987. *Molecular evolutionary genetics*. New York, Columbia University Press.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol. 3*: 418-426.

Nishioka, Y., A. Leder, and P. Leder. 1980. Unusual α-globin-like gene that has cleanly lost globin interving sequences. *Proc. Natl. Acad. Sci. USA 77*: 2806-2809.

O'Brien, S. J., and J. A. M. Graves. 1990. Report of the committee on comparative gene mapping. *Cytogenet. Cell Genet. 55*: 406-433.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol. 5*: 568-583.

Parer, J. T., and J. Metcalfe. 1967a. Respiratory studies of monotremes. I. blood of the platypus (*Ornithorynchus anatinus*). *Resp. Physiol. 3*: 136-142.

Parer, J., and J. Metcalfe. 1967b. Respiratory studies of monotremes. II. blood the echidna (*Tachyglossus setosus*). *Resp. Physiol. 3*: 143-150.

Parer, J., and J. Metcalfe. 1967c. Respiratory studies of monotremes. III. blood gas transport and homodynamic in the unanaesthetised echidna. *Resp. Physiol. 3*: 151-159.

Parsons, R. S., J. Atwood, E. R. Guiler, and R. W. L. Heddle. 1971a. Comparative studies on the blood of monotremes and marsupials-1. Haematology. *Comp. Biochem. Physiol. 39*: 203-208.

Parsons, R. S., J. Atwood, E. R. Guiler, and R. W. L. Heddle. 1971b. Comparative studies on the blood of monotremes and marsupials-2. Electrolyte organic constituents, proteins, gas analysis and enzymes. *Comp. Biochem. Physiol. 39*: 209-217.

Pascual, R., M. Archer, E. O. Jaureguizar, J. L. Prado, H. Godthelp, and S. J. Hand. 1992a. First discovery of monotremes in South America. *Nature 356*: 704-706.

Pascual, R., M. Archer, E. O. Jaureguizar, J. L. Prado, H. Godthelp, and S. J. Hand. 1992b. The first non-Australian monotreme an early Palaeocene south American platypus (Monotremata, *Ornithorhynchidae*). Pages 1-14 in M. L. Augee, ed. *Platypus and Echidnas*, 1st ed. The Royal Zoological Society of New South Wales, New South Wals.

Patterson, C. 1987. Molecules and morphology in evolution: conflict or compromise? Pages 1-22 in C. Patterson, ed. *Molecules and morphology in evolution: conflict or compromise?* NY, USA, Cambridge University.

Peacock, W. J. 1988. Functioning haemoglobin genes in non-nodulating plants. *Nature 331*: 178-180.

Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dogson. 1980. The evolution of genes: the chicken preprolinsulin gene. *Cell 20*: 555-566.

Perutz, M. F. 1978. Haemoglobin structure and respiratory transport. *Scientific American 239*: 92-125.

Perutz, M. F., H. Muirhead, J. M. Cox, and L. C. G. Goaman. 1968. Three dimensional fourier synthesis of horse oxyhaemoglobin at 2.8 angstrom units resolution: the atomic model. *Nature 219*: 131-139.

Potts, M., S. V. E. Angeloni, and D. Bassam. 1992. Myoglobin in cyanobacterium. *Science 256*: 1690-1692.

Prestridge, D. S., and Q. K. Chen. 1991. SIGSCAN Version 4.07: The databases and sequences formats. *CABIOS 12*: 157-160.

Proundfoot, N. J., A. Gil, and T. Maniatis. 1982. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell 31*: 553-563.

Pruzina, S., O. Hanscombe, D. Whyatt, F. Grosveld, and S. Philipsen. 1991. Hypersensitive site 4 of the human β globin locus control region. *Nucleic Acids Res. 19*: 1413-1418.

Reddy, P. M. S., G. Stamatoyannopoulos, T. Papayannopoulou, and C. K. J. Shen. 1994. Genomic footprinting and sequencing of human β-globin locus. *J. Biol. Chem. 269*: 8287-8295.

Reitman, M., J. A. Grasso, R. Blimenthal, and P. Lewit. 1993. Primary sequence, evolution, and repetitive elements of the *Gallus gallus* (chicken) β-globin cluster. *Genomics 18*: 616-626.

Renfree, M. B. 1993. *Ontogeny, Genetic Control, and Phylogeny of Female Reproduction in Monotremes and Therian Mammals.*

Retief, J. D., R. J. Winkfein, and G. H. Dixon. 1993. Evolution of the monotremes; The sequences of the protamine P1 genes of platypus and echidna. *Eur. J. Biochem. 218*: 457-461.

Rich, T. H. 1991. Monotremes, Placentals, and Marsupials: their record in Australia and its biases. Pages 894-924 in P. Vickers-Rich, J. M. Monaghan, R. F. Baird, and T. H. Rich, eds. *Vertebrate palaeontology of Australasia*, 1st ed. Monash University Melboren, Australia.

Richard, W. P., D. L. Daniel, R. G. S. Lee, H. E. Marshall, and A. H. Clyde. 1987. The molecular organisation of the β-globin complex of the deer mouse, *Peromyscus maniculatus. Mol. Biol. Evol. 4*: 30-45.

Richardson, B. J. 1987. A new view of the relationships of Australian and American marsupials. *Aust. Mammal. 11*: 71-73.

Richmond, R. C. 1970. Non-Darwinian evolution: a critique. *Nature 225*: 1025-1028.

Rismiller, P. D., and R. S. Seymour. 1991. The echidna. *Scientific American*: 80-87.

Roninson, I. B., and V. M. Ingram. 1982. Gene evolution in the chicken β-globin cluster. *Cell 28*: 515-521.

Rowe, T. 1993. Phylogenetic systematics and the early history of mammals. Pages 129-145 in F. S. Szalay, M. J. Novacek, and M. C. Mckenna, eds. *Mammal phylogeny*. New York, Springer-Verlag.

Ryan, T. M., R. R. Behringer, T. M. Townes, R. D. Palmiter, and R. L. Brinster. 1989. High-level erythroid expression of human α-globin genes in transgenic mice. *Proc. Natl. Acad. Sci. USA 86*: 37-41.

Saitou, N. 1988. Property and efficiency of the maximum likelihood method for molecular phylogeny. *J. Mol. Evol. 27*: 261-273.

Saitou, N., and M. Nei. 1986. The number of nucleotides required to determine the branching order of three species with special reference to the human-chimpanzee-gorilla divergence. *J. Mol. Evol. 24*: 189-204.

Saitou, N., and M. Neil. 1987. The neighbour-joining method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4*: 406-425.

Sambrook, J., E. F. Fritsch, and Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, USA.

Schimenti, J. C., and C. H. Duncan. 1985a. Concerted evolution of the cow ε2 and ε4 β-globin genes. *Mol. Biol. Evol. 2*: 505-513.

Schimenti, J. C., and C. H. Duncan. 1985b. Structure and organisation of the bovine β-globin genes. *Mol. Biol. Evol. 2*: 514-525.

Schimid, C. W., and W. R. Jelinek. 1982. The *alu* family of dispersed repetitive sequences. *Science 216*: 1065-1070.

Schon, E. A., M. L. Clearly, J. R. Haynes, and J. B. Lingrel. 1981. Structure and evolution of goat γ-, βc- and βa-globin genes: three developmentally regulated genes contain inserted elements. *Cell 27*: 359-369.

Shapiro, S. G., E. A. Schon, T. M. Townes, and J. B. Lingrel. 1983. Sequence and linkage of the goat εI and εII β-globin genes. *J. Mol. Biol. 169*: 31-52.

Sharman, G. B. 1978. Monotreme biology-Chairman's summary of the symposium. in M. Augee, ed. *Monotreme biology (Proceedings of a symposium held in Sydney)*, 1st ed. Royal Zoological Society of New South Wales, 1-7.

Shen, S. H., J. L. Slightom, and O. Smithies. 1981. A history of the human fetal globin gene duplication. *Cell 26*: 191-203.

Shroff, R. A. 1990. Cloning of a β-like globin gene in the Australian echidna, *Tachyglossus aculeatus*. Honours Thesis, University of Adelaide, Australia.

Shulman, R. G., J. J. Hopfield, and S. Ogawa. 1975. Allosteric interpretation of haemoglobin properties. *Q. Rev. Biophys. 8*: 325-420.

Sibley, C. G., J. E. Ahlquist, and B. L. J. Monroe. 1988. A classification of the living birds of the world based on DNA-DNA hybridisation studies. *AUK 105*: 409-423.

Slightom, J. L., A. Blechl, and O. Smithies. 1980. Human foetal Gγ- and Aγ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell 21*: 627-638.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol. 98*: 503.

Springer, M. S. 1990. The effect of random ranges truncation on patterns of evolution in the fossil record. *Paleobiology 16*: 512-520.

Springer, M. S. 1995. Molecular clock and the incompleteness of the fossil record. *J. Mol. Evol. 41*: 531-538.

Springer, M. S., and J. A. W. Kirsch. 1989. Rates of single-copy DNA evolution in phalangeriform marsupials. *Mol. Biol. Evol. 6*: 331-341.

Springer, M. S., and J. A. Kirsch. 1993. A molecular perspective on the phylogeny of placental mammals based on mitochondrial 12S rDNA sequences, with special reference to the problem of the paenungulata. *J. Mammal. Evol. 1 (2)*: 149-166.

Staden, R. 1980. A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res. 8*: 3673-3694.

Staden, R. 1984. Computer methods to aid the determination and analysis of DNA sequences. *Biochem. Soc. Trans. 12*: 1005-1008.

Stenzel, P., B. Brimhall, R. T. Jones, J. A. Black, A. McLachlan, and D. Gibson. 1979. Opossum haemoglobin: the amino acid sequence of the alpha and beta chains. *J. Biol. Chem. 254*: 2071-2076.

Strahan, R. 1991. *Complete book of Australian mammals*, The Australian Museum ed., Sydney.

Swofford, D. L. 1993. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1.1. Computer program distributed by the Illinios Natural History Survey, IL, USA

Swofford, D. L., and Olsen.G. J. 1990. Phylogeny reconstruction. Pages 411-501 in D. M. Hillis, and C. Moritz, eds. *Mol. syst.*. Sunderland, Sinauer Associates.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pages 407-513 in D. M. Hillis, C. Moritz, and B. K. Mable, eds. *Molecular systematics*, 2nd ed. Sunderland, Massachusetts, USA, Sinauer Asso. Inc.

Tagle, D. A., B. F. Koop, M. Goodman, J. L. Slightom, D. L. Hess, and R. T. Jones. 1988. Embryonic ε and γ globin genes of a prosimian primate (*Galago crassicaudatus*): nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J. Mol. Biol. 203*: 439-455.

Terwilliger, R. C. 1980. Structure of invertebrate haemoglobins. *Comp. Biochem. Physiol. 81*: 255-261.

Thompson, E. O. P., W. K. Fisher, and R. G. Whittaker. 1973. Studies of monotreme proteins. III. Amino acid sequence of the α- and β-globin chains of the minor haemoglobin from the echidna, *Tachyglossus Aculeatus Aculeatus. Aust. J. Biol. Sci. 26*: 1327-1335.

Thuring, R. W. J., J. P. M. Sanders, and P. Borst. 1975. A freeze-squeeze method for recovering long DNA from agarose gels. *Biochemistry 66*: 213-220.

Tipler, T. D., Y. H. Edwards, and D. A. Hopkinson. 1978. Developmental changes in the protein profiles of human cardiac and skeletal muscle. *Ann. Hum. Genet. 41*: 409-418.

Townes, T. M., and R. R. Behringer. 1990. Human globin locus activation region(LAR): role in temporal control. *TIG 6*: 219-223.

Townes, T. M., M. C. Fitzgerald, and J. B. Lingrel. 1984. Triplication of a four-gene set during evolution of the goat β-globin locus produced three genes now expressed differentially during development. *Proc. Natl. Acad. Sci. USA 81*: 6589-6593.

Trewitt, P. M., R. A. Luhm, F. Samad, S. Ramakrishnan, and W.-Y. Kao. 1995. Molecular evolutionary analysis of the YWVZ/7B globin gene cluster of the insect *Chironomus thummi. J. Mol. Evol. 41*: 313-328.

Tuan, D., W. Solomon, Q. Li, and I. M. London. 1985. The "β-like globin" gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. USA 82*: 6384-6388.

Tuan, D. Y. H., W. B. Solomon, I. M. London, and D. P. Lee. 1989. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human β-like globin genes. *Proc. Natl. Acad. Sci. USA 86*: 2554-2558.

Wainwright, B., and R. Hope. 1985. Cloning and chromosomal location of the α- and β-globin genes from a marsupial. *Proc. Natl. Acad. Sci. USA 82*: 8105-8108.

Wakabayashi, S., H. Matsubora, and D. Webster. 1986. Primary sequences of a dimeric haemoglobin from *Vitreoscilla. Nature 322*: 481-483.

Watson, J. M. 1990. Monotreme genetics and cytology and a model for sex chromosome evolution. *Aust. J. Zool. 37*: 385-406.

Watson, J. M., and J. A. M. Graves. 1992. *Clues about the evolution of mammalian sex chromosomes from monotremes gene mapping*, Augee, M. L. ed. Sydney, Australia, The Royal Zoological Society of New South Wales.

Watson, J. M., J. A. Spencer, A. D. Riggs, and J. A. M. Graves. 1990. The X chromosome of monotremes shares a highly conserved region with the eutherian and marsupial X chromosomes despite the absence of X chromosome inactivation. *Proc. Nat. Acad. Sci. USA 87*: 7125-7129.

Watson, J. M., G. Collett, M. Westerman, and J. A. Graves. 1991. Localisation of autosomal gene homologous in the monotreme mammal *Ornithorhynchus anatinus* (the platypus). *Cytogenet. Cell Genet. 58*: 2131.

Watson, J. M., J. Meyne, and J. A. M. Graves. 1992. Chromosomal composition and position in the echidna meiotic translocation chain. Pages 53-63 in M. L. Augee, ed. *Platypus and echidnas*, 1st ed. New South Wales, Sydney, The Royal Zoological Society.

Weller, D., A. J. Jeffreys, V. Wilson, and A. Blanchetot. 1984. Organisation of the human myoglobin gene. *EMBO 3*: 439-446.

Westerman, W., and D. Edwards. 1991. The divergence between echidna (Monotremata: *Tachyglossidae*) and platypus (Monotremata: *Ornithorhynchidae*) new data from DNA studies. *Aust. Mammal. 14*: 115-120.

Westerman, W., and D. Edwards. 1992. DNA-DNA hybridisation and the phylogeny of monotremes. Pages 28-34 in M. L. Augee, ed. *Platypus and Echidna*, 1st ed. Royal Zoological Society of New South Wales, Sydney.

Wheeler, D. 1997. Is there a third β-like globin gene in Marsupials? Honours Thesis, University of Adelaide, Australia.

Whittaker, R. G., and E. O. P. Thompson. 1974. Studies on monotremes proteins. V. amino acid sequences of the α-chain of haemoglobin from the platypus, *Ornithorhynchus anatinus*. *Aust. J. Biol. Sci. 27*: 591-605.

Whittaker, R. G., and E. P. O. Thompson. 1975. Studies on monotremes proteins. VI. amino acid sequences of the β-chain of haemoglobin from the platypus, *Ornithorhynchus anatinus*. *Aust. J. Biol. Sci. 28*: 353-365.

Whittaker, R. G., W. K. Fisher, and E. P. O. Thompson. 1972. Studies on monotreme proteins. I. Amino acid sequences of the β-chain of haemoglobin from the echidna. *Tachyglossus Aculeatus Aculeatus*. *Aust. J. Biol. Sci. 25*: 989-1004.

Whittaker, R. G., W. K. Fisher, and E. O. P. Thompson. 1973. Studies on monotremes proteins. II. amino acid sequences of the α-chain in haemoglobin from the echidna, *Tachyglossus Aculeatus Aculeatus*. *Aust. J. Biol. Sci. 26*: 877-888.

Whittaker, R. G., W. K. Fisher, and E. O. P. Thompson. 1978. Monotreme haemoglobin and myoglobin amino acid sequences and their use in phylogenetic divergence point estimations. *Aust. Zool. 20*: 60-71.

Whittington, R. J., and T. R. Grant. 1983. Haemoglobin and blood chemistry of the free-living platypus, *Ornithorhynchus anatinus* (Shaw) (Monotremata: *Ornithorhynchidae*). *Aust. J. Zool. 31*: 475-482.

Whittington, R. J., and T. R. Grant. 1984. Haematology and blood chemistry of the conscious platypus, *Ornithorhynchus anatinus* (shaw) (Monotremata: *Ornithorhynchidae*). *Aust. J. Zool. 32*: 631-635.

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

Wilson, A. C., H. Ochman, and E. M. Prager. 1987. Molecular time scale for evolution. *TIG 3*: 241-247.

Wittenberg, J. B. 1970. Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. *Physiol. Rev. 50*: 559-636.

Wolfe, K. H., and P. M. Sharp. 1993. Mammalian gene evolution: nucleotide sequences divergence between mouse and rat. *J. Mol. Evol. 37*: 441-456.

Wood, W. I., J. Nickol, and G. Felsenfeld. 1981. Repeated sequences organisation and RNA transcription map of the chicken adult β-globin gene region. *J. Biol. Chem. 256*: 1502-1505.

Wrigley, J. M., and J. A. M. Graves. 1984. Two monotreme cell lines, derived from female platypus(*Ornithorhynchus anatinus*; Monotremata, Mammalia). *In Vitro 20*: 321-328.

Wrigley, J. M., and J. A. M. Graves. 1988a. Karyotype conservation in the mammalian order monotremata (subclass Prototheria). *Chromosoma 96*: 231-247.

Wrigley, J. M., and J. A. M. Graves. 1988b. Sex chromosome homology and incomplete tissue-specific x-inactivation suggest that monotremes represent an intermediate stage of mammalian sex chromosome evolution. *Heredity 79*: 115-118.

Wu, C.-I., and W.-H. Li. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA 82*: 1741-1745.

Yakura, K., and S. Tanifuji. 1983. Molecular cloning and restriction analysis of *Eco*RI fragments of *Vicia faba* rDNA. *Plant Cell Physiol. 24*: 1327-1330.

Yang, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol. 10*: 1396-1401.

Yang, Z. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol. 42*: 587-596.

Young, P. R., R. W. Scott, D. H. Hamer, and S. M. Tilghman. 1982. Construction and expression in vivo of an internally deleted mouse α-fetoprotein gene: presence of a transcribed *alu*-like repeat within the first intervening sequence. *Nucleic Acids Res. 10*: 3099-3116.

Zeller, U. 1993. Ontogenetic evidence for cranial homologies in monotremes and therians, with special reference to *Ornithorhynchus*. Pages 95-107 in F. S. Szalay, M. J. Novacek, and M. C. Mckenna, eds. *Mammal phylogeny*. New York, Springer-Verlag.

Zhu-H, and A. F. Riggs. 1992. Yeast flavohemoglobin is an ancient protein related to globin and a reductase family. *Proc. Natl. Acad. Sci. USA 89*: 5015-5019.

Zitnik, G., Q. Li, G. Stamatoyannopoulos, and T. Papayannopoulou. 1993. Serum factors can modulate the developmental clock of γ- to β-globin gene switching in somatic cell hybrids. *Mol. Cell. Biol. 13*: 4844-4851.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pages 97-166 in V. Bryson, and H. J. Vogel, eds. *Evolving genes and proteins*, 1st ed. New York, Academic Press.

APPENDIX

Appendix 1 Primers used for sequencing and PCR.

	Name of Primers	Sequence (recorded 5'→ 3')
pobg 1-1	ppf 1.1	GTG ACC TGG TGT CAT GCG
	ppr 1.2	GTA ATG GTC AGA CAG GGC CAC
	ppf 1.3	GAC CTC ACC TTG TGC TTG
	ppr 1.4	GTG CAG ATC AGA CAA GCC
pobg 1-2	ppr 2.1	GAC CAC AGT CCC CTC CTG TCC
pobg 1-3	ppf 3.1	GCC TAC TCC CGC AGT GCT G
	ppf 3.2	CCA GCC CAT ACA CAA TGC
	ppf 3.3	CCA GCG CTT AGA ACA GTG C
	ppr 3.4	GAG GTC GGC AGT GAG ATA G
	ppr 3.5	GGG ACA GGG ACT GTG TCT G
	ppr 3.7	GTG ATC AGA TTG TCC ACG
pobeg 1-1	pepf 1.1	CTG CAC TGC GAC AAG CTG C
	pepr 1.2	GGT CAG TGG TGC GAG TAG
	pepf 1.3	GCT CTG GGT TGT GAC AGT C
	pepr 1.4	CTC TGC CAC GTG ACT GCT G
	pepf 1.5	GCT CTG CCT CTC TCT GTT G
	pepr 1.6	GAA TGG GAG GTG ATG ATG G
	npepf 1.5	CAG TCC AGT TTC TCC TCC
	pepr 1.8	CTC TCG GGT CAG GCC TC
	pepr 1.10	CAT GCC AGT TCC TGC AG
	pepf 1.9	GCC ACT TAC CCC AAG AGC
	pepf 1.11	CTT AGT TCA ACC ACG TGC
	pepf 1.13	CTT TAG CTG CAT CCT CTG
pobeg 2-1	pepr 21.1	CAA TCA TCT CTG AGG CC
	pepf 21.2	GCT GAG GGC GTG ACC TG
pobeg 2-2	pepf 22.1	GGG TTG TGT TGA GCT GTA G
	pepr 22.2	CCT ACA CTG CCT CAT CC
	pepf 22.3	GAC TGG GAA CCT CAC GTG
	pepf 22.5	GGC GAC ACA GAA GGA AG
	pepr 22.4	CTT CCT CCT CCC TCT CC
	pepr 22.6	CTA CTC TCC CAA GCG CTG
	pepf 22.7	GTA GGT GGA TCG GCA AAC
	pepr 22.8	CCA CTG AGC CAC GCT GTT
pTAG-3	707f	GGC CAA TCA GCT CTG AG
	Leef	GCA TTT GAG TGG TTC TGA G
	Leef2	GCA TGT GAA CGT CAA TGA ACT G
	697r	CAG GGC ATC GCC GAA GG
	696f	GCC TTA GAC TAG ACC TGG C
	706f	GGT ACA GAA GGA GAG GAG G
	713r	CCA CGA CCA GCA CGT TAC C
	Leer	TCA GTG GTA CTT GTG G
	698f	CCT GAG AGC CAG CCT GGC C
	714r	GGC CAG GCT GGC TCT CAG G
	737r	GTA CTG CGG GAG TAG GC
	711f	CTG GAT CTC AAT GTA GC
	712r	CTG GAC ATT GTG TAT GG
For PCR		
	nugf 2	CTT ACC CAC CAG GCT GCT G
	b1f	AAA GTG AAC ATC AAT GAG TTG
	ef 1	TCT GCC TCT GCC ATC ATG
	ef 2	ATG GTG CAC TTT ACT GCT GAG GGA
	er 1	CAT GAT GGC AGA GGC AGA
	er 2	GTG GGC CAG AGC ATT GGC CAC A
	er 3	AGT TTG TCA CAG TGC AGC TCA CTC AGC T
	Leef 22	CAT GTG AAC GTC AAT GAA C
	Leer 2	CTG G(G/C)C CA(C/T) AAG TAC CAC TG
	Leer3	CTC AGG ACT CAG TGG TAC
	707 F	GGC CAA TCA GCT CTG AG
	Lee F	GCA TTT GAG TGG TTC TGA G
	697 R	CCT TCG GCG ATG CCC TG
	696 F	GCC TTA GAC TAG ACC TGG C
	706 F	GGT ACA GAA GGA GAG GAG G
	713 R	GGT AAC GTG CTG GTC GTG G
	698 F/714 R	CCT GAG AGC CAG CCT GGC C
	737 R	GCC TAC TCC CGC AGT AC
	711 F	CTG GAT CTC AAT GTA GC
	712 R	CCA TAC ACA ATG TCC AG

Appendix 2 Partial sequences of λEL.1.1 subclones.

pTAG 3-1

F
GCTCTGGGGTGTGAGAGTCACTGGGCTCAGACATTCTGCCTTGGATCAGGTCCCCACTAT
TCACCCTGCCCTGTGTCCATTTCTCCCACAGCGCCTGGGTAACGTGCTGGTCGTGGTCCT
GGCCCGTCACTTCAGCAAGGAATTCACCCCCGAGGCCCAGGCTGCCTGGCAGAAGCTGGT
GTCTGGTGTTTCCCACGCCCTGGCCCACAAGTACCACTGAGTCCTGAGCTGCCCCAGTGA
CTTCTCCCCCAGGTCCCATGACCTGCCCGTGGGAGTGCAGCACCACAGGCCCCTGAGAGC
CAGCCTGGCCCAATAAAGAGCTTTAATCAGTTT

R
TAGCCTGCAGGTCAGACTCTAGAGGATCCCCAGGGCTTTGGAGTCAGAAGTCATGGGTTC
AGATCCCGGCTCTGGCAATTGTCAGCTGTGTGACTTTGGGCAAGTCACTTGACTTCTCTG
TGCCTCAGCTATCTCAATTGTAAAATGGGGATTAATACTGTGAGGCCCCTGTGGACAACC
TGATAATAATAATGATGGTATTTGTTAAGCACTTCTATGTGTAAAGCACTGTTCTAAGCG
CTGGGGGGATAAAGGTGATCAGATTGTCCCACGTGGGGGTCAGTCTTAATCCCATTTTAC
AGATGAGGGAACCAAGGCACAGAGAAGTTAAGTGACTTGTCCATAGTCACACAGCTGACA

pTAG 3-2

F
CAGATATTAATATAAATAAATAAATAAATTACAGATCTATACAAATGTGCTGTGAGGATG
GGTGGGAGGATGAATGAAGGAGTAAGTAAGAGTGGCAGAGAG

R
GGGCTCCTTCCACTGAGCCTGGGTGAGCCCTTCTTCTTGCTGGTCCGCCATGCCCAACCT
AGCCATGCCAAGGGCATTCCTGCCTCCTTCACTCTGCCCTACCCTCTATTTAGAGAATAG
GAGCAGAAGGTAACTGAGGGTGTAACACCCAATCATCACCTCCCATTCATTATCTTTGTT
CAACCACGGCCCGTGGACAACCCCCCAACTCTGCATACTCACACTGACTCAAGACTAAAT
ACCCATCAGCCCTTGACCTCACTCTCTACCTGTTTGTTCCACACATA

pTAG 3-4

F
GGGTGTGCTTGTTGTTATATTGTACTCTCCCAAGCTCTTAGCACAGTGCTTTGCACACAA
TAAACACTCAATATATACGATTGAATGAATGAATAAGTGAAAAAGGTAGACCTCAAAACC
GCAGATTTAAAGCCAGTTTGAAAAGGGACATTGGGAAGTCCTGGGTGGAGAAGGTCAAGT
TTGTGAGGAAAAGAGACCCGTTATTGGGTAGGGACCGTCTCTATATGTTAACCAATTTGT
ACTTCCCAAGCGCTTAGTACATGCTCTGCACACAGTAAGTGCTCAATAAATACCATTGAA
TGAATGATGAATAGTGGAAAAAGGAAGGCCTGGGAGTCAAGGAACTGGGATACACCATGT
GACCTTGGCAAGCCATTTGACTTCTCCGCACTTCCATTACCTCATCTATAAAATGGGATC
CCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTG

R
CAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGGATCCCATTTTATAGATG
AGGTAATGGAAGTGCGGAGAAGTCAAATGGCTTGCCAAGGTCACATGGTGTATCCCAGTT
CCTTGACTCCCAGGACCTTCCTTTTTCCACTATTCATCATTCATTCAATGGTATTTATTG
AGCACTTACTGTGTGCAGAGCATGTACTAAGCGCTTGGGAAGTACAAATTGGCCAACATA
TAGGAACGGTCCCTACCCAATAACGGGTCTTNNNTCACAAACTTGACCTTCTCCACCCAG
GACTTCCAATGGCCTTTGCACTGGCTTAATCTGCGGTTTGAGGT

pTAG-4

F
ACTGCTGGAGCTCTGTCCCCAGATCCCCCTCCATAGGCCAAAACGTTCTTTGTTCGGCTT
GTCTGATCTGCACCCCTGGTTAGAACTAAAAGCGCTGACCTCACCTTGTGCTTGGCTTCT
GGTGGGCCCTGCCAGTAGCCCCTAGACCTCACCTGAGCCCCGAGGCCATT

R
GGATCCCTTATCCTATCAGCACTCAGCCTCCTCTGCCCACCTCTCCCCTTTACCCTCTTC
AAATTACAGGTCCTCTTTCTACTGGCTCATGACTTTTTTCATTTTCATCATTGGTTTTCC
TTCAGGACTCATGGTCAAACTCTGTTTTAGGGGTCCTCTGGCACCGCCCCCCACCCCCCC
ACTTCCTCAAGATTCATGTTAAAAGTTTTCTCTTCCAGTGAACCTTCTGGACTGAAAATC
GAT

pTAG-5

F
ACTGCTGGAGCTCTGTCCCCAGATCCCCCTCCATAGGCCAAAACGTTTCTTTTTCGGCTT
GTCTGATCTGCACCCCTGGTTAGAACTAAAAATGACCTCACCTTGTGCTTGGCTTCTGGT
GGGCCCTGCAGTAGCCCCTAGACCTCACCTGAGCCCCGAGGCCATTGGGTGGGAGAGGAG
CACCTGCCAGCCTGTACCAGAGAAGGGGAAGCTGAGGG

R
ACTTTAAACTCATTGCAAGCAGGGAATTTACCTGTTTATTGTTATATTGTACTCTCCCAA
GCGCTGAGTACAGTACTCTTGCACGTGGTAAGCTCTTAATAAATATGAATGAATGAATGA
ATGAATGAATGAATGAATGAATGAATGAATGAAGGCTGTTGCAGCCTTCCTTCAATCAT

Appendix 3 Partial sequences of pobg 1-3, a subclone of λPL.1.2.

pobgr1-3

CTGCAGGTCGACTCTAGAGAGGGTGACAGATATTAATATAAATGAATAAATAAATTACAG
ATCTATACAAATGTGCTGTGAGGATGGGTGGGAGGATGAATGAAGGAGTAAGTAAGAGTG
GCAGAGAAGGGAGTGGAGAACCGGAGAGGGGGAATTAGTCAGGGAAGTCTTCTCGGAGGA
GATGTGCCTTCAATAAGGTTTTGAAGTAGGGGAGAGCAATGATAGGAGCAGAGAGGGTGT
CCCAGGCCAGAGATAGGATCTCCCCCAAGAGGTCGGCAGTGAGATAGATGAGGTCGAGGT
ACAATGAGAAGGTTAGCATTAGAGGAGCAAGGTGATTAAGTGCTTTAATGCCATGGTGAG
GAGTTTTTGTTTGATGTGGAGGTGGATGGGCAATCACTGGAGTTTCCTGAGGCTAGGGAA
AACTAGGTCTGAATGTTTTTGAAGAATGCAGCCAAAGGTGGGAGCGCAGATGGGGAGAAG
GGCAGGGGGCATGGAATACCAGGAAAGCATCTTGGCCTAGTGGAAAGAGCATGGACCTGG
AAGTCAGATCACTTAGGTTCTAATCCCAGCTCTGCCATGTGACTGCTGTGTGATCTTGGA
CAAGTCACTCAAGTTCTCTGTGCCTCTGTTTCTGCATCTGTAAAAGGGGAGGAAATACCT
GCCGTTTCTCCTACTTGGACTGTGAACGCCATATGGGACAGGGACTGTGTCTGACCTGGT
TTATTCCTATCTATCACAATTGCTTAGAACAGTGCTTGACACATAGTAATGCACTAACAA
TATCT--?--GGGACAGCGACGGGGGTGTTAATGTGTGGGAAGCGAAAACGAGGTAGAGA
GTGGAGGTCAAGGGCTGATGGGTATTTAGTCTTGAGTCAGTGTGAGTATCCAGAGTTGGG
GGGTTGTCCACGGGCCGTGGTTGAACAAAGATAATGAATGGGAGGTGATGATTGGGTGTT
ACACCCTCAGTTACCTTCTGCTCCTATTCTCTAAATAGAGGGTAGGGCCAGAGTGAAGGA
GGCAGGAATCCCTTGGCATGGCTAGGTTGGGCATGGCGGGACCAGCAAGAAGAAGGGCTC
ACCCAGGCTCAGTGGAAGGAGCCCT

Appendix 4 Sequences of Lee F2/Lee R PCR product from platypus genomic DNA.

SLee F2

TATGACTCGAGGGGAGATTTTGAATTCCTGTGTTTGGGCCCACCTAGCTCTCAAAGGGCC
TCCAGAGGGTAACCAGGTGGAAACTCCAAGTTTCTCACTCTGTCTTTCTCTCTCTTTCCT
TTGCCCTCCCTCTTACCCACCAGGCTGCTGGTCGTCTACCCCTGGACCCAGAGGTTCTTC
GAATCCTTCGGTGACCTGTCCAGCGCCGATGCTGTGATGGGAAACGCCAAGGTCAAGGCC
CATGGTGCCAAGGTGCTGACCTCCTTCGGCGATGCCCTGAAGAACCTCGACAACCTCAAG
GGCACCTTCGCCAAGCTGAGCGAACTGCACTGCGACAAGCTGCACGTGGACCCCGAGAAT
TTCAATGTGAGTCCCCCAGCCCTTAGACTAGACCTTGGCCATCTTGACCCCTGCGGGAAG
CTTGGCATTGGGAAGCTTGGAAAGGAAAAAGGGGCTTGAAACCCCCCCT

SLee R

CCGCACCCTTTGTGGTAATGTGTAGAAGACGGTCCGTGGGACCCGGAGCCCCCACTTAAG
GAACGACTTCACTGCCCGGTCCTGGTGTTGGTCGTGCAATGGGTCCGCGACACCCTCTTT
ACCTGTGTCCCGTCCCACTTATCACCCCTGGACTAGGTTCCGTTTTACAGACTCGGGTCA
CTGAGAGTGTGGGGTTTCGAGAATGGTCCGTGGTAGCGACCGGAACCGAGTTATAGTTTT
ACGGATGGTGTGGGGCAAGGATAGAGGATTTCGGAGGATTACGTATTGAACTGGTGTCAG
GGGAGGAGAGGAAGACATGGGGGTTTACTATCGGGGGTGAGACTGGCTGTGAATGAAGCA
TTACATAGGGTGTCATCGTTGGGGACCAGGAAAGAGGGAGTTTCCGAGAGTAAGGACTGA
GTAAAGGGAGTTTTACGGGGTCCGATAGAGAAACGGGTTGTGTGTGAGGTCCTTGGGTAG
TAAAGAAGGGGTCCCCC

Appendix 5 Multiple sequence alignment of four monotreme β-globin genes.

Exons

```
                Exon I
echid β1.ex     ATGGTGCATTTGAGTGGTTCTGAGAAGACTGCTGTCACCAACCTGTGGGGGCATGTGAAC
NGechid β1.ex   ATGGTGCATTTGAGTGGTTCTGAGAAGACTGCTGTCACCACCCTGTGGGGGCATGTGAAC
platy β1.ex     ATGGTGCATCTGAGTGGTGGTGAGAAGAGTGCCGTCACCAACCTGTGGGGGAAAGTGAAC
platy β2.ex     ATGGTGCATTTGAGTGGTTCTGAGAAGACTGCTGTCACCAACCTGTGGGGGCATGTGAAC
                ********* ********  ******** *** ******* ********** * ******

echid β1.ex     GTCAATGAACTCGGTGGCGAGGCCCTTGGCAG
NGechid β1.ex   GTCAATGAACTCGGTGGCGAGGCCCTTGGCAG
platy β1.ex     ATCAATGAGTTGGGTGGCGAAGCCCTTGGCAG
platy β2.ex     GTCAATGAACTCGGTGGCGAGGCCCTTGGCAG
                 *******  * ******** ***********

                Exon II
echid β1.ex     GCTGCTGGTCGTCTACCCCTGGACCCAGAGGTTCTTCGAATCCTTCGGTGACCTGTCCAG
NGechid β1.ex   GCTGCTGGTCGTCTACCCCTGGACCCAGAGGTTCTTCGAATCCTTTGGTGACCTGTCCAG
platy β1.ex     GCTGCTGGTCGTCTACCCCTGGACCCAGAGGTTCTTCGAAGCCTTTGGTGACCTGTCCTC
platy β2.ex     GCTGCTGGTCGTCTACCCCTGGACCCAGAGGTTCTTCGAATCCTTCGGTGACCTGTCCAG
                ****************************************** **** ************


echid β1.ex     CGCCGATGCTGTGATGGGAAACGCCAAGGTCAAGGCCCATGGTGCCAAGGTGCTGACCTC
NGechid β1.ex   CGCCAATGCTGTGATGGGAAACCCCAAGGTCAAGGCCCATGGTGCCAAGGTGCTGACCTC
platy β1.ex     CGCCGGCGCTGTGATGGGCAACCCCAAGGTCAAGGCCCATGGTGCTAAGGTGCTGACCTC
platy β2.ex     CGCCGATGCTGTGATGGGAAACGCCAAGGTCAAGGCCCATGGTGCCAAGGTGCTGACCTC
                ****   *********** *** ********************** **************

echid β1.ex     CTTCGGCGATGCCCTGAAGAACCTCGACAACCTCAAGGGAACCTTCGCCAAGCTGAGCGA
NGechid β1.ex   CTTCGGCGATGCCCTGAAGAACCTCGACGACCTCAAGGGCACCTTCGCCAAGCTGAGCGA
platy β1.ex     CTTCGGCGATGCCCTGAAGAACCTCGACGACCTCAAGGGCACCTTTGCCAAGCTGAGCGA
platy β2.ex     CTTCGGCGATGCCCTGAAGAACCTCGACAACCTCAAGGGCACCTTCGCCAAGCTGAGCGA
               ****************************************** ***** ***************

echid β1.ex     ACTGCACTGCGACAAGCTGCACGTGGACCCCGAGAATTTCAAT
NGechid β1.ex   ACTGCACTGCGACAAGCTGCACGTGGACCCCGAGAATTTCAAT
platy β1.ex     ACTGCACTGCGACAAGCTGCACGTGGACCCTGAGAATTTCAAT
platy β2.ex     ACTGCACTGCGACAAGCTGCACGTGGACCCCGAGAATTTCAAT
                ****************************** *************

                Exon III
echid β1.ex     CGCCTGGGTAACGTGCTGGTCGTGGTCCTGGCCCGTCACTTCAGCAAGGAATTCACCCC
NGechid β1.ex   CGCCTGGGTAACGTGCTGGTCGTGGTCCTGGCCCGTCACTTTAGCAAGGAATTCACCCC
platy β1.ex     CGTCTGGGTAACGTGCTGATCGTGGTCCTGGCCCGTCACTTCAGCAAGGACTTCAGCCC
platy β2.ex     CGCCTGGGTAACGTGCTGGTCGTGGTCCTGGCCCGTCACTTCAGCAAGGAATTCACCCC
                ** *************** ********************** ******** **** ***

echid β1.ex     CGAGGCCCAGGCTGCCTGGCAGAAGCTGGTGTCTGGTGTTTCCCACGCCCTGGCCCACA
NGechid β1.ex   CGAGGCCCAGGCTGCCTGGCAGAAGCTGGTGTCTGATGTTTCCCACGCCCTGGCCGACA
platy β1.ex     CGAGGTCCAGGCTGCCTGGCAGAAGCTGGTGTCTGGTGTTGCCCACGCCCTGGGCCACA
platy β2.ex     CGAGGCCCAGGCTGCCTGGCAGAAGCTGGTGTCTGGTGTTTCCCACGCCCTGGCCCACA
                ***** ****************************** **** ************ * ***

echid β1.ex     AGTACCACTGA
NGechid β1 ex   AGTACCACTGA
platy β1.ex     AGTACCACTGA
platy β2.ex     AGTACCACTGA
                ***********
```

Introns I and Intron II

```
                Intron I
echid β1.in1    GTATGACCTCGAGGGCAGAATTTTGAATTCCTGTGTTTGGGCCC-ACCTAGCTCTCAAAG
NGechid β1.in1  GTATGACCTCGAGGGGAGAATTTTGAATTCCTGCATTTGGGCCC-ACCTAGCTCTCAAAG
platy β1.in1    GTACGACCTCGAAGGGAGAATTTCGAATTCCTGTATTTGGGCCCCACCTAGCTCTCAAAG
platy β2.in1    GTATGACCTCGAGGGGAGAATTTTGAATTCCTGTGTTTGGGCCG-ACCTAGCTCTAAAAG
                *** ******** ** ******* *********  ********  ********** ****
```

```
echid β1.in1     GG-CCTCCAGAGGGTAACCAGGTGGAAACTCCAAGTTTCTCACTCTGTC------TTTCT
NGechid β1.in1   GG-CCTCCAGAGGGTAACCAGGTGGAAACTCCAAGTTTCTCACTCTGTC------TTTCT
platy β1.in1     GG-CCTCCAGAGCGTGACCAGGTAGAAACTCCAGGTCTCTCACTCTGTCACACTCTGTCT
platy β2.in1     GGGCCTCCAGAGGGTAACCAGGTGGAAACTCCAAGTTTCTCACTCTGTC------TTTCT
                 ** ********* ** ******* ********* ** ************      * ***

echid β1.in1     CTCTCTTTCCTTTGCCCTCCCTCTTACCCACCAG
NGechid β1.in1   CTCTCTTTCCTTTGCCCTCGCTCTTACCCACCAG
platy β1.in1     CTCTCTCTCCTTTGCCCTCGCTCTTACCCACCAG
platy β2.in1     CTCTCTTTCCTTTGCCCTCCCTCTTACCCACCAG
                 ****** ************ **************

                 Intron II
echid β1.in2     GTGAGTACCCCAGCCTTAGACTAGACCTGGCCATCTGAC--TGCGGGAGCTGGCATGGGA
NGechid β1.in2   GTGAGT-CCCCAGCCTTAGGGTAGACCTGGCCATCTGAC--TTCCGGAGCTGGCATTGGA
platy β1.in2     GTGAGT-CTCTGGCCCTGGGATAGACCTGCCCCTCTGACCCTGCAGGAACTGGCATGGGA
platy β2.in2     GTGAGT-CCCCAGCCTTAGACTAGACCTGGCCATCTGAC--TGCGGGAGCTGGCATGGGA
                 ****** * *  *** * *  ******** ** *******  * * *** ******* ***

echid β1.in2     GCTGGAAGGAAAAGGGGC-TGAACGCCTGGGGAAGAAATGATGGGG-TCCTGAGTGTGTG
NGechid β1.in2   GCTGGAAGGAAAAGGCTG-AAGCCCCCTGGGGGAGAAATGATGGGG-TTCTGAGTGTGTG
platy β1.in2     GGTGGAAGAAAAAGGTGCCGAAGCCCCTGGGGGAGAAAGGTTGGGGGTCCTGAGTGTCTC
platy β2.in2     GCTGGAAGGAAAGCTGA----AGCCCCTGGGGAAGAAATGATGGGG-TCCTGAGTGTGTG
                 * ****** ***           * ******* ***** * ***** * ******** *

echid β1.in2     TTGGGCAAAGAGATAGCCTGGGGCATCTGAAGGAAATGAGTCAGAATGAGAGCCTTTGAG
NGechid β1.in2   TTGGGCAAAGAGATAGCCTGGGGCATCTGAAGGAAATGAGTCAGAATGAGAGCCTTTGAG
platy β1.in2     AT-----------------------TCTCAGGG----G--TCAGAATGAGAGACTTTGAG
platy 2.in2      TTGGGCAAAGAGATAGCCTGGGGCATCTGAAGGAAATGAGTCAGAATGAGAGCCTTTGAG
                  *                       *** * **    *  ************ *******

echid β1.in2     GGAGAAAGGACCAGGGGTTGCTACTGTGGGATACTTTACGAAGTAAGTGTCGGTCAGAGT
NGechid β1.in2   GGAGAAAGGACCAGGGGTTGCTACCGTGGGATACTTTACGAAGTAAGTGTCACTCAGAGT
platy β1.in2     GGAGAAAGGAGCAGCAGTTGCTACTGAGGGATACTTTACCAAATGATTGTCAGTCAGAGT
platy β2.in2     GGAGAAAGGACCAGGGGTTGCTACTGTGGGATACTTTACGAAG-AAGTGTCGGTCAGAGT
                 ********** ***  ******** * ************ **   * ****  *******

echid β1.in2     GGGGGCTATCATTTGGGGGTACAGAAGGAGAGGAGGGGACTGTGGTCAAGTTATGCATTA
NGechid β1.in2   GGGGGCTATCATTTGGGGGTAGAGAAGGAGAGGAGGGGACTGTGGTCAAGTTATGCATTA
platy β1.in2     CCGGGTTATCATTTGGGGGTAAGGAAGGAGAGGAGGGGGCTGTGGTCAAGTTATGCCTTA
platy β2.in2     GGGGGCTATCATTTGGGGGTACAGAAGGAGAGGAGGGGACTGTGGTCAAGTTATGCATTA
                   *** ***************  *************** ***************** ***

echid β1.in2     GGAGGCTTTAGGAGATAGGAACGGGGTGTGGTAGGCATTTTGATATTGAGCCAAGGCCAG
NGechid β1.in2   GGAGGCTTTAGGAGATAGGGATGGGGTGTGGTAGGCATTTTGATATTGAGCCAAGGCCAG
platy β1.in2     AGAGGCTTTAGGAGATAGAGATGGG-TGTGGTAGGTGTTTTGATATTGAGCCAAGGCCAT
platy β2.in2     GGAGGCTTTAGGAGATAGGAACGGGGTGTGGTAGGCATTTTGACATTGAGCCAAGGCCAG
                  ****************  * *** **********  ****** ***************

echid β1.in2     CGATGGTGCCTGGTAAGAGCTCTGGGGTGTGAGAGTCAC-TGGGCTCAGACATTCTGCCT
NGechid β1.in2   CGATGGTGCCTGGCAAGAGCTCTGGGGTGTGAGAGTCAC-TGGGCTCAGACATTCTGCCT
platy β1.in2     GGATAGTGCCTGGCAAGAGCTCTGGGTTGTGACAGTCACCTGGGTTCAGACATTCTGACC
platy β2.in2     CGATGGTGCCTGGTAAGAGCTCTGGGGTGTGAGAGTCAC-TGGGCTCAGACATTCTGCCT
                  *** ******** ************ ***** ****** **** ************ *

echid β1.in2     TGGATCAGGTCCCCACTATTCACCC-TGCCCTGTGTCCATTTCTCCCACAG
NGechid β1.in2   TGGATCAGGTCCCCACTATTTACCC-TGCCCTGTGTCCATTTCTCCCACAG
platy β1.in2     TGGATCAGATCCCCAGTATTCACCCCTGCCCTGTGTCCATTTCTCCCACAG
platy β2.in2     TGGATCAGGTCCCCACTATTCACCC-TGCCCTGTGTCCATTTCTCCCACAG
                 ******** ****** **** **** *************************
```

'5 flanking region

```
echid β1.5       ----------GGATCCACAGCAGACTGTGCAACCATGGGGCTCTGGGAGGAGCAACATC-
platy β1.5       AGTCAAAGAAAGAAACATGCTGTGGTCTGTGGAGGGAGATCTGGGGACAGAGCTCCAGCA
platy β2.5       CTCCAGCAGTGGATCCACAGCAGACTGTGCAACCATGGGGCTTTGGGAGGAGCAACATC-
                       **  **        * **        *  **  **   ****  ** *

echid β1.5       -TTGGGCTGGGAAAGGAGGCCTGGGCCTGGTGACATGGAA--AATAATAAT----AATAA
platy β1.5       ATGGATCCATGGAAGACCACGTAACCATGGGCATCTGGGAGGAGTGACATCTTGGGCTGG
platy β2.5       -TTGGGCTGGGAAAGGA-GCCTGGGCCTGGTGACATGGAA--AATAATAAT----AATAA
                  * *  *   * ***    * *   * ***  *  *** *  * * * *       *
```

```
echid β1.5      TAATGATGGCATTTATTAAGTGCTTACAA--AGGGGGCTAAAGGTCTGGACCTGGTGGGT
platy β1.5      CAAAGGAGGACTGGGCCTGGGGATAGGGAGAAGGGGGATGAAGGTCTGGACCCGGGTGGT
platy β2.5      TAATGATGGCATTTATTAAGTGCCTACAA--AGGGGGCCAAAGGTCTGGACCTGGTGGGT
                 ** *  **  *       * *      *  ******   ************ **  ***

echid β1.5      GGGGAGTTGGGGGGGTGGTTTCTCCAGCTCTCAGAGAGGCAGATTTGCATCCCTG-CACC
platy β1.5      GGGGG--TGGGGGGGTATTTTCTCCAGCTCTCAGAGAG-CAGATTTGCATCCCCG-CACC
platy β2.5      GGGGAGTTGGGGGGGTGGTTTCCCCAGCTCTCAGAGAG-GAGATTTGCATCCCTGACACC
                ****   *********  **** ***************  ************** * ****

echid β1.5      AG-CCCTTTCAATACCTC-TCCTGCCTGCCAGGCCCGCATGACAC-AGGTCACG--CCCC
platy β1.5      A------------ACCTC-TCCTGCCTGCCAGCCCTGCATGATAC-AGGTCACG--CCCT
platy β2.5      AGACCCTTTCAATACCTCGTCCTGCCTGCCAGCCCCGCATGACACCAGGTCACAGACCCC
                *            ***** ************* ** ****** ** *******   ***

echid β1.5      CAGTGTGGACCGCACCCCTGTCTGGGCCAATCAGCTCTGAGGCTGAGAGTAGATGGAGCT
platy β1.5      CAGCG-GGACCGCACCCCAATCTGGGCCAATCATCTCTGAGGCCTAGGGTAGGTGGAGCT
platy β2.5      CAGTGTGGACCGCACCCCTGTCTGGGCCAATCAGCTCTGAGGCTGAGAGTAGATGGAGCT
                *** * ************  ************* *********  ** **** *******

echid β1.5      -CTGGGCAGGGGATAAAAAGCCAAGCTAAACCTGGAGTTGATATCCAGTCTTTGCACATC
platy β1.5      -CTGGGCAGGGGATAAAAGGCCAAGCTAAACCTAGAGTCAGTACCCAGTCTTTGCACGTC
platy β2.5      GCTGGGCAGGGGATAAAAAGCCAAGCTAAACCTGGAGTTGATATCCAGTCTTTGCACATC
                 ****************** ************** ****   ** ************ **

echid β1.5      TGAATCTTCAAGCAAACTTGCAAATCCGACAAA
platy β1.5      TGCATCTTCAAGCGAACTTGCAAATCCGACAAA
platy β2.5      TGAATCTTCAAGCAAACTTGCAAATCCGACAAA
                ** ********** ******************
```

3 flanking region

```
echid β1.3      GTCCTGAGCTGCCCCAGTGACTTCTGCGCCAGGTCCCATGACCTG-CCCGTGGGAGTGCA
platy β1.3      GTCCTGAGCTGCCCCGGTGGCTTCTGCATCAGGTCCCATGACC-----------------
platy β2.3      GTCCTGAGCTGCCCCAGTGACTTCTGCGCCAGGTCCCATGACCTGGCCCGTGGGAGTGCA
                *************** *** *******  ***************

echid β1.3      GCACCACAGGCCCCTGAGAGCC-AGCCTGGCCCAATAAAGAGCTTTCAATCCAGTTTCTC
platy β1.3      -----ACACTCCCCTGAGGCCCTATTCTGGCCCAATAAAGAGCTTTCAGTCCAGTTTCTC
platy β2.3      GCACCACAGGCCCCTGAGAGCC-AGCCTGGCCCAATAAAGAGCTTTCAATCCAGTTTCTC
                     ***  ********  ** *  ************************ ***********

echid β1 3      CTCCGCTCTGCCTCTCTCTGTGCGCCACTTTTTGGGGAAGGTCCCTACGGGATGGGGAGG
platy β1.3      CTCCGCTCTGCCTCTCTCTGTGCGCCTCTTTGTGGGGAGGGTCC-TATGGGGTGGGGAGA
platy β2.3      CTCCGCTCTGCCTCTCTCTGTGCGCCACTTTTTGGGGAAGGTCC-TACGG-ATGGGGAGG
                ************************** **** ****** ***** ** **  *******

echid β1.3      GGGAGATGGGACCCTCTCCTCC-----------AGCT-GCCTACTCCCGCAGTACTGTAG
platy β1.3      GGGAGATGGGACCCTCTCCTCCTGGCTATCTCCAGCT-GCCTACTTCCACAGGGCTGCAG
platy β2.3      GGGAGATGGGACCCCTCTCCTC----------CAGCTAGCCTACTCCCGCAGTGCTGTAG
                **************    *  *           **** ******* ** ***  *** **

echid β1.3      GGAGCTGCTCT-AAATCTCCTCCATGG-TAGCCTGGGT-GAGGAAGAGGGGCTG-AACTG
platy β1.3      GGAGCTGCTTTTAAATCCCCTCCATGG-TGGCCTGGGT-GGGGAAGAGGGGCTG-AACTG
platy β2.3      GGAGCTGCTCT--AAATTCCTCCATGGCTAGCCTGGGTTGGGGCAAGAGGGCTGCAACTG
                ********* *  **   ********* * ******** * ** *   ****** *****

echid β1.3      GTTGGGGTGAAGGAAGGGCTTTCTTGGAGGACTCTTGGGGCATCTCCAGTTTGGCTGGGA
platy β1.3      TTGGGGGCAAAGGAAGGGCTTTCTTGGAGGACTCT-------------------------
platy β2.3      GTTGGGGATAAGGAAGG-CTTTCTTGGAGGACTCTTGGGGCATCTCCAGTTTGGCTGGGA
                 * ****  ******** ******************

echid β1.3      AAAGGGAGGGCAAGGGAAGATGAGGGGGACGGTCTCTGGATCTCAATGTAGCTAGAGGGA
platy β1.3      -AAGGGAGAGAAAGGGAGGATGGCGGGGACAGTCTATGGACCTCAATATAGCTAGAGCAA
platy β2.3      AAAGGGAGGGCAAGGGAAGATGAGGGGGACGGTCTCTGGATCTCAATGTAGCTAGAGGGA
                 ******* * ****** ****  ****** **** **** ****** *********  *

echid β1.3      AGGGCCTGAA---GGCACTCCCCAAGTGCCCTGGCCCCAGACTCCCTAGAATAGGGTCTA
platy β1.3      AGGGCCTGAACTGATTACTCCCCAAGAGCCCTGACCCCAGAGTCCCCAGAATCTTGGCTA
platy β2.3      AGGGCCTGAA---GGCACTCCCCAAGTGCC-TGGCCCCAGACTCCCTAGAATAGGGTCTA
                **********      ********** *** ** ******* **** *****   * ***
```



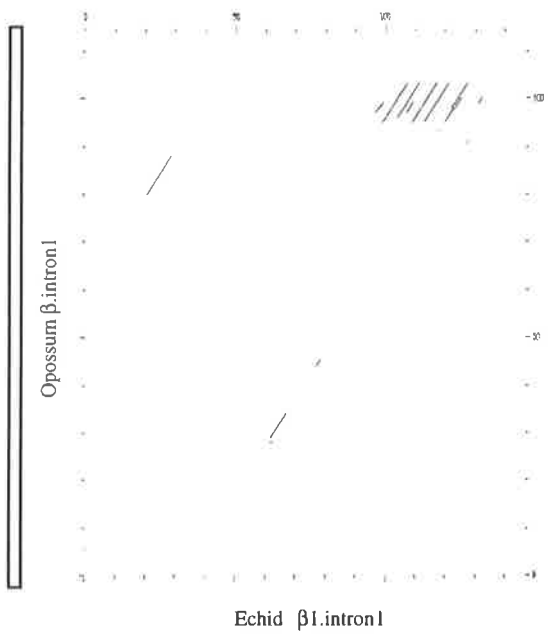
* Alignments were carried out with CLUSTALW

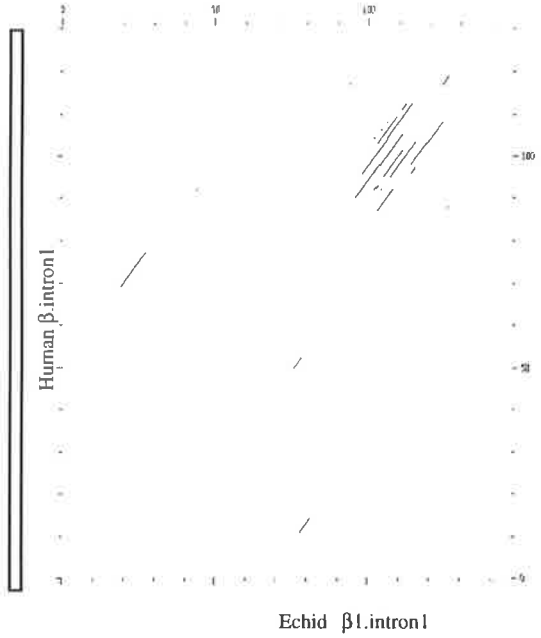
Appendix 6 Intron sizes of β-like globin genes.

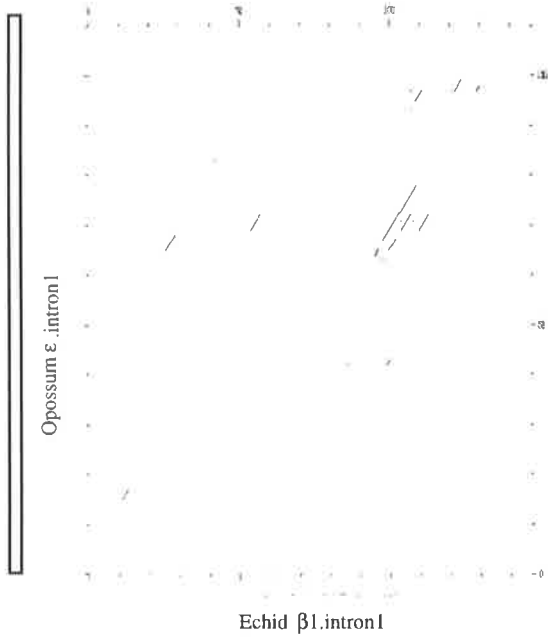
Species	β-globin	Intron I	Intron II	Species	β-globin	Intron I	Intron II
Primates							
H. sapiens	ε	122	855	*L. macaco*	ε	122	841
(Human)	^{G}γ	122	886	(Lemur)	γ	123	865
	^{A}γ	122	866		ψ	108	778
	ψ	121	851		β	130	868
	δ	128	898	*S. sciureus*	ε	128	947
	β	130	850	*S. bolivienisis*	ε	128	950
G. crassicaudatus	ε	120	814	*S. midas*	ε	127	811
	γ	128	937	*A. azarae*	ε	128	812
	δ	129	942	*A. trivirgatus*	ψ	121	876
	β	129	852	*A. belzebul*	ε	123	809
G. gorilla	ψ	121	851	*A. seniculus*	ε	123	809
	β	130	850	*A. belzebuth*	ε	123	817
	ε	122	858	*A. geoffroyi*	ε	123	819
	^{G}γ	122	906		δ	128	885
	^{A}γ	122	872		γ	123	789
P. pygmaeus	ε	122	863	*B. arachnoides*	ε	123	822
	^{G}γ	122	872	*C. goeldii*	ε	128	813
	^{A}γ	123	877	*C. albifrons*	ε	123	811
	δ	130	894		^{1}γ	123	876
H. lar	^{1}γ	123	882		^{2}γ	123	877
	^{2}γ	123	894		ψ	121	872
	ψ	120	823	*C. jacchus*	ε	128	812
H. sydactylus	ε	122	842	*C. calvus*	ε	128	823
L. lagothricha	ε	123	819	*C. kaapori*	ε	128	812
L. rosalia	ε	128	812	*C. medius*	ε	123	847
M. mulatta	ε	122	857	*C. moloch*	ε	127	820
M. murinus	ε	123	833	*C. torguatus*	ε	128	821
N. concang	ε	119	843	*C. olivaceus*	ε	128	813
P. irrorata	ε	128	822	*C. pygmaea*	ε	128	812
P. paniscus	ε	131	856	*D. madagascariensis*	ε	115	848
P. potto	ε	122	842	*C. apella*	^{1}γ	125	883
P. troglodytes	ε	122	860		^{2}γ	123	877
	ψ	121	850	*T. syrichta*	ε	123	819
	β	130	850		γψ	120	1155
	^{G}γ	122	893		ψ	160	868
	^{A}γ	122	887		δ	126	1067
					β	126	839
					γ	120	1148
The intron sizes of β-globin genes from chicken, two marsupials and other mammals							
M. musculus	y2	90	700	*O. cuniculus*	ε	121	802
(Mouse)	h0	327	851	(Rabbit)	γ	124	817
	h1	106	808		δ	102	755
	h2	80	722		β	126	573
	h3	91	701	*L. europaeus*	β	126	566
	major β	116	654	*B. taurus*	ε2	125	843
	minor β2	116	628	(Bovin)	ε4	125	840
R. norvegicus	major β	109	685		ψ1	129	570
(Rat)	minor β2	116	639		ψ2	129	460
	β	115	637		γ	420	579
C. hircus	εI	116	1038		ψ3	123	937
(Goat)	εII	126	842		β	128	898
	εV	115	1030	*O. aries*	γ	129	841
	βF	129	826	(Sheep)	βB	128	898
	βA	128	905		βC	132	843
	βC	132	847	*D. virginiana*	ε	109	1858
	ψx	125	626	(Opossum)	β	113	1465
G. gallus	γh0	108	541	*S. crassicauda*	ε	120	1438
(Chicken)	βH	94	558	(Dunnart)	β	112	1541
	βA	92	810				
	ε	108	973				

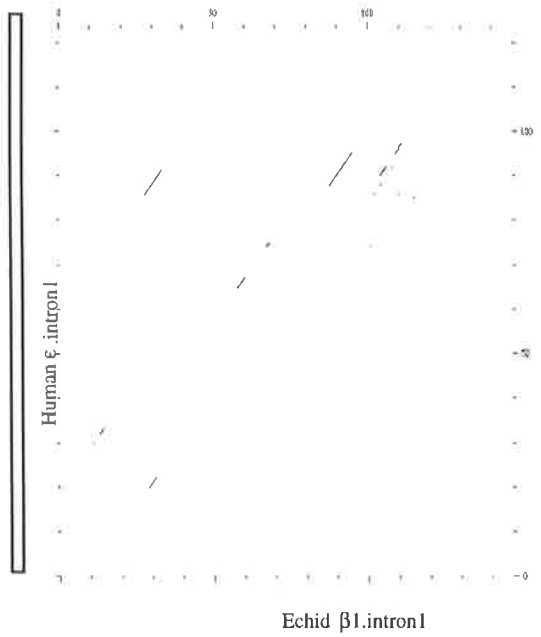
Appendix 7 Dot plot analyses.

A: Intron I comparisons between *echid* β1 and other mammalian ε- and β-globin genes.

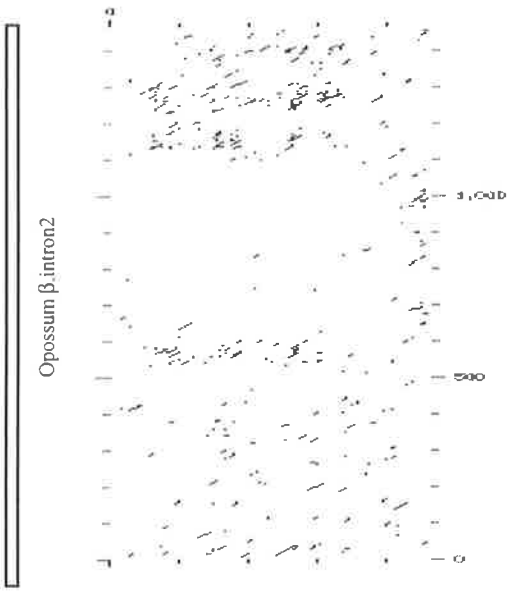


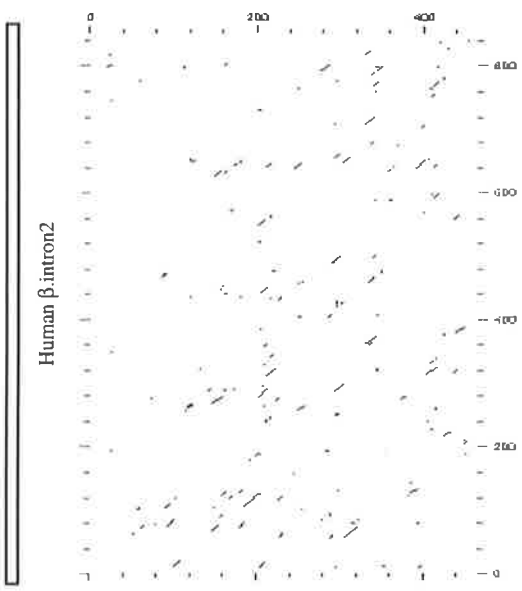


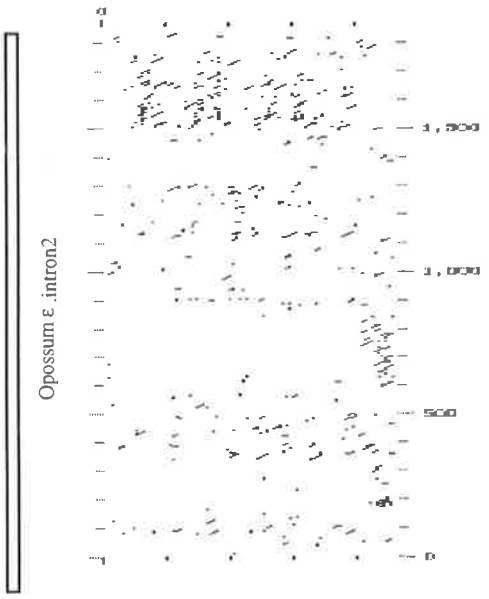


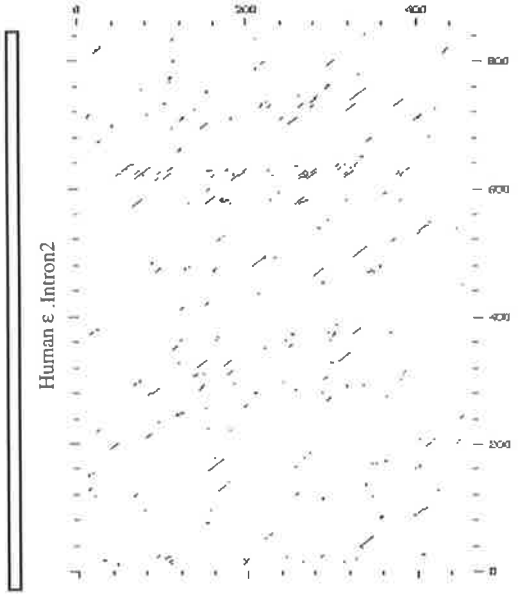


B: Intron II comparisons between *echid* β1 and othe mammalian ε- and β-globin genes.

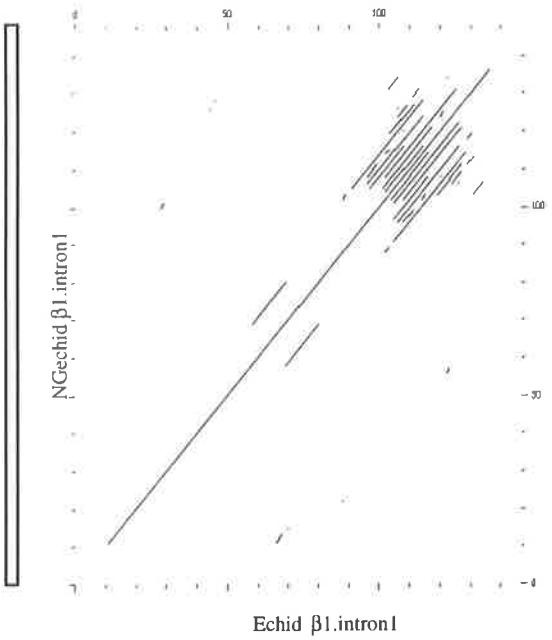


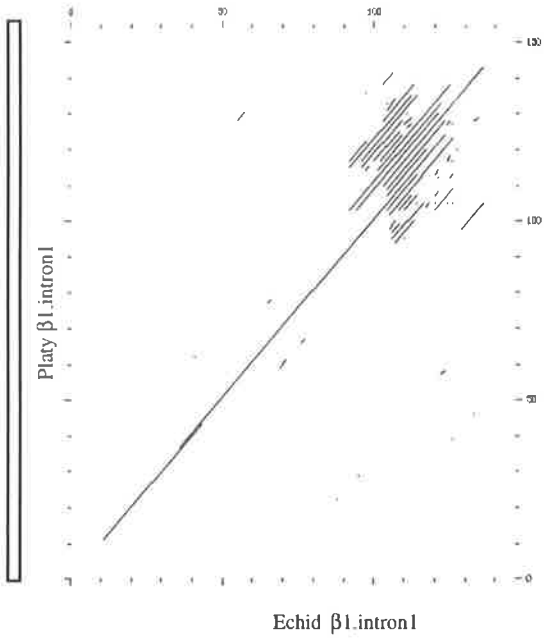


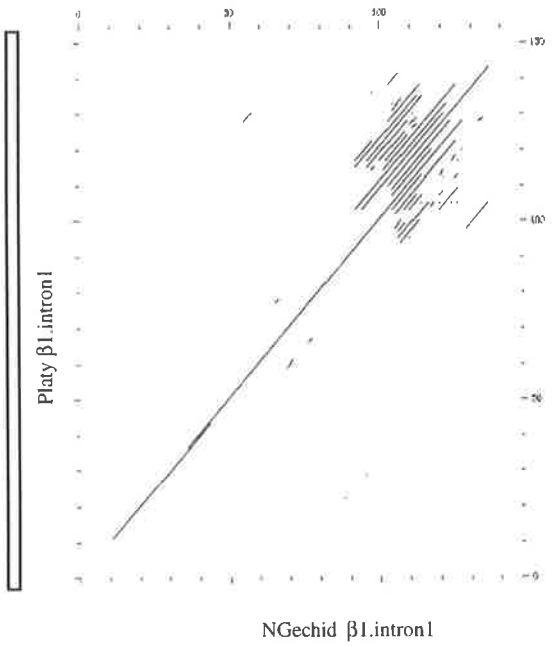


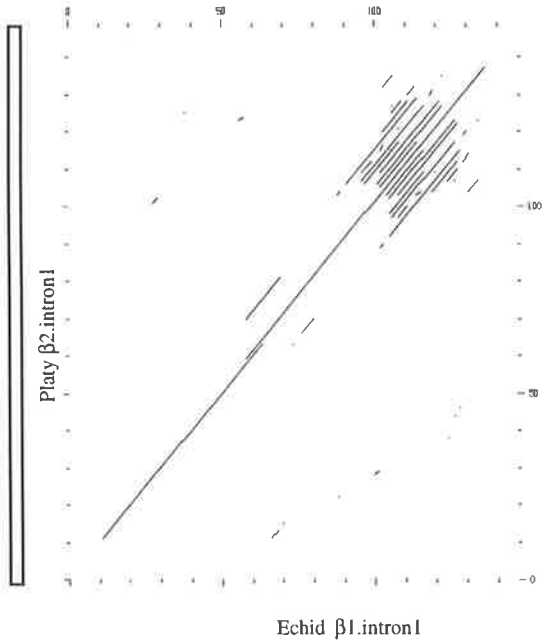


C: Intron I comparisons among the four monotreme β-globin genes.

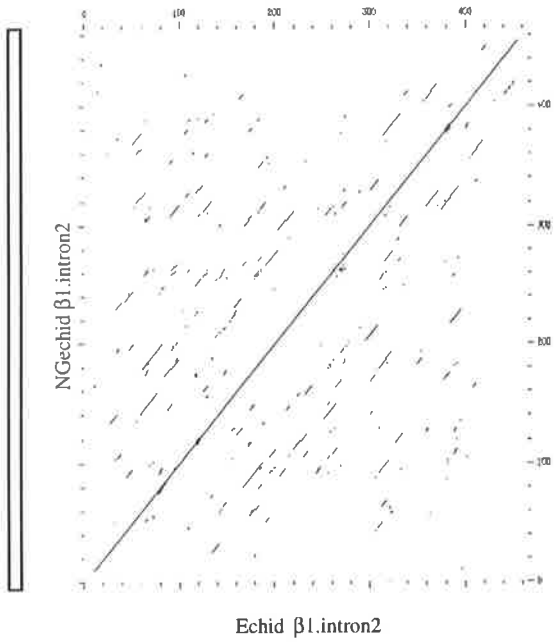


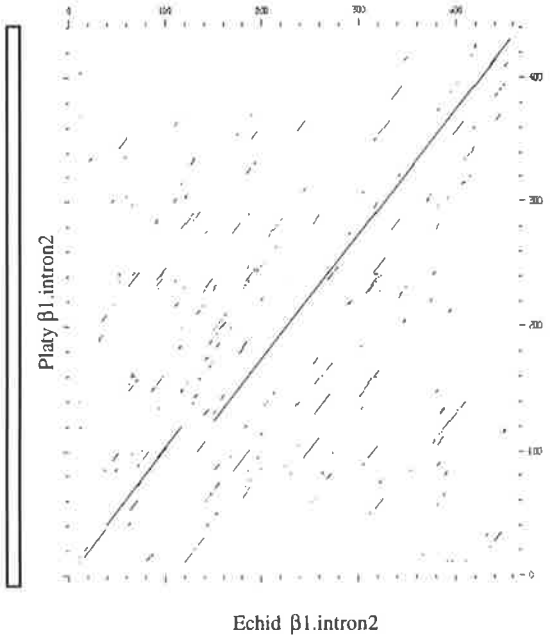


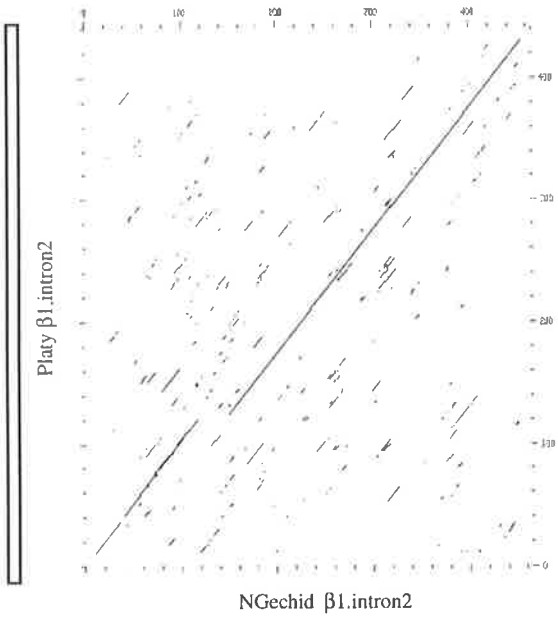


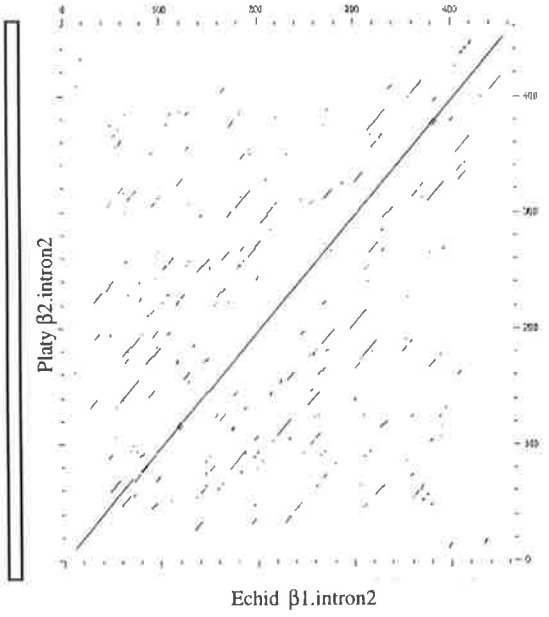


D: Intron II comparisons among the four monotreme β-globin genes.



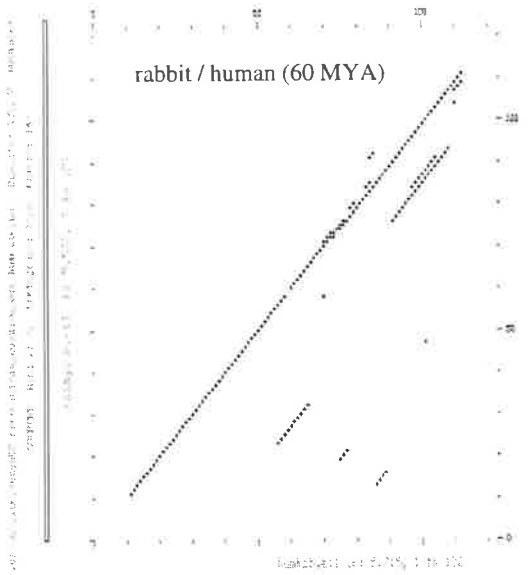




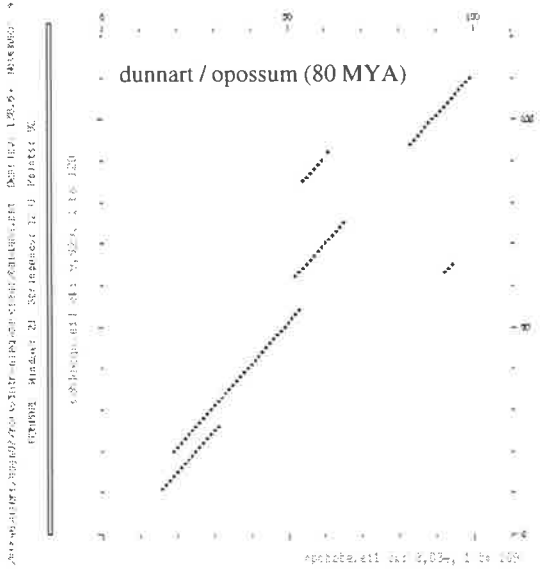


Appendix 8 : Dot plot comparisons of intron I sequences of β-like globin genes. Analyses were carried out under identical conditions (window = 21, stringency = 12) using the GCG algorithms COMPARE and DOTPLOT. The pairs of globin genes from which the intron sequences were derived, and the estimated time since each pair diverged from a common ancestral sequence (Hope, personal communication) are as follows:
A: rabbit ε-globin / human ε-globin (60 MYA) **B**: dunnart ε -globin / opossum ε -globin (80 MYA) **C**: human ε -globin / opossum ε -globin (135 MYA)
Note: i) all comparisons are between orthologous sequences; ii) after approximately 200 million years, most sequence identity between pairs of introns has dissipated

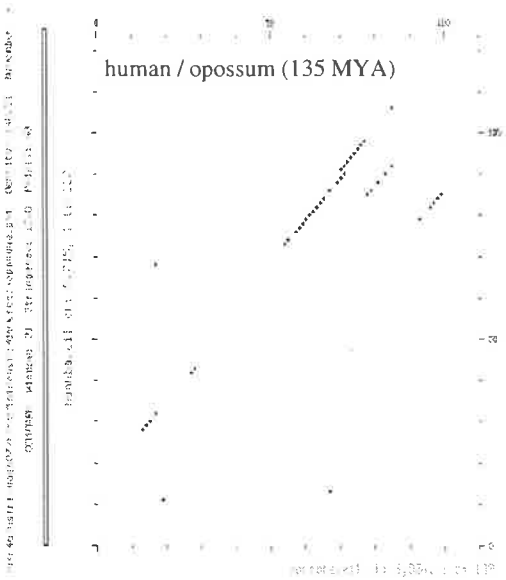
A



B



C



Appendix 9 The A+T and G+C contents of β-like globin genes.

	β-like globin gene	A	T	average of A+T	G	C	average of G+C
1	Echidna β1	89	91	90	131	133	132
2	NGechid β1	88	93	90.5	130	133	131.5
3	Platypus β1	87	92	89.5	137	128	132.5
4	Platypus β2	88	91	89.5	131	134	132.5
5	Chicken βh	92	83	87.5	130	139	134.5
6	Galago β	88	106	97	137	113	125
7	Galago δ	89	108	98.5	137	110	123.5
8	Human β	88	106	97	136	114	125
9	Human δ	89	113	101	138	104	121
10	Mouse β1	92	105	98.5	130	117	123.5
11	Mouse β2	95	102	98.5	127	120	123.5
12	Rabbit β	92	110	101	137	105	121
13	Goat βA	89	107	98	137	105	121
14	Goat βC	87	107	97	134	101	117.5
15	Opossum β	95	120	107.5	124	105	114.5
16	Dunnart β	96	117	106.5	126	105	115.5
17	Chicken ε2	101	79	90	123	141	132
18	Galago ε	108	108	108	124	104	114
19	Galago γ	112	108	110	126	98	112
20	Human ε	99	110	104.5	123	112	117.5
21	Human $^{A}\gamma$	107	99	103	126	112	119
22	Human $^{G}\gamma$	107	99	103	127	111	119
23	Mouse h0	115	112	113.5	116	101	108.5
24	Mouse h1	114	110	112	117	103	110
25	Mouse ε^{y}	107	113	110	124	100	112
26	Rabbit ε	106	112	109	118	108	113
27	Rabbit γ	113	104	108.5	121	106	113.5
28	Goat ε^{I}	96	109	102.5	129	110	119.5
29	Goat ε^{II}	107	105	106	121	108	114.5
30	Goat βF	86	112	99	136	104	120
31	Goat ε^{V}	101	110	105.5	120	113	116.5
32	Human ψ	114	111	112.5	115	103	109
33	Rabbit ψ	101	111	106	127	101	114
34	Opossum ε	98	119	108.5	124	103	113.5
35	Dunnart ε	101	116	108.5	118	109	113.5
	Average	98.2	105.7	101.9	127.3	111.8	119.6

Appendix 10 DNA sequences of coding regions of bird, monotreme marsupial and eutherian β-like globin genes used in the phylogenetic analysis and the estimation of non-synonymous and synonymous divergence values.

```
 438 GLOBIN                CHICKEN ε1
*** GTG CAC TGG TCC GCC GAG GAG AAG CAG CTC ATC ACC AGC GTC TGG AGC AAA GTC AAC
GTG GAG GAA TGC GGT GCC GAA GCC CTG GCC AGG CTG CTG ATC GTC TAC CCC TGG ACC CAG
AGG TTC TTT GAT AAC TTC GGG AAC CTC TCC AGC CCC ACC GCC ATC ATT GGT AAC CCC AAG
GTC CGT GCT CAC GGC AAA AAA GTG CTG AGC TCC TTT GGG GAA GCC GTG AAG AAC CTG GAC
AAC ATC AAG AAC ACC TAC GCC AAG CTG TCG GAG CTG CAC TGC GAG AAG CTG CAC GTG GAC
CCC GAG AAC TTC AGG CTC CTG GGG AAC ATC CTC ATC ATC GTG CTG GCC GCG CAC TTC ACC
AAG GAC TTC ACC CCG ACC TGC CAG GCT GTC TGG CAG AAG CTG GTC AGC GTG GTG GCC CAT
GCC CTG GCC TAC AAG TAC CAC
 438 GLOBIN                CHICKEN βH
*** GTG CAC TGG ACT GCT GAG GAG AAG CAG CTC ATC ACC GGC CTC TGG GGC AAG GTC AAT
GTG GCC GAA TGT GGG GCC GAA GCC CTG GCC AGG CTG CTG ATC GTC TAC CCC TGG ACC CAG
AGG TTC TTT GCG TCC TTT GGG AAC CTC TCC AGT GCC ACC GCC ATC ATT GGC AAC CCC ATG
GTC CGC GCC CAC GGC AAG AAA GTG CTG AGC TCC TTT GGG GAG GCT GTA AAG AAC CTG GAC
AAC ATC AAG AAG AGC TTT GCC CAG CTG AGC AAA CTG CAC TGT GAC AAG CTG CAC GTG GAC
CCC GAG AAC TTC AGG CTC CTG GGG GAC ATC CTG ATC ATC GTC CTG GCC TCC CAC TTC AGC
AAG GAC TTC ACT CCC GCC AGC CAG GCT GCC TGG CAG AAG ATG GTG CGT GTG GTG GCC CAC
GCG CTG GCC CAC GAG TAC CAC
 438 GLOBIN                ECHIDNA β1
*** GTG CAT TTG AGT GGT TCT GAG AAG ACT GCT GTC ACC AAC CTG TGG GGG CAT GTG AAC
GTC AAT GAA CTC GGT GGC GAG GCC CTT GGC AGG CTG CTG GTC GTC TAC CCC TGG ACC CAG
AGG TTC TTC GAA TCC TTC GGT GAC CTG TCC AGC GCC GAT GCT GTG ATG GGA AAC GCC AAG
GTC AAG GCC CAT GGT GCC AAG GTG CTG ACC TCC TTC GGC GAT GCC CTG AAG AAC CTC GAC
AAC CTC AAG GGA ACC TTC GCC AAG CTG AGC GAA CTG CAC TGC GAC AAG CTG CAC GTG GAC
CCC GAG AAT TTC AAT CGC CTG GGT AAC GTG CTG GTC GTG GTC CTG GCC CGT CAC TTC AGC
AAG GAA TTC ACC CCC GAG GCC CAG GCT GCC TGG CAG AAG CTG GTG TCT GGT GTT TCC CAC
GCC CTG GCC CAC AAG TAC CAC
 438 GLOBIN                NEW GUNIEA ECHIDNA β1
*** GTG CAT TTG AGT GGT TCT GAG AAG ACT GCT GTC ACC ACC CTG TGG GGG CAT GTG AAC
GTC AAT GAA CTC GGT GGC GAG GCC CTT GGC AGG CTG CTG GTC GTC TAC CCC TGG ACC CAG
AGG TTC TTC GAA TCC TTT GGT GAC CTG TCC AGC GCC AAT GCT GTG ATG GGA AAC CCC AAG
GTC AAG GCC CAT GGT GCC AAG GTG CTG ACC TCC TTC GGC GAT GCC CTG AAG AAC CTC GAC
GAC CTC AAG GGC ACC TTC GCC AAG CTG AGC GAA CTG CAC TGC GAC AAG CTG CAC GTG GAC
CCC GAG AAT TTC AAT CGC CTG GGT AAC GTG CTG GTC GTG GTC CTG GCC CGT CAC TTT AGC
AAG GAA TTC ACC CCC GAG GCC CAG GCT GCC TGG CAG AAG CTG GTG TCT GAT GTT TCC CAC
GCC CTG GCC GAC AAG TAC CAC
 438 GLOBIN                PLATYUS β1
*** GTG CAT CTG AGT GGT GGT GAG AAG AGT GCC GTC ACC AAC CTG TGG GGG AAA GTG AAC
ATC AAT GAG TTG GGT GGC GAA GCC CTT GGC AGG CTG CTG GTC GTC TAC CCC TGG ACC CAG
AGG TTC TTC GAA GCC TTT GGT GAC CTG TCC TCC GCC GGC GCT GTG ATG GGC AAC CCC AAG
GTC AAG GCC CAT GGT GCT AAG GTG CTG ACC TCC TTC GGC GAT GCC CTG AAG AAC CTC GAC
GAC CTC AAG GGC ACC TTT GCC AAG CTG AGC GAA CTG CAC TGC GAC AAG CTG CAC GTG GAC
CCT GAG AAT TTC AAT CGT CTG GGT AAC GTG CTG ATC GTG GTC CTG GCC CGT CAC TTC AGC
AAG GAC TTC AGC CCC GAG GTC CAG GCT GCC TGG CAG AAG CTG GTG TCT GGT GTT GCC CAC
GCC CTG GGC CAC AAG TAC CAC
 438 GLOBIN                GALAGO β
*** GTG CAC CTG ACT CCT GAT GAG AAG AAT GCC GTT TGT GCC CTG TGG GGC AAG GTG AAT
GTG GAA GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCA TGG ACC CAG
AGG TTC TTT GAC TCC TTT GGG GAC CTG TCC TCT CCT TCT GCT GTT ATG GGC AAC CCT AAA
GTG AAG GCC CAC GGC AAG AAG GTG CTG AGT GCC TTT AGC GAG GGC CTG AAT CAC CTG GAC
AAC CTC AAG GGC ACC TTT GCT AAG CTG AGT GAG CTG CAT TGT GAC AAG CTG CAC GTG GAC
CCT GAG AAC TTC AGG CTC CTG GGC AAC GTG CTG GTG GTT GTC CTG GCT CAC CAC TTT GGC
AAG GAT TTC ACC CCA CAG GTG CAG GCT GCC TAT CAG AAG GTG GTG GCT GGT GTG GCT ACT
GCC CTG GCT CAC AAA TAC CAC
 438 GLOBIN                GALAGO δ
*** GTG CAC CTG ACT CCT GAT GAG AAG AAT GCC GTT TGT GCC CTG TGG GGC AAG GTG AAT
GTG GAA GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCA TGG ACC CAG
AGG TTC TTT GAC TCC TTT GGG GAC CTG TCC TCT CCT TCT GCT GTT ATG GGC AAC CCT AAA
GTG AAG GCC CAC GGC AAG AAG GTG CTG AGT GCC TTT AGC GAG GGC CTG AAT CAC CTG GAC
AAC CTC AAG GGC ACC TTT GCT AAG CTG AGT GAG CTG CAT TGT GAC AAG CTG CAC GTG GAC
CCT GAG AAC TTC AGG CTC CTG GGC AAC GTG CTG GTG GTT GTC CTG GCT CAC CAC TTT GGC
AAG GAT TTC ACC CCG GAA GTG CAA GCT GCC TAT CAG AAG GTG GTG GCT GGT GTG GCT ACT
GCC CTG GCT CAC AAA TAT CAT
 438 GLOBIN                GALAGO ε
*** GTG CAT TTT ACT GCT GAG GAA AAG GCT ATT ATC ATG AGC CTG TGG GGA AAG GTG AAT
ATA GAA GAG GCT GGA GGA GAA GCC TTG GGC AGG CTC CTT GTT GTC TAC CCC TGG ACC CAG
AGA TTC TTT GAG ACC TTT GGA AAC CTG TCC TCT GCC TCT GCC ATC ATG GGC AAC CCC AAG
GTC AAG GCC CAT GGC AAG AAA GTG CTG ACC TCC TTT GGA GAA GCT GTC AAA AAC ATG GAC
AAC CTC AAG GGT GCC TTT GCT AAG CTG AGT GAG CTG CAC TGT GAC AAG CTG CAC GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGA AAT GTG ATG GTG ATC ATT CTG GCT ACT CAT TTT GGC
AAG GAA TTC ACT CCT GAT GTG CAG GCT GCC TGG CAG AAG CTA GTG TCT GGG GTG GCT ACT
GCT CTA GCC CAC AAG TAC CAC
```

```
 438 GLOBIN                GALAGO γ
*** GTT CAT TTT ACA GCT GAA GAG AAG GCT ATT ATC ACA AGC CTG TGG GGT AAA GTG AAT
GTA GAA GAG GAT GGA GGA GAA GCT CTG GGA AGA CTC CTG GTT GTC TAC CCA TGG ACC CAG
AGG TTC TTT GAC ACC TTT GGA AAT CTG TCC TCT GCT TCT GCC ATC ATG GGC AAC CCG AAG
GTC AAG GCC CAT GGC AAG AAG GTT CTG AGT TCC TTG GGA GAA GCT ATA AAG AAC ATG GAT
GAT CTC AAG GGT ACG TTT TCC CAT CTG AGT GAG CTG CAC TGT GAC AGG CTG CAT GTG GAT
CCT GAA AAC TTC AGG CTC CTG GGA AAT GTG CTG GTG ATT GTC TTG GCA AAG CAT TTT GGC
AAG GAA TTC ACC CCA CAG ATA CAG GCT GCC TGC CAG AAG ATG GTG GCT GGT GTG GCC ACT
GCC CTG GCC CAC AAG TAC CAC
 438 GLOBIN                HUMAN β
*** GTG CAC CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC CTG TGG GGC AAG GTG AAC
GTG GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG CTG GTG GTC TAC CCT TGG ACC CAG
AGG TTC TTT GAG TCC TTT GGG GAT CTG TCC ACT CCT GAT GCT GTT ATG GGC AAC CCT AAG
GTG AAG GCT CAT GGC AAG AAA GTG CTC GGT GCC TTT AGT GAT GGC CTG GCT CAC CTG GAC
AAC CTC AAG GGC ACC TTT GCC ACA CTG AGT GAG CTG CAC TGT GAC AAG CTG CAC GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAC GTG CTG GTC TGT GTG CTG GCC CAT CAC TTT GGC
AAA GAA TTC ACC CCA CCA GTG CAG GCT GCC TAT CAG AAA GTG GTG GCT GGT GTG GCT AAT
GCC CTG GCC CAC AAG TAT CAC
 438 GLOBIN                HUMAN δ
*** GTG CAT CTG ACT CCT GAG GAG AAG ACT GCT GTC AAT GCC CTG TGG GGC AAA GTG AAC
GTG GAT GCA GTT GGT GGT GAG GCC CTG GGC AGA TTA CTG GTG GTC TAC CCT TGG ACC CAG
AGG TTC TTT GAG TCC TTT GGG GAT CTG TCC TCT CCT GAT GCT GTT ATG GGC AAC CCT AAG
GTG AAG GCT CAT GGC AAG AAG GTG CTA GGT GCC TTT AGT GAT GGC CTG GCT CAC CTG GAC
AAC CTC AAG GGC ACT TTT TCT CAG CTG AGT GAG CTG CAC TGT GAC AAG CTG CAC GTG GAT
CCT GAG AAC TTC AGG CTC TTG GGC AAT GTG CTG GTG TGT GTG CTG GCC CGC AAC TTT GGC
AAG GAA TTC ACC CCA CAA ATG CAG GCT GCC TAT CAG AAG GTG GTG GCT GGT GTG GCT AAT
GCC CTG GCT CAC AAG TAC CAT
 438 GLOBIN                HUMAN ε
*** GTG CAT TTT ACT GCT GAG GAG AAG GCT GCC GTC ACT AGC CTG TGG AGC AAG ATG AAT
GTG GAA GAG GCT GGA GGT GAA GCC TTG GGC AGA CTC CTC GTT GTT TAC CCC TGG ACC CAG
AGA TTT TTT GAC AGC TTT GGA AAC CTG TCG TCT CCC TCT GCC ATC CTG GGC AAC CCC AAG
GTC AAG GCC CAT GGC AAG AAG GTG CTG ACT TCC TTT GGA GAT GCT ATT AAA AAC ATG GAC
AAC CTC AAG CCC GCC TTT GCT AAG CTG AGT GAG CTG CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGT AAC GTG ATG GTG ATT ATT CTG GCT ACT CAC TTT GGC
AAG GAG TTC ACC CCT GAA GTG CAG GCT GCC TGG CAG AAG CTG GTG TCT GCT GTC GCC ATT
GCC CTG GCC CAT AAG TAC CAC
 438 GLOBIN                HUMAN Aγ
*** GGT CAT TTC ACA GAG GAG GAC AAG GCT ACT ATC ACA AGC CTG TGG GGC AAG GTG AAT
GTG GAA GAT GCT GGA GGA GAA ACC CTG GGA AGG CTC CTG GTT GTC TAC CCA TGG ACC CAG
AGC TTC TTT GAC AGC TTT GGC AAC CTG TCC TCT GCC TCT GCC ATC ATG GGC AAC CCC AAA
GTC AAG GCA CAT GGC AAG AAG GTG CTG ACT TCC TTG GGA GAT GCC ATA AAG CAC CTG GAT
GAT CTC AAG GGC ACC TTT GCC CAG CTG AGT GAA CTG CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGA AAT GTG CTG GTG ACC GTT TTG GCA ATC CAT TTC GGC
AAA GAA TTC ACC CCT GAG GTG CAG GCT TCC TGG CAG AAG ATG GTG ACT GCA GTG GCC AGT
GCC CTG TCC TCC AGA TAC CAC
 438 GLOBIN                MOUSE β1
*** GTG CAC CTG ACT GAT GCT GAG AAG GCT GCT GTC TCT TGC CTG TGG GGA AAG GTG AAC
TCC GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCT TGG ACC CAG
CGG TAC TTT GAT AGC TTT GGA GAC CTA TCC TCT GCC TCT GCT ATC ATG GGT AAT GCC AAA
GTG AAG GCC CAT GGC AAG AAG GTG ATA ACT GCC TTT AAC GAT GGC CTG AAT CAC TTG GAC
AGC CTC AAG GGC ACC TTT GCC AGC CTC AGT GAG CTC CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAT ATG ATC GTG ATT GTG CTG GGC CAC CAC CTG GGC
AAG GAT TTC ACC CCC GCT GCA CAG GCT GCC TTC CAG AAG GTG GTG GCT GGA GTG GCC ACT
GCC CTG GCT CAC AAG TAC CAC
 438 GLOBIN                MOUSE β2
*** GTG CAC CTG ACT GAT GCT GAG AAG TCT GCT GTC TCT TGC CTG TGG GCA AAG GTG AAC
CCC GAT GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCT TGG ACC CAG
CGG TAC TTT GAT AGC TTT GGA GAC CTA TCC TCT GCC TCT GCT ATC ATG GGT AAT CCC AAG
GTG AAG GCC CAT GGC AAA AAG GTG ATA ACT GCC TTT AAC GAG GGC CTG AAA AAC CTG GAC
AAC CTC AAG GGC ACC TTT GCC AGC CTC AGT GAG CTC CAC TGT GAC AAG CTG CAT GTG GAT
CCT GAG AAC TTC AGG CTC CTA GGC AAT GCG ATC GTG ATT GTG CTG GGC CAC CAC CTG GGC
AAG GAT TTC ACC CCT GCT GCA CAG GCT GCC TTC CAG AAG GTG GTG GCT GGA GTG GCC ACT
GCC CTG GCT CAC AAG TAC CAC
 438 GLOBIN                MOUSE h0
*** GTT CAT TTC ACA GCT GAG GAG AAG GCA GCT ATC ACA AGC ATC TGG GAT AAA GTG GAC
TTG GAA AAA GTT GGA GGA GAA ACT CTG GGA AGG CTC CTG ATT GTT TAC CCA TGG ACT CAG
AGG TTC TTT GAC AAA TTT GGA AAC CTC TCT TCT GCC CAA GCC ATC ATG GGT AAT CCC AGA
ATC AAA GCC CAT GGC AAG AAA GTG CTG ACA TCC CTG GGC TTG GCA GTT AAG AAC ATG GAC
AAC CTC AAG GAG ACC TTT GCT CAT CTC AGT GAA CTG CAC TGT GAC AAG CTT CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGT AAC ATG CTG GTG ATT GTT CTT TCT TCT TAC TTT GGC
AAG GAA TTC ACA GCT GAG GCG CAG GCT GCC TGG CAG AAG TTG GTG GTT GGG GTG GCC ACT
GCT CTG TCC CAC AAG TAC CAC
 438 GLOBIN                MOUSE h1
*** GTT CAC TTC ACA GCT GAG GAG AAG GCA GCT ATC ACA AGC ATC TGG GAT AAA GTG GAC
TTG GAA AAA GTT GGA GGA GAA ACT CTG GGA AGG CTC CTG ATT GTT TAC CCA TGG ACT CAG
AGA TTC TTT GAC AAG TTT GGA AAC CTC TCT TCT GCC CTG GCC ATC ATG GGA AAC CCC CGG
ATT AGA GCC CAT GGC AAG AAA GTG CTG ACA TCC TTG GGC TTG GGG GTT AAG AAC ATG GAC
AAC CTC AAG GAG ACC TTT GCT CAT CTC AGT GAG CTG CAC TGT GAC AAG CTT CAT GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGC AAC ATG TTG GTG ATT GTC CTT TCT ACT CAT TTT GCC
AAG GAA TTC ACC CCA GAG GTG CAG GCT GCC TGG CAG AAG CTG GTG ATT GGA GTG GCC AAT
GCT CTG TCC CAC AAG TAC CAT
```

```
 438 GLOBIN                MOUSE ε'
*** GTG AAC TTT ACT GCT GAG GAA AAA ACC CTC ATC AAT GGC CTG TGG AGT AAG GTC AAT
GTT GAA GAG GTT GGT GGT GAA GCC TTG GGA AGG CTT CTT GTT GTG TAC CCA TGG ACC CAG
AGA TTC TTT GAC AGC TTT GGG AAC TTG TCC TCT GCC TCT GCC ATA ATG GGC AAC CCA AGG
GTC AAA GCC CAT GGC AAG AAG GTG CTG ACT GCT TTT GGA GAG TCC ATT AAG AAC CTA GAC
AAC CTC AAG TCT GCC TTG GCC AAG CTC AGT GAA CTG CAC TGT GAC AAG CTA CAT GTG GAT
CCT GAG AAC TTC AAA CTC TTG GGT AAT GTG CTG GTG ATT GTT TTG GCT AGT CAC TTC GGC
AAT GAA TTC ACA GCT GAG ATG CAG GCT GCC TGG CAG AAG CTG GTG GCT GGG GTG GCC ACT
GCG CTG TCC CAC AAG TAC CAC

 438 GLOBIN                OPOSSUM β
*** GTG CAC TTG ACT TCT GAG GAG AAG AAC TGC ATC ACT ACC ATC TGG TCT AAG GTG CAG
GTT GAC CAG ACT GGT GGT GAG GCC CTT GGC AGG ATG CTC GTT GTC TAC CCC TGG ACC ACC
AGG TTT TTT GGG AGC TTT GGT GAT CTG TCC TCT CCT GGC GCT GTC ATG TCA AAT TCT AAG
GTT CAA GCC CAT GGT GCT AAG GTG TTG ACC TCC TTC GGT GAA GCA GTC AAG CAT TTG GAC
AAC CTG AAG GGT ACT TAT GCC AAG TTG AGT GAG CTC CAC TGT GAC AAG CTG CAT GTG GAC
CCT GAG AAC TTC AAG ATG CTG GGG AAT ATC ATT GTG ATC TGC CTG GCT GAG CAC TTT GGC
AAG GAT TTT ACT CCT GAA TGT CAG GTT GCT TGG CAG AAG CTC GTG GCT GGA GTT GCC CAT
GCC CTG GCC CAC AAG TAC CAC

 438 GLOBIN                OPOSSUM ε
*** GTG CAT TTC ACC CCT GAG GAC AAG ACC AAT ATC ACT TCT GTG TGG ACC AAG GTC GAT
GTG GAA GAC GTT GGT GGG GAA TCT TTG GCC AGG CTG CTG GTA GTC TAT CCC TGG ACT CAG
AGG TTC TTT GAT TCC TTT GGC AAC CTA TCT TCT GCA TCT GCT GTT ATG GGG AAC CCC AAG
GTC AAG GCC CAT GGC AAG AAG GTG CTG ACC TCC TTT GGA GAA GGT GTC AAG AAC ATG GAC
AAC CTC AAG GGT ACC TTT GCC AAG CTG AGT GAG CTG CAC TGT GAC AAA CTG CAT GTA GAC
CCT GAG AAC TTC AGG CTC CTC GGG AAT GTG CTG ATT ATT GTT TTG GCT TCT CGT TTT GGC
AAG GAG TTC ACT CCA GAA GTG CAG GCT TCT TGG CAG AAG CTG GTG TCT GGT GTG TCC AGT
GCC TTG GGC CAC AAG TAC CAC

 438 GLOBIN                DUNNART β
*** GTG CAT CTG AGT GCT GAA GAG AAG GGC CAT ATC AAC GCC ATT TGG AGT AAG GTG TCT
GTT GAT CAG ACT GGA GCT GAG GCC CTT GGG AGG TTG CTC ATC GTC TAC CCA TGG ACC TCC
AGG TTT TTT GAC CAT TTT GGT GAC CTG TCC TCT GCC AAG GGT GTC ATG GGA AAT GCT AAG
GTC CAG GGC CAT GGT GCT AAG GTG TTG ACC TCC TTT GGT GAT GCT GTC AAG AAT ATG GAC
AAC CTG AAG GGT ACT TTT GCC AAG CTG AGT GAA CTC CAC TGT GAC AAA CTG CAT GTG GAC
CCT GAG AAC TTC CGG CTC CTG GGA AAC ATC CTT GTG ATC TGC CTG GCT GAG CAC TTT GGC
AAA GAT TTC ACC CCT GAG GTT CAG GCT GCC TGG CAG AAA CTT GTG GCT GGT GTG GCC ACT
GCC CTG GCC CAT AAG TAC CAT

 438 GLOBIN                DUNNART ε
*** GTG CAC TTC ACA GCT GAG GAG AAG AAC GCC ATC ACT ACC ATT TGG GGC AAG GTC AAT
GTG GAA GAG ACT GGT GGG GAA GCT TTA GGC AGA CTG TTG GTT GTC TAC CCC TGG ACC CAG
AGG TTC TTT GAT TCC TTT GGC AAC CTC TCT TCT GCC TCT GCT ATC CTG GGA AAC CCC AAG
GTT AAG GCC CAT GGC AAG AAG GTG CTG ACC TCC TTT GGA GAT GCT GTC AAG AAC CTG GAC
AAC CTC AAG GGT ACC TTC TCC AAG CTG AGT GAG CTC CAC TGT GAC AAG CTG CAT GTG GAC
CCT GAG AAC TTC CGG CTC CTT GGG AAT GTT CTG GTT ATT GTT ATG GCT GCT CAT TTC AAC
AAG GAA TTC ACT CCT GAA GTG CAG GCT GCT TTT CAG AAG CTG GTG ACT GGT GTG GCC AAT
GCT TTG GCT CAC AAG TAC CAC

 438 GLOBIN                RABBIT ε
*** GTG CAC TTT ACT CCT GAG GAA AAG TGC ATT ATC AGT AAG CAG TGG GGC CAG GTG AAC
ATC GAT GAG ACT GGA GGT GAA GCT TTG GGC AGG CTC CTT GTT GTC TAC CCC TGG ACC CAA
AGA TTC TTT GAC AAC TTT GGC AAC TTG TCC TCT TCC TCT GCC ATC ATG GGA AAC CCC AAG
GTC AAG GCA CAT GGC AAG AAG GTG CTG ACC TCC TTT GGA GAT GCC ATC AAG AAC ATG GAC
AAC CTT AAG GGT GCC TTT GCT AAG CTG AGT GAG CTG CAC TGT GAC AAA CTG CAT GTG GAT
CCT GAG AAC TTC AAA CTC CTG GGA AAC GTG CTG CTG ATC GTT CTG GCC ACT CAT TTT GGC
AAA GAA TTC ACT CCG GAG GTG CAG GCT GCT TGG CAG AAG CTC GTG TCT GGT GTT GCC ATT
GCT CTG GCC CAC AAG TAT CAC

 438 GLOBIN                RABBIT γ
*** GTT CAT TTC ACG GCT GAG GAA AAG GCT GCT ATC ACA AGC ACA TGG AAG CTG GTG GAC
GTA GAA GAT GCC GGA GCA GAA GCC CTC GGA AGG CTC CTG GTC GTC TAT CCT TGG ACC CAG
AGA TTC TTC GAT AGC TTT GGA AAC CTG TCC TCT TCC TCT GCC ATC ATG GGA AAC CCC AAA
GTG AAG GCC CAT GGC AAG AAG GTG TTG ACT GCC TTT GGA GAT GCT GTA AAG AAT GTG GAT
GAT CTC AAA AAC ACA TTC GCC CAC CTG AGT GAG CTG CAC TGT GAC AGG CTG CAT GTG GAT
CCC GAG AAC TTC AAA CTC TTA GGA AAT GTG CTA GTA ATT GTC TTG GCG AAG TAT TTT GGC
AAG GAA TTC ACT CCC CAG GTC CAG AGT GCC TGG CAG AAA TTG GTG GCT GGA GTG GCC ACT
GCC CTG GCC CAC AAG TAT CAC

 438 GLOBIN                RABBIT β
*** GTG CAT CTG TCC AGT GAG GAG AAG TCT GCG GTC ACT GCC CTG TGG GGC AAG GTG AAT
GTG GAA GAA GTT GGT GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCA TGG ACC CAG
AGG TTC TTC GAG TCC TTT GGG GAC CTG TCC TCT GCA CAT GCT GTT ATG AGC AAT CCT AAG
GTG AAG GCT CAT GGC AAG AAG GTG CTG GCT GCC TTC AGT GAG GGT CTG AAT CAC CTG GAC
AAC CTC AAA GGC ACC TTT GCT AAG CTG AGT GAA CTG CAC TGT GAC AAG CTG CAC GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAC GTG CTG GTT GTT GTG CTG TCT CAT CAT TTT GGC
AAA GAA TTC ACT CCT CAG GTG CAG GCT GCC TAT CAG AAG GTG GTG GCT GGT GTG GCC AAT
GCC CTG GCT CAC AAA TAC CAC

 438 GLOBIN                GOAT ε1
*** GTG CAT TTT ACT GCC GAG GAG AAG GCT GCT ATC ACT GGC CTG TGG GGC AAA GTC AAT
GTG GAA GAG GCT GGA GGC GAG GCT CTG GGC AGG CTC CTG GTT GTC TAC CCC TGG ACC CAG
AGG TTC TTT GAT AGC TTT GGC AAC CTG TCC TCT GCC TCT GCC ATA ATG GGA AAC CCC AAG
GTC AAG GCC CAC GGC AAG AAG GTG CTG ACC TCC TTT GGA GAA GCT ATT AAG AAT TTG GAC
AAC CTC AAA GGT GCC TTC GCT AAG CTG AGT GAG CTG CAC TGT GAC AAG TTG CAC GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAT GTG ATT GTG ATT ATT CTG GCT ACT CAT TTT GGC
AGA GAA TTC ACC CCT GAC GTG CAG GCT GCC TGG CAG AAG CTG GTG TCT GGT GTT GCC ACT
GCT CTG GCC CAC AAG TAC CAC
```

```
 432 GLOBIN                GOAT βF
*** *** *** CTG AGT GCT GAG GAG AAG GCT TCT GTC CTC TCC CTA TTT GCC AAG GTG AAC
GTG GAA GAA GTC GGT GGT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCC TGG ACT CAG
AGG TTC TTT GAG CAC TTT GGG GAC TTG TCC TCT GCT GAT GCC ATT TTG GGC AAC CCT AAG
GTG AAG GCC CAT GGC AAG AAG GTG TTA GAC ACC TTT AGT GAG GGC CTG AAG CAA CTT GAT
GAC CTC AAG GGT GCC TTT GCC TCA CTG AGT GAG CTG CAC TGT GAT AAG CTG CAC GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAT GTG CTG GTG GTT GTG CTG GCT CGC CGC TTT GGA
GGT GAA TTC ACC CCG GAG CTG CAG GCT AAC TTT CAG AAG GTG GTG ACT GGT GTT GCC AAT
GCC CTG GCC CAC AGA TAT CAC
 432 GLOBIN                GOAT βA
*** *** *** CTG ACT GCT GAG GAG AAG GCT GCC GTC ACC GGC TTC TGG GGC AAG GTG AAA
GTG GAT GAA GTT GGT GCT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCC TGG ACT CAG
AGG TTC TTT GAG CAC TTT GGG GAC TTG TCC TCT GCT GAT GCT GTT ATG AAC AAT GCT AAG
GTG AAG GCC CAT GGC AAG AAG GTG CTA GAC TCC TTT AGT AAC GGC ATG AAG CAT CTT GAC
GAC CTC AAG GGC ACC TTT GCT CAG CTG AGT GAG CTG CAC TGT GAT AAG CTG CAC GTG GAT
CCT GAG AAC TTC AAG CTC CTG GGC AAC GTG CTG GTG GTT GTG CTG GCT CGC CAC CAT GGC
AGT GAA TTC ACC CCG CTG CTG CAG GCT GAG TTT CAG AAG GTG GTG GCT GGT GTT GCC AAT
GCC CTG GCC CAC AGA TAT CAC
 423 GLOBIN                GOAT βC
*** *** *** CCG AAT *** *** *** AAG GCC CTA ATC ACC GGC TTC TGG AGC AAG GTG AAA
GTG GAC GAA GTT GGT GCT GAG GCC CTG GGC AGG CTG CTG GTT GTC TAC CCC TGG ACT CAG
AGG TTC TTT GAG CAC TTT GGG GAC TTG TCC TCT GCT GAT GCT GTT TTG GGC AAT GCT AAG
GTG AAC GCC CAT GGC AAG AAG GTG CTA GAC TCC TTT AGT AAC GGC GTG CAG CAT CTT GAC
GAC CTC AAG GGC ACC TTT GCT GAG CTG AGT GAG CTG CAC TGT GAT AAG CTG CAT GTG GAT
CCT GAG AAC TTC AGG CTC CTG GGC AAT GTG CTA GTG ATT GTG CTG GCT CGC CAC TTT GGC
AAG GAA TTC ACC CCG GAG CTG CAG GCT GAG TTT CAG AAG GTG GTG GCT GGT GTG GCT AGT
GCC CTG GCC CAC AGA TAT CAC
 438 GLOBIN                GOAT ε2
*** GTG CAT TTT ACT ACC GAG GAG AAG GCT GCT GTT GCT AGT CTG TGG GCC AAA GTG AAT
GTG GAG GTG GTC GGC GGT GAG AGC CTA GCA AGG CTC TTG ATC GTC TAC CCA TGG ACC CAG
AGG TTC TTT GAC AGT TTT GGT AAC TTA TGC TCT GAG TCT GCC ATA ATG GGC AAC CCC AAG
GTC AAG GCC CAC GGC AGG AAG GTG CTG AAC TCC TTT GGA AAT GCC ATT AAG CAC ATG GAT
GAT CTC AAG GGC ACC TTT GCA GAT CTA AGC GAG CTG CAC TGT GAC AAG TTG CAC GTC GAT
CCT CCG AAC TTC AGG CTC CTA GG* AA* ATG ATA TTG ATT GTC TTG GCA ACC CAC TTC AGC
AAG GAA TTT ACC CCG CAG ATG CAG GCT GCT TGG CAG AAG CT* GAC CAA CGC GTG GCT AAT
GCT CTG GCC CAC AAG TAC CAC
```

UNIVERSITY OF ADELAIDE

Molecular biology and evolution of b-globin genes in Monotremes .

Mi-Hye Lee
Department of Genetics, University of Adelaide

Thesis submitted for the Degree of Doctor of Philosophy, December 1997.

ERRATUM[1]

Page 9, line 1: Replace "subterranean or high altitude" with "sub-aqua".

Page 23, last para, second sentence: This statement is in error with regards to the names of the primates and also with regard to whether one or two γ-globin genes are present. Catarrhines consist of Old World monkeys, apes and humans, not of New World monkeys (the latter are platyrrhines). The references cited are out of date . More recent work (see following references) established that the primate γ-globin gene tandemly duplicated in the stem of the Anthropoidea (ie before the divergence of platyrrhines and catarrhines). Fitch, D.H.A., Bailey, W.J., Tagle, D.A., Goodman, M., Sieu, L. and Slightom, J. (1991). Duplication of the γ-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. *Proc. Natl. Acad. Sci.* **88**: 7396-7400. Hayasaka, K., Skinner, C.G., Goodman, M., and Slightom, J. (1993). The γ-globin genes and their flanking sequences in primates: findings with nucleotide sequences of capuchin monkey and tarsier. *Genomics* **18**: 20-28.

Page 26, para 2, first sentence: After "150 - 200 MYA", insert the following references: "(Goodman *et al.* 1987a; Koop and Goodman, 1988; Hope *et al.* 1990; Cooper and Hope, 1993; Cooper *et al.* 1996)."

Page 26, para 2, second sentence: After "marsupials and eutherians" insert the following reference: "Clemens (1989)."

Page 27, para 2, first sentence: Replace "Tachyglossus" with "TACHYGLOSSIDAE", and replace "Ornithorhynchus" with "ORNITHORHYNCHIDAE"

Page 27, Fig. 1.2.1: Alter figure caption to read: "The classification of extant mammals by Strahan (1991). (Not all eutherian orders are shown)."

Page 31, line 15: Replace "ossicles" with "ossicle".

Page 32, second para: Omit the following references: "Murray, 1984" and Kielan-Jaworowska *et al.* 1987".

Page 34, line 8: Insert "some" so that the sentence reads " ..... monotremes and some marsupials".

Page 34, line 9: Change the Gregory reference to "Gregory, 1947".

Page 37, line 24: Omit "are".

Page 38, line 4: Replace "collaborated" with "calibrated".

Page 43, line 11: Omit the words "and lower cooperatively".

Page 48, line 14: Replace "nothing" with "noting".

Page 57. Buffer 1: Replace "Tris" with "Tris/HCl pH = 7.5".

Page 65, para 2, line 6: Replace "2 volumes" with "2 ½ volumes".

Page 137, 3 lines up: Replace "based" with "basal".

Page 173, Fig. 4.2 caption, line 3: Replace the Goodman reference with Goodman *et al.* (1979).

Page 187: Insert an additional reference: "Goodman, M., Czelusniak, J., Moore, G.W., Romero-Herrera, A.E. and Matsuda, G. 1979. Fitting the gene lineage into its species lineage; a parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst. Zool.* **28**: 132-163."

Page 187, last reference: Add the reference: "Gregory, W. K. 1947. The monotremes and the palimpsest theory. *Bull. Amer. Mus. Nat. Hist.* **88**:1-52."

[1] This erratum follows errors brought to the attention of the candidate by the external examiners of the thesis.