

Atypical molecular evolution of afrotherian and xenarthran
 β -globin cluster genes with insights into the
 β -globin cluster gene organization of stem eutherians.

By

ANGELA M. SLOAN

A thesis submitted to the Faculty of Graduate Studies
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Zoology
University of Manitoba
Winnipeg, Manitoba, Canada

© Angela M. Sloan, July 2005

THE UNIVERSITY OF MANITOBA
FACULTY OF GRADUATE STUDIES

COPYRIGHT PERMISSION

Atypical molecular evolution of afrotherian and xenarthran β -globin cluster genes with insights into the β -globin cluster gene organization of stem eutherians.

BY

Angela M. Sloan

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of
Manitoba in partial fulfillment of the requirement of the degree
Of**

Master of Science

Angela M. Sloan © 2005

Permission has been granted to the Library of the University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University Microfilms Inc. to publish an abstract of this thesis/practicum.

This reproduction or copy of this thesis has been made available by authority of the copyright owner solely for the purpose of private study and research, and may only be reproduced and copied as permitted by copyright laws or with express written authorization from the copyright owner.

ABSTRACT

Our understanding of β -globin gene cluster evolution within eutherian mammals is based solely upon data collected from species in the two most derived eutherian superorders, Laurasiatheria and Euarchontoglires. Hence, nothing is known regarding the gene composition and evolution of this cluster within afrotherian (elephants, sea cows, hyraxes, armadillos, elephant shrews, tenrecs and golden moles) and xenarthran (sloths, anteaters and armadillos) mammals. To address this shortcoming, I sequenced and classified a total of 24 ' β -like' globin genes from 11 afrotherian and 2 xenarthran species, and conducted a comprehensive analysis on the molecular evolution of β -globin cluster genes within these two basal eutherian clades.

Analyses of the 'adult expressed' β - and δ -globin products suggested that the δ -globin locus of paenungulate (and presumably all afrotherian) mammals encodes the only ' β -type' chain component of their post-natal hemoglobin. This finding challenges recent views on the evolution of the eutherian β -globin cluster, and provides the first documented evidence of a dispensable β -locus within Mammalia. It appears that the δ -locus of stem afrotherians gained a functional advantage by means of a capable promoter region, donated by the neighbouring β gene via gene conversion, after which β was silenced through nonfunctionalization. Hyraxes were also found to possess a second putatively functional δ -globin gene (δ^H), which may or may not be expressed. A ML analysis of δ -globin IVS2 sequences grouped elephants and sirenians into a monophyly to the exclusion of hyraxes supporting the 'Tethytheria' hypothesis, and placed the tree hyrax as the most basal hyrax species.

Analyses of the 'embryonically expressed' ϵ -, γ - and η -globin genes revealed the presence of both γ - and ϵ -loci within members of the superorder Afrotheria, providing compelling evidence that the ancestral eutherian β -globin cluster contained at least four genes (ϵ , γ , δ , β) by the time this clade diverged from other eutherians ~107 mya. Earlier studies have revealed that hemoglobin of 5-month old elephant fetuses and manatee calves contains two ' β -type' chains. My data suggests these distinct β -chain components are encoded by the γ - and δ -loci in these species, thus providing only the second example of developmentally delayed expression of a γ locus within mammals. Sequence analyses of the armadillo β -globin cluster further indicated that the eutherian β -cluster contained five (ϵ , γ , η , δ , β) loci by the time xenarthrans diverged from eutherians ~102 mya. Remarkably, of the five armadillo loci, only the two outermost genes (ϵ and β) appeared capable of transcribing a polypeptide. Together with earlier studies that demonstrated adult armadillos express two distinct β -type chains, this finding implies that these xenarthrans may uniquely extend expression of the ϵ -locus into post-natal growth stages or, like marsupials, possess an unlinked ' β -like' gene outside their β -globin cluster.

A comparative assessment of evolution rates indicated that, with few exceptions, descendants of the ' ϵ -like' loci diverged more slowly than the ' β -like' loci in both coding and non-coding gene regions within species of all four eutherian superorders. These findings signify that rates of evolution for the paralogous genes within eutherian β -clusters are a function of selection pressure, contradicting the molecular clock theory that all genes evolve at an equal rate.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my academic advisor, Dr. Kevin Campbell, for giving me the opportunity to gain invaluable research experience, and devising a thesis project to which I am proud to attach my name. I am also very grateful for all your advice, patience, and good-natured kidding, which provided me with a pleasant work environment for the entirety of my graduate work. You took me on as a novice to the field of molecular phylogenetics and I thank you for your faith in me – it will forever be appreciated.

I also thank my other committee members, Dr. Gunnar Valdimarsson who not only donated essential lab equipment but plentiful advice concerning laboratory procedures, and Dr. Ivan Oresnik who allowed me to partake in an intriguing research project which ultimately boosted my career interest in the field of site-directed mutagenesis. I additionally appreciate the helpful feedback they both supplied throughout the course of my thesis project, which proved invaluable to the efficiency of data collection and analyses. Thanks to both for your kind and generous natures, as well as your eyes for detail!

For allowing me to work in his lab for more than two years, I extensively thank Dr. Nate Lovejoy, who was more patient and giving than I could have ever hoped for. I also thank the “Lovejoy Lab” graduate students, Kristie Lester, Stuart Willis, and Tommy Sheldon for their friendship and moral support over the years.

I must also graciously thank the providers of blood, tissue, and DNA samples. These include Andrew Taylor at the University of Pretoria in South Africa, Brenda McDonald at the University of Adelaide in Australia, Christophe Douady at Université

Claude Bernard Lyon 1 in France, Gabi Gerlach at the Marine Biological Laboratory in Woods Hole, Massachusetts, Gary Bronner at the University of Cape Town in South Africa, Galen Rathbun at the California Academy of Sciences in Cambria, California, Gordon Glover at the Assiniboine Park Zoo in Winnipeg, Manitoba, Robert Bonde at the Florida Integrated Science Centre in Gainesville, Florida, Sarita Maree at the University of Pretoria in South Africa, Terry Robinson at the University of Stellenbosch in Cape Town, South Africa, and Wendy Korver at the Bowmanville Zoo in Ontario, Canada. Without the generous donations of these skilled academics, this study could not have been the complete molecular analyses I dreamed it would be.

I could not forget my mother, Merle, and father, Gary, to whom I thank for everything they have ever given me, be it advice, support, friendship, and love. These are the best parents anyone could ask for and I owe them everything. I also thank my sister, Jennifer, for a precious friendship and an immense amount of moral support and encouragement when I needed it most. Thanks as well to Amrit Singh for all the talks and advice on lab procedures, and to Octavio Orellana for providing me with analytical equipment and helping me become a stronger and more confident individual. I love you all.

Last but not least, I would like to thank the organizations that funded my research, including the University of Manitoba Graduate Fellowship association (providing financial support to myself), and the Natural Sciences and Engineering Research Council of Canada (NSERC), The Winnipeg Foundation and the University of Manitoba Research Grants Program (URGP) (providing support to Kevin Campbell).

TABLE OF CONTENTS

	Page
Abstract	i
Acknowledgements	iii
Table of Contents	v
List of Figures	viii
List of Tables	xii
Abbreviations	xiv
 General Introduction	 1
 Chapter I	
Atypical molecular evolution of afrotherian and xenarthran 'β-like' globin genes	7
Introduction	8
Materials and Methods	13
Results	19
Discussion	38
 Chapter II	
Molecular evolution of the eutherian β-globin cluster inferred from afrotherian and xenarthran gene sequences	45
Introduction	46
Materials and Methods	49
Results	52
Discussion	69

Literature Cited	77
------------------------	----

Appendix 1

1-1	95
Primers used in the (A) amplification and (B) sequencing reactions of "β-like' globin genes.	
1-2	96
DNA sequences of the 'β-like' globin genes amplified in this study.	
1-3	125
Lengths of exons and introns for sequenced 'β-like' globin genes.	
1-4	126
Lengths of additional sequence data obtained for the 5' external region of sequenced 'β-like' globin genes, and positions of upstream features relative to the putative CAP site.	
1-5	127
Lengths of additional sequence data obtained for the 3' external region of sequenced 'β-like' globin genes, and positions of polyadenylation signals (AATAAA) relative to the termination codon.	

Appendix 2

2-1	128
Primers used in the (A) amplification and (B) sequencing reactions of "ε-like' globin genes.	
2-2	129
DNA sequences of the 'ε-like' globin genes amplified in this study.	
2-3	142
Lengths of exons and introns for 'ε-like' globin genes.	
2-4	143
Lengths of sequence data obtained for the 5' external region of 'ε-like' globin genes, and positions of upstream features relative to the putative Cap site.	

2-5	144
Lengths of sequence data obtained for the 3' external region of 'ε-like' globin genes, and positions of poly-adenylation signals (AATAAA) relative to the termination codon.	

LIST OF FIGURES

Figure	Page
1-1. The β -globin gene clusters of select mammals (see text for details; Konkel <i>et al.</i> 1979; Schon <i>et al.</i> 1981; Lingrel <i>et al.</i> 1983; Shapiro <i>et al.</i> 1983; Townes <i>et al.</i> 1984; Schimenti and Duncan 1985b; Cooper <i>et al.</i> 1996; Satoh <i>et al.</i> 1999) mapped onto a phylogram adapted from the maximum likelihood analysis (concatenated 17,736 bp data set) of Amrine-Madsen <i>et al.</i> (2003).	9
1-2. Hypothesized relationships among the three extant paenungulate orders based on 19 recent molecular phylogenetic studies.	12
1-3. Comparison of the ' β -globin' chain amino acid sequences of four afrotherian and two xenarthran species (top row in each case) with those deduced from the δ - and β -globin gene sequences determined for these species (bottom row; dots represent identity with the top sequence).	23
1-4. Dotplot comparisons (window size = 100; threshold = 80) of the putative Asian elephant δ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	24
1-5. The 5' external sequence data and features of δ - and β -globin genes amplified in this study and those of armadillo, human, tarsier and galago.	25
1-6. Dotplot comparisons (window size = 100; threshold = 80) of the putative dugong δ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	26

LIST OF FIGURES ...cont'd

Figure	Page
1-7.	27
Dotplot comparisons (window size = 100; threshold = 80) of the putative rock hyrax δ^H -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	
1-8.	28
Dotplot comparisons (window size = 100; threshold = 80) of the putative rock hyrax δ^H -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	
1-9.	30
Dotplot comparisons (window size = 100; threshold = 80) of the putative elephant shrew β -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	
1-10.	31
Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo $\psi\delta$ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	
1-11.	32
Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo β -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	
1-12.	34
Dotplot comparisons (window size = 100; threshold = 80) of the putative sloth $\psi\delta$ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	

LIST OF FIGURES ...cont'd

Figure	Page
1-13.	35
Dotplot comparisons (window size = 100; threshold = 80) of the putative sloth β -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis).	
1-14.	36
Xenarthran-rooted maximum likelihood tree constructed from homologous δ -globin IVS2 gene sequences (773 bp; see text for details) illustrating the inter-relationships among paenungulate mammals.	
2-1.	56
Dotplot comparisons (window size = 100; threshold = 80) of the putative Asian elephant γ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis).	
2-2.	57
Dotplot comparisons (window size = 100; threshold = 80) of the putative dugong γ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis).	
2-3.	59
The 5' external sequence data and features of 'ε-like' globin genes sequenced in this study, plus those of the armadillo ϵ -, $\psi\gamma$ - and $\psi\eta$ -globin genes, the human ϵ -, γ^A - and $\psi\eta$ -globin genes and the goat η -globin gene.	
2-4.	60
Dotplot comparisons (window size = 100; threshold = 80) of the putative elephant shrew ϵ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis).	

LIST OF FIGURES ...cont'd

Figure	Page
2-5.	61
Dotplot comparisons (window size = 100; threshold = 80) of the putative sloth ϵ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis).	
2-6.	62
Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo ϵ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis).	
2-7.	64
Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo $\psi\gamma$ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis).	
2-8.	65
Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo $\psi\eta$ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis).	
2-9.	70
The β -globin gene clusters of select mammals, updated using data collected for this study (see Fig. 1-1 legend for details).	

LIST OF TABLES

Table	Page
1-1. List of taxa examined in this study.	14
1-2. Genes included in the sequence analyses of 'β-like' globin genes with corresponding GenBank accession numbers.	15
1-3. List of 'β-like' globin gene products amplified in this study.	20
1-4. Percent identity matrix comparing the IVS1 and IVS2 sequences of 'β-like' globin genes sequenced in this those of human, rabbit and mouse.	22
2-1. Genes included in the sequence analyses of 'ε-like' globin genes with corresponding GenBank accession numbers.	50
2-2. List of 'ε-like' globin gene products amplified in this study.	53
2-3. Percent identity matrix comparing the IVS1 and IVS2 external sequences of 'ε-like' globin genes sequenced in this study with those of human, rabbit, mouse and goat.	54
2-4. Percent identity matrix comparing the 5' and 3' external sequences of 'ε-like' globin genes amplified in this study with those of human, rabbit, mouse and goat.	55

LIST OF TABLES ...cont'd

Table	Page
2-5.	66
Percent of synonymous (silent) and nonsynonymous (replacement) substitutions between the ' ϵ -like' and ' β -like' globin genes amplified in this study (plus those of armadillo, rabbit, mouse, goat and tarsier) and their ortholog in humans. Silent and non-silent evolution rates are also shown.	
2-6.	68
Percent of nucleotide substitutions between the IVS2 sequences of ' ϵ -like' and ' β -like' globin genes amplified in this study (plus those of armadillo, rabbit, mouse, goat and tarsier) and their ortholog in humans. The corresponding non-coding DNA mutation rates are also shown.	

ABBREVIATIONS

A	adenine
α	alpha
β	beta
bp(s)	base pair(s)
δ	delta
ddH₂O	double distilled water
dNTP	deoxynucleotide triphosphate
ε	epsilon
<i>E. coli</i>	<i>Escherichia coli</i>
EMBOSS	The European Molecular Biology Open Software Suite
η	eta
(F)	forward primer
γ	gamma
G	guanine
Hb	hemoglobin
HKY85	Haegawa-Kishino-Yano (1985) model
HS	hypersensitive site(s)
IPTG	isopropyl-β-D-thiogalactoside
IVS	intervening sequence
LB	Luria-Bertani
LCR	locus control region
MatGAT	Matrix Global Alignment Tool
MgCl₂	magnesium chloride
ML	maximum likelihood
mya	million years ago
NCBI	National Centre for Biotechnology Information
NISC	NIH (National Institute of Health) Intramural Sequencing Centre
NSERC	Natural Sciences and Engineering Research Council of Canada
ω	omega
PAUP*	Phylogenetic Analysis Using Parsimony (*and other methods)
PCR(s)	polymerase chain reaction(s)
ψ	psi (pseudo)
(R)	reverse primer
RPM	revolutions per minute
T	thymidine
T_a	annealing temperature
TAE	tris-acetate EDTA
<i>Taq</i>	<i>Thermus aquaticus</i>
TBR	tree bisection and reconnection
U	unit of enzyme activity

URGP
V
v/v
X-GAL

University of Manitoba Research Grants Program
volts
volume per volume
5-bromo-4-chloro-3-indolyl-B-D-galactoside

GENERAL INTRODUCTION

The crown-group Mammalia is defined as the most recent common ancestor of the existing mammalian clades and includes the egg-laying platypuses and echidnas (Monotremata (or Prototheria)), the pouched marsupials (Marsupialia (or Metatheria)) and the placental mammals (Eutheria) (McKenna and Bell 1997; Nowak 1999). Both the Monotremata and Marsupialia are relatively small assemblages, containing one and seven orders respectively (McKenna and Bell 1997; Nowak 1999), whereas the placental mammals are a diverse group, consisting of eighteen orders recently divided into four superorders: (1) Euarchontoglires (or Supraprimates) [rodents, lagomorphs, primates]; (2) Laurasiatheria [carnivores, cetaceans, ungulates, bats, shrews, moles]; (3) Xenarthra [armadillos, anteaters, sloths]; and (4) Afrotheria (Eizirik *et al.* 2001; Madsen *et al.* 2001; Murphy *et al.* 2001b; Lin *et al.* 2002; Springer *et al.* 2003). Afrotherians represent a clade of placentals thought to have arisen on the African continent and comprise six orders: Proboscidea (elephants), Sirenia (sea cows), Hyracoidea (hyraxes), Tubulidentata (aardvarks), Macroscelidea (elephant shrews), and Afrosoricida (golden moles and tenrecs) (Stanhope *et al.* 1998). Although they do not share a single morphological synapomorphy, the members of this group were first grouped into a monophyly by Springer (1997) and Stanhope *et al.* (1998) based on molecular data. Further support for this clade comes from many more recent comparative studies utilizing mitochondrial, nuclear and concatenated gene sequences (Liu and Miyamoto 1999; Springer *et al.* 1999; Eizirik *et al.* 2001; Murphy *et al.* 2001a,b; Lin *et al.* 2002) and from a common nine base pair deletion in exon 11 of the BRCA1 gene found only in afrotherians (Madsen *et al.* 2001). Recently, van Dijk *et al.* (2001) also provided independent support for the clade

by identifying unique amino acid substitutions in three proteins of these animals relative to other eutherians, while Nikaido *et al.* (2003) isolated and characterized a novel family of SINEs (short interspersed repetitive elements) termed “AfroSINEs”, specific to species within Afrotheria.

Molecular clock data based on gene studies indicate that eutherians have been present and evolving into present-day clades since the mid Cretaceous (~100 mya) (Eizirik *et al.* 2001; Murphy *et al.* 2001b; Springer *et al.* 2003). The deepest split is thought to have occurred either between afrotherians + xenarthrans (making them ‘sister groups’) and other placental mammals (Madsen *et al.* 2001; Delsuc *et al.* 2002; Lin *et al.* 2002) or simply between afrotherians and other placentals (Stanhope *et al.* 1998; Liu and Miyamoto 1999; Eizirik *et al.* 2001; Murphy *et al.* 2001a,b; van Dijk *et al.* 2001; Springer *et al.* 2003). Both hypotheses place this divergence at ~107 mya (Eizirik *et al.* 2001; Murphy *et al.* 2001b; Madsen *et al.* 2002; Springer *et al.* 2003), a date coinciding with a major plate tectonic separation event (Smith *et al.* 1994). At roughly this time, Africa and South America separated from the southern hemispheric super-continent Gondwana, which also incorporated Antarctica, Australia, India and Madagascar before it broke apart (Smith *et al.* 1994). The ‘Afrotheria hypothesis’ suggests that six of the twenty placental orders evolved on Africa during the period when the continent was isolated from others via plate tectonics (~107 to 40 mya) (Eizirik *et al.* 2001; Murphy *et al.* 2001b; Springer *et al.* 2003). After differentiating into various ecological niches, these mammals were able to distribute themselves throughout Europe and Asia, as Africa began to collide with these continents ~30 mya (Hedges 2001). Xenarthran mammals are thought to have evolved solely on South America (Murphy *et al.* 2001b). Laurasiatheria

and Euarchontoglires have been deemed sister taxa (constituting the clade Boreoeutheria) (Murphy *et al.* 2001b) with a northern hemispheric origin on the ancient supercontinent Laurasia (Murphy *et al.* 2001a,b). Undoubtedly, the early divergence of Afrotheria and Xenarthra from other placentals provides a framework for ancient eutherian evolution studies, while the collection of molecular data from these basal superorders could have many important implications on the early evolution of eutherian genomes.

An ideal candidate to address these intriguing possibilities is provided by the respiratory pigment hemoglobin (Hb). Hemoglobin is one of four classes of functional porphyrin-containing proteins capable of reversible oxygen-binding (Perutz 1979; Dickerson and Geis 1983). Myoglobin (supplying O₂ to mitochondria within muscle tissues), neuroglobin (supplying O₂ to the brain; Burmester *et al.* 2000) and the ubiquitous cytoglobin (whose function is unknown; Burmester *et al.* 2002; Trent and Hargrove 2002) are monomeric proteins comprising one protein chain. However, Hb (binding oxygen in the lung capillaries and facilitating its delivery to respiring cells) is a tetrameric protein comprising two α - and two β -helices (Dickerson and Geis 1983; Weber and Wells 1989; Poyart *et al.* 1992; Weber 1995). It is hypothesized that the α - and β - globin genes encoding hemoglobin arose from the duplication of a primordial globin gene roughly 600-800 mya (Antoine and Niessing 1984). This pair of globins then underwent further duplication events over time to produce the various α -like and β -like globin genes found in extant vertebrate species (Goodman and Moore 1975; Czelusniak *et al.* 1982; Proudfoot *et al.* 1982; Goodman *et al.* 1987). Both the α - and β -globin gene clusters of derived eutherians have undergone extensive scrutiny. However,

it is the β -globin cluster, comprising the genes encoding the ' β -type' polypeptide chains of hemoglobin, that is the focus of this study.

Functional β -globin cluster genes are composed of three coding regions separated by two introns (intervening sequences or IVS). Exon sequences are highly conserved in length, (normally comprising 89, 223 and 126 bp in exons 1, 2 and 3, respectively) and share strong similarity between orthologs of different species (>60%) and within paralogs of the same species (Dawson and Yamamoto 1998). These findings are most likely the consequence of purifying selection, as conserved sequences have important implications for the structure and function of the protein. The two IVS regions are variable in size, normally ranging from 122 to 130 bp in the first intron and from 850 to 914 bp in the second (Dickerson and Geis 1983). Based on sequence alignments, it is assumed that the vast amount of variability in both size and composition is the result of numerous base substitution and insertion/deletion events over time (Konkel *et al.* 1979). Exon-intron boundaries show a high degree of conservation, however, as both introns almost always possess the 5'GT – AG 3' sequence normally found at splice junctions (Nishioka *et al.* 1980). It is therefore quite possible that introns are necessary for proper transcription, especially since one mouse pseudogene, containing only exon sequences, is not expressed under normal circumstances (Dickerson and Geis 1983).

Mammalian genomes often employ regulating mechanisms to ensure that the genes within multi-gene families are expressed at appropriate times, including proximal promoters and distal regulatory elements working in conjunction with environmental cues (Fritsch *et al.* 1980; Hill *et al.* 1984). The locus control region (LCR) is a large segment of DNA found in mammals whose activity has extreme effects on the expression of genes

within a multigene family (Tuan *et al.* 1985; Forrester *et al.* 1987; Grosveld *et al.* 1987). The β -globin LCR is roughly 12 kb in length, located from ~6 to 20 kb upstream region of the cluster, and is divided into three sites (200-300 bp long) hypersensitive (HS) to DNase I digestion (Tuan *et al.* 1985; Forrester *et al.* 1987; Grosveld *et al.* 1987). These sites act as recognition sequences for DNA-binding proteins, which directly control the chromatin structure of the 60 kb functional β -globin domain. The mechanisms by which these proteins regulate chromatin structure, however, are currently unknown. Experiments in which the LCR of humans was fully deleted resulted in failure to express any of the β -globins, even though the genes themselves were structurally intact (Drisocoll *et al.* 1989; Feng *et al.* 2005). Interestingly however, the presence of only one HS site within the LCR is sufficient for full gene expression in transgenic mice (Ryan *et al.* 1989). The results of these studies confirm the utmost importance of the LCR in the expression of each globin gene within the cluster. Certain sequences of non-coding DNA, located both upstream and downstream of the 5' (CACCC, CCAAT, ATA) and 3' (AATAAA) flanking end of each globin gene, respectively, also play important roles in the control of both transcription and translation (Proudfoot and Brownlee 1976; Efstratiadis *et al.* 1980; Grosveld *et al.* 1982; Dierks *et al.* 1983; Kosche *et al.* 1985; Myers *et al.* 1986), and are discussed in detail within the upcoming chapters.

Surprisingly, our knowledge of eutherian β -globin clusters has been deduced from data collected on only the two most derived orders, Laurasiatheria and Euarchontoglires, and therefore, the goal of this study was to obtain genetic information on the β -clusters of the two most basal clades, Afrotheria and Xenarthra. This information should provide much needed insight into the β -cluster gene organization of

stem eutherians and ultimately lead to a better understanding of molecular evolution of the β -cluster within the whole of Eutheria.

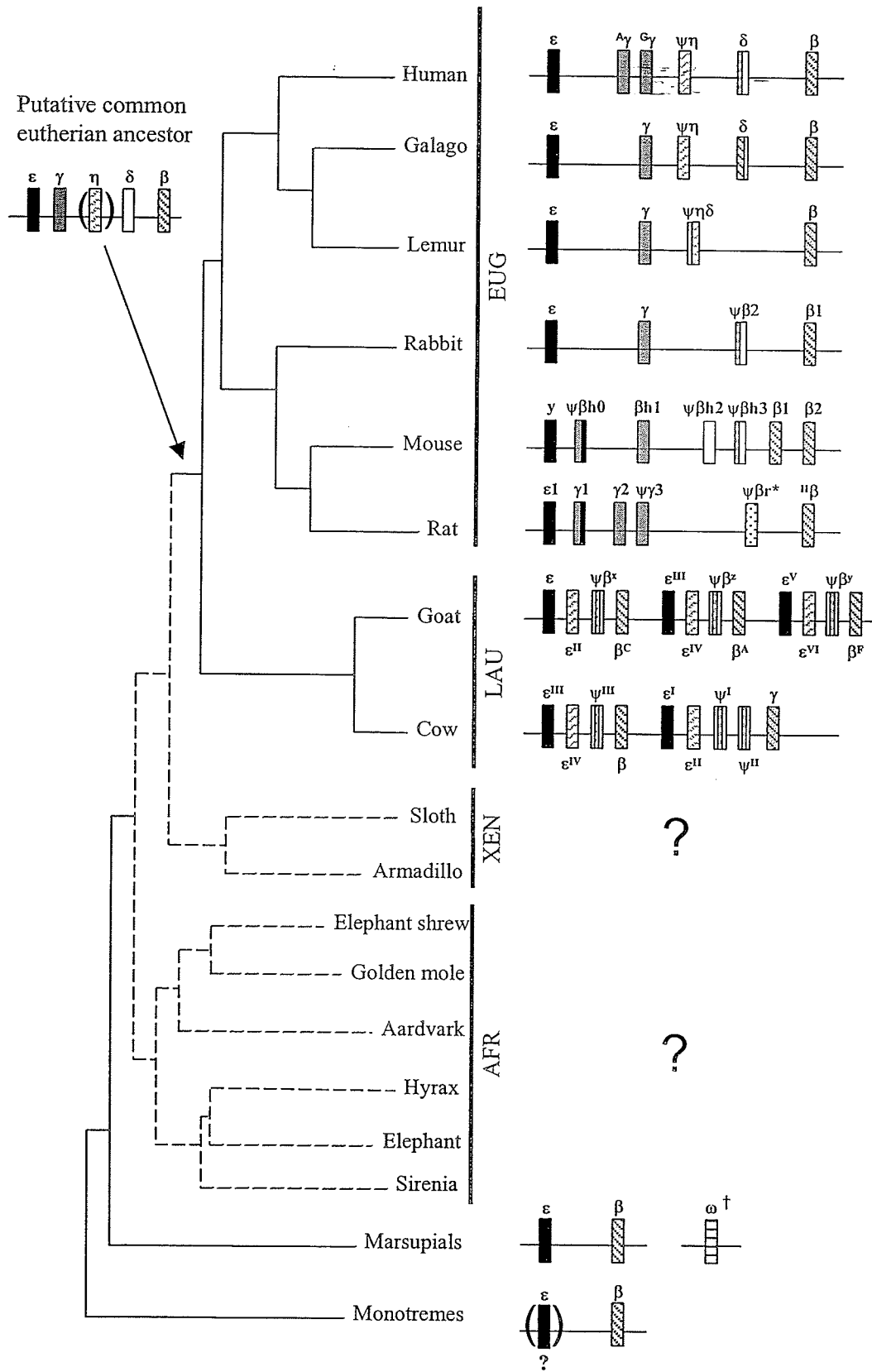
CHAPTER I

ATYPICAL MOLECULAR EVOLUTION OF AFROTHERIAN AND XENARTHROAN 'β-LIKE' GLOBIN GENES.

INTRODUCTION

The mammalian β -globin gene cluster is thought to have originated from the tandem duplication of a single proto- β -globin gene about 200 mya (Efstratiadis *et al.* 1980; Czelusniak *et al.* 1982). Differential selection on the upstream regulatory elements of these paralogous loci subsequently confined expression of the 5' ϵ -globin gene to early embryonic stages, while the 3' β -globin gene became developmentally suppressed (Efstratiadis *et al.* 1980; Czelusniak *et al.* 1982). This gene cluster organization and expression pattern is still found in marsupials (Koop and Goodman 1988; Cooper and Hope 1993; Cooper *et al.* 1996; but see Wheeler *et al.* 2001), and presumably monotremes (Lee *et al.* 1999). However, as indicated by the four-to-twelve gene clusters of rodents, lagomorphs, artiodactyls and primates (Fig. 1-1), the composition and regulation of the ancestral mammalian β -cluster has been substantively altered over the eutherian radiation (Efstratiadis *et al.* 1980; Jeffreys *et al.* 1982; Martin *et al.* 1983; Goodman *et al.* 1984; Hardies *et al.* 1984; Scott *et al.* 1984; Townes *et al.* 1984; Schimenti and Duncan 1985a,b; Hardison and Gelinas 1986; Tagle *et al.* 1988). Despite their compositional differences, remarkably congruent (though independent) gene conversion and inactivation events are seen within the β -clusters of these four orders, leading to the development of several widely accepted paradigms pertaining to the evolution of this cluster (Jeffreys *et al.* 1982; Goodman *et al.* 1984; Hardison 1984; Hardison and Margot 1984; Harris *et al.* 1984; Tagle *et al.* 1991; Satoh *et al.* 1999; Koop *et al.* 1989a; Prychitko *et al.* 2005).

For instance, the presence of a δ -globin locus in each of the four examined placental orders (Rodentia, Lagomorpha, Artiodactyla, Primates; Fig. 1-1) has led to the

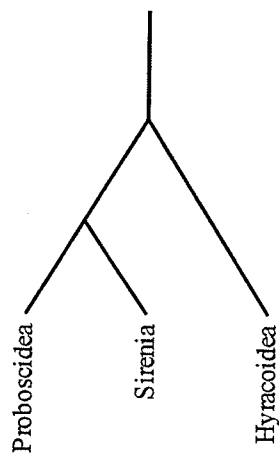


general acceptance that this gene arose via duplication of the adult-expressed β -locus (Efstratiadis *et al.* 1980; Czelusniak *et al.* 1982) soon after the divergence of Eutheria from Marsupialia (148-159 mya; Archibald 2003). Since this event, the orthologous δ genes from each of these four orders were converted (in some cases repeatedly) by the β -locus, with β preferentially converting δ in its upstream external and 5' exonic regions (Jeffreys *et al.* 1982; Martin *et al.* 1983; Hardison and Margot 1984; Hardies *et al.* 1984). The δ gene thus proved to be largely expendable, with β -globin becoming the major (and generally only) functional adult gene in the cluster (Martin *et al.* 1983; Hardies *et al.* 1984; Koop *et al.* 1989a; Prychitko *et al.* 2005). Indeed, δ -globin is only known to be expressed in New World monkeys, hominoids, tarsiers and galagos (Martin *et al.* 1983; Hardies *et al.* 1984; Prychitko *et al.* 2005) and, with the exception of the latter two groups (Tagle *et al.* 1988; Koop *et al.* 1989a), encodes for only a small fraction of the hemoglobin molecules of post-natal erythrocytes (Boyer *et al.* 1971; Martin *et al.* 1983; Ross and Pizarro 1983). The distribution of expressed δ -globin genes among primates suggests these functional loci arose (possibly via the resurrection of an anciently silenced δ -globin gene; Martin *et al.* 1983) independently in stem anthropoids, tarsiiformes and loriformes via recombination with the 5' β -globin promoter region necessary for transcription (Koop *et al.* 1989a; Tagle *et al.* 1991; Prychitko *et al.* 2005). The δ -locus was subsequently silenced in the lineage leading to Old World monkeys (Martin *et al.* 1983), or hybridized by an unequal crossover event with the $\psi\eta$ -locus in the ancestry of lemurs (Jeffreys *et al.* 1982). Accordingly, δ -globins have been little more than vestigial genes since their inception, except where their integrity has been maintained through sporadic, fortuitous events of gene conversion (Hardies *et al.* 1984).

Recent molecular studies have provided compelling support for the presence of four placental 'superorders' (Madsen *et al.* 2001; Murphy *et al.* 2001a,b; Amrine-Madsen *et al.* 2003). Because these (and other) studies place the superorders Xenarthra and Afrotheria at the base of the eutherian clade (Fig. 1-1), members of these two groups provide ideal model systems to test hypotheses regarding the molecular evolution of the β -globin gene cluster. While protein sequences have been documented for the ' β -type' chains of several afrotherian and xenarthran species, prior to the initiation of this study no nucleotide data was available. Consequently, the identity (i.e. β or δ) and evolutionary history of the genes encoding these polypeptide chains have not been established. To fill this void, the primary aim of this study was to determine whether the β/δ duplication event predates the eutherian radiation. If so, a second aim was to explicate the evolutionary events leading to the present-day post-natal β -like globin genes of both superorders to test whether the β -globin locus exhibited a competitive advantage over the 5' δ -locus in members of these basal clades.

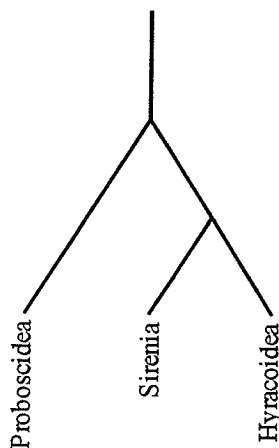
The final goal of this study was to examine the phylogenetic relationships among the afrotherian orders Proboscidea, Hyracoidea and Sirenia. Since their grouping into the superorder Paenungulata (Simpson 1945), the inter-relationships among these orders have remained controversial. For example, studies examining morphological characters consistently place sirenians and elephants as sister taxa ('Tethytheria hypothesis'; McKenna 1975; Shoshani and McKenna 1998; Liu and Miyamoto 1999; Liu *et al.* 2001), while molecular analyses recover only limited support for this relationship (Fig. 1-2). To help resolve this issue, phylogenetic relationships amongst all seven extant paenungulate genera were examined using non-coding, homologous gene sequences.

Fig. 1-2. Hypothesized relationships among the three extant paeunungulate orders based on 19 recent molecular phylogenetic studies. Letters below each hypothesis correspond to studies providing support for the above topology.



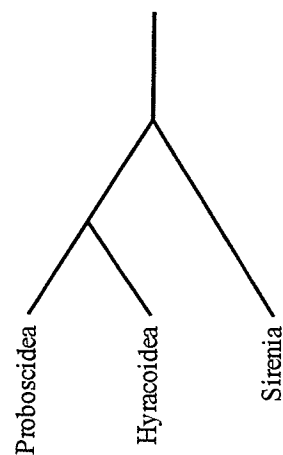
A

Lavergne *et al.* 1996
 Eizirik *et al.* 2001
 Murphy *et al.* 2001a
 Madsen *et al.* 2002
 Murata *et al.* 2003



B

Stanhope *et al.* 1998
 Liu and Miyamoto 1999
 Amrine and Springer 1999
 Murphy *et al.* 2001b
 Delsuc *et al.* 2002
 Douady *et al.* 2002
 Malia *et al.* 2002
 Douady *et al.* 2003
 Springer *et al.* 2003
 Waddell and Shelley 2003



C

Springer *et al.* 1997
 Madsen *et al.* 2001
 Amrine-Madsen *et al.* 2003
 Fleming *et al.* 2003

MATERIALS AND METHODS

Data collection

DNA was extracted from whole blood (100 μ l) and tissue (10-25 mg) samples from 11 afrotherian and two xenarthran species (Table 1-1) using a QIAGEN DNeasy[®] Tissue Kit. The optical density of eluates was measured with an Ultrospec 3100 *pro* spectrophotometer (Biochrom), the concentration and purity of the DNA was calculated from absorbance readings at 260 and 280 nm, and 50 ng/ μ l working solutions prepared.

To design primers for polymerase chain reactions (PCRs), degenerate nucleotide sequences were first deduced from the ' β -globin' polypeptide chains of the African (*Loxodonta africana*) and Asian elephant (*Elephas maximus*), Amazon manatee (*Trichechus inunguis*), rock hyrax (*Procavia capensis*), nine banded armadillo (*Dasypus novemcinctus*) and three-toed sloth (*Bradypus tridactylus*) (see Fig. 1-3 for references). Resulting sequences were aligned with primate, rodent, lagomorph and artiodactyls β -globin gene sequences (see Table 1-2), and highly conserved regions used for primer design (Primer Premier 5.0; Premier Biosoft International). PCRs were conducted on 100 ng of template DNA (2 μ l) in 0.2 ml tubes containing 48 μ l of reaction mixture (1.0 μ l of each dNTP (2.5 mM), 5.0 μ l 10x PCR Reaction Buffer, 3.0–4.5 μ l of 50 mM MgCl₂, 5.0 μ l of each primer (10.0 pmol/ μ l), 0.5 μ l Taq polymerase (5 U/ μ l) and 24.0–25.5 μ l ddH₂O), using PTC-0220 DNA Engine Dyad[®] (MJ Research) and Mastercycler[®] Gradient (Eppendorf) thermocyclers. Initial amplifications entailed a 30 cycle protocol (94°C for 30 s; 50°C for 15 s; 72°C for 30–75 s) followed by a 10 min interval at 72°C. PCR products were isolated using Montage[™] PCR Centrifugal Filter Devices

Table 1-1. List of taxa examined in this study.

Superordinal clade	Family	Genus	Species	Common name	Sample type	Original source	Gender
Afrotheria	Elephantidae	<i>Loxodonta</i>	<i>africana</i>	African elephant	blood	not specified	female
		<i>Elephas</i>	<i>maximus</i>	Asian elephant	blood	not specified	male
	Dugongidae	<i>Dugong</i>	<i>dugon</i>	dugong	DNA	Mabuiag Island, Torres Straits	female
	Trichechidae	<i>Trichechus</i>	<i>manatus</i>	West Indian manatee	skin	Crystal River, Florida	female
	Procaviidae	<i>Procavia</i>	<i>capensis</i>	rock hyrax	blood	not specified	not specified
		<i>Heterohyrax</i>	<i>brucei</i>	yellow-spotted hyrax	ear	not specified	not specified
		<i>Dendrohyrax</i>	<i>dorsalis</i>	western tree hyrax	DNA	not specified	not specified
	Orycteropodidae	<i>Orycteropus</i>	<i>afer</i>	aardvark	liver	Karoo, South Africa	female
	Macroscelididae	<i>Elephantulus</i>	<i>intufi</i>	bushveld elephant shrew	liver	not specified	not specified
	Chrysochloridae	<i>Chrysochloris</i>	<i>asiatica</i>	cape golden mole	spleen	Cape Town, South Africa	male
<i>Amblysomus</i>		<i>hottentotus</i>	Hottentot golden mole	spleen	Palm Springs, South Africa	male	
Xenarthra	Bradypodidae	<i>Bradypus</i>	<i>tridactylus</i>	pale-throated three-toed sloth	not specified	not specified	not specified
	Dasypodidae	<i>Dasypus</i>	<i>novemcinctus</i>	nine-banded armadillo	not specified	not specified	not specified

Table 1-2. Genes included in the sequence analyses of 'β-like' globin genes with corresponding GenBank accession numbers. Hyrax-specific-δ-globin genes (δ^H) are denoted by asterisks (*).

Eutherian superorder	Species	Common Name	Globin gene	Accession No.	Reference
Afrotheria	<i>Loxodonta africana</i>	African elephant	δ	DQ091201	This study
	<i>Elephas maximus</i>	Asian elephant	δ	DQ091202	This study
	<i>Trichechus manatus</i>	West Indian manatee	δ	DQ091203	This study
	<i>Dugong dugon</i>	dugong	δ	DQ091204	This study
	<i>Procavia capensis</i>	rock hyrax	δ δ ^H *	DQ091208 DQ091205	This study
	<i>Heterohyrax brucei</i>	yellow-spotted hyrax	δ ^H *	DQ091206	This study
	<i>Dendrohyrax dorsalis</i>	western tree hyrax	δ ^H *	DQ091207	This study
	<i>Elephantulus intufi</i>	bushveld elephant shrew	δ β	DQ091211 DQ091212	This study
Xenarthra	<i>Dasypus novemcinctus</i>	nine-banded armadillo	ψδ β	DQ091209 DQ091213	This study
	<i>Bradypus tridactylus</i>	pale-throated three-toed sloth	ψδ β	DQ091210 DQ091214	This study
	<i>Dasypus novemcinctus</i>	nine-banded armadillo	ψδ, β	AC151518	Green 2004, unpublished
Euarchotoglires	<i>Homo sapiens</i>	human	δ, β	U01317	Efstratiadis <i>et al.</i> 1980
	<i>Mus musculus</i>	mouse	βh2, βh3, β1,β2	X14061	Shehee <i>et al.</i> 1989
	<i>Oryctolagus cuniculus</i>	rabbit	ψβ2, β1	M18818	Margot <i>et al.</i> 1989
	<i>Otolemur crassicaudatus</i>	galago	δ, β	U60902	Tagle <i>et al.</i> 1992
	<i>Tarsius syrichta</i>	tarsier	δ β	J04428 J04429	Koop <i>et al.</i> 1989a
Laurasiatheria	<i>Capra hircus</i>	goat	β ^A	M15387	Schon <i>et al.</i> 1981

(Millipore), and small samples of purified template (0.8 μ l) pipetted into 19.2 μ l aliquots of reaction mixture for subsequent amplification experiments. To increase template-primer binding specificity, a nested PCR protocol utilizing an annealing temperature (T_a) gradient of 50–60°C was employed. Nucleotide sequences of the resulting products were used to design species-specific primers to amplify the flanking regions of each gene via a walking reaction (APAgeneTM Genome Walking Kit; Bio S&T Inc.).

A 15 μ l sample of each final PCR product was mixed with 3.5 μ l loading buffer and 1.85 μ l SYBR Green (Fisher Scientific), pipetted into gels prepared from 0.75 g of low melting temperature agarose dissolved in 50 ml 1x TAE buffer, and electrophoresed at 100 V for 1 hr. Target bands were excised, and purified using a MinEluteTM Gel Extraction Kit (QIAGEN). These extracts were ligated using a QIAGEN[®] PCR Cloning^{plus} Kit and transformations performed. Transformation products were incubated on agar plates containing 0.1% v/v ampicillin (100 mg/ml), X-GAL (40 mg/ml) and IPTG (100 mM) for 18 hr at 37°C. Positive colonies were transferred to culture medium containing 0.1% v/v ampicillin (100 mg/ml), and incubated (37°C for 18 hr) in 50 ml LB culture tubes placed on a platform shaker set to 225 RPM. Finally, a 3 ml subsample of this bacterial culture was pelleted by centrifugation (4 min at 5,000 RPM) and plasmid DNA extracted (QIAprep[®] Spin Miniprep Kit; QIAGEN).

To verify that the plasmids contained target bands, 2.0 μ l (350–400 ng) of purified plasmid DNA was digested in a restriction enzyme mixture (1.5 μ l 10x REact[®] 3 Buffer, 0.5 μ l *Eco*R1 (10 U/ μ l) and 11.0 μ l ddH₂O) for 1-2 hr at 37°C, then electrophoresed (80 V; 1 hr) on a 1% agarose gel. Positive samples were sequenced in both directions with a 3739 ABI PRISM Genetic Analyzer (University of Calgary DNA

core laboratory) and BigDye Sequencing Kit using the universal sequencing primers, M-13(F)-40 and M-13(R) (Appendix 1-1).

Data analyses

Consensus gene alignments were constructed for each species using SequencherTM (Version 4.2.2) software. Overlapping gene fragments incorporated between 3 and 14 cloned fragments amplified from 2 to 7 individual PCRs, respectively. Dotmatcher (EMBOSS) software was used to identify homology, and MatGAT (Version 2.02) employed to calculate percent identity, between sequences of the 'β-like' genes amplified in this study and the δ- and β-globin genes of human, rabbit and mouse (see Table 1-2 for GenBank accession numbers). For comparative purposes, the δ- and β-globin gene flanking regions of the nine-banded armadillo (*Dasypus novemcinctus*) were first identified in a working draft BAC sequence available on GenBank (accession number AC151518, representing clone VMRC5-69F10 from the NISC Comparative Vertebrate Sequencing Project) and included in these analyses. To assess promoter regions, the 5' flanking sequences of sequenced genes were aligned and compared with the 'β-like' globin genes of armadillo, human, galago and tarsier (Table 1-2).

To examine the inter-relationships amongst paenungulates, a phylogenetic analysis was conducted on intron 2 (IVS2) sequences of the δ-globin genes amplified from these species, with homologous xenarthran sequences serving as outgroups. A 773 bp alignment was created using Clustal X (Version 1.8) (Thompson *et. al* 1994) using gap opening and extension penalty values of 16 and 6.66, respectively, and refined by eye. A ML analysis was conducted on the alignment using PAUP* (Version 4.0b10)

software, applying the HKY85 two-parameter model variant for unequal base frequencies. All positions were weighted equally and the transition:transversion ratio set to 2.00. A heuristic search was executed with 1000 bootstrap replicates. The final tree was constructed using random stepwise addition with weighted least-squares and TBR algorithms.

RESULTS

A total of 10 complete and 3 incomplete ' β -like' gene products were amplified and sequenced from the 11 afrotherian species (Table 1-3). 'Complete' products possessed gene regions running from at least the initiation to termination codons, while 'incomplete' products were missing the gene region containing either one or both signals. All putative gene products possessed the three exon/two intron arrangement (Appendix 1-3) characteristic of other β -globin cluster genes. For genes with no frameshift mutations, IVS1 divided codon 30 between the second and third nucleotide, while IVS2 separated codons 104 and 105. Comparative sequence examinations revealed that both gene products obtained from the armadillo and the single product from each of the golden and Hottentot moles were artificial ' δ/β -globin' chimeras, with the conceptual point of conversion occurring at exactly the same base position within each "gene"; these PCR artifacts were thus excluded from all subsequent analyses. MatGAT and dotplot analyses (see below) suggested that all but one gene product from each of the remaining afrotherian species shared the highest degree of sequence homology with the δ -globin genes of human, rabbit and mouse (Table 1-2). Accordingly, these genes were classified as δ -globins. Conversely, one of the amplified gene fragments from the elephant shrew was shown to share highest sequence similarity with other placental mammal β -globin sequences (Fig.1-9; Table 1-4) and classified as β -globin. Complete δ - and β -globin gene sequences were obtained from both xenarthran species included in this study.

Classification of afrotherian ' β -like' globin genes

Dotplot analyses of the Asian (Fig. 1-4) and African elephant (data not shown) genes (which shared 99.2% sequence identity from initiation to termination codons)

Table 1-3. List of ' β -like' globin gene products amplified in this study.

Taxon	Gene identity	Number of base pairs sequenced	Amplified gene region
African elephant	δ	1,744	complete
Asian elephant	δ	2,119	complete
dugong	δ	2,639	complete
West Indian manatee	δ	1,696	complete
rock hyrax	δ	1,180	partial
	δ^H	1,609	complete
yellow-spotted hyrax	δ^H	1,604	complete
western tree hyrax	δ^H	1,591	complete
aardvark	δ/β	1,177	complete, artifact
	δ/β	1,192	complete, artifact
bushveld elephant shrew	β	1,840	partial
	δ	180	partial
cape golden mole	δ/β	1,713	complete, artifact
Hottentot golden mole	δ or β ?	426	partial
pale-throated sloth	β	1,318	complete
	$\psi\delta$	1,590	complete
nine-banded armadillo	β	1,383	complete
	$\psi\delta$	1,296	complete

recovered the highest sequence similarity with δ -globin genes of human and rabbit in the IVS2 and the 3' flanking regions, while 5' flanking regions revealed homology with other β -globins. Examination of the Asian elephant 5' promoter region further revealed the presence of all transcriptional control motifs characteristic of functional β -globin genes (Fig. 1-5; Appendix 1-4). The translated protein chains of the African and Asian elephant δ -globin genes were identical to the ' β -globin' protein chain determined for these species (Fig. 1-3), suggesting these genes are expressed.

Dotplot analyses of the dugong (Fig. 1-6) and manatee (data not shown) gene products (which were 96.2% identical from initiation to termination codons) revealed homology with other mammalian δ -globin genes in the IVS2 and 3' flanking regions. Like the elephant gene products, the 5' flanking regions of both sirenian δ -globin genes were β -like. However, both species possessed ATA $\rightarrow$ GTA mutations at the ATA transcriptional control motif (Fig. 1-5, Appendix 1-4). While these base substitutions may have an effect on the transcriptional regulation of these genes, the conceptual protein chain of the Florida manatee δ -globin gene precisely matched the amino acid sequence of the ' β -globin' protein chain determined from the Amazon manatee, *Trichechus inunguis* (Fig. 1-3).

Dotplot comparisons of the first rock hyrax gene fragment (Fig. 1-7) also suggested this gene was δ -like within the IVS2 and 3' flanking region. This gene fragment (designated δ -globin) encoded for a protein identical to codons 82-146 of the ' β -globin' protein chain documented for the Abyssinian hyrax (*Procavia capensis habessinica*) (Fig. 1-3). Notably, the second rock hyrax gene product was also δ -like in the IVS2 and 3' flanking region (Fig. 1-8). However, the translated protein sequence was

Table 1-4. Percent identity matrix comparing the IVS1 and IVS2 sequences of ' β -like' globin genes sequenced in this study and those of human, rabbit and mouse. Gap opening and extension penalties were set to 16 and 4, respectively. Numbers in gray represent IVS1 comparisons while values in black denote IVS2 comparisons.

	1 (δ)	2 (δ)	3 (δ)	4 (δ)	5 (δ)	6 (δ)	7 (δ)	8 (δ)	9 ($\psi\delta$)	10 ($\psi\delta$)	11 (β)	12 (β)	13 (β)	14 (δ)	15 (δ)	16 (δ)	17 (δ)	18 (β)	19 (β)	20 (β)	21 (β)
1. African elephant δ			84.0	84.5	73.4	75.9	76.3	77.2	67.6	65.6	48.7	49.5	47.8	54.9	61.5	54.4	56.4	47.6	46.4	48.6	48.4
2. Asian elephant δ	99.2		83.8	84.2	73.4	75.9	76.3	77.2	67.2	65.3	49.1	48.2	47.3	54.9	61.2	54.0	56.0	48.0	46.8	48.2	48.2
3. dugong δ	92.3	91.5		97.4	77.0	76.0	76.8	77.8	67.1	65.8	46.9	50.4	48.5	56.8	61.7	54.4	56.6	48.5	46.2	48.2	47.9
4. manatee δ	91.5	90.8	96.9		77.2	76.3	77.2	78.2	69.0	66.6	47.0	49.6	48.0	57.6	62.1	54.9	57.0	49.3	46.8	47.1	47.5
5. rock hyrax δ	--	--	--	--		72.3	73.2	74.4	64.0	60.3	45.8	49.6	47.7	52.4	58.5	52.3	53.8	48.2	47.7	47.1	47.2
6. rock hyrax δ^H	75.4	74.6	77.6	78.4	--		97.7	95.5	63.6	62.5	47.2	49.1	47.9	52.6	57.5	52.6	55.5	48.7	47.8	47.9	48.5
7. yellow-spotted hyrax δ^H	78.5	77.7	80.0	80.0	--	95.5		95.9	64.2	62.5	48.1	49.8	48.7	53.2	57.7	53.7	56.7	48.7	47.4	48.6	48.0
8. tree hyrax δ^H	77.7	76.9	80.0	80.8	--	93.3	94.6		65.1	62.3	48.8	50.3	48.5	53.2	58.0	52.9	56.0	47.5	47.5	48.0	48.7
9. sloth $\psi\delta$	79.2	78.5	81.5	81.5	--	68.7	68.5	70.0		73.0	47.6	50.0	49.0	55.5	57.3	52.9	54.7	46.8	47.3	49.0	50.8
10. armadillo $\psi\delta$	74.8	74	77.9	76.3	--	67.4	68.7	68.4	80.0		49.3	49.4	49.5	53.1	56.1	54.2	54.6	47.4	48.5	48.2	49.3
11. elephant shrew β	71.8	71.0	74.0	74.0	--	63.4	64.6	63.8	70.0	65.4		55.9	56.5	41.6	48.2	48.8	50.0	47.3	53.2	47.6	51.4
12. armadillo β	76.3	75.6	77.9	79.4	--	65.2	65.6	65.6	80.8	87.4	63.4		70.8	42.2	48.0	47.5	49.7	48.6	58.4	51.5	53.4
13. sloth β	79.2	78.5	81.5	81.6	--	68.7	68.5	70.0	100.0	80.0	70.0	80.8		43	45.9	45.8	49.2	46.6	60.5	51.5	51.8
14. human δ	68.9	68.9	69.9	68.7	--	63.8	64.2	62.3	70.5	71.8	63.4	66.9	70.5		58.2	52.4	51.3	49.4	41.0	43.0	44.5
15. rabbit $\psi\beta 2$ (δ -like)	54.2	54.2	55.3	56.1	--	51.1	50.4	49.6	61.5	59.8	48.8	58.3	61.6	56.2		55.6	56.3	48.7	46.9	47.4	47.7
16. mouse $\beta H 2$ (δ -like)	43.4	43.4	42.3	43.1	--	42.1	41.1	43.8	45.7	44.1	42.9	44.9	45.7	43.5	44.7		59.0	47.7	46.4	48.0	48.7
17. mouse $\beta H 3$ (δ -like)	51.9	51.9	46.9	47.7	--	45.1	47.3	42.6	49.6	48.8	43.8	52.4	49.6	48.5	56.5	53.8		45.3	46.5	49.0	47.5
18. human β	68.9	68.9	69.9	68.7	--	63.8	64.2	62.3	70.5	71.8	63.4	66.9	70.5	100.0	56.2	43.5	48.5		47.2	48.6	47.3
19. rabbit $\beta 1$	64.1	63.4	61.1	61.4	--	54.3	54.8	57.1	65.4	69.0	56.0	66.4	65.4	70.0	57.1	43.7	47.2	70.0		49.5	52.5
20. mouse $\beta 1$	62.6	61.8	62.9	62.1	--	57.1	57.4	58.1	65.9	62.2	55.0	58.7	65.9	63.1	52.1	48.3	53.4	63.1	59.2		64.6
21. mouse $\beta 2$	55.4	54.6	54.6	53.1	--	51.5	52.3	50.4	59.7	54.3	51.2	53.5	59.7	55.7	45.5	53.8	47.0	55.7	51.9	85.5	

Fig. 1-3. Comparison of the 'β-globin' chain amino acid sequences of four afrotherian and two xenarthran species (top row in each case) with those deduced from the δ- and β-globin gene sequences determined for these species (bottom row; dots represent identity with the top sequence). *Florida manatee (*T. manatus*); †rock hyrax (*P. capensis*); §conceptual amino acid sequence of the armadillo β-globin gene (GenBank accession number AC151518). The degenerate amino acid symbol B corresponds to D or N, while Z represents E or Q.

African elephant (*Loxodonta africana*); Braunitzer *et al.* (1982)

VNLTAAEKTQVTNLWGKVNKELGGEALSR LLVVYPWTRRFFEHFGDLSTAEAVLHNAKVLAGEKVLTSFGGEGLKHLNLTGTFADLSELHCDKLHVDPENFR LLGNVLVIVLARHFGKEFTPDVQAAAYEKVVAGVANALAHKYH (β-chain)
 (δ-globin)

Asian elephant (*Elephas maximus*); Braunitzer *et al.* (1984)

VNLTAAEKTQVTNLWGKVNKELGGEALSR LLVVYPWTRRFFEHFGDLSTADAVLHNAKVLAGEKVLTSFGGEGLKHLNLTGTFADLSELHCDKLHVDPENFR LLGNVLVIVLARHFGKEFTPDVQAAAYEKVVAGVANALAHKYH (β-chain)
 (δ-globin)

Amazon manatee (*Trichechus inunguis*); Kleinschmidt *et al.* (1986)

VHLPPEEKALVIGLWAKVNVKEYGGEALGR LLVVYPWTQRFFEHFGDLSSASAIMNPKVKAHGEKVLTSFGGGLKHLEDLKGAFAELSELHCDKLHVDPENFR LLGNVLVIVLARHFGKEFTPDVQAAAYQKVVAGVANALAHKYH (β-chain)
 * (δ-globin)

Rock (Abyssinian) hyrax (*Procavia capensis habessinica*); Kleinschmidt and Braunitzer (1983)

VHLTDAEKA AVTGLWGKVKVDEYGEALGR LLVVYPWTQRFFEHFGDLSDADAIMHNPVKLAHGKVLTSFGGGLNHLNLTGTFQQLSELHCDKLHVDPENFR LLGNVLVIVLARHFGKEFTPDVQAAAFQKVVAGVANALAHKYH (β-chain)
 † (δ-globin)
 † .R...S...E..S..S..DEKII.S.... ..I....R.....T...LK..R.Q.....E.....R.....E..... ..I.....YGK...LE....C..F.A.M..... (δ^H-globin)

Pale-throated three-toed sloth (*Bradypus tridactylus*); Kleinschmidt *et al.* (1989)

VHLADDEKAAVSALWNKVHVEEFGGEALGR LLVVYPWTSRFFESFGDLSSADAVFSNAKVKAHGKVLTSFGGGLKHLDLKGTYAHLSELHCDKLHVDPENFR LLGNVLVIVLARHFGKEFTPDVQAAAYQKVVAGVANALAHKYH (β-chain)
 (β-globin)

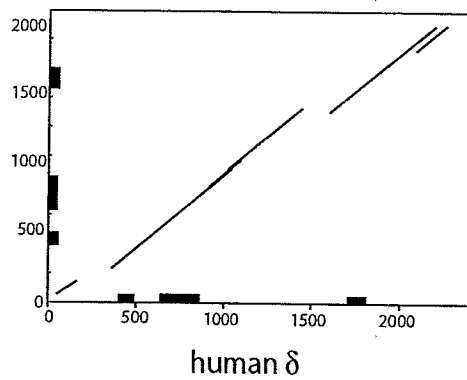
Nine-banded armadillo (*Dasypus novemcinctus*); de Jong *et al.* (1981)

VNLTSDDEKTA VLALWNKVVBZBHGGZALGR LLVVYPWTQRFFESFGDLSTPAAVFANAKVKAHGKVLTSFGGGMNHLNLTGTFAKLSELHCDKLHVDPENFR LLGNMLVVVMARHFGKEFDHWMHACFQRVVAGVANALAHKYH (β-chain)
 §D.EDC..E..... (β-globin)
D.EDC..E.....KL.....K..... (β-globin)

Fig. 1-4. Dotplot comparisons (window size = 100; threshold = 80) of the putative Asian elephant δ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Asian elephant δ -globin gene comparisons

δ -type genes



β -type genes

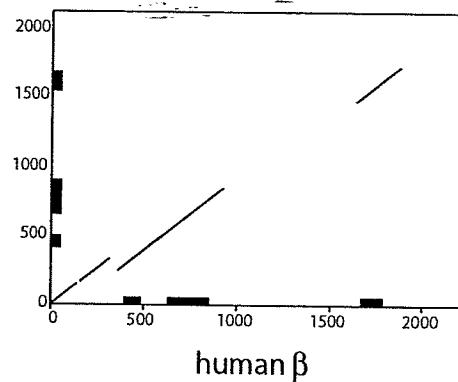
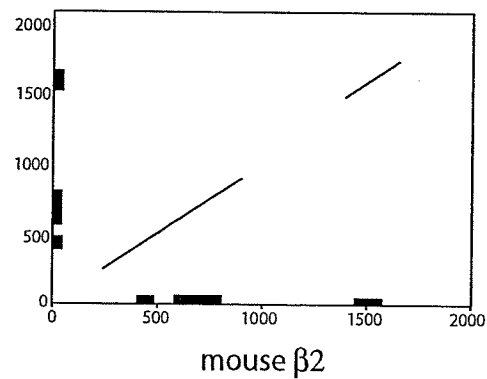
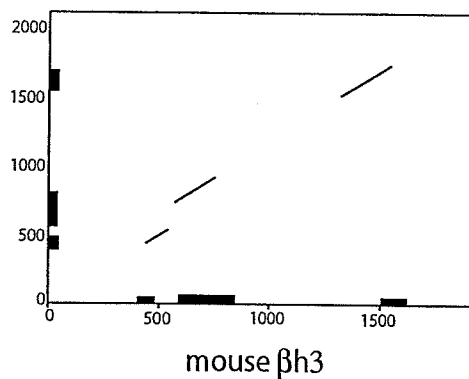
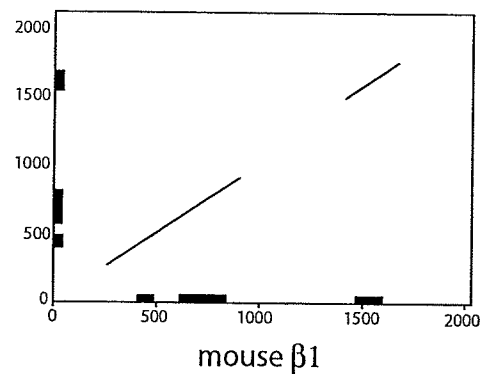
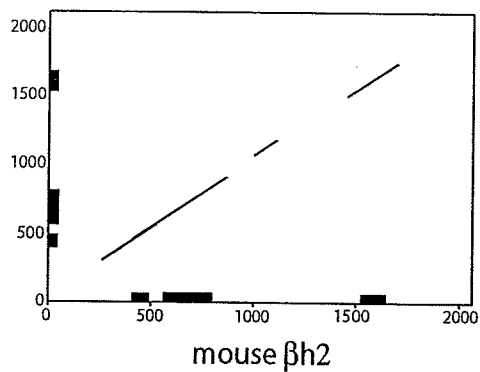
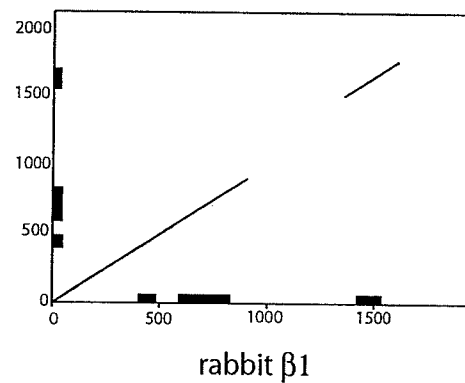
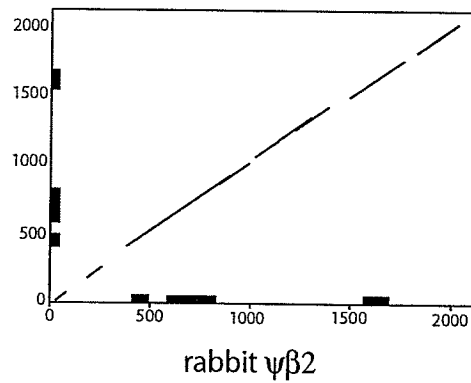
Asian elephant δ 

Fig. 1-5. The 5' flanking sequence data and features of δ - and β -globin genes amplified in this study and those of armadillo, human, tarsier and galago. Data runs from seven bases upstream of the distal "CACCC" site to the initiation ("ATG") codon. The putative cap site ("CAP") is underlined and lies 52 bp (51 bp in the sloth) upstream of the initiation codon. Sequences considered important for transcriptional regulation (CACCC, CCAAT, and ATA) are denoted by bold CAPITAL letters. Nucleotides surrounding the ATA motif correspond to a consensus sequence ("GGGCATAAAAG") found in many postnatally expressed β -globin genes (Hardison 1983) and are shaded in gray. The consensus sequence "CTTPyTG" (Baralle and Brownlee 1978), found seven bases downstream of the cap site, is also marked in gray. Colons (:) denote base deletions. Underlined bases represent those which deviate from the consensus sequence normally recovered in β -globin genes, but should not be considered significant unless they compose a sequence element involved in transcriptional regulation.

	"CACCC"	"CACCC"	"CCAAT"					
	+110	+100	+90	+80	+70	+60	+50	+40
Asian elephant δ	caggcct	CACCC tgcagaacca	CACCC tggc:::	cttgg	CCAAT ctgctcacaagagcaaaaagggcagg:accagggtt			
dugong δ	cagacct	CACCC tgcagaacca	CACCC tggc:::	ctygg	CCAAT ctgytcccaggagcaagaagggcagaaaccagggtt			
manatee δ	cagacct	CACCC tgcagaacca	CACCC tggc:::	cttgg	CCAAT ctgctcccaggagcaagaagggcagaaaccagggtt			
armadillo $\psi\delta$	tottacaaatcaatgaacatggatatatgctatgccttagg	ACAAT ctgcttcaaggagcagagaggggaagggaccaaggct						
human δ	tttcattctcacaaactaatga	AACCC tg:cttatcttaaa	CCAAC ctgctcac:::					
tarsier δ	ttttattctcacaaaccaatcaattttctgccatatacctagg	TTAAT ctgctcacatgagcagggagggcaagagtcagggct						
galago δ / galago β	tcagccctcctccgcacagcca	CACCC tgag:::	ttcag	CCTAT ctcctcataggtacagggagggcaggaaccagggca				

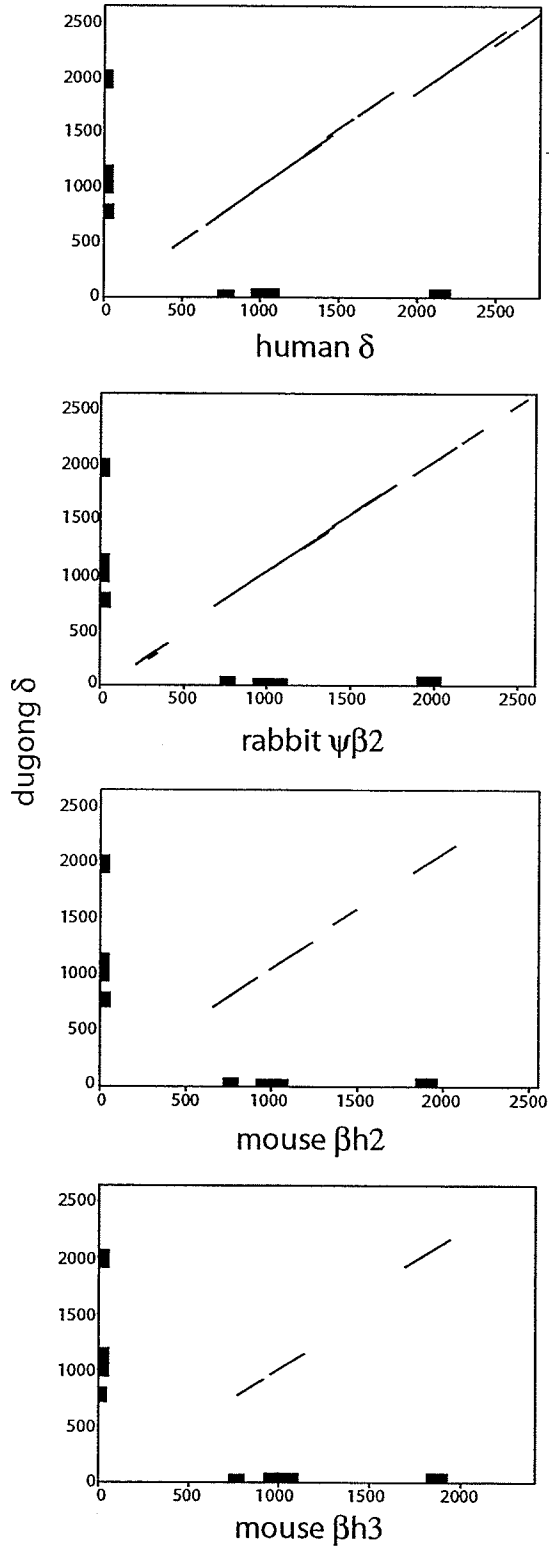
elephant shrew β	ctccgggctctttctgcagagt	CAC Tctggc:::	ctggg	CCAAT ctgcttcagaaagcaccatgggcaggaccagggct				
sloth β								
armadillo β	cagacct	CACCC tatggaacca	CACCC aggt:::	tgtga	CCAAT ctgctcagaggagcagagagggcagggaccagggct			
human β	tagacct	CACCC tgtggagcca	CACCC tagg:::	gttgg	CCAAT ctactcccaggagcagggagggcaggagccagggct			
tarsier β	caggcct	CACCC tgcagagccc	CACCC tggg:::	cttgg	CCAAT ctgctcccaggagcagtaagggcagtagccaaggct			

	"ATA"		"CAP"						"ATG"
	+30	+20	+10	0	-10	-20	-30	-40	-50
Asian elephant δ	gggc	ATA taagggaagagtagtgccagctgctgttt	A cactcacttctgacacaactgtgttgactagcaactacccaatcagacaccatg						
dugong δ	gggc	GTA aaagggaagrgcaggactagctgctgtt	A ctcttaacttctggcacaactgtgttgactagcaactacacaatcagacwcc						
manatee δ	gggc	GTA aaagggaagagcagggctagctgctgtt	A cacttaacttctggcacaactgtgttgactaacaactacacaatcagacaccatg						
armadillo $\psi\delta$	gggc	ATA aaag:::	agcaaaagccagctgttgctt	A tacttgctttggacaaaactgtgctctctagcaaccatacaaaagagacacfatg					
human δ	cagc	ATA aaaggcagggcagagtcgactgttgctt	A cactttcttctgacataaacagtgttcactagcaacctcaaa:::	cagacaccatg					
tarsier δ	gggc	ATA aaagggtagggcagagtcgactgctgtc	A cgcttgcttctgacacaacogtgttcaccagcaaccacaaa:::	cagacaccatg					
galago δ / galago β	gggc	ATA aaaggcagggcagggaaactgctgctgtc	A tacttgcttttgacacaactgtgttcactagcaacca:::	cagacatcatg					
elephant shrew β	gggc	ATA gaag:aagggcaggaccagcaaggctt	A cacttgcttctgacacaagtgtgttcactagcaactactcaaacagtcaagatg						
sloth β	tggc	ATA aaaggaggagaaagggccagctgctactc	A cacttgcttctgacacagctgtgttcactagcaaccacaaaa:::	cagacaccatg					
armadillo β	gggc	ATA aaagtaagagtaaggccagctgctgtt	A caattgcttctgacacatttgtgttcactagcaaccatataaagagacaccatg						
human β	gggc	ATA aaagtcagggcagagccatctattgtt	A catttgcttctgacacaactgtgttcactagcaacctcaaa:::	cagacaccatg					
tarsier β	gggc	ATA aaaggcagggcagggccagctgctgtc	A cgcttgcttctgacacaacogtgttcaccagcaacctcaaa:::	cagacaccatg					

Fig. 1-6. Dotplot comparisons (window size = 100; threshold = 80) of the putative dugong δ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Dugong δ -globin gene comparisons

δ -type genes



β -type genes

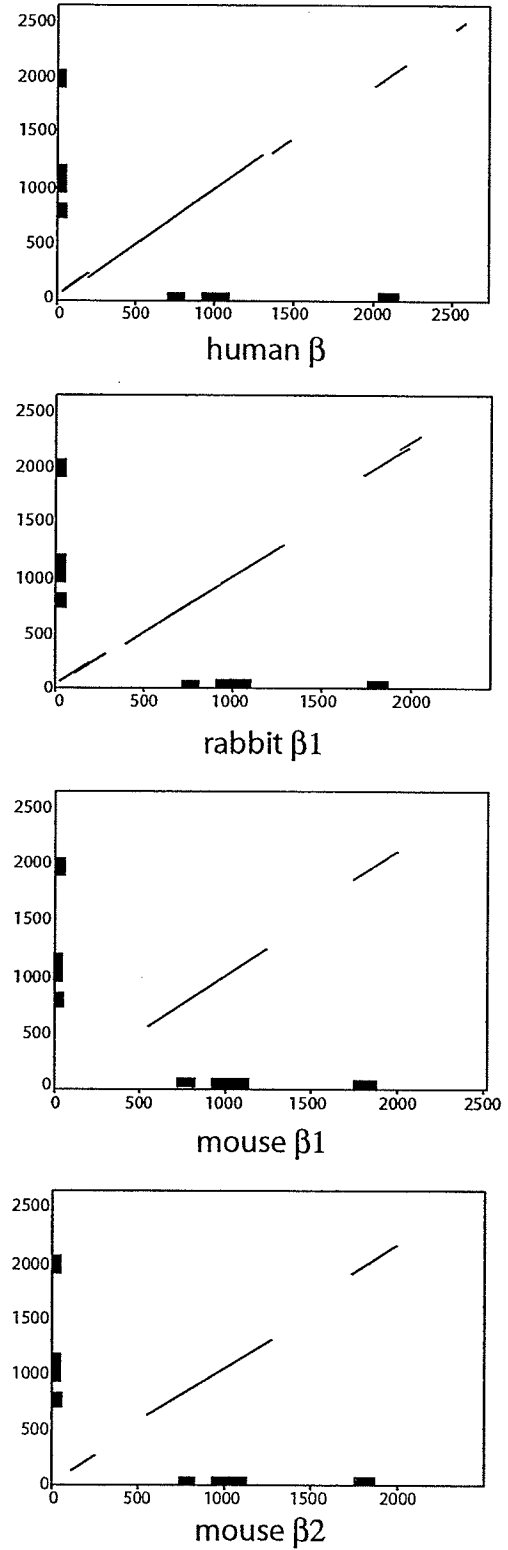


Fig. 1-7. Dotplot comparisons (window size = 100; threshold = 80) of the putative rock hyrax δ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 2 (partial sequence) and 3 for each gene are denoted by black boxes.

Rock hyrax δ -globin partial gene comparisons

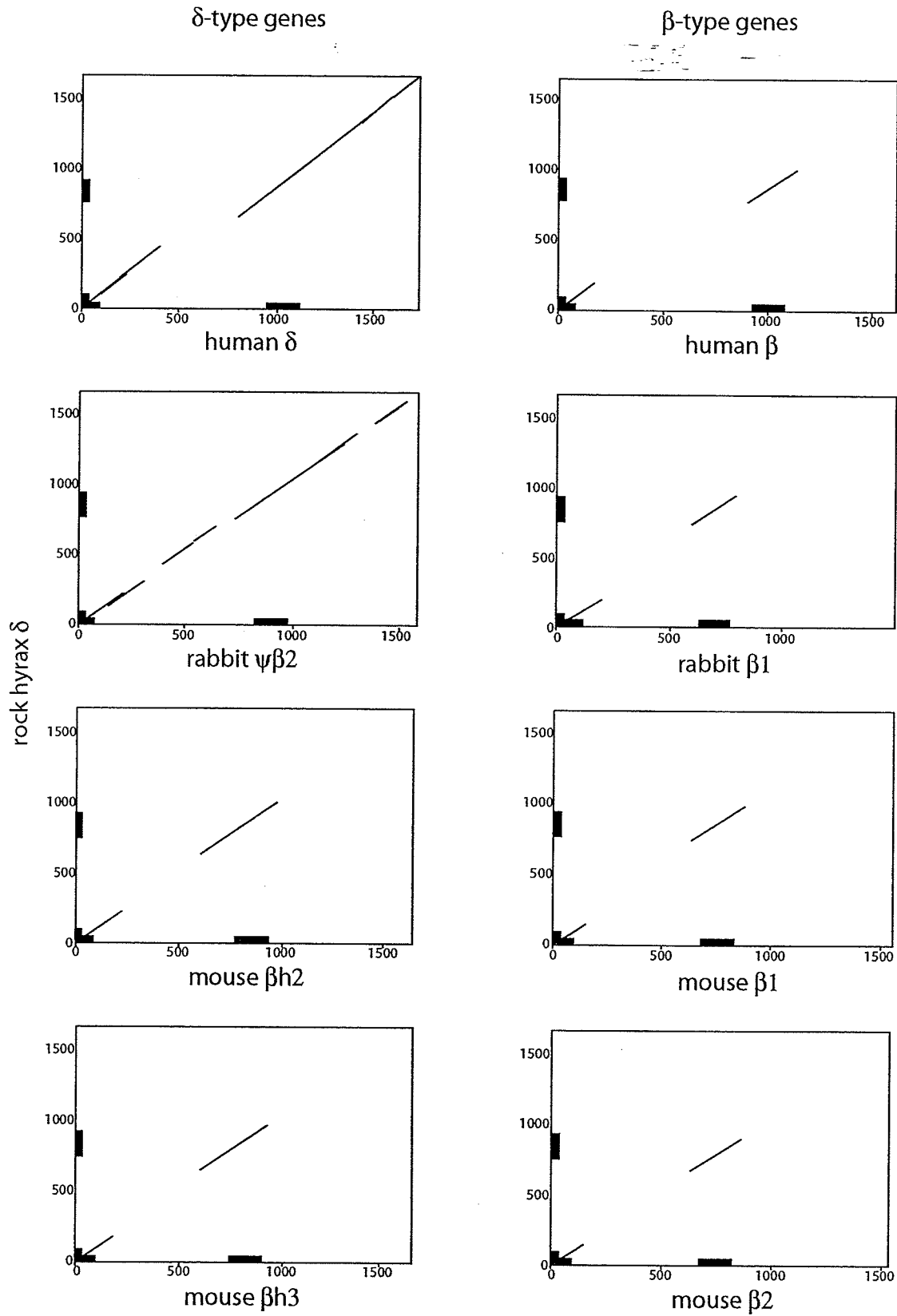
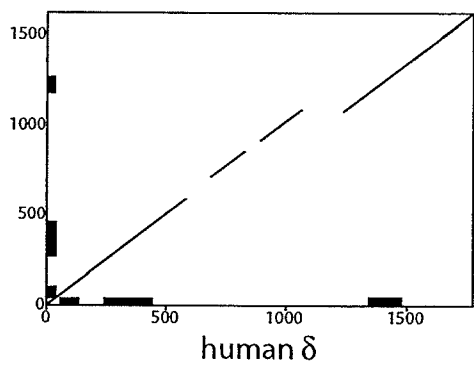


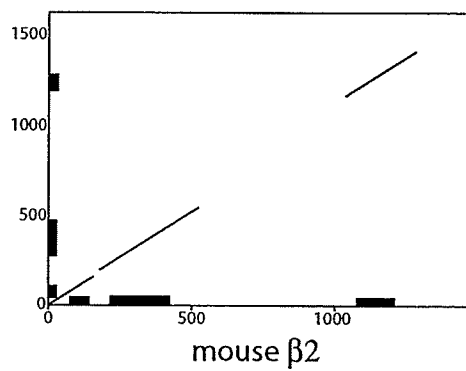
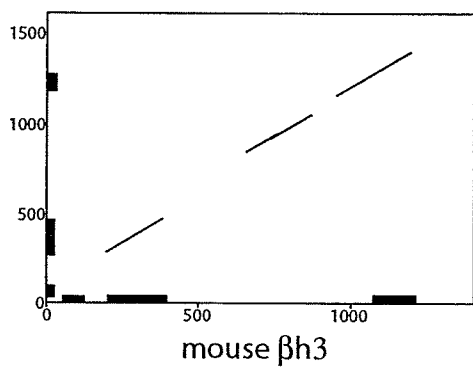
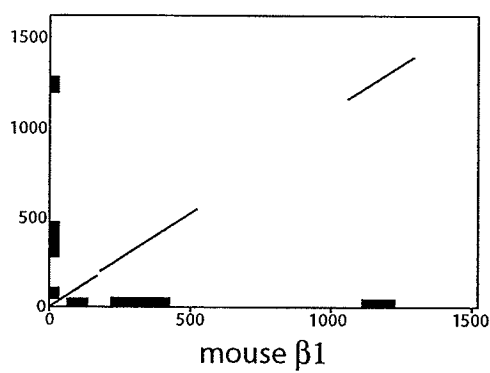
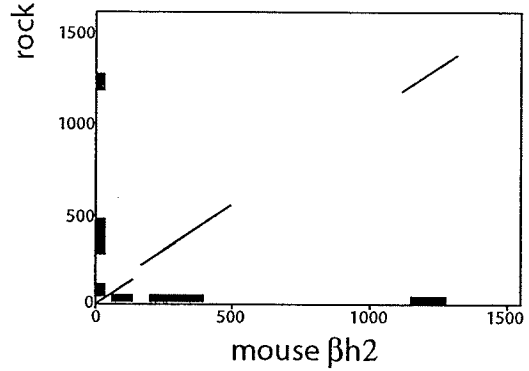
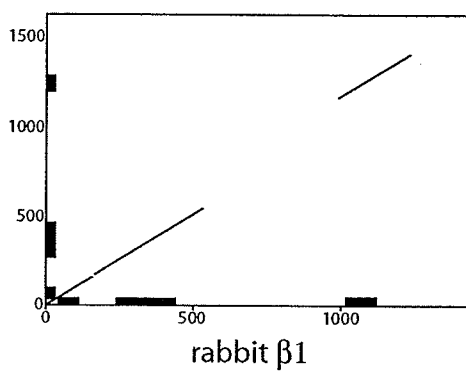
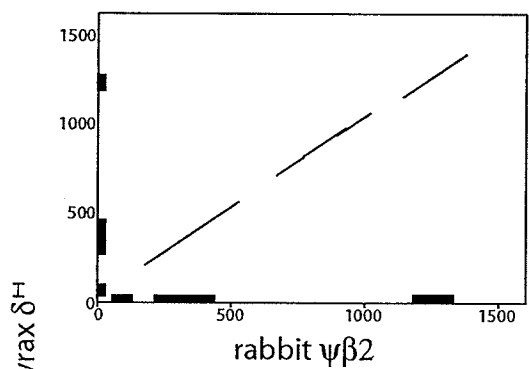
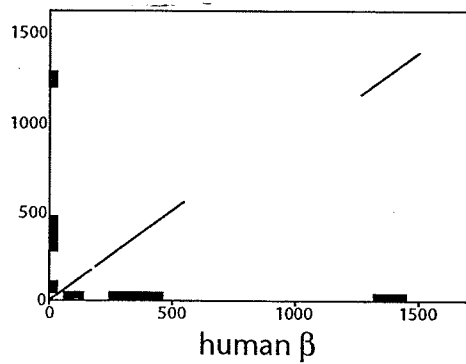
Fig. 1-8. Dotplot comparisons (window size = 100; threshold = 80) of the putative rock hyrax δ^H -globin gene (Y-axis) with the δ -, and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Rock hyrax δ^H -globin gene comparisons

δ -type genes



β -type genes



different at 31 of 146 residues (21.2%) from the ' β -chain' of this species (Fig. 1-3). The amplified gene products of the yellow-spotted and tree hyraxes shared 97.9% and 95.8% sequence identity with the second rock hyrax δ -like gene, respectively, and grouped closely together on the ML tree (Fig. 1-14) suggesting the three loci are orthologous. These apparent hyrax-specific loci were thus designated δ^H -globins.

The first gene fragment amplified from the elephant shrew ran from +670 bp upstream of the initiation codon to the end of IVS2. Dotplot analyses recovered the highest degree of sequence similarity between this fragment and the β -globin genes of human, rabbit and mouse (Fig. 1-9). Mutations at both CACCC transcriptional control motifs in the 5' promoter region of this gene were detected (Fig. 1-5, Appendix 1-4). The second gene fragment, which ran from codon position 136 in exon 3 to +180 downstream from the stop codon, shared the highest degree of sequence similarity with eutherian δ -globin genes (data not shown).

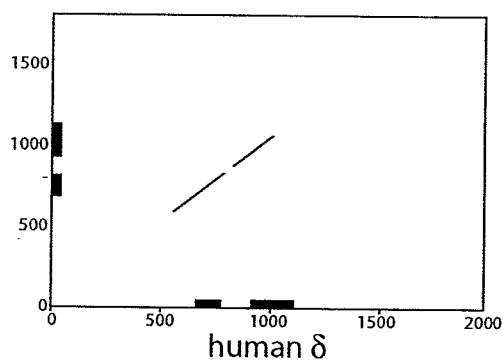
Classification of xenarthran ' β -like' globin genes

Dotplot (Figs. 1-10) and MatGAT (Table 1-4) analyses suggested that one armadillo ' β -like' gene product was homologous with eutherian δ -globin genes in IVS2. This gene was presumed a pseudogene ($\psi\delta$) due to a premature termination signal at codon 22. The second armadillo gene product showed highest sequence similarity with other mammalian β -globin genes in the IVS2 and 3' regions (Fig. 1-11). Conspicuously, the deduced protein sequence of this gene differed from the β -globin protein sequence at 4 of 146 positions (2.7%) throughout exons 1, 2 and 3 (Fig. 1-3). From the initiation to

Fig. 1-9. Dotplot comparisons (window size = 100; threshold = 80) of the putative elephant shrew β -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 1 and 2 for each gene are denoted by black boxes.

Elephant shrew β -globin partial gene comparisons

δ -type genes



β -type genes

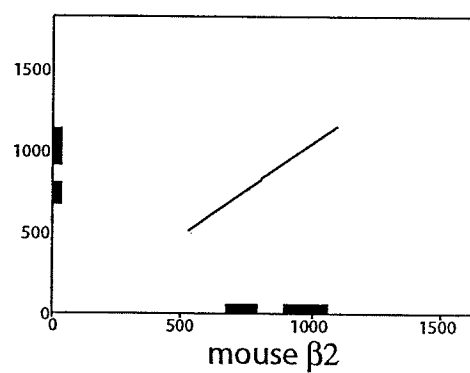
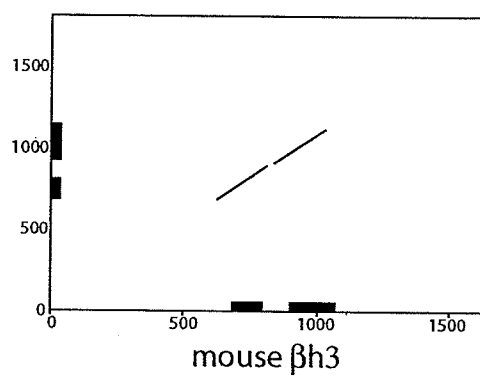
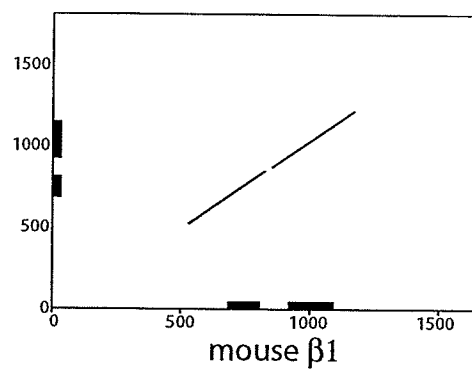
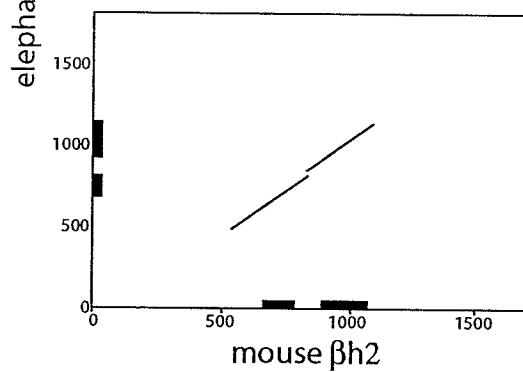
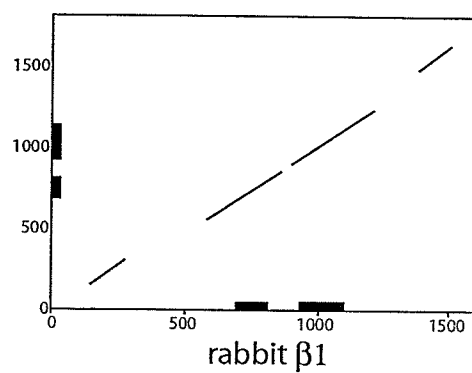
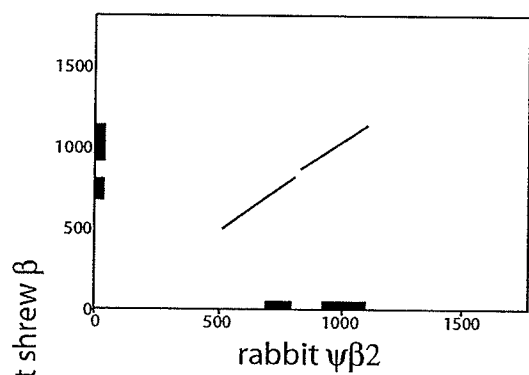
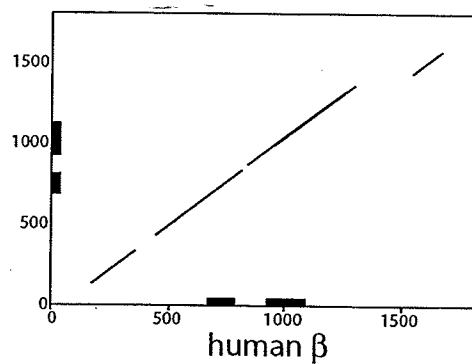
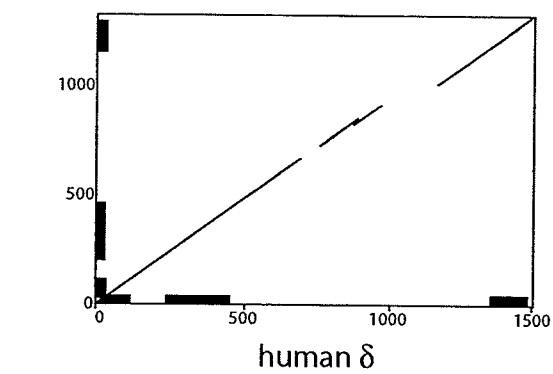


Fig. 1-10. Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo $\psi\delta$ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Armadillo $\psi\delta$ -globin gene comparisons

δ -type genes



β -type genes

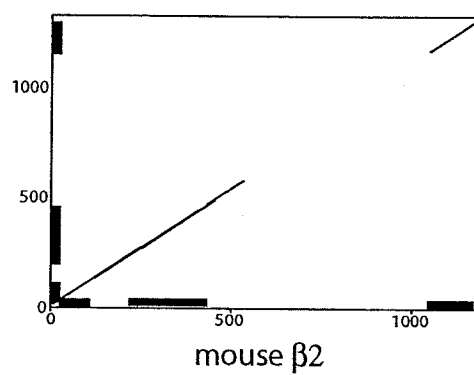
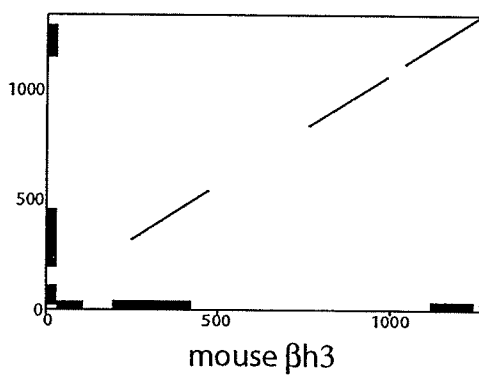
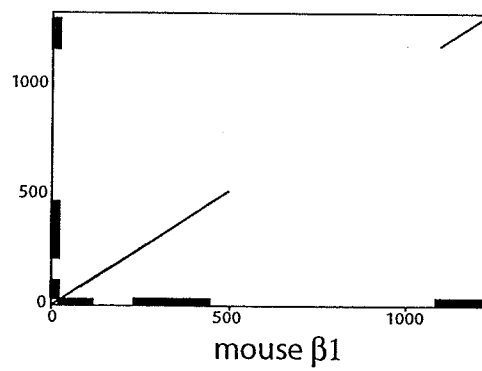
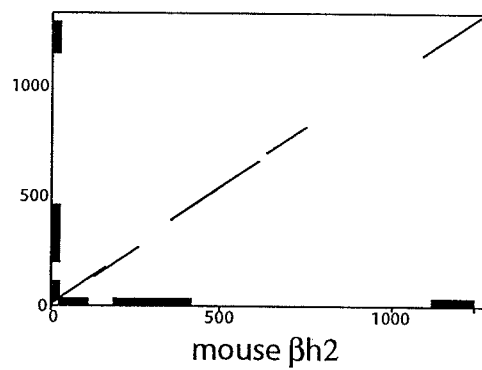
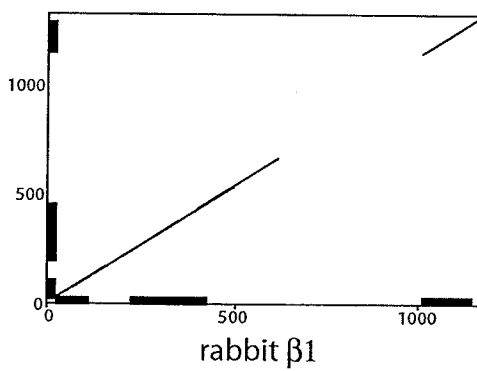
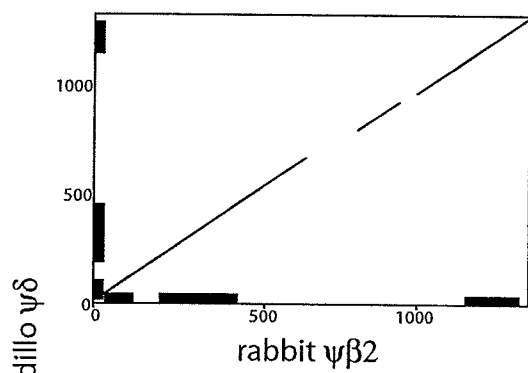
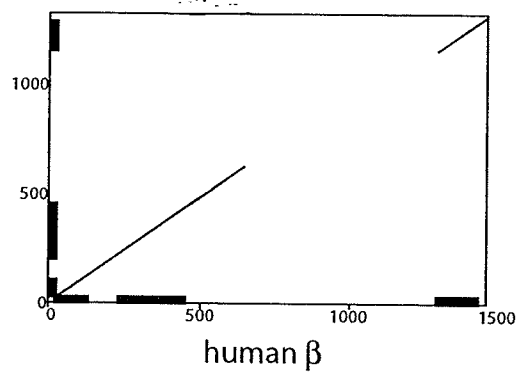
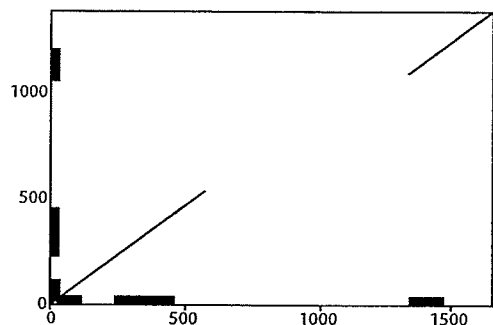


Fig. 1-11. Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo β -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

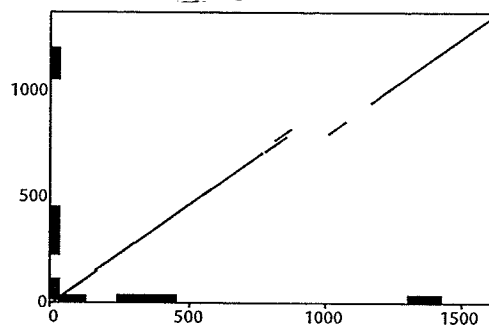
Armadillo β -globin gene comparisons

δ -type genes

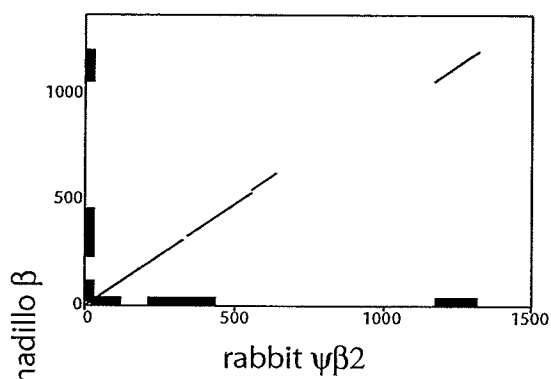


human δ

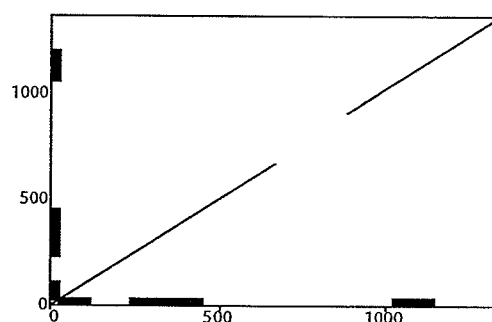
β -type genes



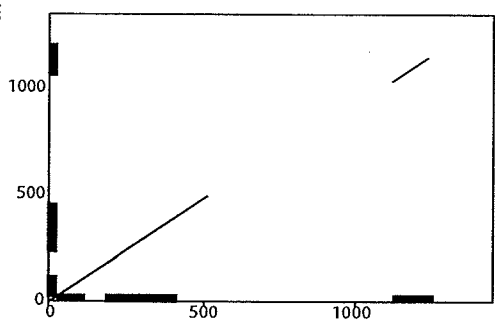
human β



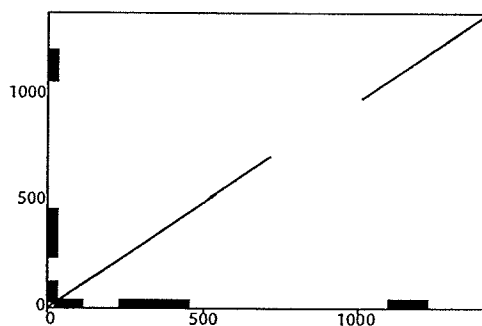
rabbit $\psi\beta 2$



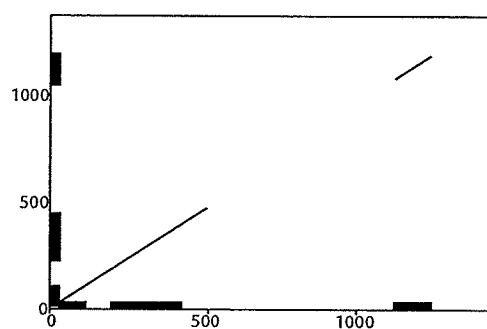
rabbit $\beta 1$



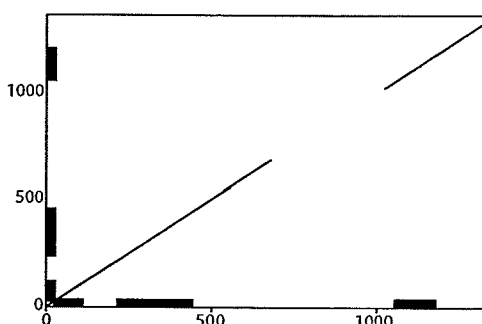
mouse $\beta h 2$



mouse $\beta 1$



mouse $\beta h 3$



mouse $\beta 2$

conceptual stop codons, nucleotide sequences of the armadillo $\psi\delta$ -globin and β -globin genes were found to share a high degree (99.4% and 99.3%) of sequence-identity with the two genes identified at the 3'-end of the *Dasypus novemcinctus* β -globin cluster. This 3'-most β -like gene seemingly possessed all the 5' and 3' control elements necessary for transcription. Conversely, the δ -like gene was missing both CACCC and CCAAT 5' transcriptional control motifs and contained no 3' poly-adenylation signal (Fig. 1-5; Appendices 1-4 and 1-5). Dotplot analyses showed the δ -globin gene to be δ -like and the β -globin gene to be β -like in their respective 5' and 3' flanking regions (data not shown).

Dotplot (Figs. 1-12, 1-13) and MatGAT (Table 1-4) analyses suggested that one sloth ' β -like' gene product shared homology with eutherian δ -globin genes, while the second gene product was homologous with other β -globin genes. The δ -globin gene was presumed a pseudogene ($\psi\delta$) due to a two-base deletion at codon 77 resulting in a premature termination signal at codon 87 in exon 2. The conceptual protein chain of the sloth β -globin gene was identical to the amino acid sequence of the ' β -globin' chain (Fig. 1-3).

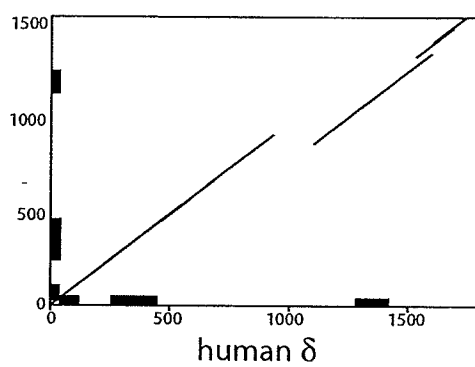
Phylogenetic analysis of δ -globin IVS2 regions

The ML analysis of δ -globin IVS2 sequences (773 bp) grouped elephants and sirenians together to the exclusion of hyraxes (Fig. 1-14). However, bootstrap support for this assemblage was relatively low (56%). Bootstrap support was strong (87–100%) for all other groupings. The three δ^H -globin genes were closely clustered together and placed

Fig. 1-12. Dotplot comparisons (window size = 100; threshold = 80) of the putative sloth $\psi\delta$ -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Sloth $\psi\delta$ -globin gene comparisons

δ -type genes



β -type genes

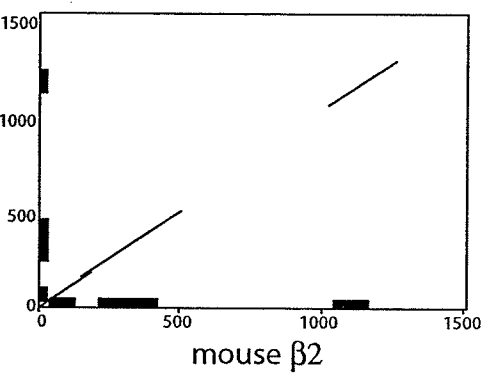
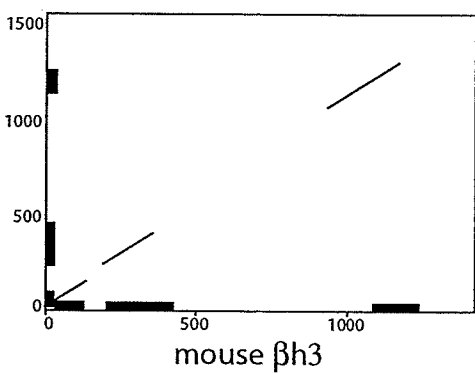
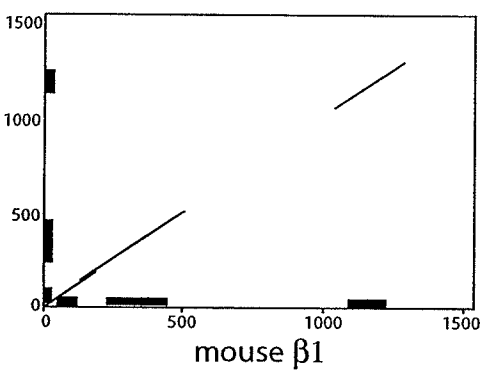
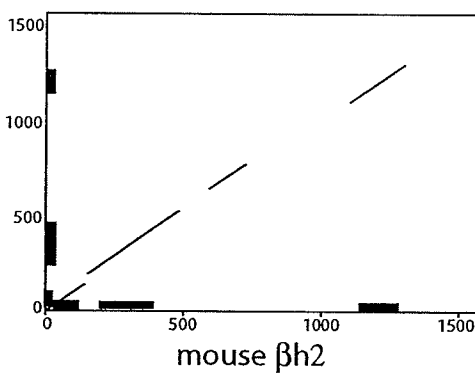
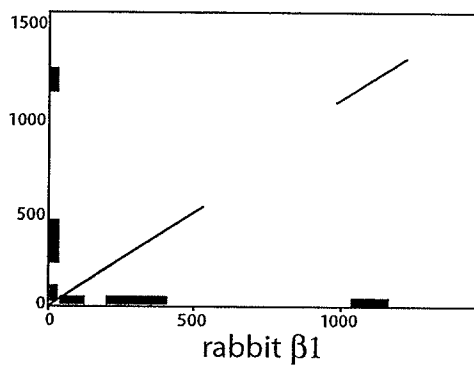
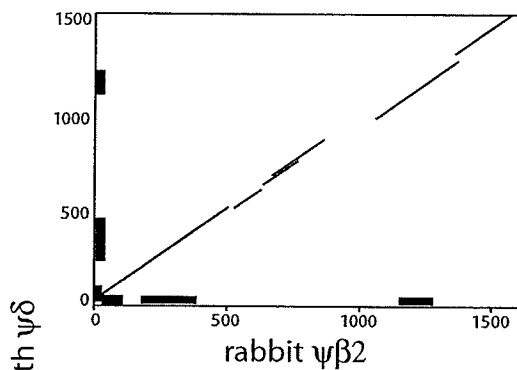
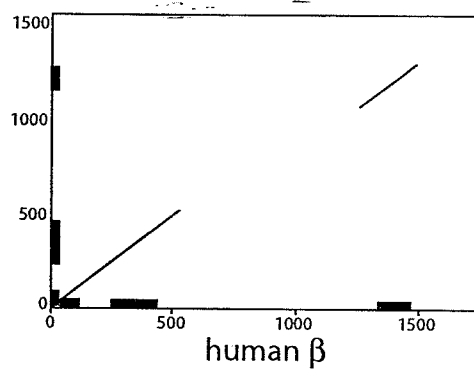


Fig. 1-13. Dotplot comparisons (window size = 100; threshold = 80) of the putative sloth β -globin gene (Y-axis) with the δ - and β -globin genes of humans, rabbits and mice (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Sloth β -globin gene comparisons

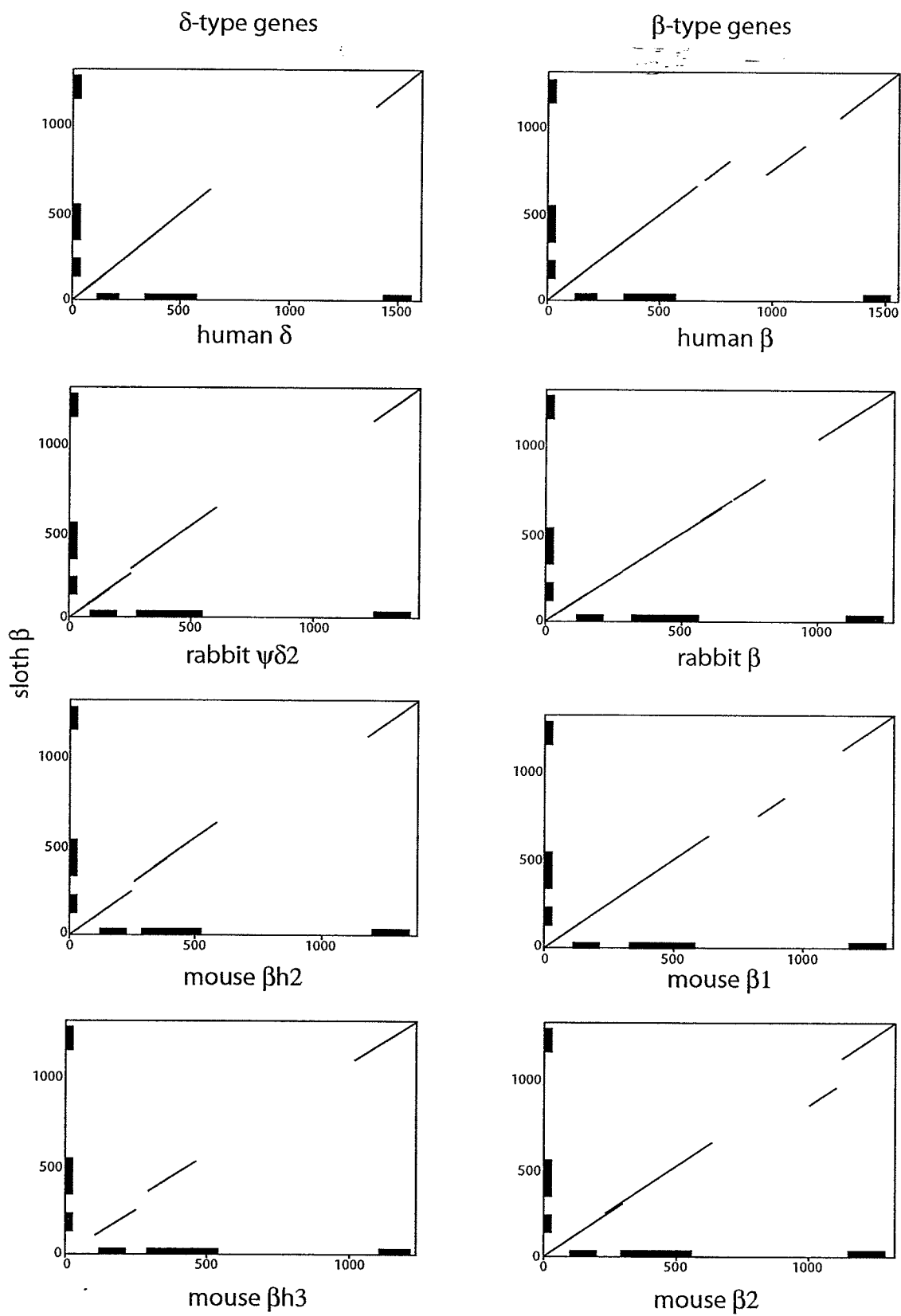
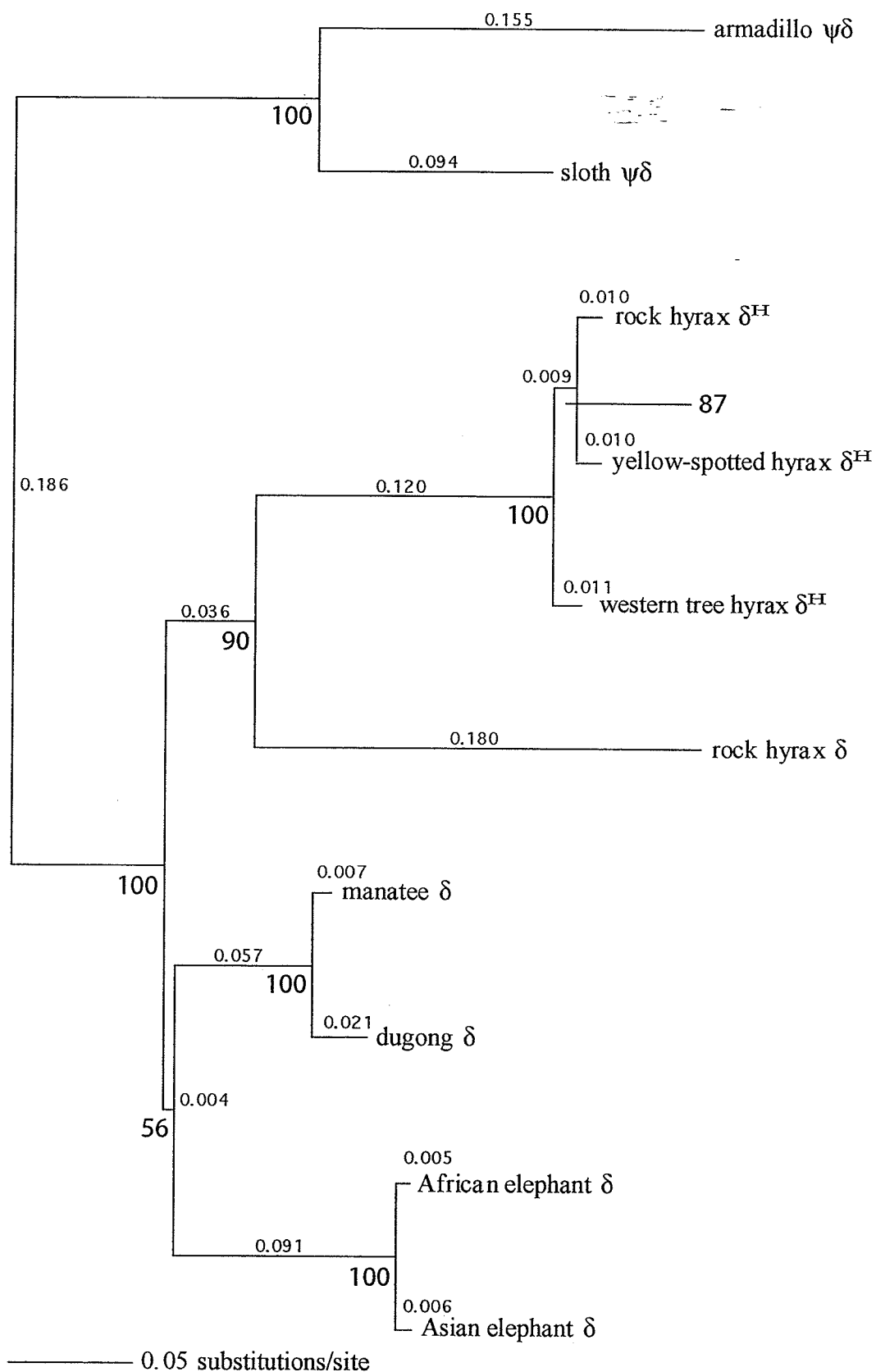


Fig. 1-14. Xenarthran-rooted maximum likelihood tree constructed from homologous δ -globin IVS2 gene sequences (773 bp; see text for details) illustrating the inter-relationships among paenungulate mammals. Values along each branch denote the degree of sequence divergence. Bootstrap support (1000 replicates) for each node are indicated in large whole numbers.



as a sister group to the rock hyrax δ -globin gene with high bootstrap support (87%).

Finally, the tree hyrax was placed as the most basal branch of the hydracoidean lineage.

DISCUSSION

The discovery of a δ -like globin gene within members of the superordinal clades, Afrotheria and Xenarthra, supports the theory that the ancestral mammalian proto- β -globin gene duplicated prior to the eutherian radiation (Hardies *et al.* 1984; Hardison 1984; Hardison and Margot 1984; Hutchison *et al.* 1984). However, the data moreover suggest that, contrary to other placental mammals, δ -like genes are the only post-natally expressed loci of the β -globin cluster found within the red blood cells of paenungulate species. This finding challenges the widely perceived notion (Martin *et al.* 1983; Hardies *et al.* 1984; Hardison 1984; Hardison and Margot 1984; Koop *et al.* 1989; Prychitko *et al.* 2005) that δ -globins have been little more than relic genes (i.e. 'failed experiments') since their inception. It is further proposed that modifications within the β -cluster leading to the sole expression of the δ -locus were already present in stem afrotherians prior to the radiation of the paenungulate clade.

These assertions are well supported by several lines of evidence. The first is based on the observation that the 5' promoter (Fig. 1-5) and 3' poly-A signals (Appendix 1-5) necessary for transcription were detected flanking the δ -globin gene of all paenungulate species for which this region was obtained. However, it is noteworthy that the 5' external ATA box of both sirenian species possessed an $\underline{\text{A}}\text{TA} \rightarrow \underline{\text{G}}\text{TA}$ substitution (Fig. 1-5). This site is thought to be crucial for the transcription of eukaryotic genes by RNA polymerase II (Dierks *et al.* 1983; Kosche *et al.* 1985), leading Satoh *et al.* (1999) to speculate that a similarly altered ATA box might diminish or prevent activity of the rat $\gamma 1$ -globin gene. In contrast, data collected for this study suggest that the mutated 5' $\underline{\text{G}}\text{TA}$ sequence does not adversely affect expression of the δ -locus in sirenians. Support for

this contention is provided by the observation that the conceptual polypeptide chain encoded by the δ -globin gene of the Florida manatee (*Trichechus manatus*) precisely matches the ' β -globin' chain of the closely related Amazon manatee, *T. inunguis* (Kleinschmidt *et al.* 1986). In addition, putative protein chains translated from the δ -globin gene products of the African elephant, Asian elephant and rock hyrax corresponded to the respective ' β -globin' protein chain sequences of these species (Braunitzer *et al.* 1982; Kleinschmidt and Braunitzer 1983; Braunitzer *et al.* 1984). Because the hemoglobin molecules of adult proboscideans (Kleihauer *et al.* 1965; Braunitzer *et al.* 1982; Braunitzer *et al.* 1984) and sirenians (White *et al.* 1976; Kleinschmidt *et al.* 1986) examined to date possess a single component (i.e. are composed of only one type of ' α -like' and one type of ' β -like' chain), this implies that the β -locus is either non-functional or absent in these groups.

Interestingly, hemoglobin of the rock hyrax possesses two chromatographically distinct (i.e. HbI and HbII) components (Kleinschmidt and Braunitzer 1983). However, amino acid analysis suggested that like the two ' α -like' chains, the sequences of the two ' β -like' chains of this species were identical, a finding the authors could not explain. The discovery of a second highly conserved δ -like (δ^H) gene in each of the three genera raises the possibility that this locus is expressed in hyraxes. Clearly, studies aimed at detecting δ^H mRNA transcripts in these species are required to resolve this issue.

Because so many δ -globin pseudogenes have been reported to date, the expression of a fully functional δ -locus in paenungulate (and likely other afrotherian) mammals is intriguing, and raises the question: how did the δ -locus gain an upper hand over the largely dominant β -locus in this clade? Traditionally, inactivation or diminished

expression of the δ -locus has been attributed to the absence of intragenic enhancers (Antoniou *et al.* 1988) and unstable mRNA transcripts (Wood *et al.* 1978; Ross and Pizarro 1983; Kosche *et al.* 1985; Steinberg and Adams 1991). However, it is now primarily thought that an inefficient promoter region, typically lacking (or possessing mutated versions of) regulatory elements crucial for proper transcription has prevented this locus from attaining the transcriptional success of its predecessor (Vincent and Wilson 1989; Koop *et al.* 1989; Tagle *et al.* 1991; Prychitko *et al.* 2005). Indeed, the region immediately upstream of the δ -loci generally contains mutations in two important structural motifs (the CCAAT and proximal CACCC boxes) upstream of the Cap site that interact with specific proteins integral for gene activation (Efstratiadis *et al.* 1980; Grosveld *et al.* 1982; Dierks *et al.* 1983; Kosche *et al.* 1985; Myers *et al.* 1986). The δ -locus of galagos notably contains an intact CACCC control sequence and is expressed at much higher levels (40% of erythrocytic ' β -like' chains; Tagle *et al.* 1991) compared to other hominoids (2–18%; Boyer *et al.* 1971; Ross and Pizarro 1983; Tagle *et al.* 1988). In addition, otherwise non-expressed δ -globin genes can be activated in both erythroid and non-erythroid cells by introduction of intact CCAAT and CACCC control sequences into the 5' flanking region of this locus (Donze *et al.* 1996; Tang *et al.* 1997; Ristaldi *et al.* 1999). Normal expression of ' β -like' globins thus appears to largely reside in these motifs. Significantly, these important transcriptional control elements, plus a second distal CACCC box (a feature characteristic of functional ' β -like' globin genes; Grosveld *et al.* 1982; Dierks *et al.* 1983; Myers *et al.* 1986) (Fig. 1-5), were detected in the upstream region of both sirenian and proboscidean δ -globin genes.

Hints into the mechanism and probable scenario of the evolutionary events leading to this never before documented attribute can be inferred from the β -clusters of other eutherian mammals (see Fig. 1-1; Konkel *et al.* 1979; Schön *et al.* 1981; Lingrel *et al.* 1983; Shapiro *et al.* 1983; Hardison and Margot 1984; Townes *et al.* 1984; Schimenti and Duncan 1985b; Cooper *et al.* 1996; Satoh *et al.* 1999), which have demonstrated that the δ -locus has been independently converted by the neighbouring β -locus in most species examined (see Fig. 1-1). However, with the exception of the galago (Tagle *et al.* 1988), these recombination events were insufficient to alter the promoter regions of the δ -locus. Dotplot analyses of the paenungulate δ -globin genes, which indicated their 5' transcriptional control regions are β -like, suggest their presumably capable promoters may have arose via conversion by the 5' region of the β -locus.

An alternate scenario is that the region downstream of exon 2 of the β -globin locus was converted by δ . This contention seems unlikely, however, since the amplified β -like globin gene of the bushveld elephant shrew (*Elephantulus intufi*) shows no evidence of conversion by a δ -like locus (Fig. 1-9). Unfortunately, the hemoglobin of *E. intufi* has not been examined, though that of the East African elephant shrews, *Petrodromus sultani* and *Rhynchocyon chrysopygus*, possess two distinct Hb components expressed in relatively equivalent proportions (Buettner-Janusch and Buettner-Janusch 1963). While this observation could signify the presence of two different ' β -like' protein chains in these species, the β -locus of *E. intufi* possesses a mutated 5' CACCC transcriptional control motif ($\rightarrow$ CACTC) and lacks the second distal 5' CACCC box common to transcribed ' β -like' genes (Grosveld *et al.* 1982; Dierks *et al.* 1983; Myers *et al.* 1986; Fig. 1-5), and is therefore probably not expressed. Together with data collected

from the paenungulates, this finding implies that, within afrotherians, up-regulation of the δ -locus preceded silencing of β , and that both events early in the evolution of this basal eutherian clade. Following gene duplication, the failure of a paralog to adopt a new function or expression pattern often leads to its nonfunctionalization (Ohno 1970). Thus, once the equally apt δ -locus was co-expressed with its predecessor (β -globin) in stem afrotherians and structurally configured for the same physiological role (oxygen transport), natural selection was likely too weak to prevent silencing mutations from accumulating in one of these superfluous genes and randomly, the β -locus was inactivated.

The hemoglobin of adult nine-banded armadillos consists of one type of ' α -chain' and two types of electrophoretically distinct ' β -chains' (de Jong *et al.* 1981). While the amino acid sequence of the minor ' β -globin' component is unknown, the major ' β -globin' chain differed from the translated protein sequence of the amplified armadillo β -globin gene at four residues (Fig. 1-3). However, because the amplified β gene shared 99.4% sequence identity with the 3'-most β -globin cluster gene (identified in the *Dasyus novemcinctus* β -globin cluster; Genbank accession number AC151518) which encoded a protein sequence 99.3% identical (145 of 146 residues) to the major ' β -like' polypeptide sequence, these residue differences possibly reflect nucleotide misincorporations or, more likely, the occurrence of two alleles within the population. Additionally, both sequences differ from the major " β -globin" chain at position 23 (histidine vs. cysteine), suggesting that the published amino acid sequence is incorrect at this site. Because the armadillo δ -globin locus appears non-functional, the gene encoding the minor ' β -like' globin component of this species is currently unknown, but speculated

to either be ϵ -globin or represent a locus situated outside of the β -cluster (see Chapter II). Unlike other eutherians, the $\psi\delta$ -locus of armadillos does not appear to have undergone conversion by β -globin (Fig. 1-10). The only other eutherian δ -globin gene thought to have remained unaltered by conversion is the $\beta h2$ pseudogene of mice (Phillips *et al.* 1984; Hardies *et al.* 1984). However, these loci show little, if any, sequence homology at the 5' flanking region (data not shown), possibly signifying a high degree of sequence differentiation or lack of orthology between the two loci.

The hemoglobin of three-toed sloths also possesses two chromatographically distinct components (Kleinschmidt *et al.* 1989). While the analysis indicated the major 'β-chain' is encoded by the β -locus, the amplified $\psi\delta$ -globin gene possessed premature termination signals within the coding region and appeared to have undergone conversion of IVS1 and exon 2 by the β -locus. This recombination event was evident from the high sequence identity shared by these two gene regions (96.4% and 97.3%, respectively), compared to 79.3% for exon 1 and 85.3% for exon 3. It seems unlikely that the frameshift mutation in exon 2 led to the silencing of this gene since the $\psi\delta$ -globin gene of armadillos appears to have been silenced in the absence of conversion events. Hence, the δ -locus was likely already transcriptionally silent prior to the divergence of these two species.

Paenungulate inter-relationships have been extremely difficult to resolve using molecular data (Amrine and Springer 1999; Waddell and Shelley 2003; Nishihara *et al.* 2005). The present analysis of δ -globin IVS2 sequences is the first to incorporate molecular data from all 7 extant genera to address this question. The ML tree (Fig. 1-14) placed elephants and sirenians as sister taxa to the exclusion of hyraxes; however, this

clade was resolved with only 56% bootstrap support. While providing support for the traditional 'Tethytheria hypothesis' (McKenna 1975), the meager support value adds to a growing body of literature suggesting that the three orders diverged over a short time interval (Murata *et al.* 2003; Waddell and Shelley 2003; Nishihara *et al.* 2005). Finally, the tree hyrax was placed as the most basal hydracoidean (Fig. 1-14), a result concordant with the conclusion of Prinsloo (1993) based on restriction fragment length polymorphism (RFLP) and phylogenetic analyses of the mitochondrial genes, cytochrome *b* and 12S rRNA.

In conclusion, data from this study reveal that the $\beta \rightarrow \delta$ duplication event within the mammalian β -globin cluster occurred prior to the eutherian radiation. The data additionally provide evidence that xenarthrans, like euarchontoglires and laurasiatherian mammals, express the β -globin gene while suppressing the δ -locus. However, it appears that the δ -locus was converted in its 5' promoter region by β , then uniquely gained a functional advantage over the normally dominant 3' locus early in the evolution of afrotherian mammals. The subsequent silencing of the β -locus in this clade thus challenges the widely accepted view that this locus is indispensable and has remained unaltered throughout eutherian evolution (Hardison and Margot 1984; Tagle *et al.* 1991; Prychitko *et al.* 2005).

CHAPTER II

MOLECULAR EVOLUTION OF THE EUTHERIAN β-GLOBIN CLUSTER INFERRED FROM AFROTHERIAN AND XENARTHROAN GENE SEQUENCES.

INTRODUCTION

The respiratory pigment hemoglobin, which binds oxygen in the lungs and delivers it to the tissues, consists of four heme groups bound to four globin polypeptide strands: two 'α-like' and two 'β-like' chains, comprising 141 and 146 amino acid residues respectively (Weber and Wells 1989; Poyart *et al.* 1992; Weber 1995). Phylogenetic studies based on nucleotide and protein sequence comparisons have suggested that the progenitor genes encoding the α- and β-globin components arose from the duplication of a single ancestral vertebrate globin gene roughly 450 mya (Czelusniak *et al.* 1982; Goodman *et al.* 1987). The proto-β-globin gene is thought to have duplicated prior to the radiation of mammals (~200 mya), producing a two-gene (proto-ε and proto-β) cluster (Efstratiadis *et al.* 1980; Goodman 1981; Czelusniak *et al.* 1982). Comparative analyses examining the β-globin clusters from four of 18 extant placental orders (i.e. the orders Rodentia, Lagomorpha, Artiodactyla, Primates) further suggested these paralogs underwent additional duplication events early in the evolution of eutherians, with the last common placental ancestor likely possessing either a four- (Hardison 1983; Hardison 1984) or five-gene cluster (Hardies *et al.* 1984; Harris *et al.* 1984; Goodman *et al.* 1984; Hardison and Gelinas 1986; Tagle *et al.* 1988) in the following orientation: 5'-ε-γ-(η)-δ-β-3'. The 'ε-like' globin genes (ε, γ and η) are thought to have arisen from the proto-ε-globin locus, while δ and β are descendants of the proto-β-globin locus (Goodman *et al.* 1987; Koop and Goodman 1988; Cooper and Hope 1993; Cooper *et al.* 1996). Chromosome mapping studies have further revealed that these genes are typically arrayed 5' to 3' on the chromosome in the order which they are expressed (Tuan *et al.* 1985; Forrester *et al.* 1987). Hence, expression of the 'ε-like' genes are generally limited to

embryonic erythrocytes (Hardies *et al.* 1984; Hardison 1984; Hill *et al.* 1984), with γ -locus expression extending to the fetal red blood cells only within anthropoid primates (Fitch *et al.* 1991).

Compared to post-natal stages of eutherian development, embryonic and fetal growth takes place in a relatively static gestational environment and consequently, the genes encoding post-natal hemoglobin chains are generally less conserved than those comprising pre-natal hemoglobins (Shapiro *et al.* 1983; Hardies *et al.* 1984; Harris *et al.* 1986; Satoh *et al.* 1999; Koop and Goodman 1988). Oddly enough, comparative studies have suggested the non-coding regions of 'ε-like' globin genes diverge more slowly than those of their adult-expressed counterparts as well (Shapiro *et al.* 1983; Hardison 1984; Hardies *et al.* 1984; Giebel *et al.* 1985; Slightom *et al.* 1985; Koop *et al.* 1989b). While the non-coding evolution rates of β -globin cluster genes are comparable to the neutral mutation rate of pseudogenes ($4 - 5 \times 10^{-9}$ substitutions/site/year; Hayashida and Miyata 1983; Kimura 1983; Li *et al.* 1985) for members of Laurasiatheria and Euarchontoglires, they are markedly low in primates ($\sim 2.9 \times 10^{-9}$ substitutions/site/year; Giebel *et al.* 1985; Harris *et al.* 1986; Koop *et al.* 1989b). This latter observation, known as the 'hominid slowdown' (Li *et al.* 1987; Hasegawa *et al.* 1989; Bailey *et al.* 1991), has been attributed to the longer generation time of primates relative to other eutherians (Giebel *et al.* 1985; Wu and Li 1985).

As noted earlier, it is generally recognized that the progenitor of placental mammals possessed either a four- or five-gene β -globin cluster. However, comprehensive molecular phylogenetic studies (Madsen *et al.* 2001; Murphy *et al.* 2001a,b; Amrine-Madsen *et al.* 2003) suggest this hypothesis is based on data from only

the two most derived (Euarchontoglires and Laurasiatheria) of the four major placental clades. It is therefore unknown whether the four- or five-gene cluster arose before or after the divergence of the superorders Afrotheria and Xenarthra (Fig. 1-T). Accordingly, the primary goal of this study was to obtain sequence data on the 'ε-like' globin genes from members of the superorders Xenarthra and Afrotheria and, together with nucleotide data obtained from their 'β-like' globin genes (see Chapter I), derive the composition of the ancestral β-globin cluster of stem eutherians. Furthermore, because nothing is known regarding the rates of evolution of the 'ε-like' and 'β-like' globin genes of xenarthran and afrotherian mammals, the second goal was to compare the rates of DNA evolution for the coding and non-coding domains of the globin genes found within their β-clusters. Finally, the above noted mutation rates were contrasted with those of other eutherian β-globin cluster genes in order to gain insight into the postulated link between generation time and the rate of 'neutral' evolution.

MATERIALS AND METHODS

Data collection

PCR, cloning and sequencing protocols were identical to those outlined in Chapter I, with the exception that PCR primers used for initial amplification reactions were designed from a 411 bp 'ε-like' globin gene fragment sequenced from both African and Asian elephants (Singh 2002). In addition, 128,121 bp of consensus sequence containing the β-globin gene cluster of the nine-banded armadillo (clone VMRC5-69F10; NISC Comparative Vertebrate Sequencing Project), together with the complete ε-, γ- and η-globin genes of human, rabbit, mouse and goat (Table 2-1) were downloaded from GenBank and included in the analyses.

Data analyses

Consensus gene alignments were constructed from each species using Sequencher™ (Version 4.2.2). Overlapping gene fragments incorporated between 4 and 7 cloned fragments amplified from 3 to 5 individual PCRs, respectively. Dotmatcher (EMBOSS) was used to determine homology between the sequenced 'ε-like' genes and those identified from the armadillo β-cluster (see above; Table 2-1) with the ε-, γ- and η-globin genes of human, rabbit, goat and mouse (Table 2-1). Additionally, MatGAT (Version 2.02) software was employed to calculate the percent identity between the non-coding regions (IVS, 5' and 3' external) of the above noted genes. To assess promoter regions, the 5' external sequences of sequenced genes were aligned with the 'ε-like' globin genes of armadillo, human and goat (Table 2-1).

Table 2-1. Genes included in the sequence analyses of 'ε-like' globin genes with corresponding GenBank accession numbers.

Eutherian superorder	Species	Common Name	Globin gene	Accession No.	Reference
Afrotheria	<i>Loxodonta africana</i>	African elephant	γ	DQ091215	This study
	<i>Elephas maximus</i>	Asian elephant	γ	DQ091216	This study
	<i>Trichechus manatus</i>	West Indian manatee	γ	DQ091217	This study
	<i>Dugong dugon</i>	dugong	γ	DQ091218	This study
	<i>Elephantulus intufi</i>	bushveld elephant shrew	ε	DQ091219	This study
Xenarthra	<i>Bradypus tridactylus</i>	pale-throated three-toed sloth	ε	DQ091220	This study
	<i>Dasypus novemcinctus</i>	nine-banded armadillo	ε, ψγ, ψη	AC151518	Green 2004, unpublished
Euarchotoglires	<i>Homo sapiens</i>	human	ε, γ ^A , ψη	U01317	Efstratiadis <i>et al.</i> 1980
	<i>Mus musculus</i>	mouse	γ, βh1	X14061	Shehee <i>et al.</i> 1989
	<i>Oryctolagus cuniculus</i>	rabbit	ε, γ	M18818	Margot <i>et al.</i> 1989
	<i>Tarsius syrichta</i>	tarsier	ε, γ, ψη	M81411 M33973	Koop <i>et al.</i> 1989a
Laurasiatheria	<i>Capra hircus</i>	goat	εI εII	X01912 X01913	Shapiro <i>et al.</i> 1983

The coding regions of sequenced and representative eutherian 'ε-like' and 'β-like' globin genes devoid of frameshift mutations (Table 2-5) were analyzed at synonymous (silent) and nonsynonymous (replacement) sites. The percentage of synonymous and nonsynonymous substitutions between the sequenced genes and their human ortholog, was calculated using the Nei and Gojobori (1986, equations 1-3) method with DnaSP software (Version 4.0). Corrections for superimposed (e.g. A→T→C) and back mutations (e.g. A→T→A) occurring at the same site were determined according to the method of Jukes and Cantor (1969). Mutation rates were calculated by dividing the percent substitution values by the estimated time of divergence between each of the compared species (with afrotherians, xenarthrans, laurasiatherians, atriodyctylans and prosimian primates splitting from the main eutherian tree at 107, 102, 87, 85 and 77 mya, respectively; Springer *et al.* 2003).

The evolution rates of non-coding DNA were calculated for eutherian 'ε-like' and 'β-like' globin genes by first subtracting the percent similarity between the IVS2 regions of the sequenced gene and its human ortholog (taken from the MatGAT analysis of IVS2 sequences, see Tables 1-4 and 2-3) from 100. The resulting value (representing % divergence) was divided by the estimated time of divergence between the sequenced and compared (human) species to infer the non-coding mutation rate.

RESULTS

A single complete 'ε-like' globin gene was amplified from the African elephant, Asian elephant, dugong, manatee and sloth, while a partial 'ε-like' gene product was amplified from the elephant shrew (Table 2-2). The genes were classified as ε-globin or γ-globin based on homology with other eutherian 'ε-like' globin genes (see below).

Classification of γ-globin genes

Dotplot analyses of the Asian (Fig. 2-1) and African (data not shown) elephant 'ε-like' genes (which were 98.7% similar from initiation to termination codons) suggested they were homologous with eutherian γ-globin genes in the IVS2 regions. Moreover, MatGAT (Tables 2-3, 2-4) analyses recovered highest sequence similarity with other placental γ-globin genes in all non-coding regions. Within the Asian elephant's 3' gene flanking region, two putative poly-adenylation signals (AATAAA) were found at +195 bp and +213 bp from the termination codon (Appendix 2-5), ~140 bp downstream from the site typical of functional eutherian γ-globin genes (i.e. +62 bp from the stop codon; Efstradiatis *et al.* 1980; Margot *et al.* 1989; Shehee *et al.* 1989). The absence of frameshift and splice junction mutations suggested both elephant γ-globin genes are capable of transcribing polypeptides.

Dotplot analyses of the dugong (Fig. 2-2) and manatee (data not shown) 'ε-like' globin gene products (which were 95.9% identical from initiation to termination codons) revealed homology with the γ-globin genes of other eutherians in the IVS2 region. Similarly, MatGAT (Tables 2-3, 2-4) analyses of the dugong and manatee 'ε-like' genes recovered highest sequence similarity with placental γ-globin genes in all non-coding

Table 2-2. List of 'ε-like' globin gene products amplified in this study.

Taxon	Gene identity	Number of base pairs sequenced	Amplified gene region
African elephant	γ	1,455	complete
Asian elephant	γ	1,798	complete
dugong	γ	2,041	complete
manatee	γ	1,578	complete
elephant shrew	ε	1,129	partial
sloth	ε	2,122	complete

Table 2-3. Percent identity matrix comparing the IVS1 and IVS2 external sequences of 'ε-like' globin genes sequenced in this study with those of human, rabbit, mouse and goat. Gap opening and extension penalties were set to 16 and 4, respectively. Numbers in gray represent IVS1 comparisons while values in black denote IVS2 comparisons.

	1(ε)	2(ε)	3(ε)	4(ε)	5(ε)	6(ε)	7(ε)	8(γ)	9(γ)	10(γ)	11(γ)	12(ψγ)	13(γA)	14(γ)	15(γ)	16(ψη)	17(ψη)	18(η)
1. elephant shrew ε		55.2	56.3	51.8	55.0	48.8	48.7	47.0	47.5	48.3	49.0	47.4	47.9	47.6	50.3	49.6	48.3	48.6
2. sloth ε	--		75.5	56.2	61.7	51.9	49.0	47.8	47.7	49.1	48.7	50.1	49.8	48.4	51.5	48.5	48.9	48.6
3. armadillo ε	--	80.2		57.7	63.8	51.4	50.5	49.7	49.3	49.5	50.7	50.7	48.4	49.2	51.9	50.8	49.6	50.4
4. human ε	--	70.7	67.5		60.4	48.6	57.9	48.1	49.0	49.5	49.0	48.3	47.5	47.3	50.4	47.5	48.7	48.0
5. rabbit ε	--	71.8	64.8	72.1		52.7	50.7	51.6	51.2	48.8	49.9	49.1	48.4	47.9	51.1	49.4	49.4	49.1
6. mouse γ (ε-like)	--	49.6	47.7	46.3	50.4		42.9	47.3	47.3	47.3	46.8	47.0	46.6	47.8	49.2	47.9	47.8	47.5
7. goat εI	--	66.4	64.3	62.1	62.3	48.3		46.9	47.2	48.3	48.2	47.6	47.1	47.0	48.6	45.8	47.1	46.3
8. African elephant γ	--	65.6	60.0	59.2	59.2	46.3	62.4		98.6	84.2	85.0	63.5	59.9	54.2	48.0	49.4	48.0	47.7
9. Asian elephant γ	--	64.8	59.2	58.4	58.4	46.3	62.4	99.2		84.4	85.2	63.2	59.9	53.6	46.6	49.5	47.3	47.9
10. dugong γ	--	53.8	54.2	54.3	55.2	40.7	60.5	82.9	83.7		95.1	62.6	60.2	55.0	47.5	47.4	48.6	48.2
11. manatee γ	--	56.9	54.3	53.6	56.0	42.3	58.7	83.7	84.6	95.9		63.5	60.1	56.2	47.5	49.7	47.1	48.9
12. armadillo ψγ	--	34.4	38.3	37.6	41.5	40.7	39.2	41.7	41.4	39.8	40.6		58.5	53.1	49.7	48.0	49.2	48.5
13. human γA	--	57.6	55.7	58.4	57.6	42.6	56.8	66.7	65.9	64.2	65.0	45.3		57.8	47.8	47.9	48.5	48.5
14. rabbit γ	--	58.7	59.2	53.2	59.8	41.1	55.2	62.9	62.1	66.1	66.1	43.4	71.0		47.2	48.2	49.0	48.1
15. mouse bh1 (γ-like)	--	52.4	48.9	50.4	50.0	46.2	53.3	49.6	48.8	52.8	53.5	43.0	54.0	52.0		49.6	49.4	49.0
16. armadillo ψη	--	34.4	38.3	37.6	41.5	40.7	39.2	41.7	41.4	39.8	40.6	100.0	45.3	43.4	43.0		57.0	54.9
17. human ψη	--	53.0	51.9	50.8	55.3	44.6	52.0	58.9	59.7	57.4	56.6	46.9	57.4	52.0	50.8	46.9		56.0
18. goat εII (η-like)	--	54.1	56.9	57.8	54.5	46.8	56.4	57.5	57.5	58.3	57.3	43.5	56.4	54.0	51.6	43.5	71.4	

Table 2-4. Percent identity matrix comparing the 5' and 3' external sequences of 'ε-like' globin genes amplified in this study with those of human, rabbit, mouse and goat. Gap opening and extension penalties were set to 16 and 4, respectively. Numbers in gray represent 3' comparisons while values in black denote represent 5' comparisons.

	1(ε)	2 (ε)	3 (ε)	4 (ε)	5 (ε)	6 (ε)	7 (ε)	8 (γ)	9 (γ)	10 (ψγ)	11 (γA)	12 (γ)	13 (γ)	14(ψη)	15(ψη)	16 (η)
1. elephant shrew ε		—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2. sloth ε	54.7		80.6	74.8	70.3	69.3	73.9	—	61.8	44.7	54.5	57.7	57.3	—	59.0	61.3
3. armadillo ε	52.6	71.3		73.4	70.9	72.4	73.4	—	64.0	47.2	56.6	58.8	57.4	—	57.5	63.4
4. human ε	59.8	66.1	66.0		74.3	71.8	73.9	—	57.3	43.2	49.1	59.5	53.9	—	54.5	62.4
5. rabbit ε	51.9	58.9	59.9	66.7		71.5	73.2	—	60.9	45.3	55.8	60.6	57.6	—	58.5	63.3
6. mouse γ (ε-like)	53.2	51.5	50.4	54.2	49.9			—	65.8	46.6	54.4	57.9	56.3	—	57.8	61.8
7. goat ε	50.0	61.0	43.2	63.6	58.3	55.7		—	45.1	45.6	59.0	60.9	59.5	—	60.0	64.4
8. Asian elephant γ	49.4	47.1	47.7	48.7	47.2	48.9	40.3		—	—	—	—	—	—	—	—
9. dugong γ	47.8	49.7	49.6	48.2	46.5	46.6	40.8	82.8		45.1	61.6	58.4	61.9	—	57.5	60.5
10. armadillo ψγ	36.5	44.0	42.5	43.8	45.4	45.0	33.3	53.5	53.0		54.3	46.8	47.6	—	47.0	48.3
11. human γA	56.8	46.3	45.6	45.4	47.1	48.3	40.0	49.9	52.6	46.6		69.3	59.8	—	58.1	57.1
12. rabbit γ	50.0	46.8	46.7	46.4	45.8	48.4	38.7	51.7	51.4	47.9	62.1		59.6	—	60.1	57.5
13. mouse bh1 (γ-like)	50.6	54.0	55.1	56.8	54.4	54.4	55.7	50.4	45.9	43.2	46.1	45.2		—	51.8	60.7
14. armadillo ψη	43.8	45.1	45.6	49.1	48.4	50.1	40.3	48.8	48.0	46.3	44.6	48.4	48.5		—	—
15. human ψη	48.0	45.4	49.9	50.1	51.3	47.5	54.4	48.1	47.0	43.7	46.0	44.9	48.7	52.7		62.2
16. goat εII (η-like)	52.5	49.9	47.1	50.6	47.3	48.3	43.4	44.6	45.5	43.9	48.9	44.8	50.1	51.3	55.3	

Fig. 2-1. Dotplot comparisons (window size = 100; threshold = 80) of the putative Asian elephant γ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Asian elephant γ -globin gene comparisons

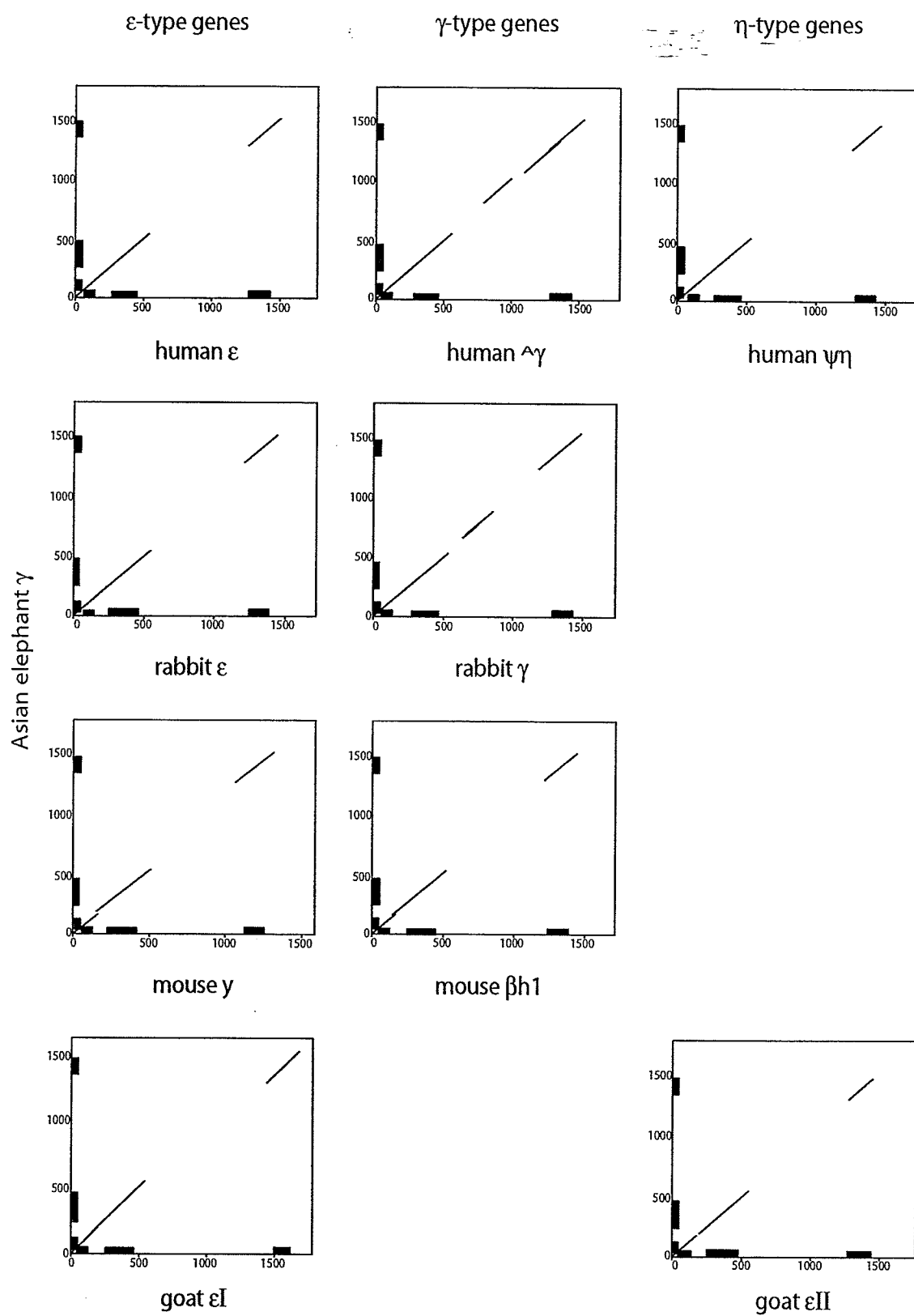
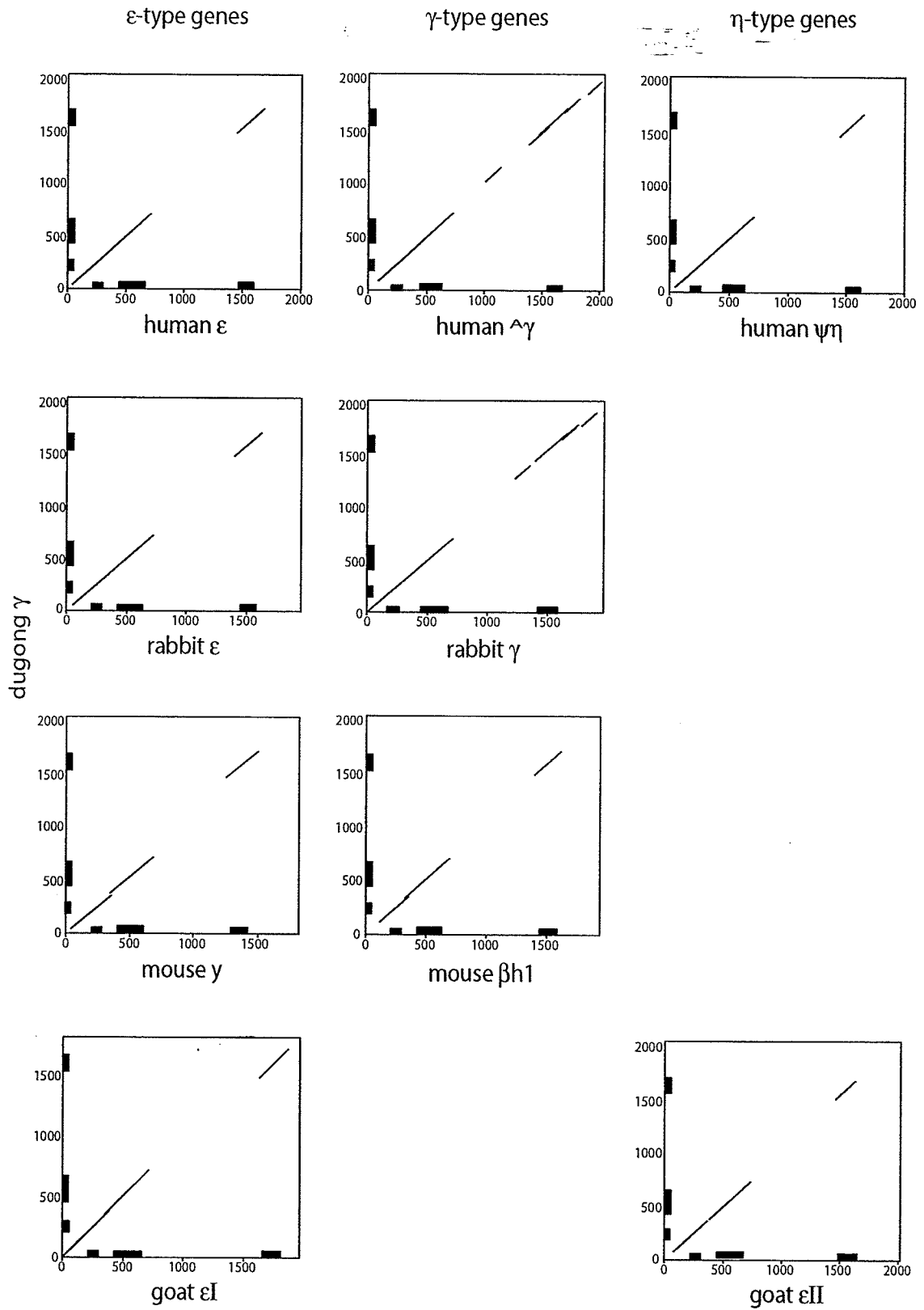


Fig. 2-2. Dotplot comparisons (window size = 100; threshold = 80) of the putative dugong γ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Dugong γ -globin gene comparisons



regions. The dugong γ -globin gene did not possess the 5' external CACCC transcriptional control motif typically located between -145 and -112 bp upstream from the eutherian γ -globin gene Cap site (Efstradiatis *et al.* 1980; Margot *et al.* 1989; Shehee *et al.* 1989), though a CACCG sequence was detected -120 bp upstream of the Cap (Fig. 2-3, Appendix 2-4). Two putative poly-adenylation signals were also found in the dugong gene's 3' gene flanking region – the first ~20 bp before the usual site (see above) and the second at the appropriate position (+62 bp from the termination codon; Appendix 2-5). Finally, a GT → AT splice junction mutation was found at the exon 2/IVS2 boundary of both sirenian γ -globin genes. Nonetheless, the absence of frameshift mutations indicate the two genes are still capable of transcribing polypeptides.

Classification of ϵ -globin genes

Dotplot (Fig. 2-4) and MatGAT (Tables 2-3, 2-4) analyses indicated the elephant shrew gene fragment, which ran from codon 60 in exon 2 to +70 bp downstream of the stop codon, shared highest sequence similarity with eutherian ϵ -globins in the IVS2 and 3' flanking region. The sloth gene product was also classified as ϵ -globin based on sequence homology with eutherian ϵ -globin genes in the IVS2 and gene flanking regions (Fig. 2-5, Tables 2-3, 2-4). This xenarthran gene possessed all the external 5' and 3' transcriptional control motifs characteristic of functional ' ϵ -like' globin genes (Fig. 2-3, Appendices 2-4 and 2-5).

Examination of the *Dasyus novemcinctus* β -globin cluster revealed the presence of three ' ϵ -like' globin genes. The 5'-most gene was homologous with other eutherian ϵ -globin genes (Fig. 2-6, Tables 2-3, 2-4) and was found to possess all the control elements

Fig. 2-3. The 5' flanking sequence data and features of 'ε-like' globin genes amplified in this study, plus those of the armadillo ε-, ψγ- and ψη-globin genes, the human ε-, γ^A- and ψη-globin genes, and the goat η-globin gene. Data runs from seven bases upstream of the "CACCC" site to the initiation ("ATG") codon. The putative cap site ("CAP") is underlined and lies 52 bp (53 bp in the sloth) upstream of the initiation codon. Sequences considered important for transcriptional regulation (CACCC, CCAAT, and ATA) are denoted by bold CAPITAL letters. Nucleotides surrounding the ATA motif correspond to a consensus sequence ("AAGAATAAAAG") found in many pre-natally expressed β-globin genes (Hardison 1983) and are shaded in gray. The consensus sequence "CTTPyTG" (Baralle and Brownlee 1978), found seven bases downstream of the cap site, is also marked in gray. Colons (:) denote base deletions. Underlined bases represent those which deviate from the consensus sequence normally recovered in β-globin genes, but should not be considered significant unless they compose a sequence element involved in transcriptional regulation. Two large inserts were omitted from the gene sequences of human γ^A-globin (5'- ctggagtcacaacccacccttgaccaat-3' containing a second CCAAT box, positioned immediately after the CCATT sequence) and ψη-globin (5'-actgggagaggcaaaggctggggggcca gagagg-3', located immediately before the "ATA" consensus sequence).

	"CACCC"		"CCAAT"						
	+120	+110	+100	+90	+80	+70	+60	+50	+40
sloth ε	ccagctc	CACCC	ctgaggg	acacagctc	cagccctg	CCAAT	gactgtg	:aagtaccaggggaacaaggggcc:::	agaggtacacagtg
armadillo ε	ctagctc	CACCC	ctgaagg	acacagctc	cagccctg	CCAAT	cactgca	:aagtaacagtgtacaagaggcc:::	agaatgtccacagt
human ε	ctgactc	CACCC	ctgagg	:acacaggtc	cagccctg	CCAAT	gactttt	:aagtaccatgggagaacagggggcca:	gaacttcgggcagta
dugong γ	ccgcccccccgcccg	gctgatag	cctagc	gttg	CCAAT	agcctcatag	caaaaggaagaaca	aaagggcc:::	agtggccagggata
manatee γ						TCATAGCAAAAGGAAGAACA	AAAGGGCC:::	AGTGGCCAGGGATA	
armadillo ψγ	atccttc	CCTTC	gtgaact	ccacaact	tagctttg	accaatagc	ctttt:::	tggaagaggcc:::	aggggacag:aaat
human Ay	caaacc	CACCC	atgggtt	ggccagc	cttgccttg	CCAAT	agattcatttc	cactgagggaggcaaaagggctggtcaatagattcatttc	
human ψη	caagttc	CACCC	ctggcagt	gaccac	tagctttg	CCAAT	agtctcatttt	tattgggggaagggaagggcct:::	ggggcagcagatg
goat η	taaactc	CACCC	:agccttg	acaaggc	aaacttg	CCAAT	agtcttagagt	atccagtgaggccaggggcggcgctggctagggatg	

	"ATA"		"CAP"						"ATG"
	+30	+20	+10	0	-10	-20	-30	-40	-50
sloth ε	aaga	ATA	aaaag	gccaca:tcttct	agaagcagcac	Agctctg	cttctg	acacgtctgtgatc	accagccagcatttagatcctacatcatg
armadillo ε	aaga	ATA	aaaag	gccaca:cctttt	agaagtagcac	Agacttg	cttctg	acaccactgtatcaccagcaag:ctctctgaatcttcgtcatg	
human ε	aaga	ATA	aaaag	gccagacagagaggca:g	cagcac	Atatctg	cttccg	acacagctgcaatcactagcaagctctcaggcctggcatcatg	
dugong γ	aaga	ATA	aaaag	gccaca:tgttcc	agttgcagcac	Atacatc	cttctg	acacatctgtgatcaccacaaa:ctccaagacoggacacpctg	
manatee γ	aaga	ATA	aaaag	gccaca:tgttcc	agttgcagcac	Atacatc	cttctg	acacatctgtgatcaccacaaa:ctccaagacoggacacpctg	
armadillo ψγ	gaag	ATA	aaaag	gccaca:aatttc	agtagcagtac	Aaatctg	cttctg	aca:::tatg	
human Ay	aaga	ATA	aaaag	gaagcacccttcagca:g	ttccac	Acactcg	cttctg	gaacgtctgaggttatcaataagctcctagtcagacgccatg	
human ψη	agga	GTA	aaaag	gccaca:catgaagca:g	caatgc	Aggc	atgcttctg	gctcatctgtgatcaccaggaaactcccagatctgacactatg	
goat η	agga	ATA	aaaag	gccatg:cagtgaagcagcggcac	Agacttg	cttctg	gccattatggatcaccagtaagctcccaga:::	caccatg	

Fig. 2-4. Dotplot comparisons (window size = 100; threshold = 80) of the putative elephant shrew ϵ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis). Sequences run from the 5' to 3' direction. Exons 2 (partial sequence) and 3 for each gene are denoted by black boxes.

Elephant shrew ϵ -globin partial gene comparisons

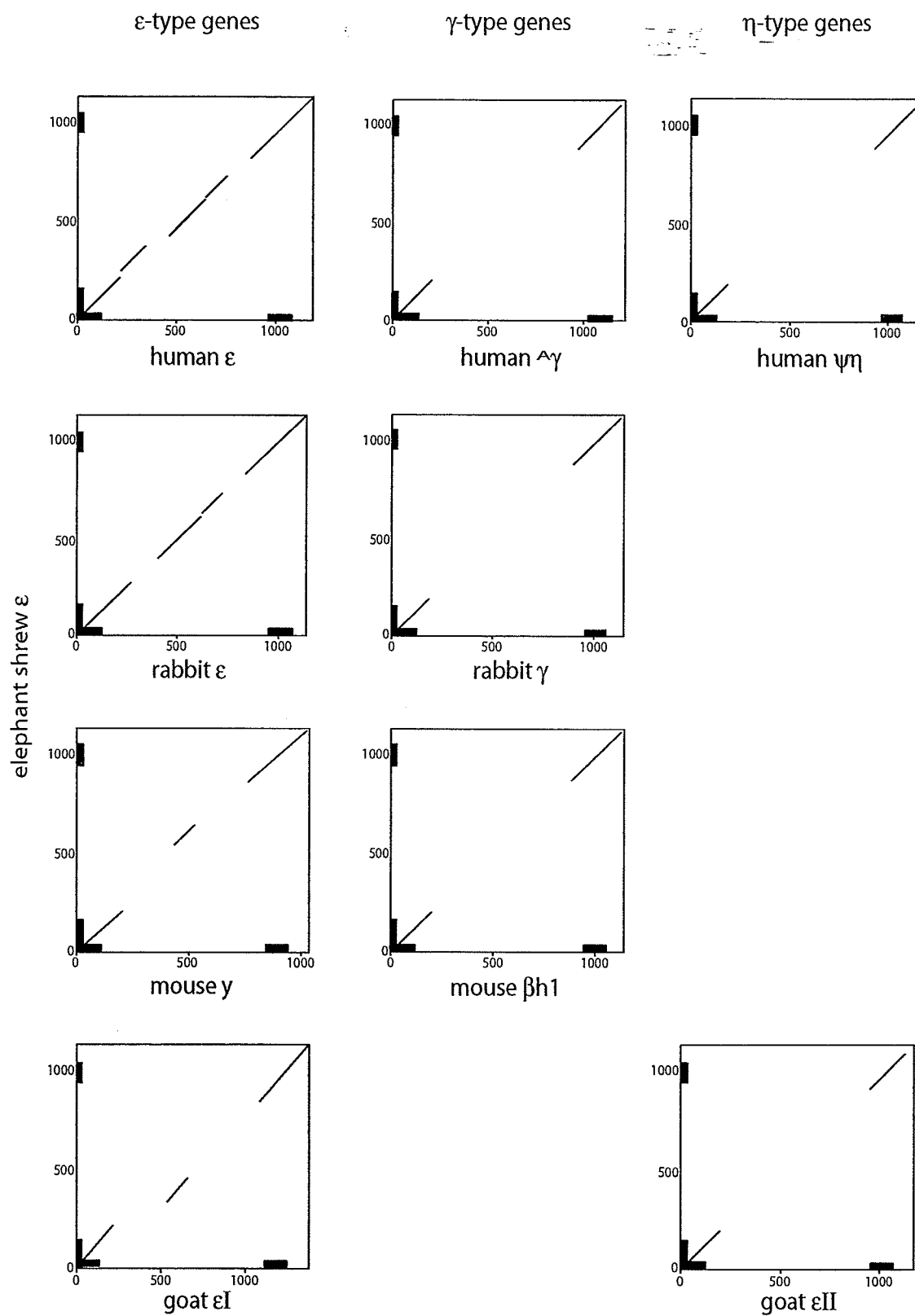


Fig. 2-5. Dotplot comparisons (window size = 100; threshold = 80) of the putative sloth ϵ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Sloth ϵ -globin gene comparisons

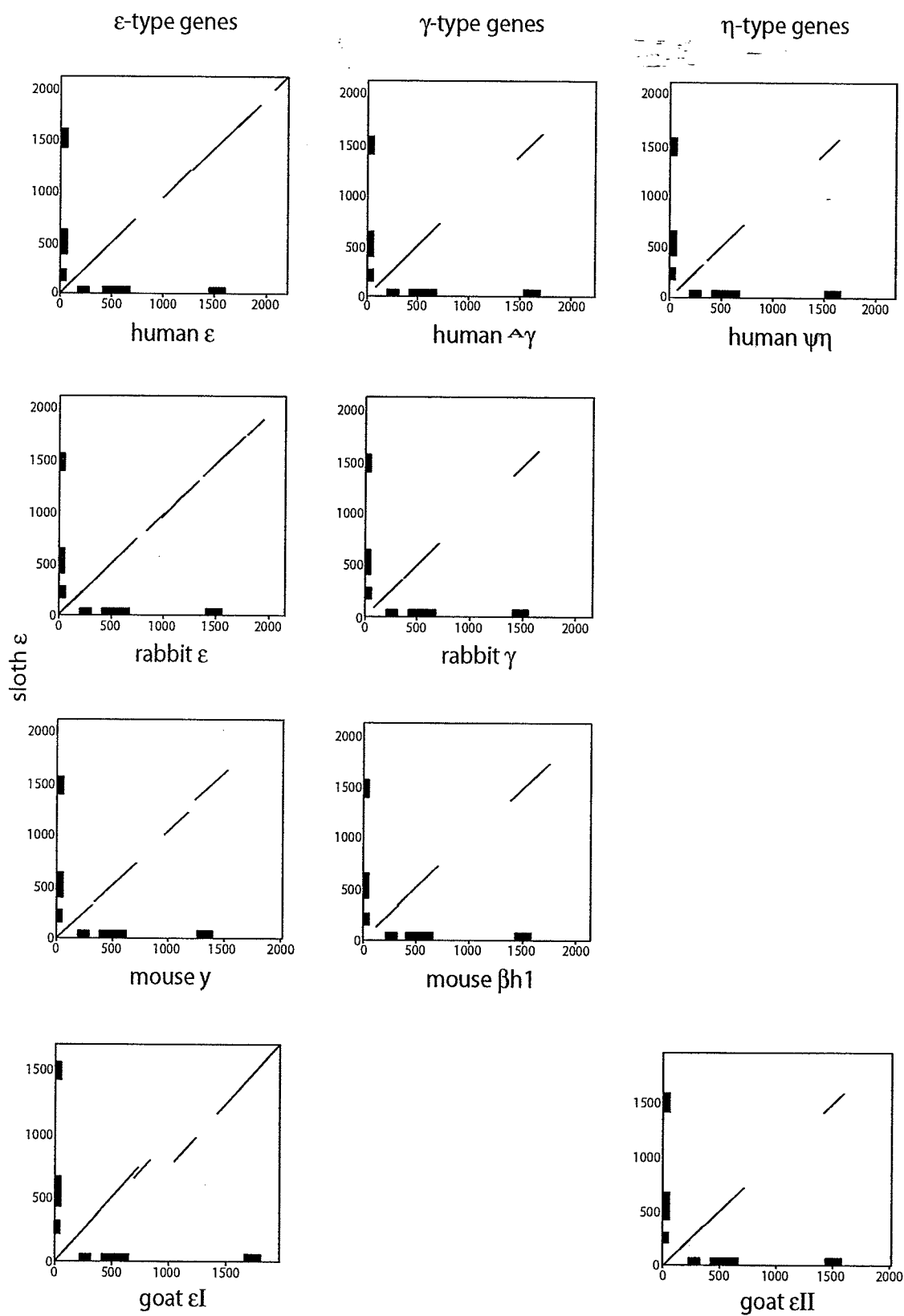
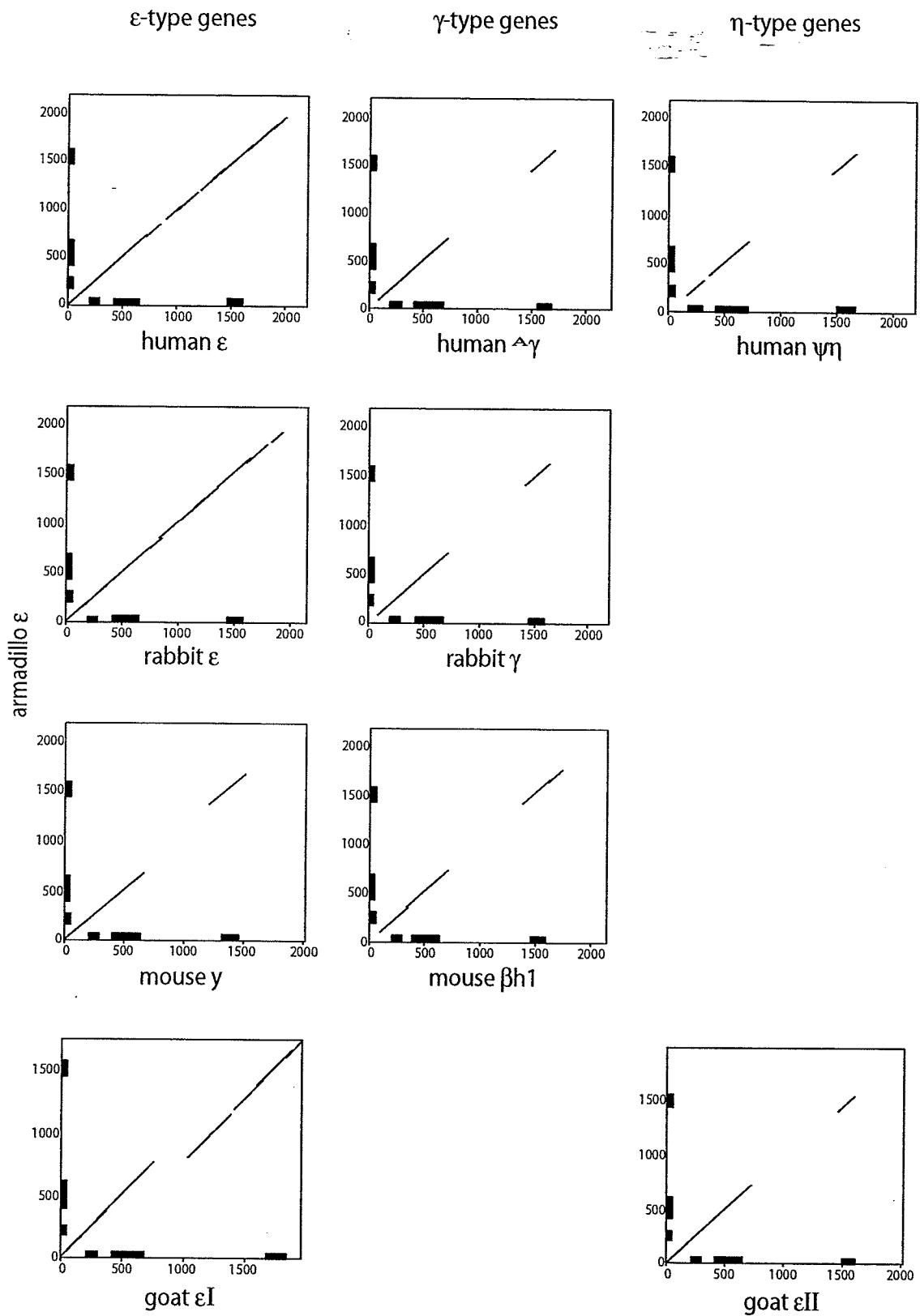


Fig. 2-6. Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo ϵ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Armadillo ϵ -globin gene comparisons



necessary for transcription (Fig. 2-3, Appendices 2-4, 2-5). A γ -like gene (Fig. 2-7, Tables 2-2, 2-3), was located 16,667 bp downstream of the ϵ -globin stop codon, the intergenic distance representing the number of nucleotides between termination codon of the ϵ -locus and the initiation codon of the γ -locus. This γ -like globin gene contained frameshift mutations in both exons 1 and 2 causing premature termination signals in each coding block. Examination of the 5' gene flanking region of this gene also revealed a CACCC box mutation ($\rightarrow$ CCTTC) and a 36-bp deletion immediately upstream of the initiation codon (Fig. 2-3). Additionally, the 3' end contained a putative polyadenylation signal +32 bp from the stop codon, roughly 30 bp before the usual site of functional γ -globin genes (Appendix 2-5). A 2,163-bp ' η -like' gene fragment was found 1,753 bp downstream from the putative $\psi\gamma$ -globin termination codon, the intergenic distance corresponding to the number of bases between the $\psi\gamma$ -gene's final stop signal and point where homology began with other eutherian η -globin genes. This fragment was found to share highest sequence homology within only the IVS2, exon 3 and 3' flanking regions of the human $\psi\eta$ - and goat η -globin genes (Fig. 2-8, Tables 2-3, 2-4).

Examination of mutation rates

Nonsynonymous replacement rates amongst the eutherian β -globin cluster genes were typically found to be lowest for the ' ϵ -like' loci and highest for the ' β -like' loci within each species (Table 2-5). Compared to the average nonsynonymous rate of mammalian β -globin cluster genes (1.0×10^{-9} substitutions/site/year; Efstratiadis *et al.* 1980), notably high replacement rates were observed for the γ -loci of rabbit and mouse. Replacement rates for ' β -like' globin genes of all examined species were similar to the

Fig. 2-7. Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo $\psi\gamma$ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice and goats (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Armadillo $\psi\gamma$ -globin gene comparisons

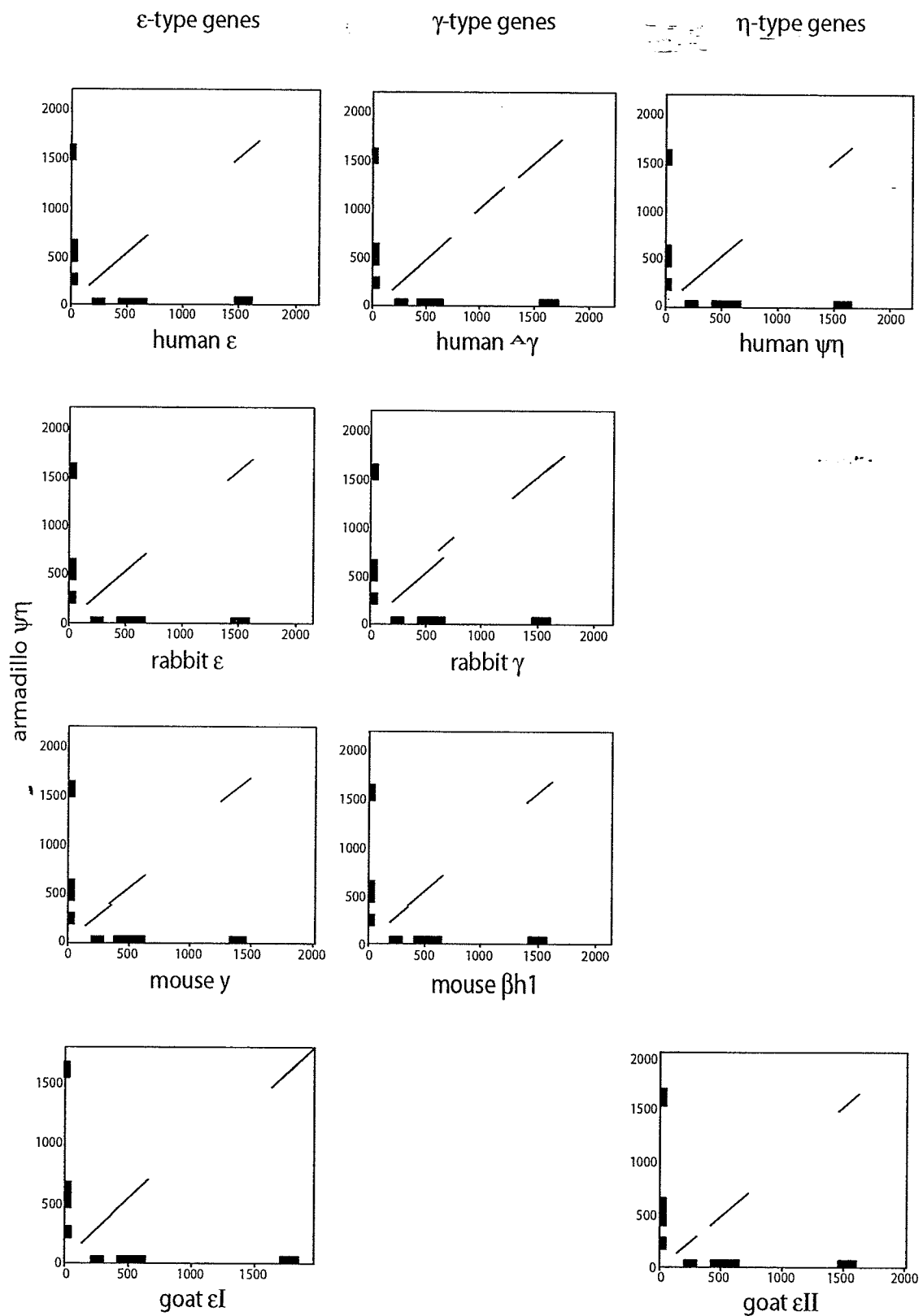


Fig. 2-8. Dotplot comparisons (window size = 100; threshold = 80) of the putative armadillo $\psi\eta$ -globin gene (Y-axis) with the ϵ -, γ - and η -globin genes of humans, rabbits, mice, and goats (X-axis). Sequences run from the 5' to 3' direction. Exons 1, 2 and 3 for each gene are denoted by black boxes.

Armadillo $\psi\eta$ -globin partial gene comparisons

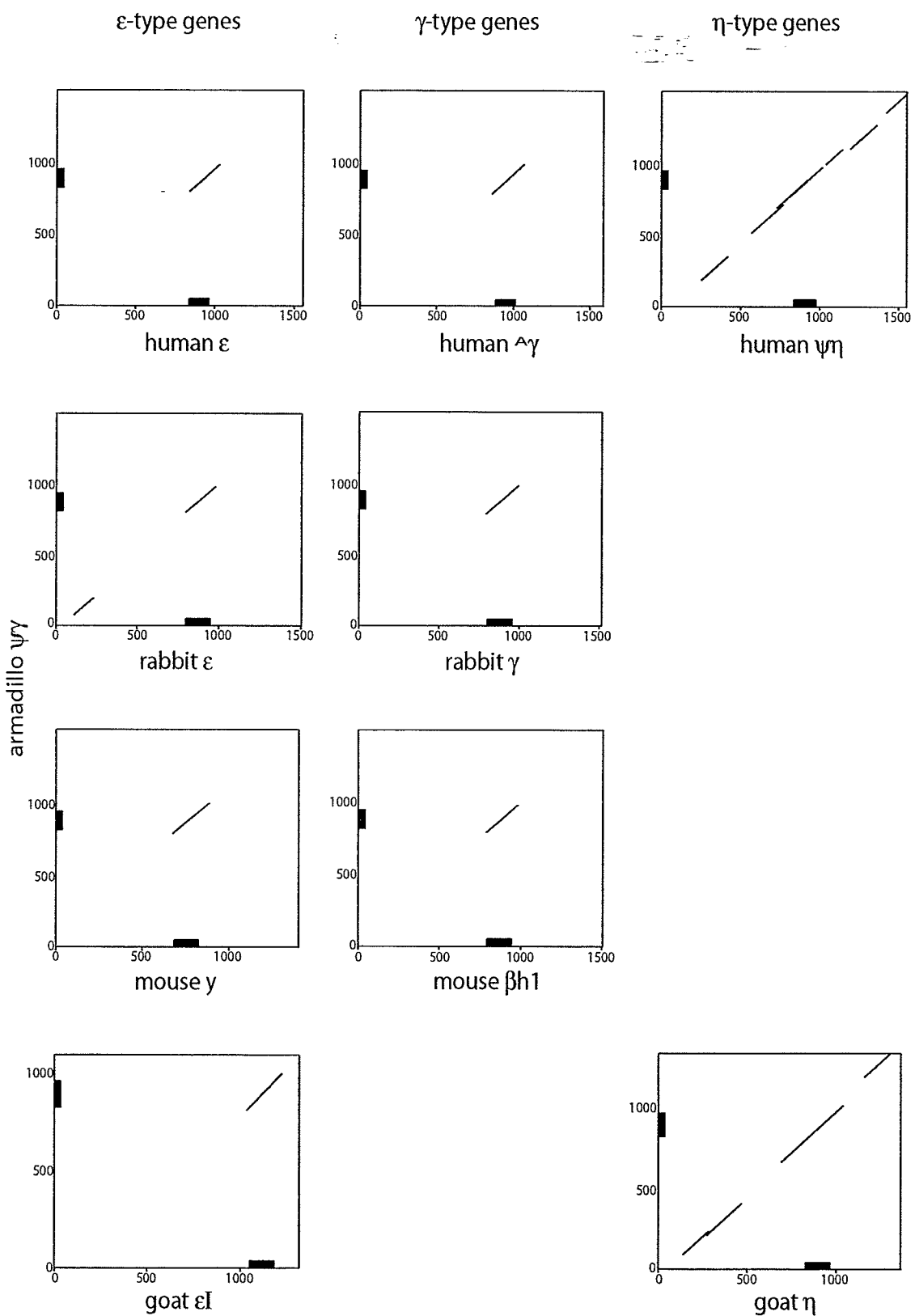


Table 2-5. Percent of synonymous (silent) and nonsynonymous (replacement) substitutions between the 'ε-like' and 'β-like' globin genes amplified in this study (plus those of armadillo, rabbit, mouse, goat, and tarsier) and their ortholog in humans. Silent and non-silent evolution rates are also shown.

Taxon	% synonymous substitutions (dS)	Silent mutation rate (no. silent mutations/site/yr)	% nonsynonymous substitutions (dN)	Non-silent mutation rate (no. non-silent mutations/site/yr)	dN / dS
African/Asian elephant γ	43.37	4.05×10^{-9}	8.07	7.54×10^{-10}	0.186
African/Asian elephant δ	37.40	3.50×10^{-9}	14.71	1.37×10^{-9}	0.393
dugong γ	41.39	3.87×10^{-9}	10.93	1.02×10^{-9}	0.264
dugong δ	38.25	3.57×10^{-9}	15.34	1.43×10^{-9}	0.401
manatee γ	37.47	3.50×10^{-9}	10.75	1.00×10^{-9}	0.287
manatee δ	36.31	3.39×10^{-9}	12.58	1.18×10^{-9}	0.346
rock hyrax δ ^H	56.89	5.32×10^{-9}	19.73	1.84×10^{-9}	0.347
yellow-spotted hyrax δ ^H	57.77	5.40×10^{-9}	18.88	1.76×10^{-9}	0.327
tree hyrax δ ^H	55.56	5.19×10^{-9}	19.29	1.80×10^{-9}	0.347
sloth ε	49.83	4.89×10^{-9}	7.70	7.55×10^{-10}	0.155
sloth β	51.34	5.03×10^{-9}	15.32	1.50×10^{-9}	0.298
armadillo ε	60.82	5.96×10^{-9}	7.01	6.87×10^{-10}	0.115
armadillo β	37.80	3.71×10^{-9}	14.77	1.45×10^{-9}	0.391
rabbit ε	55.14	6.34×10^{-9}	7.98	9.17×10^{-10}	0.145
rabbit γ	55.79	6.41×10^{-9}	13.14	1.51×10^{-6}	0.236
rabbit β1	34.39	3.95×10^{-9}	5.35	6.15×10^{-10}	0.156
mouse γ (ε-like)	57.79	6.64×10^{-9}	9.16	1.05×10^{-9}	0.159
mouse γ	66.93	7.69×10^{-9}	14.63	1.68×10^{-6}	0.219
mouse β2	48.41	5.56×10^{-9}	13.28	1.53×10^{-9}	0.274
goat εI	42.18	4.96×10^{-9}	5.11	6.01×10^{-10}	0.121
goat β ^A	34.10	4.01×10^{-9}	11.94	1.40×10^{-9}	0.350
tarsier ε	33.79	4.39×10^{-9}	5.25	6.82×10^{-10}	0.155
tarsier γ	40.73	5.30×10^{-9}	12.14	1.58×10^{-9}	0.298
tarsier δ	27.29	3.54×10^{-9}	7.46	9.69×10^{-10}	0.273
tarsier β	40.91	5.31×10^{-9}	4.69	6.09×10^{-10}	0.115

average replacement rate (1.0×10^{-9} substitutions/site/year) in each species except rabbit and tarsier, whose 'β-like' loci showed markedly low rates (Table 2-5). For each β-globin cluster gene examined, the rate of synonymous (silent) evolution was found to be higher than the nonsynonymous mutation rate (Table 2-5).

With few exceptions, non-coding evolution rates were lowest for the 'ε-like' loci and highest for the 'β-like' loci (Table 2-6). These 'neutral' rates were also found to be relatively similar between the orthologs of different species (Table 2-6). The non-coding rates of examined genes were comparable to the non-coding mutation rates of other eutherian α- and β-globin cluster genes ($4 - 5 \times 10^{-9}$ substitutions/site/year; Li *et al.* 1981; Hayashida and Miyata 1983; Kimura 1983) (Table 2-6).

Table 2-6. Percent of nucleotide substitutions between the IVS2 sequences of 'ε-like' and β-like' globin genes amplified in this study (plus those of armadillo, rabbit, mouse, goat and tarsier) and their ortholog in humans. The corresponding non-coding DNA mutation rates are also shown.

Taxon	% Non-coding DNA (IVS2) substitutions	Non-coding DNA substitution rate
African/Asian elephant γ	40.1	3.75 x 10 ⁻⁹
African/Asian elephant δ	45.1	4.21 x 10 ⁻⁹
dugong γ	39.8	3.72 x 10 ⁻⁹
dugong δ	43.2	4.04 x 10 ⁻⁹
manatee γ	39.9	3.73 x 10 ⁻⁹
manatee δ	42.4	3.96 x 10 ⁻⁹
rock hyrax δ	47.6	4.45 x 10 ⁻⁹
rock hyrax δ ^H	47.4	4.43 x 10 ⁻⁹
yellow-spotted hyrax δ ^H	46.8	4.37 x 10 ⁻⁹
tree hyrax δ ^H	46.8	4.37 x 10 ⁻⁹
sloth ε	43.8	4.29 x 10 ⁻⁹
sloth ψδ	44.5	4.36 x 10 ⁻⁹
sloth β	53.4	5.24 x 10 ⁻⁹
armadillo ε	42.3	4.15 x 10 ⁻⁹
armadillo γ	41.5	4.07 x 10 ⁻⁹
armadillo ψη	43.0	4.22 x 10 ⁻⁹
armadillo ψδ	46.9	4.60 x 10 ⁻⁹
armadillo β	51.4	5.04 x 10 ⁻⁹
rabbit ε	39.6	4.55 x 10 ⁻⁹
rabbit γ	42.2	4.85 x 10 ⁻⁹
rabbit β1	52.8	6.07 x 10 ⁻⁹
mouse γ	51.4	5.91 x 10 ⁻⁹
mouse γ	52.2	6.00 x 10 ⁻⁹
mouse β2	52.7	6.06 x 10 ⁻⁹
goat εI	42.1	4.95 x 10 ⁻⁹
goat εII (η)	44.0	5.18 x 10 ⁻⁹
goat β ^A	49.8	5.86 x 10 ⁻⁹
tarsier ε	35.2	4.57 x 10 ⁻⁹
tarsier γ	46.4	6.03 x 10 ⁻⁹
tarsier ψη	29.6	3.84 x 10 ⁻⁹
tarsier δ	41.9	5.44 x 10 ⁻⁹
tarsier β	29.8	3.87 x 10 ⁻⁹

DISCUSSION

The discovery of ϵ - and γ -loci (this study) together with δ - and β -loci (see Chapter I) within members of the basal eutherian clade Afrotheria (Fig. 2-9) provides compelling support that at least a four-gene cluster pre-dates the divergence of this group from the placental mammal tree (~107 mya; Springer *et al.* 2003). In addition, the presence of five globin gene loci (ϵ , $\psi\gamma$, $\psi\eta$, $\psi\delta$, β) within the β -globin cluster of the nine-banded armadillo (Fig. 2-9) further suggests that a five-gene set situated 5'- ϵ - γ - η - δ - β -3' on the chromosome was already established prior to the radiation of the three remaining superorders (~102 mya; Springer *et al.* 2003). It is noteworthy that molecular phylogenetic studies have been unable to reject a basal Afrotheria/Xenarthra clade (Madsen *et al.* 2001; Delsuc *et al.* 2002; Lin *et al.* 2002). Hence if these two clades diverged concurrently from other placentals, the data suggests that the five-gene β -globin cluster does indeed predate the radiation of eutherian mammals.

Although no η -globin gene products were amplified from the afrotherian templates, a ϵ -locus was amplified from the elephant shrew, and γ -globin gene products were obtained from the African elephant, Asian elephant, dugong and manatee. The absence of frameshift mutations and high sequence similarity between the two elephant (98.7%) and sirenian (95.9%) γ -globin genes, respectively, suggested the γ -locus is under conservation and possibly expressed within these paenungulate species. Support for this contention arises from the observation that blood from a 5-month old African elephant fetus possesses two distinct hemoglobin components, while blood from a 12-month old fetus and an adult each contain a single hemoglobin component which migrate at the same rate (Kleihauer *et al.* 1965; Braunitzer *et al.* 1984). These studies thus suggest that

Fig. 2-9. The β -globin gene clusters of select mammals, updated using data collected for this study (see Fig. 1-1 legend for details). Genes that presumably exist within each cluster are denoted by grey shading and broken-lined edges.

the minor (<15%) and major hemoglobin components of early fetal elephants are encoded by the γ - and δ -loci, respectively, with δ -globin assuming sole expression prior to 12 months gestation (the gestational period being 21 months). In contrast to elephants, the blood of Amazonian manatee (*Trichechus inunguis*) calves possess one major and one minor (<5%) band (Farmer *et al.* 1979), while adult blood contains a single chain constituent (Kleinschmidt *et al.* 1986; Kleinschmidt and Braunitzer 1988). Thus within sirenians, fetal hemoglobin presumably also contains two components encoded by the γ - (minor) and δ - (major) loci, with γ -globin expression extending into the early stages of post-natal life. Support for this inference is provided by the 5' flanking region of the dugong γ -globin gene, which contained no CACCC transcriptional control motif (typically located ~90 bases upstream from the Cap site of transcribed ' β -like' globin genes; Myers *et al.* 1986). This γ -globin promoter sequence has been found to be crucial for suppressing β -globin expression during pre-natal development *in vitro* (Sargent *et al.* 1999) and during fetal development of transgenic mice (Perez-Stable and Constantini 1990). Hence, the 'lingering existence' of γ -globin in the hemoglobin of Amazonian manatee calves may occur because γ -expression is not immediately suppressed after birth as it does not significantly affect expression of the globin gene encoding the major post-natal hemoglobin component (i.e. δ -globin; see Chapter I). Because anthropoid primates are the only eutherians known to express the γ -locus during post-embryonic stages of development, these findings are of considerable interest.

The tandem duplication events producing the four or five gene β -globin cluster of placental mammals is thought to have enabled the evolution of specialized forms of hemoglobin physiologically-adapted for different developmental stages (Barrie *et al.*

1981; Goodman *et al.* 1984). For instance, as an adaptive strategy to an extended fetal life (Koop and Goodman 1988) mice and rabbits express their γ - and ϵ -loci during early and late embryogenesis (Hansen *et al.* 1982; Hardison 1981; Hardison 1983; Hill *et al.* 1984), respectively. Similarly, higher primates have extended their γ expression to fetal growth stages (Shen *et al.* 1981), while artiodactyls have recruited their 3'-most β -locus for fetal development (Townes *et al.* 1984). Within the armadillo, however, only the two outermost loci (ϵ and β) appear to be capable of encoding a complete polypeptide (Fig. 2-9), and thus possess a functional arrangement similar to the ancestral β -cluster of marsupials and (presumably) monotremes (Lee 1997; Lee *et al.* 1999). Notably, hemoglobin of adult armadillos contains two different ' β -like' polypeptide chains expressed at ratios of 3:7 (de Jong *et al.* 1981) to 1:9 (Dementi and Burke 1972), while only one ' β -like' gene capable of encoding a protein exists within its β -cluster. This consequently raises the possibility that the second chain is encoded by the ϵ -locus, though significantly, no ϵ -globin gene of any eutherian is known to be expressed after the completion of embryogenesis. Interestingly, marsupials possess a ' β -like' globin gene (ω -globin) within their α -globin cluster, which constitutes <25% of both their late pre- and early post-natal hemoglobin (Wheeler *et al.* 2001; Wheeler *et al.* 2004). Thus, the possibility that armadillos may express an unlinked ' β -like' globin gene, which may be orthologous with marsupial ω -globin, can not be discounted.

The discovery of a putatively functional ϵ -globin gene within the pale-throated sloth suggests that orthologs of the ancient ϵ -globin gene have been maintained and expressed in all members of Xenarthra. While I have documented $\psi\delta$ - and β -globin loci

from this species (see Chapter I), it is still unknown whether they have retained functional γ - and η -loci.

The comparative analyses of evolution rates amongst the 'ε-like' and 'β-like' globin genes of the four eutherian superorders supports the contention that descendants of the ancestral β-globin gene family have evolved at different rates throughout evolutionary time. Furthermore, these differential rates can be primarily attributed to the different selective pressures placed upon the genes therein (Aguileta *et al.* 2004). Nonsynonymous (replacement) evolution rates were found to be lower in the 'ε-like genes" compared to the 'β-like' genes within each species (Table 2-5), reflecting stabilizing selection as little room exists for variation of ambient conditions during gestational development (Koop and Goodman 1988). Replacement rates for the γ -loci of dugong and manatee (1.02 and 1.00×10^{-9} , respectively; Table 2-5) were nearly identical to the average nonsynonymous replacement rate determined for β-globin cluster genes (1×10^{-9} substitutions/site/year; Efstradiatis *et al.* 1980), while replacement rates were slightly lower for the γ -orthologs of both elephant species (7.54×10^{-10} ; Table 2-5). The former finding supports the earlier conjecture that sirenian γ -globin genes are expressed both pre- and post-natally, as these genes would have diverged faster in order to encode a protein structure suited for two (pre- and early post-natal) developmental stages, while the latter suggests stabilizing selection. The high degree of sequence divergence observed between the γ -globin gene sequences of humans and laurasiatherians (Table 2-5) is likely due to the structural differences allowing the orthologs to function at different developmental times (i.e. during fetal and embryonic development, respectively; Hardison 1984). Thus, it is significant that replacement rates observed between the γ -loci of paenungulates and

humans are relatively similar to the average replacement rate of orthologous (and typically analogous) gene sequences (1×10^{-9} substitutions/site/year), lending further support to the notion that paenungulate γ -loci are expressed during fetal growth stages.

Because the δ - and β -globin genes of paenungulates and xenarthrans encode their major post-natal hemoglobin components, respectively (see Chapter I), it is not surprising the mutation rates among these genes were fairly similar. However, these rates (which were also similar to those of the mouse and goat ' β -like' loci) were slightly higher than the average β -globin cluster gene replacement rate (Efstradiatis *et al.* 1980) while those of rabbit and tarsier were much lower (Table 2-5). This large degree of variation most likely reflects lineage-specific factors such as population structure and generation time (Wu and Li 1985; Britten 1986; Li and Tanimura 1987).

The discovery of differing evolution rates for each gene in the β -globin family lends credence to the notion that phylogenetic studies using molecular clocks (which assume all hemoglobin genes evolve at equal rates) are based upon unsound assumptions (Goodman 1981; Hardison 1984; Giebel 1985; Li *et al.* 1987; Bailey *et al.* 1991). I speculate that, although the branching patterns derived from clock-based analyses of mammalian β -globin gene families are most likely in the correct order, dates estimating divergence times may need to be revised as information is gained on these various rates of evolution.

In accordance with earlier findings (Shapiro *et al.* 1983; Hardison 1984; Hardies *et al.* 1984; Giebel *et al.* 1985; Slightom *et al.* 1985; Koop *et al.* 1989b), the data further demonstrate marked heterogeneity between the non-coding evolution rates of sequenced ' ϵ -like' and ' β -like' globin genes of the same species. Specifically, descendants of the

' ϵ -like' loci were shown to diverge at neutral sites more slowly than the ' β -like' loci within every eutherian species but the tarsier (Table 2-6). The slow rate of neutral evolution previously observed in various primate β -globin cluster genes (Giebel *et al.* 1985; Harris *et al.* 1986; Koop *et al.* 1989b) was not fully supported by the analyses, as only the $\psi\eta$ - and β -loci of tarsier diverged at a rate comparatively lower than the neutral mutation rate of pseudogenes ($4 - 5 \times 10^{-9}$ substitutions/site/year; Hayashida and Miyata 1983; Kimura 1983; Li *et al.* 1981). Further challenging the 'hominid slowdown' theory (Li *et al.* 1987; Hasegawa *et al.* 1989; Bailey *et al.* 1991) is the observation that no links were found between species generation time and rate of neutral evolution for any loci but γ . As the 'neutral' mutation rates were relatively similar between orthologs of different species (with the exception of tarsier), it is notably apparent that underlying mechanisms (possibly owing to species-specific differences in DNA mutation rates, localized structural constraints, etc.; Wu and Li 1985; Koop *et al.* 1989b) are acting on these 'unconstrained' gene sequences within each species. Consequently, the referral to non-coding mutation rates as 'paradigms of neutral evolution' (Li *et al.* 1981; Kimura 1983) should clearly be reconsidered.

In conclusion, the results of this study confirm that the β -globin cluster of stem eutherians consisted of at least four, and probably five genes. And, with the exception of primates, descendants of the 5'-most ϵ -locus have typically evolved at a slower rate than genes derived from the β -locus in both coding and non-coding regions throughout evolutionary time. Furthermore, the data suggest that the γ -globin gene is expressed until at least 1 year of age in sirenians, but silenced relatively early (<5 months) in elephant development. Finally, the progenitor of modern day nine-banded armadillos atavistically

regressed to a two-gene cluster (5'- ϵ - β -3'), suggesting the potential flexibility afforded by the five-gene expansion did not provide any selective advantages within this lineage.

LITERATURE CITED

- Aguileta, G., J. P. Bielawski, and Z. Yang. 2004. Gene conversion and functional divergence in the β -globin gene family. *J. Mol. Evol.* **59**:177-189.
- Amrine, H., and M. Springer. 1999. Maximum-likelihood of the tethythere hypothesis based on a multigene data set and a comparison of different models of sequence evolution. *J. Mamm. Evol.* **6**:161-176.
- Amrine-Madsen, H., K. P. Koepfli, R. K. Wayne, and M. Springer. 2003. A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. *Mol. Phylogenet. Evol.* **28**:225-240.
- Antoine, M., and J. Niessing. 1984. Intronless globin genes in the insect *Chironomus thumi thumi*. *Nature* **310**:795-798.
- Antoniou, M., E. deBoer, G. Habets, and F. Grosveld. 1988. The human β -globin gene contains multiple regulatory regions: identification of one promoter and two downstream enhancers. *EMBO J.* **7**:377-384.
- Archibald, D. 2003. Timing and biogeography of the eutherian radiation: fossils and molecules compared. *Mol. Phylogenet. Evol.* **28**:350-359.
- Bailey, W. J., D. H. A. Fitch, D. A. Tagle, J. Czelusniak, J. L. Slightom, and M. Goodman. 1991. Molecular evolution of the $\psi\eta$ -globin gene locus: gibbon phylogeny and the hominid slowdown. *Mol. Biol. Evol.* **8**:155-184.
- Baralle, F. E., and G. G. Brownlee. 1978. AUG is the only recognizable signal sequence in the 5' non-coding regions of eukaryotic mRNA. *Nature* **274**:84-87.
- Barrie, P. A., A. J. Jeffreys, and A. F. Scott. 1981. Evolution of the β -globin gene cluster in man and the primates. *J. Biol. Evol.* **149**:319-336.

- Boyer, S. H., E. F. Crosby, A. N. Noyes, G. F. Fuller, S. E. Leslie, L. J. Donaldson, G. R. Vrablik, E. W. Schaefer Jr., and T. F. Thurmon. 1971. Primate hemoglobins: some sequences and some proposals concerning the character of evolution and mutation. *Biochem. Genet.* **5**:405-448.
- Braunitzer, G., W. Jelkmann, A. Stangl, B. Schrank, and C. Krombach. 1982. Hemoglobins, XLVIII: the primary structure of hemoglobin of the Indian elephant (*Elephas maximus*, Proboscidea): $\beta 2 = \text{Asn}$. Hoppe-Seyler's Z. Physiol. Chem. **363**:683-691.
- Braunitzer, G., A. Stangl, B. Schrank, C. Krombach, and H. Wiesner. 1984. Phosphate-hemoglobin interaction. The primary structure of the hemoglobin of the African elephant (*Loxodonta africana*, Proboscidea): asparagine in position 2 of the β -chain. Hoppe-Seyler's Z. Physiol. Chem. **365**:743-749.
- Britten, R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. *Science* **231**:1393-1398.
- Buettner-Janusch, J., and V. Buettner-Janusch. 1963. Hemoglobins of two elephant shrews. *Nature* **199**:918-919.
- Burmester, T., B. Ebner, B. Weich, and T. Hankeln. 2002. Cytooglobin: a novel globin type ubiquitously expressed in vertebrate tissues. *Mol. Biol. Evol.* **19**:416-421.
- Burmester, T., B. Weich, S. Reinhardt, and T. Hankeln. 2000. A vertebrate globin expressed in the brain. *Nature* **407**:520-523.
- Cooper, S., and R. 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., R. Murphy, G. Dolman, D. Hussey, and R. 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.
- Czelusniak, J., D. Goodman, M. Hewett-Emmett, M. Weiss, P. Venta, and R. Tashian. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian hemoglobin genes. *Nature* **298**:297-300.
- de Jong, W., A. Zweers, and M. Goodman. 1981. Relationship of armadillo to elephants, hyraxes, and sea cows from α -crystallin sequences. *Nature* **292**:538-540.
- Delsuc, F., M. Scally, O. Madsen, M. Stanhope, W. de Jong, F. Catzeflis, M. Springer, and E. Douzery. 2002. Molecular phylogeny of living xenarthans and the impact of character and taxon sampling on the placental tree rooting. *Mol. Biol. Evol.* **19**:1656-1671.
- Dementi, P. L., and J. D. Burke. 1972. Oxyhemoglobin affinity in the armadillo. *Amer. J. Anat.* **134**:509-514.
- Dickerson, R.E., and I. Geis. 1983. Hemoglobin: structure, function, evolution, and pathology. Benjamin/Cummings, Menlo Park, California.
- Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissmann. 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β -globin gene in mouse 3T6 cells. *Cell* **23**:695-706.
- Donze, D., P. H. Jeancake, and T. M. Townes. 1996. Activation of δ -globin gene expression by erythroid Kruppel-like factor: a potential approach for gene therapy of sickle cell disease. *Blood* **88**:4051-4057.

- Douady, C., F. Catzeflis, D. Kao, M. Springer, and M. Stanhope. 2002. Molecular evidence for the monophyly of Tenrecidae (Mammalia) and the timing of the colonization of Madagascar by Malagasy tenrecs. *Mol. Phylogenet. Evol.* **22**:357-363.
- Douady, C., F. Catzeflis, J. Raman, M. Springer, and M. Stanhope. 2003. The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). *Proc. Natl. Acad. Sci. USA* **100**:8325-8330.
- Efstratiadis A, J. W. Posakony, T. Maniatis et al. (15 co-authors). 1980. The structure and evolution of the human β -globin gene family. *Cell* **21**:654-668.
- Eizirik, E., W. Murphy, and S. O'Brien. 2001. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* **92**:212-219.
- Farmer, M., R. Weber, J. Bonaventura, R. Best, and D. Domning. 1979. Functional properties of hemoglobin and whole blood in an aquatic mammal, the Amazonian manatee (*Trichechus inunguis*). *Comp. Biochem. Physiol.* **62**:231-238.
- Feng, Y. Q., R. Warin, T. Li, E. Olivier, A. Besse, A. Lobell, H. Fu, C. M. Lin, M. I., Aladjem, and E. E. Bouhassira. 2005. The human β -globin locus control region can silence as well as activate gene expression. *Mol. Cell Biol.* **25**:3864-3874.
- Fitch, D. H., W. J. Bailey, D. A. Tagle, M. Goodman, L. Sieu, and J. L. Slightom. 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. USA* **88**:7396-7400.

- Fleming, M., J. Potter, C. Ramirez, G. Ostrander, and E. Ostrander. 2003. Understanding missense mutations in the BRCA1 gene: an evolutionary approach. *Proc. Natl. Acad. Sci. USA* **100**:1151-1156.
- Forrester, W. C., S. Takegawa, T. Papayannopoulou, G. Stamatoyannopoulos, and M. Groudine. 1987. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucl. Acids Res.* **15**:10159-10177.
- Fritsch, E. F., R. M. Lawn, and T. Maniatis. 1980. Molecular cloning and characterization of the human β -like globin gene cluster. *Cell* **19**:959-972.
- Giebel, L. B., V. L. van Santen, V. L., Slightom, J. L., and R. A. Spritz. 1985. Nucleotide sequence, evolution, and expression of the fetal globin gene of the spider monkey, *Ateles geoffroyi*. *Proc. Natl. Acad. Sci. USA* **82**:6985-6989.
- Goodman, M. 1981. Globin evolution was apparently very rapid in early vertebrates: a reasonable case against the rate-constancy hypothesis. *J. Mol. Evol.* **17**:114-120.
- Goodman, M., B. Czelusniak, D. Koop, D. Tagle, and J. Slightom. 1987. Globins: a case study in molecular phylogeny. *Cold Spring Harb. Symp. Quant. Biol.* **52**:875-890.
- Goodman, M., B. Koop, J. Czelusniak, M. Weiss, and J. 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., and G. Moore. 1975. Darwinian evolution in the genealogy of hemoglobin. *Nature* **253**:603-608.

- Grosveld, G. C., A. Rosenthal, and R. A. Flavell. 1982. Sequence requirements for the transcription of the rabbit β -globin gene in vivo: the -80 region. *Nucl. Acids Res.* **10**:4951-4971.
- Grosveld, F., G. B. van Assendelft, D. R. Greaves, and G. Kollias. 1987. Position-independent, high-level expression of the human β -globin gene in transgenic mice. *Cell* **51**:975-985.
- Hansen, J. N., D. A. Konkel, and P. Leder. 1982. The sequence of a mouse embryonic β -globin gene: evolution of the gene and its signal region. *J. Biol. Chem.* **257**:1048-1052.
- Hardies, S., H. Edgell, and C. Hutchison III. 1984. Evolution of the mammalian β -globin gene cluster. *J. Biol. Chem.* **259**:3748-3756.
- Hardison, R. C. 1981. The nucleotide sequence of rabbit embryonic globin gene $\beta 3$. *J. Biol. Chem.* **256**:11780- 11786.
- Hardison, R. C. 1983. The nucleotide sequence of the rabbit embryonic globin gene $\beta 4$. *J. Biol. Chem.* **258**:8739-8744.
- 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, and R. 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 $\psi\beta 2$ is a hybrid of δ - and β -globin gene sequences. *Mol. Biol. Evol.* **1**:302-316.

- Harris, S., P. Barrie, M. Weiss, and A. Jeffreys. 1984. The primate $\psi\beta 1$ 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. Nucleotide sequence analysis of the lemur β -globin gene family: evidence for major rate fluctuations in globin polypeptide evolution. *Mol Biol. Evol.* **3**:465-484.
- Hasegawa, M. H., H. Kishino, and T. Yano. 1989. Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominidae. *J. Hum. Evol.* **18**:461-476.
- Hayashida, H., and T. Miyata. 1983. Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA* **80**:2671-2675.
- Hedges, S. 2001. Afrotheria: plate tectonics meets genomics. *Proc. Natl. Acad. Sci. USA* **98**:1-2.
- Hill, A., S. C. Hardies, S. J. Phillips, M. G. Davis, C. A. Hutchison III, and M. H. Edgell. 1984. Two mouse early embryonic β -globin gene sequences: evolution of the nonadult β -globins. *J. Biol. Chem.* **259**:3739-3747.
- Hutchison III, C. A., S. C. Hardies, R. W. Padgett, S. Weaver, and M. H. Edgell. 1984. The mouse globin pseudogene $\beta h3$ is descended from a premammalian δ -globin gene. *J. Biol. Chem.* **259**:12881-12889.
- Jeffreys, A. J., P. A. Barrie, S. Harris, D. H. Fawcett, Z. J. Nugent, and A. C. Boyd. 1982. Isolation and sequence analysis of a hybrid δ -globin pseudogene from the brown lemur. *J. Mol. Biol.* **156**:487-503.

- Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21-132 in H. N. Munro, ed. Mammalian protein metabolism. Academic Press, New York.
- Kimura, M. 1983. Rate variant alleles in the light of the neutral theory. *Mol. Biol. Evol.* **1**:84-93.
- Kleihauer, E., I. O. Buss, C. P. Luck, and P. G. Wright. 1965. Hemoglobins of adult and fetal African elephants. *Nature* **207**:424-425.
- Kleinschmidt, T., and G. Braunitzer. 1983. The primary structure of hemoglobins of the rock hyrax (*Procavia habessinica*, Hyracoidia): insertion of glutamine in the α -chains. *Hoppe Seylers Z. Physiol. Chem.* **364**:1303-13.
- Kleinschmidt, T., and G. Braunitzer. 1988. The primary structure of the hemoglobin of the Brazilian manatee (*Trichechus inunguis*, Sirenia). *Hoppe-Seyler's Z. Physiol. Chem.* **389**:427-435.
- Kleinschmidt, T., J. Czelusniak, M. Goodman, and G. Braunitzer. 1986. Paenungulata: a comparison of the hemoglobin sequences from elephant, hyrax, and manatee. *Mol. Biol. Evol.* **3**:427-435.
- Kleinschmidt, T., J. Marz, and G. Braunitzer. 1989. The primary structure of pale-throated three-toed sloth (*Bradypus tridactylus*, Xenarthra) hemoglobin. *Biol. Chem. Hoppe-Seyler* **370**:303-308.
- Konkel, D. A., J. V. Maizel Jr., and P. Leder. 1979. The evolution and sequence comparison of two recently diverged mouse chromosomal β -globin genes. *Cell* **18**:865-873.

- Koop, B., and M. Goodman. 1988. Evolutionary and developmental aspects of two hemoglobin β -chain genes (ϵ M and β M) of opossum. *Proc. Natl. Acad. Sci. USA* **85**:3893-3987.
- Koop, B., F. D. Siemieniak, J. L. Slightom, M. Goodman, J. Dunbar, P. C. Wright, and E. L. Simons. 1989a. Tarsius δ - and β -globin genes: conversions, evolution, and systematic implications. *J. Biol. Chem.* **265**:68-79.
- Koop, B. F., D. A. Tagle, M. Goodman, and J. L. Slightom. 1989b. A molecular view of primate phylogeny and important systematic and evolutionary questions. *Mol. Biol. Evol.* **6**:580-612.
- Kosche, K. A., C. Dobkin, and A. Bank. 1985. DNA sequences regulating human β -globin gene expression. *Nucl. Acids Res.* **13**:7781-7793.
- Lacy, E., R. C. Hardison, D. Quon, and T. Maniatis. 1979. The linkage arrangement of four rabbit β -like globin genes. *Cell* **18**:1273-1283.
- Lavergne, A., E. Douzery, T. Stichler, F. Catzeflis, and M. Springer. 1996. Interordinal mammalian relationships: evidence for the paenungulate monophyly is provided by complete mitochondrial 12S rRNA sequences. *Mol. Phylogenet. Evol.* **6**:245-258.
- Lee, M.-H. 1997. Molecular biology and evolution of β -globin genes in monotremes. Doctoral thesis, University of Adelaide, Australia.
- Lee, M.-H., R. Scroff, S. Cooper, and R. Hope. 1999. Evolution and molecular characterization of a β -globin gene from the Australian echidna *Tachyglossus aculeatus* (Monotremata). *Mol. Phylogenet. Evol.* **12**:205-214.
- Li, W.-H., Gojobori, T., and M. Nei. 1981. Pseudogenes as a paradigm of neutral evolution. *Nature* **292**:237-239.

- 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.-H., M. Tanimura, and P. M. Sharp. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* **25**:330-342.
- Lin, Y.-H., P. McLenachan, A. Gore, M. Phillips, R. Ota, M. Hendy, and D. Penny. 2002. Four new mitochondrial genomes and the increased stability of evolutionary trees of mammals from improved taxon sampling. *Mol. Biol. Evol.* **19**:2060-2070.
- Lingrel, J. B., T. M. Townes, S. G. Shapiro, S. E. Spence, P. A. Liberator, and S. M. Wernke. 1983. Organization, structure, and expression of the goat globin genes. *Prog. Clin. Biol. Res.* **134**:131-139.
- Liu, R., and M. Miyamoto. 1999. Phylogenetic assessment of molecular and morphological data for eutherian mammals. *Syst. Biol.* **48**:54-64.
- Liu, F. G., M. M. Miyamoto, N. P. Freire, P. Q. Ong, M. R. Tennant, T. S. Young, and K. F. Gugel. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. *Science* **291**:1786-1789.
- Madsen, O., M. Scally, C. Douady, D. Kao, R. DeBry, R. Adkins, H. Amrine, M. Stanhope, W. de Jong, and M. Springer. 2001. Parallel adaptive radiation in two major clades of placental mammals. *Nature* **409**:610-614.
- Madsen, O., D. Willemssen, B. Ursing, U. Arnason, and W. de Jong. 2002. Molecular evolution of the mammalian alpha 2B adrenergic receptor. *Mol. Biol. Evol.* **19**:2150-2160.

- Malia, M. J. Jr., R. M. Adkins, and M. W. Allard. 2002. Molecular support for Afrotheria and the polyphyly of Lipotyphla based on analyses of the growth hormone receptor gene. *Mol. Phylogenet. Evol.* **24**:91-101.
- Margot, J. B., G. W. Demers, and R. C. Hardison. 1989. Complete nucleotide sequence of the rabbit β -like globin gene cluster. Analysis of intergenic sequences and comparison with the human β -like globin gene cluster. *J. Mol. Biol.* **205**:15-40.
- Martin, S., K. Vincent, and A. Wilson. 1983. Rise and fall of the δ -globin gene. *J. Mol. Biol.* **164**:513-528.
- McKenna, M.C. 1975. Toward a phylogenetic classification of the Mammalia. Pp. 1-4 in W.P. Luckett and F.S. Szalay, eds. *Phylogeny of the primates*. Plenum Press, New York.
- McKenna, M., and S. Bell. 1997. *Classification of Mammals Above the Species Level*. Columbia Univ. Press, New York.
- Murata, Y., M. Nikaido, T. Sasaki, Y. Cao, Y. Fukumoto, M. Hasegawa, and N. Okada. 2003. Afrotherian phylogeny as inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* **28**:253-260.
- Murphy, W., E. Eizirik, W. Johnson, Y. Zhang, O. Ryder, and S. O'Brien. 2001a. Molecular phylogenetics and the origins of placental mammals. *Nature* **409**:614-618.
- Murphy, W., E. Eizirik, S. O'Brien, O. Madsen, M. Scally, C. Douady, E. Teeling, O. Ryder, M. Stanhope, W. de Jong, and M. Springer. 2001b. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* **294**:2348-2351.
- Myers, R., M. K. Tilly, and T. Maniatis. 1986. Fine structure genetic analysis of a β -globin promoter. *Science* **232**:613-618.

- Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **11**:613-619.
- Nikaido, M., H. Nishihara, Y. Hukumoto, and N. Okada. 2003. Ancient SINEs from African endemic mammals. *Mol Biol. Evol.* **20**:522-527.
- Nishihara, H., Y. Satta, M. Nikaido, J. G. M. Thewissen, M. J. Stanhope, and N. Okada. 2005. A retroposon analysis of afrotherian phylogeny. *Mol. Biol. Evol.* (in press).
- Nowak, R. 1999. Walker's mammals of the world, Sixth Edition (Vol.2). The Johns Hopkins University Press. Baltimore.
- Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin, New York.
- Perez-Stable, C., and F. Constantini. 1990. Roles of fetal γ -globin promoter elements and the adult β -globin 3' enhancer in the stage-specific expression of globin genes. *Mol. Cell. Biol.* **10**:1116-1125.
- Pesce, A., M. Bolognesi, and A. Bocedi. 2002. Neuroglobin and cytoglobin. *EMBO* **3**:1146-1151.
- Phillips, S., J. S. C. Hardies, C. L. Jahn, M. H. Edgell, and C. A. Hutchison III. 1984. The complete nucleotide sequence of a β -globin-like structure, β h2, from the [Hbb]^d mouse BALB/c. *J. Biol. Chem.* **259**:7947-7954.
- Prinsloo, P. 1993. Molecular and chromosomal phylogeny of the Hyracoidea. Doctoral thesis, University of Pretoria, South Africa.
- Poyart, C., H. Wajcman, and J. Kister. 1992. Molecular adaptation of hemoglobin function in mammals. *Resp. Physiol.* **90**:3-17.

- Proudfoot, N. J., and G. G. Brownlee. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**:211-214.
- Proudfoot, N., A. Gil, and T. Maniatis. 1982. The structure of the human ζ -globin gene and a closely linked, nearly identical pseudogene. *Cell* **31**:553-563.
- Prychitko, T., R. M. Johnson, D. E. Wildman, D. Gumucio, and M. Goodman. 2005. The phylogenetic history of New World monkey β -globin reveals a platyrrhine β to δ gene conversion in the atelid ancestry. *Mol. Phylogenet. Evol.* **35**:225-234.
- Ristaldi, M. S., S. Casula, S. Porcu, M. F. Marongiu, M. Pirastu, and A. Cao. 1999. Activation of the δ -globin gene by the β -globin gene CACCC motif. *Blood Cells Mol. Dis.* **25**:193-209.
- Ross, J., and A. Pizarro. 1983. Human β and δ globin messenger RNAs turn over at different rates. *J. Mol. Biol.* **167**:607-617.
- Sargent, T. G., C. C. DuBois, A. M. Buller, and J. A. Lloyd. 1999. The roles of 5'-HS2, 5'-HS3, and the γ -globin TATA, CACCC, and stage selector elements in suppression of β -globin expression in early development. *J. Biol. Chem.* **274**:11229-11236.
- Satoh, H., N. Inokuchi, N. Yasuhiro, and T. Okazaki. 1999. Organization, structure, and evolution of the nonadult rat β -globin gene cluster. *J. Mol. Evol.* **49**:122-129.
- Schimenti, J. C., and C. H. Duncan. 1985a. Concerted evolution of the cow ϵ^2 and ϵ^4 β -globin genes. *Mol. Biol. Evol.* **2**:505-513.
- Schimenti, J. C., and C. H. Duncan. 1985b. Structure and organization of the bovine β -globin genes. *Mol Biol. Evol.* **2**:514-525.

- Schon, E. A., M. L. Cleary, J. R. Haynes, and J. B. Lingrel. 1981. Structure and evolution of goat γ -, β^C - and β^A -globin genes: three developmentally regulated genes contain inserted elements. *Cell* **27**:359-369.
- Scott, A. F., P. Heath, S. Trusko, S. H. Boyer, W. Prass, M. Goodman, J. Czelusniak, L. Y. Chang, and J. L. Slightom. 1984. The sequence of the gorilla fetal globin genes: evidence for multiple gene conversions in human evolution. *Mol. Biol. Evol.* **1**:371-389.
- Shapiro, S. G., E. A. Schon, T. M. Townes, and J. B. Lingrel. 1983. Sequence and linkage of the goat ϵ^I and ϵ^{II} β -globin genes. *J. Mol. Biol.* **169**:31-52.
- Shehee, W. R., D. D. Loeb, N. B. Adey, F. H. Burton, N. C. Casavant, P. Cole, C. J. Davies, R. A. McGraw, S. A. Schichman, and D. M. Severynse. 1989. Nucleotide sequence of the BALB/c mouse β -globin complex. *J. Mol. Biol.* **205**:41-62.
- Shen, S. H., J. L. Slightom, and O. Smithies. 1981. A history of the human fetal globin gene duplication. *Cell* **26**:191-203.
- Shoshani, J., and M. McKenna. 1998. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Mol. Phylogenet. Evol.* **9**:572-584.
- Simpson, G.G. 1945. Principles of classification and a classification of mammals. *Bull. Am. Mus. Nat. Hist.* **85**:1-350.
- Singh, A. 2002. Phylogenetic relationships of paenungulate mammals inferred from α - and β -globin gene sequences. Honours thesis, B.Sc., University of Manitoba, Canada.

- Slightom, J. L., L. Y. Chang, B. F. Koop, and M. Goodman. 1985. Chimpanzee fetal G_{γ} - and A_{γ} -globin gene nucleotide sequences provide further evidence of gene conversions in hominid evolution. *Mol. Biol. Evol.* **2**:370-389.
- Smith, A., D. Smith, and B. Funnell. 1994. *Atlas of Mesozoic and Cenozoic coastlines*. Cambridge Univ. Press. Cambridge, U.K.
- Springer, M. 1997. Molecular clocks and the timing of the placental and marsupial radiations in relation to the Cretaceous-Tertiary boundary. *J. Mammal. Evol.* **4**:285-302.
- Springer, M., H. Amrine, A. Burk, and M. Stanhope. 1999. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition. *Syst. Biol.* **48**:65-75.
- Springer, M., G. Cleven, O. Madsen, W. de Jong, V. Waddell, H. Amrine, and M. Stanhope. 1997. Endemic African mammals shake the phylogenetic tree. *Nature* **388**:61-64.
- Springer, M., W. Murphy, E. Eizirik, and S. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci. USA* **100**:1056-1061.
- Stanhope, M., O. Madsen, V. Waddell, G. Cleven, W. de Jong, and M. Springer. 1998. Highly congruent molecular support for a diverse superordinal clade of endemic African mammals. *Mol. Phylogenet. Evol.* **9**:501-508.
- Steinberg, M. H., and J. G. Adams III. 1991. Hemoglobin A2: origin, evolution, and aftermath. *Blood* **78**:2165-2177.

- Tagle, D. A., B. F. Koop, M. Goodman, J. L. Slightom, D. Hess, and R. 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.
- Tagle, D. A., J. L. Slightom, R. T. Jones, and M. Goodman. 1991. Concerted evolution led to high expression of a prosimian primate δ -globin gene locus. *J. Biol. Chem.* **266**:7649-7480.
- Tagle, D. A., M. J. Stanhope, D. R. Siemieniak, P. Benson, M. Goodman, and J. L. Slightom. 1992. The β -globin gene cluster of the prosimian primate *Galago crassicaudatus*: nucleotide sequence determination of the 41-kb cluster and comparative sequence analyses. *Genomics* **13**:741-760.
- Tang, D. C., D. Ebb, R. C. Hardison, and G. P. Rodgers. 1997. Restoration of the CCAAT box or insertion of the CACCC motif activates δ -globin gene expression. *Blood* **90**: 421-427.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673-4680.
- 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.
- Trent III, J.T., and M. S. Hargrove. 2002. A ubiquitously expressed human hexacoordinate hemoglobin. *J. Biol. Chem.* **277**:19538-19545.

- 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.
- van Dijk, M., O. Madsen, F. Catzefflis, M. Stanhope, W. de Jong, and M. Page. 2001. Protein sequence signatures support the African clade of mammals. *Proc. Natl. Acad. Sci. USA* **98**:188-193.
- Vincent, K. A., and A. C. Wilson. 1989. Evolution and transcription of old world monkey globin genes. *J. Mol. Biol.* **207**:465-480.
- Waddell, P., and S. Shelley. 2003. Evaluating placental inter-ordinal phylogenies with novel sequences including RAG1, -fibrinogen, ND6, and mt-tRNA, plus MCMC-driven nucleotide, amino acid, and codon models. *Mol. Phylogenet. Evol.* **28**:197-224.
- Weber, R. 1995. Hemoglobin adaptations to hypoxia and altitude – the phylogenetic perspective. Pp. 31-44 *in* J. Sutton, C. Houston, and G. Coates, eds. *Proceedings from the ninth international hypoxia symposium, Lake Louise, Canada, Burlington, Vermont.*
- Weber, R., and R. Wells. 1989. Hemoglobin structure and function. Pp. 279-300 *in* S. Wood, ed. *Comparative pulmonary physiology.* Marcel Dekker, Inc. New York.
- Wheeler, D., R. M. Hope, S. J. B. Cooper, G. Dolman, G. C. Webb, C. D. K. Bottema, A. A. Gooley, M. Goodman, and R. A. B. Holland. 2001. An orphaned mammalian β-globin gene of ancient evolutionary origin. *Proc. Natl. Acad. Sci. USA* **98**:1101-1106.

- Wheeler, D., R. M. Hope, S. J. B Cooper, A. A. Gooley, and R. A. B. Holland. 2004. Linkage of the β -like omega-globin gene to α -like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters. *J. Mol. Evol.* **56**:642-652.
- White, J., D. Harkness, R. Isaacks, and D. Duffield. 1976. Some studies on blood of the Florida manatee, *Trichechus manatus latirostris*. *Comp. Biochem. Physiol.* **55**:413-417.
- Wood, W. G., J. M. Old, A. V. Roberts, J. B. Clegg, and D. J. Weatherall. 1978. Human globin gene expression: control of β , δ and $\delta \beta$ chain production. *Cell* **15**:437-446.
- 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.

Appendix 1-1. Primers used in the (A) amplification and (B) sequencing reactions of 'β-like' globin genes.

(A) Primers used in polymerase chain reactions.

Primer name	Sequence
HBB-3(F)	5' TTG CTT CTG ACA CAA CTG TGT T 3'
HBB-4(F)	5' AAA CAG ACA CCA TGG TGC 3'
HBB-5(R)	5' AGT GGT ACT TGT GGG CC 3'
HBB-6(F)	5' GGT TGT CTA CCC CTG GAC 3'
HBB-7(R)	5' TTC TCG GGA TCC ACG TGC 3'
HBB-9	5' CTG ACT GCT GCT GAG AAG ACT CAG G 3'
HBB-9(F)-ARM	5' CTG ACT TCT GAT GAG AAG ACT GCG G 3'
HBB-10	5' CCC TTG AGG TTG TCC AGG TGC 3'
HBB-11(F)	5' CTG TCC TGC ACA ACG CTA AAG TGC 3'
HBB-12(R)	5' GCC ACC ACC TTC TCA TAG GCA GC 3'
HBB-13(R)	5' GGG CTT CCG ATG GAT ACC 3'
HBB-15(F)	5' CAC CCT GGC CTC GGC CAA TCT 3'
HBB-16(R)	5' TTA TGC CTT TGA AGT AGA ATT GTT C 3'
HBB-18(F)	5' TTC TGA CAC AAC TGT GTT CAC TAG CAA CTA 3'
HBB-19(R)	5' GAG CTG AAA GTT CTT TAT TAG GCA 3'
HBB-20(F)-STH	5' GCT GAC GAT GAG AAG GCT GGT GTA 3'
HBB-ARM-B(R)	5' TAC ACC CAA GCA ACG TAG TAG TGA ACC 3'
HBB-ARM-B/D(R)	5' GGG CTT TTA CCT TAG CAT TTG CGA AC 3'
HBB-ARM-D(R)	5' ATG TGG AGG TAA AGC CAA CAG GTT A 3'
HBB-CGM-1(F)	5' CAG GCA GAC AAA TGAAAC TCA CTG 3'
HBB-A1-II16	5' CAG AGG AAG TCA GTG CCT GTC AGA AAC TGG 3'
HBB-B1-II45	5' ACT GTT CTG TCT CCA CAT GCC CAG CC 3'
HYR-B1-R	5' ACC GTT TTG TCT CCA CAG GCT TGG CC 3'
HYR-C1-R	5' CCT AGA TAC AAA CCT GCC AAG TGC CTC A 3'
HBB-C1-II74	5' CCC ACT CAA CCC TCC TTA AGT CTA CCT TGC 3'
ELE-A2-2I	5' TGT GTG GGC AAC ATA AAT GGG TTT C 3'
ELE-B2-2I	5' CCC ATT GTC TCA GAG GCC AAG CT 3'
HYR-2B2-2I	5' ATG TCT GTG TAT CTG TCT GTC TCT TCT CCA 3'
ELE-C2-2I	5' CTC CTG GGC AAT GTG CTG GTG A 3'
ESH-A(R)	5' TCA ATA CCT GTG AGA AGC CTG GAA AAC GG 3'
ESH-B(R)	5' CTC CAT ATT CCA GTC CCA TTC AAG CAT CC 3'
ESH-C(R)	5' CCT GAC CAC CAA GTT TAT CTA CGT CCA CC 3'
ESH-A(F)-2	5' TGC ACA AGA GTC AAG TGT GAG TGC AGC TTG 3'
ESH-A(F)	5' TGA TGC CCT AAT CTC GAT CAC TCT GCC 3'
ESH-C(F)	5' AGG CTG CTT TCC AGA AGG TGG TGA CT 3'

(B) Primers used in sequencing reactions at the University of Calgary DNA core laboratory.

Primer name	Sequence
M-13(F)-40	5' GTT TTC CCA GTC ACG AC 3'
M-13(R)	5' AAC AGC TAT GAC CAT G 3'

Appendix 1-2. DNA sequences of the ' β -like' globin genes amplified in this study. Exons are identified by CAPITAL letters, while introns and external gene sequences are represented by lower case letters. The 5' upstream transcriptional control motifs (CACCC, CCAAT, ATA) and 3' poly-adenylation signal (AATAAA) are underlined within the 5' and 3' flanking regions of each gene sequence, respectively. Within each coding block, translated amino acids are listed below their corresponding base triplet.

Appendix		Page
1-2A	African elephant δ -globin gene sequence	97
1-2B	Asian elephant δ -globin gene sequence	99
1-2C	Dugong δ -globin gene sequence	101
1-2D	West Indian manatee δ -globin gene sequence	104
1-2E	Rock hyrax δ -globin partial gene sequence	106
1-2F	Rock hyrax δ^H -globin gene sequence	108
1-2G	Yellow-spotted hyrax δ^H -globin gene sequence	110
1-2H	Western tree δ^H -globin gene sequence	112
1-2I	Bushveld elephant shrew β -globin partial gene sequence	114
1-2J	Bushveld elephant shrew δ -globin partial gene sequence	116
1-2K	Nine-banded armadillo $\psi\delta$ -globin gene sequence	117
1-2L	Nine-banded armadillo β -globin gene sequence	119
1-2M	Pale-throated sloth $\psi\delta$ -globin gene sequence	121
1-2N	Pale-throated sloth β -globin gene sequence	123

Appendix 1-2A
African elephant delta-globin gene sequence

1 gactagcaac tacccaatca gacaccATGG TGAATCTGAC TGCTGCTGAG AAGACACAAG
V N L T A A E K T Q
61 TCACCAACCT GTGGGGCAAG GTGAATGTGA AAGAGCTTGG TGGTGAGGCC CTGAGCAGgt
V T N L W G K V N V K E L G G E A L S R
121 ttgtatctag gttgcaagg agacttaagg aggggttgagt ggggctgggc atgtggagac

181 agaacagtct cccagtttct gacaggcact gacttcctct gcaccctgtg gtgctttcac

241 cttcagGCTG CTGGTGGTCT ACCCATGGAC CCGGAGGTTC TTTGAACACT TTGGGGACCT
L L V V Y P W T R R F F E H F G D L
301 GTCCACTGCT GAAGCTGTCC TGCACAACGC TAAAGTGCTG GCCCATGGCG AGAAAGTGTT
S T A E A V L H N A K V L A H G E K V L
361 GACCTCCTTT GGTGAGGGCY TGAAGCACCT GGACAACCTC AAGGGCACCT TTGCCGATCT
T S F G E G L K H L D N L K G T F A D L
421 GAGCGAGCTG CACTGTGACA AGCTGCACGT GGATCCTGAG AATTCAGGg tgagtctagg
S E L H C D K L H V D P E N F R
481 agacattcta ttttttcttt tcactttgta gtctttcact gtgattattt tgctttattg

541 aatttcctct gtatctcttt ttactcgact atgtttcatc atttagtgct ttttcaactt

601 ataccatttt gtattacttt tctttcaata ttcttccttt tttcctgact cacattcttg

661 ctttatatca tgctctttat ttaatttcct gcgtttttgc tcttgctctc cttttctct

721 aatttccttc cctctgaaca gtaccagat tgtgcatacc acctctcatc cactatttct

781 gcactggggc aaatccccac ccctcctcca tatgaggggt ggaaaggact gaatcaaaga

841 ggagaggatc atcgtgctgt tctagagtct gtgattcatt tcagacttga aggataactt

901 gaataatata aaatcaggag taaatggaga ggaaagtcag tatctgagaa tgaagatca

961 gaaggtcata gacgagatgg ggagcagaag ttactaagaa actgaccatt gtggctataa

1021 ttaatcactt aattagttaa ttaatatggt tgttatttat tcacgttttt catthttggtg

Appendix 1-2A

1081 ggagtaaatt tgggctagtg tgtgggcaac ataaatgggt ttcaccccat tgtctcagag

1141 gccaaagctgg attgctttgt taaccatgtc tgtgtatgta tctacctctt ccccatagCT
L

1201 CCTGGGCAAT GTGCTGGTGA TTGTCCTGGC CCGCCACTTT GGCAAGGAAT TCACCCAGAG
L G N V L V I V L A R H F G K E F T P D

1261 TGTTCAAGCT GCCTATGAGA AGGTTGTGGC TGGTGTGGCG AATGCCCTGG CTCACAAATA
V Q A A Y E K V V A G V A N A L A H K Y

1321 CCACTGAgat cctggcctgt tcctggtatc catcggaagc cccatttccc gagatgctat
H .

1381 ctctgaattt gggaaaataa tgccaactct caagggcatc tcttctgcct aataaagtac

1441 tttcagctca actttctgat tcattttattt ttttctcagt cactcttgtg gtgggggaag

1501 ttcccaaggc tctatggaca gagagctctt gtgccttata ggaaaagtgc aaggggaatt

1561 ggaaaataaa gggaaccata cacagatatt aatgggaaca attctacttc aaaggcataa

1621 agattgggat ggtttggcaa ataggatact ggtactacag ggattccatg ggcctcaggc

1681 ctaagacata gcccagggc taactttcag attcaattcc agaaattact cacaaaaataa

1741 tgga

Appendix 1-2B
Asian elephant delta-globin gene sequence

1 ttctgggcct cagtttcctc atttgtataa taacagaatt ggagagtaaa ttcttaagag

61 gcttaccagg ctgtaattct aaaaaaatg cataaataaa cttgccaaagg cagatgtttt

121 tagcagcaat tcctgaaaga aacgggacca ggagataagt agagaaagag tgaaggctctg

181 aatcaaact aataagacag tcccagactg tcaaggagag gtatggctgt catcattcag

241 gcctcaccct gcagaaccac accctggcct tggccaatct gtcacaaga gcaaaaaggg

301 caggaccagg gttgggcata taaggaagag tagtgccagc tgctgtttac actcattct

361 gacacaactg tgttgactag caactaccca atcagacacc ATGGTGAATC TGACTGCTGC
V N L T A A

421 TGAGAAGACA CAAGTCACCA ACCTGTGGGG CAAGGTGAAT GTGAAAGAGC TTGGTGGTGA
E K T Q V T N L W G K V N V K E L G G E

481 GGCCCTGAGC AGgtttgtat ctaggttgca aggtagactt aaggaggggtt gagtggggct
A L S R

541 gggcatgtgg agacagaaca gtctcccagt ttctgacagg cactgacttc ctctgcaccs

601 tgtggtgctt tcaccttcag GCTGCTGGTG GTCTACCCAT GGACCCGGAG GTTCTTTGAA
L L V V Y P W T R R F F E

661 CACTTTGGGG ACCTGTCCAC TGCTGACGCT GTCCTGCACA ACGCTAAAGT GCTGGCCCAT
H F G D L S T A D A V L H N A K V L A H

721 GGCGAGAAAG TGTTGACCTC CTTTGGTGAG GGCCTGAAGC ACCTGGACAA CCTCAAGGGC
G E K V L T S F G E G L K H L D N L K G

781 ACCTTTGCCG ATCTGAGCGA GCTGCACTGT GACAAGCTGC ACGTGGATCC TGAGAATTTT
T F A D L S E L H C D K L H V D P E N F

841 AGGgtgagtc taggagacac tctatTTTT cttttcactt tgtagtcttt cactgtgatt
R

901 attttgctta tttgaatttc ctctgtatct ctttttactc gactatgttt catcatttag

961 tgttttttca acttatacca ttttgtatta cttttctttc aatattcttc cttttttcct

1021 gactcacatt cttgctttat atcatgctct ttatttaatt tcctacgttt ttgctcttgc

Appendix 1-2B
Asian elephant delta-globin gene sequence

1081 tctccctttc tcctagtttc cttccctctg aacagtaccc aaattgtgca taccacctct

1141 cgtccactat ttctgcactg gggcaaatcc ccaccctcc tccatatgag gggtggaag

1201 gactgaatca aagaggagag gatcatgggtg ctgttctaga gtatgtgatt catttcagac

1261 ttgaaggata acttgaataa tataaaatca ggagtaaag gagaggaaag tcagtatctg

1321 agaatgaaag atcagaaggc catagacgag atggggagca gaagttacta agaaactgac

1381 cattgtggct ataattaatc acttaattag ttaattaata tgtttgttat ttattcacgt

1441 ttttcatttt ggtgggagta aatttgggct agtgtgtggg caacataaat gggtttcacc

1501 ccattgtctc agaggccaag ctggattgct ttgttaacca tgtctgtgta tgtatctacc

1561 tcttcccat agCTCCTGGG CAATGTGCTG GTGATTGTCC TGGCCCGCCA CTTTGGCAAG
L L G N V L V I V L A R H F G K

1621 GAATTCACCC CAGATGTTCA GGCTGCCTAT GAGAAGGTTG TGGCAGGTGT GGCGAATGCC
E F T P D V Q A A Y E K V V A G V A N A

1681 CTGGCTCACA AATACCACTG Agatcctggc ctgttcctgg tatccatcgg aagccccatt
L A H K Y H .

1741 tcccgagatg ctatctctga atttgggaaa ataatgccaa ctctcaaggg catctcttct

1801 gcctaataaa gtactttcag ctcaactttc tgattcattt attttttttc tcagtcactc

1861 ttgtggtggg ggaagttccc aaggctctat ggacagagag ctcttgtgcc ttataggaaa

1921 agttcaaggg aaattggaaa ataaaggga ccatacacag atattaatgg gaacaattct

1981 acttcaaagg cataaagatt gggaagggtt ggcaaatagg atactggtac tacagggtt

2041 ccatgggcct caggcctaag acatagcccc agggctaact ttcatgattca attccagaaa

2101 ttactcacia aataatgga

Appendix 1-2C
Dugong delta-globin gene sequence

1 tcccctttga aaccactgct ttgaaacaaa agtgttattc geaacttgtg cttataattg
61 atttttactt gtttggaccc ccaaagccta atatggtgac cttgcatata gcattcttca
121 tttatattta ttcagttttc agcgatcatg tatttggta aacgacaaaa aaaagaaaca
181 aacaaacaaa caaatgaatg aatattaatt ttcttaccag gagaatttaa tccaaatcag
241 gagaaagaag aaatgtttag aattgtcaca gatgtttcat tcattctgtc cttgtaatta
301 tcttggatat tctggaaaca caggaatagc tccatcctaa tactcctaaa ctgaaccaca
361 gtgggcaatt ctttctgcc ctctgggcct tggctcctc attcgtgtac taagagaact
421 ggagagtaga tgtctaagag accaggcagt tattctaaaa aaaacctcat aaataaactt
481 gccaaaggag atgttttcag aagtaattcc tgaagcaat gggacaaaa gataagtaga
541 ggaagagtga gggctctgaa tcaaactact aagavgggtgc cagactgcta aggacaggta
601 cggtgtcat cattcagacc tcacctgca gaaccacacc ctggcctygg ccaatctgyt
661 cccaggagca agaagggcag aaaccagggg tgggcgtaaa aggaagrgca ggactagctg
721 ctgcttactc ttacttctgg cacaactgtg ttgactagca actacacaat cagacwccAT
781 GGTGCATCTG ACTGCTGATG AGACGGCTTT GGTCACCGGC CTGTGGGCCA AGGTGAACGT
V H L T A D E T A L V T G L W A K V N V
841 GAAAGAATAC GGTGGTGARG CCCTGGGCAG gtttgtgtct aggttgaag gcaggcttaa
K E Y G G E A L G R
901 ggaggggttga ttggggctgg gcatgtggag acagaacagt ctcccagttt ctgacagtca
961 ctgactccct ctgtcccctg tgggtgctttc acccttcagG CTGTTGGTTG TCTACCCATG
L L V V Y P W
1021 GACCCAGAGG TTCTTTGAGC ACTTTGGGGA CCTGTCCTCT GCTTCGGCTG TCATGCACAA
T Q R F F E H F G D L S S A S A V M H N

Appendix 1-2C
Dugong delta-globin gene sequence

1081 CCCTAAAGTG AAGGCCCATG GCGAGAAAGT GTTGGCCTCC TTTGGTGACG GCCTGAAGCA
P K V K A H G E K V L A S F G D G L K H

1141 CCTGGACGAC CTCAAGGGCG CCTTTGCTGA GCTGAGTGCG CTGCACTGTG AGAAGTCGCA
L D D L K G A F A E L S A L H C E K S H

1201 CGTGGATCCT CAGAACTTCA AGgtgagtct aggagactct ctactttttt cttttcactt
V D P Q N F K

1261 tgtagtcttt cactgtgggt attttactta cttgcatttc ctccacatct ctttacttga

1321 ctgtatttga gaatttgggtg ctttttcaac ttacaccatt ttgtattatt ctttctttca

1381 atgttcttcc ttttctcttg tctcacagtc ttgctttata tcatgccctt tattttaattt

1441 cctgcctttt gcctctgtct ctctctcgat ctcctttttt cctaatttcc ttccccttaa

1501 acaataccca agttatgcat accacctctc atccactact tctgcgcttg ggcaaatcct

1561 actgctcttc catatgaggg ttagaaagga ctgaatcaaa gaggggagga tcatgggtgtt

1621 gttctagagt ctgagactcg ttttgaaatt gaaggataat ttgaataata taaaatcagg

1681 agtaaatggg aaggaaagtc tatatctgag aatggaagat cagaaggaca tattggagct

1741 gaggagcaga agttactaag agacagacca ttattactct aattaatcaa ttaattagtt

1801 aatgagtatg cttgctaatt tatttacttg ttttttaatt ttagtgggaa taagtttggg

1861 ccgtcagttt gggctactgt gtgggcaaca taaatgggtg tcagcccatt ttctcagagg

1921 ccaagctgga ttgctttggt aaccatgtct gtgtatctat ctacctcttc cccacagCTC
L

1981 CTGGGCAATA TGCTGGTGTG TGTGCTGTCC CGCCACTTGG GCAAGGAATT CTCTCCACAG
L G N M L V C V L S R H L G K E F S P Q

2041 GCACAGGCTG CCTATGAGAA GGTTGTGGCT GGGGTGGCGA ACGCCTTGGC TCACAAATAC
A Q A A Y E K V V A G V A N A L A H K Y

2101 CACTGAgatt ctggcctatt tcctgggtatc cactggaagc cctgtttccc tagatgctat
H .

Appendix 1-2C
Dugong delta-globin gene sequence

2161 ctctgaattt ggggaaataa tgcccactct caagggtata gcttctgcct aataaagaac
2221 tttcagctca actttctgat tcatttcatt tattttatctt cttactcttg aggtatcgaa
2281 agttccccaa ggctatatgg ataaggagtt attgtgtctc ataggaagag ttcaagtga
2341 gttagaaaat gaagggaccc atacacagct attaattgga ataattctac ttctaaggca
2401 taatggttgg ggagggttgg caaatagaga ggatgctggc cctacatgga ttccatagac
2461 cttaaaccta agacatagcc ccagggctaa cttccagatt ctattccgac aattactcac
2521 aaaataatgg gctcaattag agaagtcatt atgattgaaa acattgtttt gttttcttgc
2581 ttctacttc tcttgctgt acatttttagc ccacaatcta taccctttct agtctttct

Appendix 1-2D
West Indian manatee delta-globin gene sequence

1 aaaacagtgc cagactgctc aggacgggta cggtgtcat cattcagacc tcacctgca
61 gaaccacacc ctggccttgg ccaatctgct cccaggagca agaagggcag aaaccagggg
121 tgggcgtaaa aggaagagca gggctagctg ctgcttacac ttacttctgg cacaactgtg
181 ttgactaaca actacacaat cagacaccAT GGTGCATCTG ACTCCTGAAG AGAAGGCTTT
 V H L T P E E K A L
241 GGTCATCGGC CTGTGGGCCA AGGTGAACGT GAAAGAATAT GGTGGTGAGG CCCTGGGCAG
 V I G L W A K V N V K E Y G G E A L G R
301 gtttgatatct aggttgtaag gcaggcttaa ggaggggttg ttggggctgg gcatgtggag
361 acagaacagc ctcccagttt ctgacagtca ctgactccct ctgtcccctg tggtgctttc
421 acccgtcagG CTGTTGGTTG TCTACCCATG GACCCAGAGG TTCTTTGAGC ACTTTGGGGA
 L L V V Y P W T Q R F F E H F G D
481 CCTGTCCTCT GCTTCGGCTA TCATGAACAA CCCTAAAGTG AAGGCCCATG GCGAGAAAAGT
 L S S A S A I M N N P K V K A H G E K V
541 GTTCACCTCC TTTGGTGACG GCCTGAAGCA CCTGGAAGAC CTCAAGGGTG CTTTTGCTGA
 F T S F G D G L K H L E D L K G A F A E
601 GCTGAGTGAG CTGCACTGTG ACAAGTTGCA CGTGGATCCT GAGAACTTCA GGgtgagtct
 L S E L H C D K L H V D P E N F R
661 aggagacgct ctactttttt cttttcactt tgtagtcttt cactgtgggt attttactta
721 cttgcatttc ctccacatct ctttacttga ctgtatttga gaatttagtg ctttttcaac
781 ttacaccatt ttctattatt ttttctttca atattcttcc ttttttctg tctcacagtc
841 ttgctttata tcatgccctt tatttaattt cctgcctttt gcctctgtct ctctctccat
901 ctcttttttt tcctaatttt ctttcccctt aaacaatacc caaattatgc ataccaccte
961 tcatccacta cttctgcgct tgggcaaadc ctactgctct tccatatgag ggtagaag
1021 gactgaatca aagaggggag gatcatgggtg ttgttctaga gtctgagact catTTTTGAAA

Appendix 1-2D
West Indian manatee delta-globin gene sequence

1081 ttgaaggata atttgaataa tttaaaatca ggagtaaag gggaggaaag tctatatctg
1141 agaatgaaag atcagaagga catattgcmc tgaggagcag aagttactaa gagacagacc
1201 attattactc taattaatca attaattagt taatgaatat gcttgctaatt ttatttactt
1261 gttttttaat tttagtggga ataagtttgg gccatcagtt tgggctactg tgtgggcaac
1321 ataaatgggt gtcaccccat tttctcagag gccaaactgg attgctttgt taacctatgc
1381 tgtgtatcta tctacctctt cccacagCT CCTGGCAAT GTGCTGGTGT GTGTGCTGGC
L L G N V L V C V L A
1441 CCGCCACTTT GGCAAGGAAT TCTCTCCAGA GGCACAGGCT GCCTATCAGA AGGTTGTGGC
R H F G K E F S P E A Q A A Y Q K V V A
1501 TGGTGTGGCG AACGCCTTGG CTCACAATA CCACTGAgat tctggcctat ttcctggtat
G V A N A L A H K Y H .
1561 ccattggaag ccctgtttcc ctagatgcta tctctgaatt cggggaaata atgcccactc
1621 tcaagggtat agcttctgcc taataaagaa ctttcagctc agctttctga ttcattctcat
1681 ttattttatt cct

Appendix 1-2E
Rock hyrax delta-globin partial gene sequence

1 AAGGGCACCT TTGCCCAGCT GAGCGAGCTG CACTGTGACA AGCTGCATGT GGACCCTGAG
 K G T F A Q L S E L H C D K L H V D P E
61 AACTTCAGGg tgagtccagg ggatgctcta ctcttttcac tttgttttca ttccttgtag
 N F R
121 ttattttaact tccttgcatc tcctctgtat ctctacttac ttgattgcat tgcattcattt
181 agtacttttc caacttatac agcttgcatc attttttcatt caaacttcct ctttctttcc
241 tgactcacia tcttgcttta catcatgccc ttgtataaa cttatacctt tcgcctctct
301 ctgtctctct ctctctctct ctcttttttt cctaatttcc ttttctccat tatgcatggg
361 ccctctcact gactgcttct gcacttaggc caatcctact tgcctccat atcagggttg
421 ggaaggactg aatcaaaggg gacacgatca tgatgttatt ttagagtctg tgacacattt
481 tgtaactgaa gaataattgg aataaaacia aatcatgggt aaatggaagc gctaattgcat
541 atctgaggat gaaagatcag aaggctatat aacagatggg gagcagaagt ttttaagaga
601 cagatagcta ttactctagt caactaatga gtttattata tgtttgctta tttatttaca
661 tgtttttaat tttgggtggga ataagtttgg gctgtcagtt tgagctagtg tgctgtcagc
721 atagatggat gtcaccccat ttctacagat gccaaagctgg atcgctttgt taaccatgcc
781 tctgtatctg tctatccctt cccacagCT CCTGGGCAAT GTCTTGGTGG TTGTCCTGGC
 L L G N V L V V V L A
841 CCGGCACTTC CGTGAGGAGT TCACCCAGA TGTTCAAGCT GCCTTTCAGA AGGTTGTGAC
 R H F R E E F T P D V Q A A F Q K V V T
901 TGGTGTGGCA AATGCTTTGG CTCACAAATA CCAAGTAgat cctgttttct ggtatctatc
 G V A N A L A H K Y H .
961 agaagcactg tttccctagc tgtgatctct gaattcgggg aaataatgcc cactctcagg
1021 ggcatcactt ctgcctaata aagaactttc agctcaactt tatgattcat tttattttatt

Appendix 1-2E
Rock hyrax delta-globin partial gene sequence

1081 ttatatttatt tctactcact tttgaaatat gggagtgtcc tgaaaggctt acagataggg
1141 aacgcttggt tccaacagga agagttcaag ggaaattgag aaatgaaggg taccttacac
1201 agttgtaaat ttaatgggaa taacttgact cgtatgacat aaagattggg aaatttggca
1261 aatagagagg atgctgctac tcctgaaatt ccatggaagt caaacctta cacacagcat
1321 gaaggctact ttcaaattc ataatcaaca attatttaca aaataatagg ttcaataagg
1381 ggagtcatga ttattgaaac tattgttttc cttccttcct tcctacttct cttgtacagt
1441 ttaggtcata agctatgtcc ttgtaagctt tcctactcta tggaaccatc ctcttctggt
1501 gttagtccta aaactgttct gaaaataatc gtttctcctg ttgttttcat ttctctggga
1561 gttcagcgct catgtatcac ctgtaacttc tcttacctca ttccaatgga tttcctagta
1621 gtgttttaaa aatatttcc

Appendix 1-2F
Rock hyrax deltaH-globin gene sequence

```

1      gactagcaac tacatgctct ggcaccATGG TGCGCCTGAC TGATAGTGAG AAAGCTGAGG
                                     V  R  L  T  D  S  E  K  A  E
61     TCGTCTCCTT GTGGAGCAAA GTGGATGAGA AAATAATCGG CAGTGAGGCA CTTGGCAGgt
      V  V  S  L  W  S  K  V  D  E  K  I  I  G  S  E  A  L  G  R
121    ttgtatctag gccgtacgta aggcaggctg aacgaggatt gccccagggcc aagcctgtgg

181    agacaaaacg gtctcccagt ttctgacaga cactgagtcc ctctgccac tctgtatattt

241    caccctcag GCTGCTGGTC ATCTACCCAT GGA CTGAGGAG GTTCTTTGAA CACTTTGGGG
                                     L  L  V  I  Y  P  W  T  R  R  F  F  E  H  F  G
301    ACCTTTCCAC TGCTGACGCT ATCCTGAAAA ACCCTCGGGT ACAGGCCCAT GGAGAGAAAG
      D  L  S  T  A  D  A  I  L  K  N  P  R  V  Q  A  H  G  E  K
361    TGCTGTCCTC CTTTGGGGAG GGCCTGAATC ACCTGGACAA CCTCAGGGGC ACCTTTGCCC
      V  L  S  S  F  G  E  G  L  N  H  L  D  N  L  R  G  T  F  A
421    AGCTGAGCGA GCTGCACTGT GACGAGCTGC ATGTGGACCC TGAGAACTTC AGGgtgagtc
      Q  L  S  E  L  H  C  D  E  L  H  V  D  P  E  N  F  R
481    caggggatgc tctactcttt tcttctcact ttgttttcat tcagtgtggt tatttgacct

541    acttacttga atttctttca tgtcactttt tttcttgact gtatctcatc atttagtgct

601    ttttcrgctt atactacttt gcactgtttt tttctttcaa tattctttat ttttgtctta

661    taatcctgct ttatatcaag cccttttagct ttctacattt tgccctctctc tcttttccta

721    acaattccct tccccctaaa cagtactcaa attacgaatg ttacctctca tccagtgttc

781    atgtacttag gcaaatacctt ctggtcctcc gtatgagagg tggcaaggaa tgaatcaaag

841    atgcaaagag tgttatgttg ttctagagtt tgtgactcat tttgaaatta aaagataatt

901    tgaataatat aaaatcagga gtaaaaggaa aggaaaatca gtatctgggg atggaaagat

961    cagaagatca tataggtgat ggagagcaga agtttctaag aaacagacca ctattgctgc

1021   aactaagcaa ttagttagtt aatatacttg cttatttggt tccatgtttt tacttttgat

```

Appendix 1-2F
Rock hyrax deltaH-globin gene sequence

1081 gggartaaat tcgggccatc agtttggggc aacagaaatg ggtatcatcc cattttcaca

1141 gatgccaagc tggattgttt tattaacctat gtctgtatat ctatctgtct cttctccaca

1201 gCTCCTGGGC AATATCCTGG TGGTTGTCCT GGCCCGCCAT TATGGCAAGG AATTCACCCT
 L L G N I L V V V L A R H Y G K E F T L

1261 AGAGGTTTCTG GCTGCCTGTC AGAAGTTTGT GGCTGGTATG GCAAATGCCT TGGCTCACAA
 E V Q A A C Q K F V A G M A N A L A H K

1321 ATACCACTGA grtcctggac tgtttcctgg tatccactgg aagccctgtt tccctagatg
 Y H .

1381 tgacctctga gtttgtaaata agtgctcatt cgcaaggcca ttgcttctgc ctaataaaga

1441 accttcagct cgactttctg attcatttta tttattttat ttcttatgca ctttaggagt

1501 atgggagtgt cctgaaaagc ttaccgatag cgatctcttg tgtccacacag gcagagtccc

1561 agggaaattg gaaacggacg gataccttat gcagttgtta atttaattgg

Appendix 1-2G
Yellow-spotted hyrax deltaH-globin gene sequence

```

1      gactagcaac tacatgctct ggcaccATGG TGGCCTGAC TGATAGTGAG AAAGCTGAGG
                                     V  R  L  T  D  S  E  K  A  E
61     TCGTCTCCTT GTGGAGCAAA GTGGATGAGA AAATAATCGG CAGTGAGGCA CTTGGCAGgt
      V  V  S  L  W  S  K  V  D  E  K  I  I  G  S  E  A  L  G  R
-121   ttgtatctag gccgcaaggc aggtgaacg aggattgccc agggccaagc ctgtggagac

181    aaaacggtct cccagtttct gacagacact gagtcctct gcccactttg tattttcacc

241    cctcagGCTG CTGGTCATCT ACCCATGGAC TCAGAGGTTC TTTGAACACT TTGGGGACCT
      L  L  V  I  Y  P  W  T  Q  R  F  F  E  H  F  G  D  L
301    TTCCACTGCT GACGCTATCA TGAAAAACCC TCGAGTACAG GCCCATGGAG AGAAAGTGCT
      S  T  A  D  A  I  M  K  N  P  R  V  Q  A  H  G  E  K  V  L
361    GTCTTCCTTT GGGGAGGGCC TGAATCACCT GGACAACCTC AGGGGCACCT TTGCCAGCT
      S  S  F  G  E  G  L  N  H  L  D  N  L  R  G  T  F  A  Q  L
421    GAGCGAGCTG CACTGTGACG AGCTGCATGT GGACCCTGAG AACTTCAGGg tgagtccacg
      S  E  L  H  C  D  E  L  H  V  D  P  E  N  F  R
481    ggatgctcta ctcttttctt ctcactttgt ttcatcac tgtggttatt tgacctactt

541    acttgaattt ctttcatgtc actttttttc ttgactgtat ttcacatatt agtgcttttt

601    cagcttatac tactttgcac tgtttttttc tttcaatatt ctttattttt gtcttataat

661    cctgctttat atcaagccct ttagctttct acattttgcc tctctctctt ttcctaacaa

721    ttcccttccc cctaaacagt actcaaatta cgaatgttac ctctcatcca gtgtttatgt

781    acttaggcaa atccttcttg tcctccgtat gagagggtggc aaggaatgaa tcaaggatgc

841    aaagagtatt atgttgttct agagtttgtg actcattttg aaattaaaag ataatttgaa

901    taatataaaa tcaggagtaa aaggaaagga aatcaatat ctggggatga aagatcagaa

961    ggtcatatag gtgacggaga gcagaagttt ctaagaaaca gaccactatt gctgcaacta

1021   agcaattagt tagttaatat acttgcttat ttgtttacat gtttttactt ttgatgggaa

```

Appendix 1-2G
Yellow-spotted hyrax deltaH-globin gene sequence

1081 taaattcggg ccatcagttt ggggtaacag aaatgggtat catcccattt tcacagatgc

1141 caagctggat tgttttatta accatgtctg tgtatctgtc tgtctcttct ccacagCTCC
L

1201 TGGGCAATAT CCTGGTGGTT GTCCTGGCCC GCCATTATGG CAAGGAATTC ACCCTAGAGG
L G N I L V V V L A R H Y G K E F T L E

1261 TTCAGGCTGC CTGTCAGAAG TTTGTGGCTG GTATGGCAAA TGCCTTGGCT CACAAATACC
V Q A A C Q K F V A G M A N A L A H K Y

1321 ACTGAgatcc tggactgttt cctggtatcc attggaagcc ctgtttccct agatgtgacc
H .

1381 tctgagtttg taaatagtgc tcattcgcaa gggcattgct tctgcctaat aaagaacctt

1441 cagctcgact ttctgattca ttttatttat tttatttctt atgtacttta ggagtatggg

1501 agtgtcctga aaagcctacc gatagcgatc tcctgtgtcc cacaggcaga gtcccagggg

1561 aattggaaac ggacggatac cttatgcagt tgттаattta atgg

Appendix 1-2H
Western tree hyrax deltaH-globin gene sequence

```

1   catgctctgg caccATGGTG CGCCTGACTG ATAGTGAGAA AGCTGAGGTC GTCTCCTTGT
      V R L T D S E K A E V V S L
61  GGAGCAAAGT GGATGAGAAA ATAATCGGCA GTGAGGCACT TGGCAGgttt gtatctaggc
      W S K V D E K I I G S E A L G R
121 cgtaaggcag gctgaacgag gactgccagg gccaaagcctg tggagacaaa acagtctccc

181 agtttctgac agacacttag tccctctgcc cactctggta ttttcacccc tcagGCTGCT
      L L
241 GGTCATCTAC CCATGGACTC AGAGGTTCTT TGAACACTTT GGGGACCTTT CCACTGCTGA
      V I Y P W T Q R F F E H F G D L S T A D
301 CGCTATCATG AAAAACCTC GAGTACAGGC CCATGGCGAG AAAGTGCTGT CCTCCTTTGG
      A I M K N P R V Q A H G E K V L S S F G
361 GGAGGGCCTG AATCACCTGG GCGACCTCAA GGGCACCTTT GCCCAGCTGA GCGAGCTGCA
      E G L N H L G D L K G T F A Q L S E L H
421 CTGTGACGAG CTGCATGTGG ATCCTGAGAA CTTCAAGgtg agtccagggg atgctctact
      C D E L H V D P E N F K
481 ctcttcttct cactttgttt tcattcgctg tggttatttg acttacttga atttctttca

541 tgtcactttt ttcttgactg tatttcatca ttagtgctt tttcagctta tactactttg

601 cactattttt ttctttcaat attctttgtt tttgtcttat aatcctgctt tatatcaagc

661 ccttttagctt tctacatttt gcctctctct ttctcttttc ctaacaattc cttccccct

721 aaatagtact caaattacga atgttacctc tcatccaatg tttatgtact taggcaaatc

781 cttctgggtcc tccatatgag aggtggcaag gagtgaatca aagatgcaaa gagtattatg

841 ttgttctaga gcttgtgact cattttgaaa ttaaaagata atttgaataa tataaaatca

901 ggagtaaaag gaaaggaaaa tcagtatctg gggatgaaag atcagaagggt catacagggtg

961 atggagagca gaagtttcta agaaacagac cactattgct gcaactaagc aattagttag

1021 ttaatatact tgcttatttg tttacatggt ttttaatttg atgggaataa gtttgggccca

```

Appendix 1-2H
Western tree hyrax deltaH-globin gene sequence

```

1081   tcagtttggg gcaacagaaa tgggaatcat cccattttca cagatgccaa gctggattgt
1141   tttattaacc atgtctgtat atctatctgt ctcttctcca cagCTCCTGG GCAATATCCT
                                     L L G N I L
1201   GGTGGTTGTC CTGGCCCGCC ATTATGGCAA GGAATTCACC CTAGAGGTTT AGGCTGCCTA
      V V V L A R H Y G K E F T L E V Q A A Y
1261   TCAGAAGTTT GTGGCTGGTA TGGCAAATGC CTTGGCTCAC AAATACCACT GAgatcctgg
      Q K F V A G M A N A L A H K Y H .
1321   cctgtttcct ggtatccatt ggaagccctg tttccctaga cgtgacctct gagtctgtaa
1381   atagtgtca ttcgcaaggg cattgcttct gcctaataaa gaaccttcag ctcaactttc
1441   tgattcattt tgtttatttt atttcttatg cactttagaa gtatgggagt gtcctgaaaa
1501   gcttaccgat agcgatctct tgtgtcccat aggcagagtc caagggaaac tggaaaatga
1561   cgggtacctt atacagttgt taatttaatg g

```

Appendix 1-2I
Bushveld elephant shrew beta-globin partial gene sequence

```

1      ctttaaagt aaatttttgc tcagtaactg tgactgtaat tcactatcct tgttggtgaa
61     gagtcaatgc caaggatata aacaaataaa tcaatggata tttcaggaaa tttttatcag
121    ggatctttaa tataagacat ggtaaaggaa gaatatttgg aatagtggct gattacttgt
181    tcattttatc tttgaaataa ttttgtatag tccaagcat agtacttact ggctctgtcc
241    aaacattggt aaactgcccc atagtggtaa agtatctccc acagtcaaga tctcaatggt
301    atcatttatt tactaaatga atgacataag aaatacctaa gaggccttagc agcttgtttt
361    ttttttaa atacatactt gctaaaaaga tgtttttcgt ttcaatttct gagaggaatg
421    tgattagaga taagtaggag agtgagggcc tgaaatcaaa ctcaaatgac agtgccagcc
481    tgccaatgac agtcagagct gtgatcactc cgggctcttt ctgcagagtc actctggcct
541    gggccaatct gcttgcagaa gcaccatggg caggaccag ggctgggcat agaagaaggg
601    caggaccagc aaaggcttac acttgcttct gacacaagtg tgttcactag caactactca
661    aacagtcaag ATGGTGCATC TTACTGACGG CGAGAAGGCT CTGGTCAATG GCATTTGGTC
           V H L T D G E K A L V N G I W S
721    CAAGGTGGAC GTAGATAAAC TTGGTGGTCA GGCTCTGGGC TGgtgagtat tttgattaag
           K V D V D K L G G Q A L G W
781    gcaggtttaa ggatgcttga atgggactgg aatatggaga cagccgtttt ccaggcttct
841    cacaggtatt gactctttct ggcccctgtg tttcttttac ccatcagTCT GCTGATTGTA
           L L I V
901    TACCCACGGA CTCAGAGGTT CTTTGAATCC TTTGGGGACC TGTCCTCTGC TGATGCTATC
           Y P R T Q R F F E S F G D L S S A D A I
961    ATAAAGAATC CCAAGGTGGC AGCCCATGGC AAGAAAGTAG TGAACTCCTT CAGTGAGGGC
           I K N P K V A A H G K K V V N S F S E G
1021 ATGAAGCATC TGGATGACCT CAAGGGCACC TTTGCCCAGC TGAGTGAGCT GCACTGTGAC
           M K H L D D L K G T F A Q L S E L H C D

```

Appendix 1-2I
Bushveld elephant shrew beta-globin partial gene sequence

1081 AAGCTCCATG TGGATCCTGA GAACTTCAGG gtgagtctat gggaataaag agtttccttc
 K L H V D P E N F R
1141 tgctctctat ttttgtcaaa ttctagcagg gggagggggg tccaggtcag aattgaaaca

1201 aagtgttctg actacaaagt aggaactctc cagggccttt ttaggaagta taatctcttc

1261 ctctacaaa tctttattgc ccctgttgca tggttcagtg ttcctttgtc ttcttcttct

1321 aactacttcc tcctctactt actttttatc acttcaaaaa tgatttaaaa aattaaacct

1381 tcttctttaa atctactagg ggtttcccct tctattctta gactttcaga attcatatth

1441 ttaaagcatg ggggaataaaa tggttctttt tgacattaca tacgtttttt aaaaggtcac

1501 taataaaatc tgaactaaga tggatggagg aaaatattca catgcataaa ttcagactga

1561 tatgacaaca ctgacatccg tagtagaact aagtgtcaat tatcacacca ctcatattta

1621 gtgcacaaga gtcaagtgtg agtgcagctt gtgatgccct aatctcgatc actctgccag

1681 ccaaattaat gtgtcttata tttctctca g

Appendix 1-2J

Bushveld elephant shrew delta-globin partial gene sequence

1 AATGTGGCCA ACGCCCTGGC TCACAAGTAC CACTAAgatc ctgccctttt ccttgggatc catg
 N V A N A L A H K Y H .
71 ctctgtctcc ctagtctga cccctaaatt gggagaagaa ccaaagtgt aatatttgct taat

141 ctttcagctc aaatTTTTTA ttaatttctt ttttttctt attcatccta ctgtatagga gaga

211 aggaca

Appendix 1-2K
Nine-banded armadillo pseudo-delta-globin gene sequence

```

1   taaaaagaga caccATGGTG AATCTGACTT CTGTGGAGAA GTCTGCTGTC ACAGGTCTGT
      V N L T S V E K S A V T G L
61   GGAGAAAGGT GAACGTGAGA TAAtgtggtg gtggctctgg gcaggtttgt gtggaggtta
      W R K V N V R .
121  caaggcagct taagggtggga agatgcaagc tgggcatgtg gagacatagc agtctcctgg

181  gtttatgaca ggtactgaca gctgtcccct ttgctgtttt caccctcag GCTGCTGGTT
      L L V
241  GTGAACCCTT GGGCCGAAAG TTTCTTTGAA ACCTTGGGGA ACTTGTCTCTC TCCTTCTGCC
      V N P W A E S F F E T L G N L S S P S A
301  ATATTTGGCA ATAGTAAAGT GAAGGCTCAT TTCAAGAAGG TTCTGACTTC CTTTAGTGAT
      I F G N S K V K A H F K K V L T S F S D
361  GGTGTGATGA AGCACCTGGA CAACCTCAAG GGTACCTATG CTCATCTGAG TGAACTGCAC
      G V M K H L D N L K G T Y A H L S E L H
421  TGTGACAAGC TGCACGTGGA TCCTGAGAAC TTCAAAgtga gtctaggaaa tattccaatt
      C D K L H V D P E N F K
481  ttttcatttc taatttttga ctgtgcctct ttaacctgtt ggctttacct ccacatycca

541  ttttactttt tctatatctt acavcttaat gcttttctaaa ctttatgtca atctttcctt

601  tcaatattct catgagctta tatttaatca agctctttta ttaatttcct agcttttgcc

661  cctctctccc tttttcttaa ttttctttcc attaacctgt actcaaatta tgcctaccag

721  cttagatctc atctgctact tctgccttta agaaaatact tttttcctt caaatgtagg

781  ttggtaaagt ttgaattaaa gaaaagaggc agacaatgtt tttctagaaa ttgtgcatca

841  ttttaagatt tgaattgttt aaagtcagag aggaaaatta atatctgagc atgaaacatc

901  agaagatcat ataggagggtg ggatagcaaa agttaatagg agacagccca tatcacatta

961  aataatcaat gtatcagtta attaatgttt atttatagat ttttaatttt ttttttggtc

1021 ggaatagcct gaggatctgt tgtggctaag ggatagttag aataaatggg gaacacccca

```

Appendix 1-2K
 Nine-banded armadillo pseudo-delta-globin gene sequence

```

1081  gtgtctcagg agtcaagctg gagtcctttc tttactatgt ctctgtgtca acctacctct

1141  tcccttcagC TCCTTAGCAA CATGCTGGTG ATTGTGCTGG CCTGTCACCT TGGCAAGGAC
          L  L  S  N  M  L  V  I  V  L  A  C  H  F  G  K  D
1201  TTCACCCTGG AGTTGCACGC TGCCTTTCAG AAGGTCGTAG TTGGAGTGGC GAATGCCCTG
          F  T  L  E  L  H  A  A  F  Q  K  V  V  V  G  V  A  N  A  L
1261  GCTCACAACT ACTATTGAga ttctggtctg tttgct
          A  H  N  Y  Y  .
  
```

Appendix 1-2L
Nine-banded armadillo beta-globin gene sequence

```

1   tataaagaga caccATGGTG AACCTGACCT CTGACGAGAA GACTGCCGTC CTTGCCCTGT
      V N L T S D E K T A V L A L
61  GGAACAAGGT GGACGTGGAA GACTGTGGTG GTGAGGCCCT GGGCAGgttt gtatggaggt
      W N K V D V E D C G G E A L G R
121 tacaaggctg cttaaggagg gaggatggaa gctgggcatg tggagacaga ccacctcctg

181 gatztatgac aggaactgat tgctgtctcc tgtgctgctt tcaccctca gGCTGCTGGT
      L L V
241 CGTGTATCCC TGGACCCAGA GGTTCITTGA AAGCTTTGGG GACTTGCCA CTCCTGCTGC
      V Y P W T Q R F F E S F G D L S T P A A
301 TGTGTTGCA AATGCTAAGG TAAAAGCCCA TGGCAAGAAG GTGCTAACTT CCTTTGGTGA
      V F A N A K V K A H G K K V L T S F G E
361 AGGTATGAAT CACCTGGACA ACCTCAAGGG CACCTTTGCT AAAGTGAAGT AGCTGCACTG
      G M N H L D N L K G T F A K L S E L H C
421 TGACAAGCTG CACGTGGATC CTGAGAATTT CAAGgtgagt caatattctt cttcttcctt
      D K L H V D P E N F K
481 ctttctatgg tcaagctcat gtcatgggaa aaggacataa gagtcagttt ccagttctca

541 atagaaaaaa aaattctggt tgcactactg tggactcctt gggaccattc atttctttca

601 cctgctttgc ttatagttat tgtttcctct ttttcctttt tctcttcttc ttcataagtt

661 tttctctctg ttttttttta acacaatctt ttaattttgt gcctttaaat tttttttaag

721 ctttcttctt ttaattacta ctcgtttcct ttcatttcta tactttctat ctaatcttct

781 cctttcaaga gaaggagtgg ttcactacta ctttgcttgg gtgtaaagaa taacagcaat

841 agcttaaatt ctggcataat gtgaataggg aggacaattt ctcataaag ttgaggctga

901 tattggagga tttgcattag tagtagaggt tacatccagt taccgtcttg ctcataattt

961 gtgggcacaa cacagggcat atcttggaac aaggctagaa tattctgaat gcaaactggg

1021 gacctgtggt aactatgttc atgcctgttg tctcttcctc ttcagCTCCT GGGCAATATG
      L L G N M

```

Appendix 1-2L
Nine-banded armadillo beta-globin gene sequence

1081 CTGGTGGTTG TGCTGGCTCG CCACTTTGGC AAGGAATTCG AETGGCACAT GCACGCTTGT
L V V V L A R H F G K E F D W H M H A C
1141 TTTCAGAAGG TGGTGGCTGG TGTGGCTAAT GCCCTGGCTC ACAAGTACCA TTGAgctcct
F Q K V V A G V A N A L A H K Y H .
1201 ctcccaactt tccagttcct acaaaagggtg cttttgtcct cagagtccaa ctactgaatg

1261 tgggaaaatt atatagagcc ttgggaatct ggttgtgcct aataaagaac atttatttcc

1321 actgcatttg tgtattttaa ttatttctgc atatctcact cagatgggca tatgggaggg

1381 aag

Appendix 1-2M
Pale-throated sloth pseudo-delta-globin gene sequence

```

1   tcactagcaa ctacaaaaca gacaccATGG TGTATCTGTT TGCTCATGAG GTGTCTGCCG
                                   V   Y   L   F   A   H   E   V   S   A
61  TCTCCGGCCT GTGGGGCAAG GTGAATGTGC AACAACCTGG TGGCGAGGCC CTGGGCAGgt
    V   S   G   L   W   G   K   V   N   V   Q   Q   L   G   G   E   A   L   G   R
121  tggtactgag gttataaagc agattaagga gggagggttg aagctgggct cctggagaca

181  gagcagtctc ctgggtttct gacaggcact gactccttct gtccctctgt ctactttcac

241  ccttcagGCT GCTGGTCGTG TACCCCTGGA CCCAGAGGTT CTTTGAAAGC TTTGGGGACT
                                   L   L   V   V   Y   P   W   T   Q   R   F   F   E   S   F   G   D
301  TGTCCTCTGC TGATGCTGTG TTTTCCAATG CTGAAGTGAA GGCCCACGGC AAGAAGGTGC
    L   S   S   A   D   A   V   F   S   N   A   E   V   K   A   H   G   K   K   V
361  TGACCTCCTT CGGTGAGGGT CTGAGCACCT GGACAATCTC AAGGGCACCT ATGCTCACCT
    L   T   S   F   G   E   G   L   S   T   W   T   I   S   R   A   P   M   L   T
421  GAgcgagctg cactgtgaca agctgcacgt ggatcctgag aacttcaagg cgagttcagg
    .
481  agatgttcca atTTTTTTTca TTTTcttgtc tttactgta cttctttaac ttactagctt

541  tgcctccaca ttccttttca ctttttctat atTTTatcca ttaaatgctt tttagaattt

601  acatccttgt tttcttttcaa tattctttgt ttcctatctc atgaccttat tttacaatcaa

661  gctcttactt tactagcctt tgcctctctc tccctatttc ttaatgtttt ttccactaac

721  caatactcaa attatgcaga ccagctatta tctgctaatt ctacccttgg gaaattcttc

781  atTTTctcca aattgggggt ggtaaagcct gaatcaaaga gaagaggcac atgatactct

841  agaaactgtg cataatTTTtg aagtttgaat tgccagaagt caggaataaa ttgggaggag

901  agtcagtatc tgagcatgaa agatcagaag gtcatatagg aagtgtgaga cagcccatat

961  cacactaact aatcaatgaa tttgTTaatt aatttataga ttcactTTTT taaagtttgg

1021 tgggaataaa attgggatcc gttttggcta gggtgtgggt agaataaatg ggcattaccc

```

Appendix 1-2M
Pale-throated sloth pseudo-delta-globin gene sequence

1081 cagtttctca gaagtcaagc tggattcgtc tgtgaaccat gtetatgtgt ctacctacct

1141 ctgccctcag CTCCTGGGTA ATGTACAGGT GATTGTGCTG GCCTGCTGCT TTGGCAGGGA
L L G N V Q V I V L A C C F G R E

1201 ATTCACCCCG CAGTTGCAGG CTTCCTGTCA GAAGATGGTG ACTGCTGTGG CTATTGCCCT
F T P Q L Q A S C Q K M V T A V A I A L

1261 GGCTCACAAG TACCATTGAc atcctggcct gtttgctggc atccatcgga agtccatggt
A H K Y H .

1321 tccgtatatt ctgtcctcag aacttgggga aaaaatgttc accatcaaga acattacttc

1381 tgcctttttt ttaattctct ttatttaaatt ttatttttgt tatttctgcc taataaagat

1441 ctttcattctc aacctgatga ttcattctcac ttattttctt tcatttttgc tcagtctagt

1501 ggtatgggaa gatcccttag ggtctacaaa tagggaactc atgtgtctta tgaaaagatg

1561 tcaaggggaaa tggaaagatg aaggagggtct

Appendix 1-2N
Pale-throated sloth beta-globin gene sequence

```

1   actccctgga aggaagggga gggcccaggg cttggcataa agggaggaga agggccagct
61  gctactcaca cttgcttctg acacagctgt gttcactagc aaccacaaaa cagacaccAT
121 GGTGCACCTG GCTGACGATG AGAAGGCTGC TGTATCCGCC CTGTGGCACA AGGTGCATGT
    V H L A D D E K A A V S A L W H K V H V
181 GGAAGAATTT GGTGGCGAGG CCCTGGGCAG gttggtactg aggttataag gcagattacg
    E E F G G E A L G R
241 gagggatgat ggaagctggg ctcctggaga cagagcagtc tcctggtttc tgacaggcac
301 tgactccttc tgtcccctgt gctactttca cccttcagGC TGCTGGTCGT GTACCCCTGG
                                     L L V V Y P W
361 ACCAGCAGAT TCTTTGAAAG CTTTGGGGAC TTGTCCTCTG CTGATGCTGT GTTTTCCAAT
    T S R F F E S F G D L S S A D A V F S N
421 GCTAAAGTGA AGGCCACCGG CAAGAAGGTG CTGACCTCCT TCGGTGAGGG TCTGAAGCAC
    A K V K A H G K K V L T S F G E G L K H
481 CTGGATGATC TCAAGGGCAC CTATGCTCAC CTGAGCGAGC TGCACTGTGA CAAGCTGCAC
    L D D L K G T Y A H L S E L H C D K L H
541 GTGGATCCTG AGAACTTCAA Ggtgagcctg cgggcacttc agtgtttctcc ttcctccttc
    V D P E N F K
601 tttctatgat caagcttggtg tcatgggaaa agggcacagt atccagggtc cagtttggaa
661 aagaaaaaat cttctggttt tntacctatg gactccttgg agctatttat tttctttacc
721 tgctttgttc acaatcattg ttttctcatt tcattcttct ttttcttttc cataatttct
781 tctgccttta tttttattta aacttttcat tttgtgcctt taaattattt ttaaaccttc
841 ttcttttaat ccactatttt tcttatctct atattttctt tctctaagtgt tttctttcct
901 agcgaaggaa cggacagctg ctactttgca taggtctaaa gaatcactat gttaaagctc
961 taagttgagc tgggtgggga gagccatttc tgcattgtaca ctcaggctgg tgtgggagga
1021 gcagcgtcag tggcagaggt aacatctggc tatcatcctg ctctngattt gtgggcaaac

```

Appendix 1-2N
Pale-throated sloth beta-globin gene sequence

```

1081  cctagncaca gtttanatga tgctggaata ttctgattcc agtttggggc cctctgttaa
1141  ctatgttctt gcctcttttc tcttcccctc agCTCTTGGG TAACGTACTG GTGATTGTGC
                                     L L G  N V L  V I V
1201  TGGCCCGCCA CTTTGGAAAG GAATTCATC CGCAGTTGCA GGCTGCCTAT CAGAAGGTGA
      L A R H  F G K  E F T P  Q L Q  A A Y  Q K V
1261  CGACTGGTGT GTCTACTGCC CTGGCCCACA AGTACCACTG Agcaccactt tctgcttt
      T T G V  S T A  L A H K  Y H .

```

Appendix 1-3. Lengths of exons and introns for sequenced 'β-like' globin genes. Values do not include those of the initiation and termination codons. "--" denotes missing data.

Taxon	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	Gene length
African elephant δ	89 bp	128 bp	223 bp	729 bp	126 bp	1,295 bp
Asian elephant δ	89 bp	128 bp	223 bp	729 bp	126 bp	1,295 bp
dugong δ	89 bp	129 bp	223 bp	755 bp	126 bp	1,322 bp
manatee δ	89 bp	129 bp	223 bp	756 bp	126 bp	1,323 bp
rock hyrax δ	--	--	--	739 bp	126 bp	--
rock hyrax δ ^H	89 bp	132 bp	223 bp	728 bp	126 bp	1,298 bp
yellow-spotted hyrax δ ^H	89 bp	128 bp	223 bp	727 bp	126 bp	1,293 bp
tree hyrax δ ^H	89 bp	128 bp	223 bp	726 bp	126 bp	1,292 bp
elephant shrew β	89 bp	125 bp	223 bp	601 bp	--	--
sloth ψδ	89 bp	129 bp	223 bp?	680 bp	126 bp	1,247 bp
sloth β	89 bp	128 bp	223 bp	611 bp	126 bp	1,177 bp
armadillo ψδ	87 bp	126 bp	226 bp	693 bp	126 bp	1,258 bp
armadillo β	89 bp	125 bp	223 bp	611 bp	126 bp	1,174 bp

Appendix 1-4. Lengths of additional sequence data obtained for the 5' external region of sequenced 'β-like' globin genes, and positions of upstream features relative to the putative CAP site. The positions of putative cap sites from the initiation codon ("ATG") are also shown. Lengths of collected 5' end sequence data do not include that of the initiation codon. "--" denotes missing data.

Taxon	CACCC site	CCAAT site	ATA site	Putative CAP site	Length of obtained 5' external data
African elephant δ	--	--	--	--	-26 bp
Asian elephant δ	-104 bp, -89 bp	-75 bp	-31 bp	-52 bp from "ATG"	-400 bp
dugong δ	-105 bp, -90 bp	-76 bp	-31 bp mutated into "GTA"	-52 bp from "ATG"	-778 bp
manatee δ	-105 bp, -90 bp	-76 bp	-31 bp mutated into "GTA"	-52 bp from "ATG"	-208 bp
rock hyrax δ ^H	--	--	--	--	-26 bp
yellow-spotted hyrax δ ^H	--	--	--	--	-26 bp
tree hyrax δ ^H	--	--	--	--	-14 bp
elephant shrew β	distal absent; proximal at -89 bp, mutated into "CACTC"	-75 bp	-30 bp	-52 bp from "ATG"	-670 bp
sloth ψδ	--	--	--	--	-26 bp
sloth β	--	--	-31 bp	-51 bp from "ATG"	-118 bp
armadillo ψδ	--	--	--	--	-14 bp
armadillo β	--	--	--	--	-14 bp

Appendix 1-5. Lengths of additional sequence data obtained for the 3' external region of sequenced 'β-like' globin genes, and positions of poly-adenylation signals (AATAAA) relative to the termination codon. Lengths of collected 3' end sequence data do not include that of the termination codon. "--" denotes missing data.

Taxon	AATAAA signal	Length of obtained 3' external data
African elephant δ	+103 bp	417 bp
Asian elephant δ	+103 bp	418 bp
dugong δ	+104 bp	533 bp
manatee δ	+104 bp	156 bp
rock hyrax δ	+99 bp	702 bp
rock hyrax δ ^H	+102 bp	279 bp
yellow-spotted hyrax δ ^H	+102 bp	279 bp
tree hyrax δ ^H	+102 bp	279 bp
elephant shrew δ	+95 bp	180 bp
sloth ψδ	+152 bp	311 bp
sloth β	--	17 bp
armadillo ψδ	--	18 bp
armadillo β	+106 bp	189 bp

Appendix 2-1. Primers used in the (A) amplification and (B) sequencing reactions of 'ε-like' globin genes.

(A) Primers used in polymerase chain reactions.

Primer Name	Sequence
HBB-12(R)	5' GCC ACC ACC TTC TCA TAG GCA GC 3'
HBG-1(F)	5' CCA TCA TGG GCA ACC CCA GG 3'
HBG-4(F)	5' CTG TCA TCA CCA CGA ACT 3'
HBG-5(F)	5' CCA AGA CCG GAC ACC ATG 3'
HBG-6(R)	5' TTA ATT GTG CTG TCA TGT GAA 3'
HBG-7(R)	5' AAC AAG CCT TCT CCC TAT T 3'
MNT-GA(F)	5' CCA TGG CGA GAA TGT GCT GAA C 3'
DUG-GA(R)	5' GAT GGT CCT ACT TAT GCT CAG ATC AGT GCT 3'
DUG-GB(R)	5' TGA GGC TAA ACA CAA GAA AGC ACA GAC ATA 3'
ELE-GB(R)	5' TGA GAC AAA ACA CAA GAG AGC ACA AAC ATA 3'
DUG-GC(R)	5' GTG CAG CTC ACT CAG CTT ACG AAA GGT 3'
DUG-GA(F)	5' ACT GGG TAA GAC CAC AGC CTT TGA G 3'
DUG-GB(F)	5' GGC CAG ACT GGA ACT CTC TGC TCA C 3'
DUG-GC(F)	5' GGG AAA CGT GAT AGT GAT TGT CTT GGC 3'
GGW-A(F)-RH	5' CTA CAG ATG CCA AGC TGG ATC GCT TTG TTA 3'
GGW-A(F)-STH	5' AGG AGT GGG TAG GCA GTA ATG GTA TTT AGC 3'
GGW-B(F)	5' CCC AAC AGC TCC TGG GCA ATG T 3'
GGW-C(F)-ESH/STH	5' TGG CTT CAC ATT TTG GCA AGG AGT T 3'
GGW-C(F)-RH	5' GCT CGG CAC TTC CGT GAG GAG TTC 3'
GGW-A(R)-RH	5' AAT GAC GCA ACG CAA TCA AGT AAG TAG AGA 3'
GGW-A(R)-STH	5' CTC CTA ACA CTG AAC TTG ACC CAT TAT T 3'
GGW-B(R)-RH/ESH	5' TGT CCA TGT TCT TAA CAG CTT CTC CA 3'
GGW-B(R)-STH	5' TGA GGT CAT CCA TGT GCT TAA CAG CAT CT 3'
GGW-C(R)	5' GTC AGC ACC TTC TTG CCA TGA GCC TTG 3'

(B) Primers used in sequencing reactions at the University of Calgary DNA core laboratory.

Primer Name	Sequence
M-13(F)-40	5' GTT TTC CCA GTC ACG AC 3'
M-13(R)	5' AAC AGC TAT GAC CAT G 3'

Appendix 2-2. DNA sequences of the 'ε-like' globin genes amplified in this study. Exons are identified by CAPITAL letters, while introns and external gene sequences are represented by lower case letters. The 5' flanking transcriptional control motifs (CACCC, CCAAT, ATA) and 3' poly-adenylation signal (AATAAA) are underlined within the 5' and 3' flanking regions of each gene sequence, respectively. Within each coding block, translated amino acids are listed below their corresponding base triplet.

Appendix		Page
2-2A	African elephant γ-globin gene sequence	130
2-2B	Asian elephant γ-globin gene sequence	132
2-2C	Dugong γ-globin gene sequence	134
2-2D	West Indian manatee γ-globin gene sequence	136
2-2E	Bushveld elephant shrew ε-globin partial gene sequence	138
2-2F	Pale-throated sloth ε-globin gene sequence	140

Appendix 2-2A
African elephant gamma-globin gene sequence

```

1      GTGCATTTTA CTGCCGAAGA GAAGGCTGCT ATCGCAAGCC TATGGGGCCA GGTGAATGTG
      V H F T A E E K A A I A S L W G Q V N V
61     GAAGAGACGG GAGGCGAGGC CCTGGGCAGg tacattctgg gaagcagcgg aggggagggga
      E E T G G E A L G R
121    ataaaggaag ctgagcctgg caagaatgca gaccagttac caagttttgt gactaggtct

181    gattttccat ctgttatggt tccatcccat agGCTTCTGG TTGTCTACCC TTGGACCCAG
                        L L V V Y P W T Q
241    AGGTTTTTTG ACACCTTTGG CAACCTATCC TCTGCCTCTG CCATCATGGG CAACCCCAGG
      R F F D T F G N L S S A S A I M G N P R
301    GTCAAGGCCC ATGGYAAGAA GGTGCTGACC TCCTTTGGAG ATGCTGTAA GAACCTGGAC
      V K A H G K K V L T S F G D A V K N L D
361    AACCTCAAGG GCACCTTTGC TAAGCTGAGC GAGCTGCACT GTGACAAGCT GCACGTGGAT
      N L K G T F A K L S E L H C D K L H V D
421    CCTGAGAACT TCAGGgtgag tccaggatat gtttggtgctc tcttggtgtt tgtctcaggc
      P E N F R
481    aacttggaca accaagcact gacctgagca caagtaggac catctatggg ctgtgaggta

541    tttggagatc ttgggttagt aataaagcaa ctccagagag ctgaattcaa actaggtggt

601    acagtaacaa ggctcaagaa gtgctttagt aagtctacag tggtgacttg tctaaactct

661    gttggctttt aatggagact tgtgtacaga tgagaaaaat gtcttttggg aattggaatt

721    ccgtggaaag gagaactttg cttttcttcc ctccctcatc tatcaccgcc tgcccctcat

781    atgttggagg tagaaaaaga atgtcatggg tgacgatata tgcagacatg tattgggttaa

841    ttataacaaa gaaacctggg tggatcatgtg tatattacta aggctattcc tatacttggg

901    ggccatttgt acaaaaatgg taatacctcc tttcaaacct ttttccgaaa actcgaggag

961    taagacaaaa tgtgtctgag ggagggaaac aactggaata tgtgaccaag gagggatatg

1021   aatagaaagc tgtgaagttt tatttttgct ttttctctcc caaaaaagca tatggagata

```

Appendix 2-2A

1081 gtgctggtag gaagatcada aaaatgaaac cgtgagtgtg gggcagtaea aaaagagaag
1141 caaattagct atgttaataa ctgggcaaga ccatagcctt tgaggaagct aacatagact
1201 tagttaggtc tgagtggcag ctgggtggcc agatactcgg aggccagact ggagctctcg
1261 gctcacaaca tgaatccata ggaatgcata tgttgtctgt ttctacctcc acagCTCCTG
L L
1321 GGAAATGTGC TAGTGATTGT CTTGGCAAAC CACTTTGGCA ARGAATTCAC CCCCCAGGTG
G N V L V I V L A N H F G K E F T P Q V
1381 CAGGCTGCCT GGCAGAAGAT GGTGACGGGC GTGGCCAATG CCCTGGCCTA CAAGTATCAC
Q A A W Q K M V T G V A N A L A Y K Y H
1441 TGAgtcctc gt

Appendix 2-2B
Asian elephant gamma-globin gene sequence

```

1      gtggtcacca cgaactccca agcccagaca ccATGGTGCA TTTTACTGCC GAAGAGAAGG
                                     V H F T A E E K
61     CTGCTATCAC AAGCCTATGG GGCCAGGTGA ATGTGGAAGA GACCGGAGGC GAGGCCCTGG
      A A I T S L W G Q V N V E E T G G E A L
121    GCAGgtacat tctgggaagc agcggagggg aggggaataaa ggaagctgag cctggcaaga
      G R
181    atgcagacca gttaccaagt tttgtgacta ggtctgattt tccatctgtt atgattccat

241    cccatagGCT TCTGGTTGTC TACCCTTGA CCCAGAGGTT TTTTGACACC TTTGGCAACC
                                     L L V V Y P W T Q R F F D T F G N
301    TATCCTCTGC CTCTGCCATC ATGGGCAACC CCAGGGTCAA GGCCCATGGC AAGAAGGTGC
      L S S A S A I M G N P R V K A H G K K V
361    TGACCTCCTT TGGAGATGCT GTTAAGAACC TGGACAACCT CAAGGGCACC TTTGCTAAGC
      L T S F G D A V K N L D N L K G T F A K
421    TGAGCGAGCT GCACTGTGAC AAGCTGCACG TGGATCCTGA GAAGTCAGG gtgagtccag
      L S E L H C D K L H V D P E N F R
481    gatatgtttg tgctctcttg tgttttgtct caggcaactt ggacaaccaa gcactgatct

541    gagcacaagt aggaccatct atgggctgtg agatatttgg agatcttggg ttagtaataa

601    agcaactcca gagagctgaa ttcaaactag gtgttacagt aacaaggctc aagaagtgct

661    tgtagaagtc tacagtgggtg acttgtctaa actctgttgg cttttaatgg agacttgtgt

721    acagatgaga aaaatgtctt ttggtaattg gaattccgtg gaaaggagaa ctttgctttt

781    ctttcctccc tcatctatca ccgcctgccc ctcatatgtt ggaggtagaa aaagaatgtc

841    atgggtgacg atacatgcag acatgtattg gttaattata acaaagaaac ctgggtggtc

901    atgtgtatat tactaaggct attcctatat ctgggtggcca tttgtacaaa aatagtaata

961    tctcctttca aacttttttc tgaaaactcg aggagtaaga caaatgtgt ctgagggagg

1021   gaaacaactg gaatatgtga ccaaggaggg tatggaatag aaagctgtga agttttat

```

Appendix 2-2B
Asian elephant γ -globin gene sequence

1081 ttgctttttt cctcccaaaa cagcgtatgg agatggtgct ggtaggaaga tcaaaaaaat
1141 gaaaccgtgg gtgtaggcca gtacaaaaag agaagcaaatt tagctatggt aataactggg
1201 caagaccata gcctttgagg aagctaacaat agacttagtt aggtctgagt ggcatgtggg
1261 tggccagata ctcgagggcc agactggagc tctcggctca caacatgaat ccataggaat
1321 gcatctgttg tctgtttcta cctccacagC TCCTGGGAAA TGTGCTAGTG ATTGTCTTGG
L L G N V L V I V L
1381 CAAACCACTT TGGCAAAGAA TTCACCCCC AGGTGCAGGC TGCCTGGCAG AAGATGGTGA
A N H F G K E F T P Q V Q A A W Q K M V
1441 CGGGCGTGGC CAATGCCCTG GCCTACAAGT ATCACTGAgc tcctcgcaat aggaagaaga
T G V A N A L A Y K Y H .
1501 cttatttttt acatgacaac acaattaatt aaaagtattc tgtaagaaa taagatgtaa
1561 tggactcctg ttgtttcttt ttcatgtggc cttaagtaaa cgaatttcca ggggttttat
1621 gttgggggtgt gtgtgtgctc cctattcact tttggcaaaa ggtaagaatt ttgataaataa
1681 aagaacaagg caataaagga acatacactg ggagtttctga aaggaaaaga aaatatctctg
1741 ggatagtttt ggtgggagag aaagggcttt attggacagg gactcctcag aactagat

Appendix 2-2C
Dugong gamma-globin gene sequence

1 atcacagatg tgaatgtcca ttttaactct tccatgccta-aacaccgccc ccccgccccc
61 gcctgatagc ctagcgttga ccaatagcct catagcaaaa ggaagaacaa agggccagtg
121 gccagggata aagaataaaaa agccacatgt tccagttgca gcacatacat ctttctgaca
181 catctgtgat caccacaaac tccaagaccg gacaccATGG TGTATTTTAC TGCTGAAGAG
241 AAGGCTGCTA TCACAAGCCT GTGGGGCAAG GTGAATGTGG AAGAGGCTGG AGGCAAGGCC
K A A I T S L W G K V N V E E A G G K A
301 CTAGGCAGgt agattctgga gggtagggga agggagggaa taaaggaagc tgagcttggc
L G R
361 aggaatgcag gtctgttacc aagtcttgtg acaagctctg atttaccatc tgctatgatt
421 ctatcctgta gGCTCCTGAT TGTCTACCCT TGGACCCAGA GGTTTTTTGA CAAATTTGGC
L L I V Y P W T Q R F F D K F G
481 AACCTATCCT CTGCCTCTGC TATCATGGGC AACCCCAAGA TCAAGGCCCA TGGCAAGAAG
N L S S A S A I M G N P K I K A H G K K
541 GTGCTGAACT CCTTTGGCGA TGCCGTTGAG AACCCGGACA ACCTCAAGGG TACCTTTGCT
V L N S F G D A V E N P D N L K G T F A
601 AAGCTGAGTG AGCTGCACTG TGACAAGCTG CTTGTGGATC CTGAGGACTT CAGGatgagt
K L S E L H C D K L L V D P E D F R
661 ctaggatatg tctgtgcttt cttgtgttta gcctcagaca gcttagacaa ctgagcactg
721 atctgagcat aagtaggacc atctatggcc tgtgagatat ttggagatct tgggttagta
781 acaaaacaat tccagagggc tgaattcaaa ctaagtgtta tggttaacaag gctcgagaag
841 tgcttctagg agtctggagt gacttgtcta aactctgttg gcttttagtg gagactcgtg
901 tagagatgag gacgtctttt ggtaattgga atcctgtaga aagaagaatt ttgcctttct
961 ttctctacgt atctatcact gcctgcccc catatgtagg aggtagaaaa agaatgtcca
1021 gtgggacaat acacgcagac atgtattggg taattataac aaagaaactg ggtggtcatg

Appendix 2-2C
Dugong gamma-globin gene sequence

1081 catatattac taaggctatt cctgtactgg gcaggcattt gtgcaaaatc tactaatgtc

1141 tcctttcaaa cttatttcct aaaactctag gagtaaggca aagtgtctct gaggggaaggg

1201 aaaccaactg gaatatgtgg cagaggaggg tatggaatag aaagctgtga gggttttatt

1261 tttgtccccc acccccaccc cccaccaaaa aaaaaaagta tgtggagatg gtgctggtag

1321 gaagatagaa agagtgaat catggatgtg ggtcagtaca gaaaggggaag caaattagct

1381 atgtttaata actgggtaag accacagcct ttgagaaagc taacatagac ttaggcctta

1441 gtggcagttg ggtggccaga ttcttgagg ccagactgga actctctgct cactatatga

1501 atccatagga atgtgtctgt tatctctttc taccctaca gCTCCTGGGA AACGTGATAG
L L G N V I

1561 TGATTGTCTT GGCAAACCAC TTTGGCAAAG AATTTACCCC CCAGGTGCAG GCTGCCTGGC
V I V L A N H F G K E F T P Q V Q A A W

1621 AGAAGATGGT GA CTGGTGTG GCCAGTGCCC TGGCCCGCAA GTATCACTGA gtatctcata
Q K M V T G V A S A L A R K Y H .

1681 atagggagaa gccttgttct tcacatgaca gcacaattaa taaaattatt ctgttaataa

1741 ataagattta ctagactcct attgtttctt ttcatgtaa cttaagtaa atgaattccc

1801 agggatttta tgttgggggg gtgtgtgtgt gctccccatt cactgtttgc aaaagatgaa

1861 gattttgaca agaaaataag agaacaggac gataaaggaa catccctggg agttctgaaa

1921 gggaaagaaa aatctctggg agagttttgg tgggaaagga agggctttat gggacaagga

1981 ctctcagaa ctaaataatt ttgagcagtg ggaaaagact tcagggaag tagcagaggc

2041 a

Appendix 2-2D
West Indian manatee gamma-globin gene sequence

1 tcatagcaaa aggaagaaca aagggccagt ggccagggat ~~aaagaataaa~~ aagccacatg

61 ttccagttgc agcacataca tccatctgac acatctgtga tcaccacaaa ctccaagacc

121 ggacaccATG GTGGATTTTA CTGCTGAAGA GAAGGCTGCT ATCACAAGGC TGTGGGGCAA
 V D F T A E E K A A I T R L W G K

181 GATGAATGTG GAAGAGGCTG GAGGCAAGGC CCTGGGCAGg tagattctgg gggatagggg
 M N V E E A G G K A L G R

241 aagggagggg ataaaggaag ctgagctagg caggaatgca ggtctgttac caagttttgt

301 gacaagctct gattttccat ctgctatgat tctatcctgt agGCTCCTGA TTGTCTACCC
 L L I V Y P

361 TTGGACCCAG AGGTTTTTTG ACAACTTTGG CAACCTATCC TCTGCCTCTG CCATCATGGG
 W T Q R F F D N F G N L S S A S A I M G

421 CAACCCCAAG GTCAAGGCC ATGGCAAGAA GGTGCTGAAC TCCTTTGGAG ATGCCGTTAA
 N P K V K A H G K K V L N S F G D A V K

481 GAACCCGGAC AACCTCAAGG GTACCTTTGC TAAGCTGAGT GAGCTGCACT GTGACAAGCT
 N P D N L K G T F A K L S E L H C D K L

541 GCTTGTGGAT TCTGAGAACT TCAGGatgag tctagggat gtttgtgctt tcttgtgttt
 L V D S E N F R

601 agcctcagac agcttagaca actgagcact gatctgagca taagtaggac catctatggc

661 ctgtgagata ttgggagatc ttgggttagt gaaaaaaca ttccagaagg ctgaattcaa

721 actaagtgtt atggtaacaa ggctcaagaa gtgcttctag gagtctagag tgacttgtct

781 aaactctgtt ggcttttaat ggagactcgt gtagagatga gaacaatgtc ttttggtaat

841 tggaaatcctg tagaaggaag aattttgcct ttctttcctc acgcatctat cactgcctga

901 ccccatatg taggaggtag aaaaagaatg tccagtggga cgatacacgc aggcattgtat

961 tggttaatta taacaaagaa actgggtggt catgcatatg tcactaaggc tattcctgta

1021 ctgggcaggc atttgtgcaa aatctactaa tgcctccttt caaatattata tcctaaaact

Appendix 2-2D
West Indian manatee gamma-globin gene sequence

1081 ctgggagtag gacaaagtgt ctctgagggg gggaaacaac tggaatatgt ggcagaggag

1141 ggtatggaat agaaagctat gaggttttat ttttgttccc cccaccaaaa aaaaaaagta

1201 tgcggagatg gtgctggtag gaagatagaa agagtgaat catggatgtg ggtcagtaca

1261 gaaagggag caaattagct atgtttaata actgggtaag accacagctt ttgagaacgc

1321 taacatagac ttaggcctta gtggcagttg ggtggccaga ttcttggagg ccagactgga

1381 actctctgct cactatatga atccatagga atgtgtctgt tatctctttc taccctaca

1441 gCTCCTGGGA AACGTGCTAG TGATTGTCTT GGCAAACCAC TTTGGCAAAG AATTACCCC
L L G N V L V I V L A N H F G K E F T P

1501 CCAGGTGCAG GCTGCCTGGC AGAAGATGGC GACTGGTGTG GCCAGTGCCG TGGCCCGCAA
Q V Q A A W Q K M A T G V A S A V A R K

1561 GTATCACTGA gtacctca
Y H .

Appendix 2-2E
Bushveld elephant shrew epsilon-globin partial gene sequence

```

1   GTCAAGGCTC ATGGAAAGAA GGTGCTGACC TCCTTTGGAG ATGCAGTTAA GAACATAGAC
    V K A H G K K V L T S F G D A V K N I D
61  AATCTCAAGG GAGCTTTTGC TAAACTAAGT GAAGTGCAGT GTGACAAGTT ACATGTGGAT
    N L K G A F A K L S E L H C D K L H V D
121 CCTGAAAAC TCCGGgtaag tccagaagat gttcacttgg tttggttttt tttttaattt
    P E N F R
181 ttattatgga ataataatta ataataataa taacaacaat aatagataac cctttgacta

241 gaaagccaaa accaacaanaa ctctaaaata ataatttggg gttaaaagag taagcttctg

301 gttggcataa ttatagccta ttaggetcag aggtgaagtc aataagggat gttagatacc

361 tcagtttaaa aaagaaccca aacatttaac agaatcagat gtattcatcc taaagaatac

421 tgaggaagaa caggtaaact ttatttctctg tgaatctaaa agatttcagt atgtcaataa

481 tatggaggaa aatgataaat attttgtttc ttgaagaaaa agataatata atttataaaa

541 tatattaaat aatctatcaa tatcatatcc aaagagactt ttttctttct gcattgattg

601 atgactttta caaaattatg cctaaggagc tcttgtgtcc ctgaaatcct acaattataa

661 gcataaggta atttataaat tagaaagggg ggcaaacaaa atagtacagt ggatagaaaa

721 ggatgtagct aagaggaggc ataatgttga acagagagtt taaaataatg ttgaatatag

781 atgctatctt acttgagtat aatggagaag tttgagatag gaaatgatca aggatatggg

841 tattcagttt taccctctcc cccaaaaaaa attattttta gaactctcta ctcaatatgt

901 ctctttgtca ttttgtcttt tccccaatag CTCCTGGGCA ATGTGATCGT TATCATCATG
                                L L G N V I V I I M
961 GCTTCGCATT TTGGCAAGGA GTTCACCCCT GAAGTGCAGG CTGCTTGGCA GAAACTGGTG
    A S H F G K E F T P E V Q A A W Q K L V
1021 GCTGGTGTG CCACTGCTCT GGCTACAAG TACCATTGAg tcctcttgct ctcatgcaag
    A G V A T A L A H K Y H .

```

Appendix 2-2E
Bushveld elephant shrew epsilon-globin partial gene sequence

1081 tgccccgtgt tcccaccacc ttattcttc acatgacagc acaattaaa

Appendix 2-2F
Pale-throated sloth epsilon-globin gene sequence

1 caaaggcggt gtataccctt cactgctgac cctctgctga cccagctcca cccctgaggg

61 acacagctca gccctgacca atgactgtga agtaccaggg gaacaagggg ccagaggtac

121 acagtgaaga ataaaaagcc acatcttcta gaagcagcac agctctgctt ctgacacgtc

181 tgtgatcacc agccagcatt tagatcctac atcATGGTGC ATTTTACTGC TGAGGAGAAG
V H F T A E E K

241 GCTACTGTGG CGAGCCTGTG GGGCAAGGTG AACGTGGAGG AGGCTGGCGG CGAGGTGCTG
A T V A S L W G K V N V E E A G G E V L

301 GGCAGgtagg catttgggtc tcaacgcagg ggaagaagg tgaatgcaag cctggaaaat
G R

361 tgacaagaaa tagccaaagg atttctgtgt ctctgatttt ccattttcta tgggtctcatc

421 tcatagACTT CTGGTTGTCT ACCCCTGGAC CCAGAGATTT TTTGACAACT TTGGTAACCT
L L V V Y P W T Q R F F D N F G N L

481 GTCTTCTTCC TCTGCAATCA TGGGCAACCC CAAGGTCAAG GCCCACGGCA AGAAGGTGCT
S S S S A I M G N P K V K A H G K K V L

541 GACCTCCTTT GGAGATGCTG TTAAGCACAT GGATGACCTC AAGGGCACCT TTGCCCATCT
T S F G D A V K H M D D L K G T F A H L

601 GAGCGAGCTG CACTGTGACA AGCTGCACGT GGATCCCGAG AACTTCCGGg tgagtccagg
S E L H C D K L H V D P E N F R

661 aaaggagcat gtgcccattt ttgcttttta cttattctga aataatgggt caagttcagt

721 gttaggagga cagaactcta gttggcataa ccaataccaa ctagtctcag aagtgggaacc

781 agtatgggtc tactaccagc ctgattttcc ctaaaacctt ttagggactc tagccaatat

841 tagttgtgtt tatcctggag agtattgggtg agtagcagaa ctgtgcaagg agaaatcaaa

901 tttggctact tatagatgaa ggactatata aaagatggaa taagccttat tttgttgagt

961 actaaggatt gatttggaga aaaatttgag acagattttt tcaaagaaaa atacaaaaat

1021 ttcctaatat atattaaatt ccctgtcagt attgtgttca agaaaaggct tgtccctata

Appendix 2-2F
Pale-throated sloth epsilon-globin gene sequence

1081 ttggctggag attttaactc cctctgagaa accttgcagc-actgaactcc tttgattaca
1141 ggagtgagat aactggtaaa tgagtgaagg aagtaaataa aacagctgaa ttgggcaaag
1201 gaggtgggaa tatacatagg cagaatactg gactgatggc tcataaaagt aatattgaac
1261 ccaggatcta gctaatttga gcacattgga taggcacaac tcttggagca gtttgagggtg
1321 aggagtgggt aggcagtaat ggtatttagc catTTTTatct gaaaattctt tttggaaact
1381 tctgttcaca tgcctgtgtg ttgtctgcct ttcccctaac agCTCCTGGG CAATGTGATG
L L G N V M
1441 GTGGTTGTTT TGGCTTCTCA TTTTGGCAAG GAATTCACTC CCGAAGTGCA GGCTGCTTGG
V V V L A S H F G K E F T P E V Q A A W
1501 CAGAAGTTGG TGGGTGGTGT TGCCAATGCT CTGGCTCACA AGTACCACTG Agccttctgt
Q K L V G G V A N A L A H K Y H .
1561 ccacctcgtc agggccccctg tgtccccagc catccttctg cacgtgttta atgggttttg
1621 gctttgagag cacagcttct ccttaataaaa gtgcattcta ttcagtaatt aatattttat
1681 ttcccttcate ttttgttctt gttttaaaagg aggaagggtt cattggctga ggggtgggaa
1741 gagacataag cataaacaag tgctcttttg aggagggtt gagttctcat caaaggaagg
1801 tagaagttaa cggaacgtt ctaggaggcc aaggggcata ctaagatgca gaaggagttt
1861 tctaaggcgt ggaaaaggct tctgtgaagt ggacaggtagc ttgtggggcg gtgtggcaaa
1921 aacttacaca aaggaaggaa ggggatggta agcagtgtt gagaaagaag aaagcagatt
1981 tgtggtaaat acatgcagac atacatacat aaaataataa aaaataagct ttgtagggat
2041 ccaagttcta gcctagagct actatttcaa tttgctctaa cacatatttt taagcagcca
2101 gcttgtcaga atgtttgtga gg

Appendix 2-3. Lengths of exons and introns for 'ε-like' globin genes. Values do not include those of the initiation and termination codons. "--" denotes missing data.

Taxon	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	Gene length
African elephant γ	89 bp	123 bp	223 bp	879 bp	126 bp	1,440 bp
Asian elephant γ	89 bp	123 bp	223 bp	879 bp	126 bp	1,440 bp
dugong γ	89 bp	123 bp	223 bp	887 bp	126 bp	1,448 bp
manatee γ	89 bp	123 bp	223 bp	876 bp	126 bp	1,437 bp
elephant shrew ε	--	--	--	795 bp	126 bp	--
sloth ε	89 bp	121 bp	223 bp	773 bp	126 bp	1,332 bp
armadillo ε	89 bp	125 bp	223 bp	819 bp	126 bp	1,382 bp
armadillo ψγ	116 bp	122bp	200 bp	854 bp	126 bp	1,418 bp
armadillo ψη	absent	absent	absent	~817 bp	125 bp	n/a

Appendix 2-4. Lengths of sequence data obtained for the 5' external region of 'ε-like' globin genes, and positions of upstream features relative to the putative Cap site. The positions of putative Cap sites from the initiation codon ("ATG") are also shown. Lengths of collected 5' end sequence data do not include that of the initiation codon. "--" denotes missing data.

Taxon	CACCC site	CCAAT site	ATA site	Putative Cap site	Length of obtained 5' external data
Asian elephant γ	--	--	--	--	-32 bp
dugong γ	absent	-84 bp	-30 bp	-52 bp	-216 bp
manatee γ	--	--	-30 bp	-52 bp	-127 bp
sloth ε	-112 bp	-83 bp	-30 bp	-53 bp	-213 bp
armadillo ε	-111 bp	-82 bp	-32 bp	-54 bp	n/a
armadillo ψγ	-103 bp, mutated into CTCCC	-69 bp	-28 bp	-9 bp	n/a
armadillo ψη	absent	absent	absent	absent	n/a

Appendix 2-5. Lengths of sequence data obtained for the 3' external region of 'ε-like' globin genes, and positions of polyadenylation signals (AATAAA) relative to the termination codon. Lengths of collected 3' end sequence data do not include that of the termination codon. "--" denotes missing data.

Taxon	AATAAA signal	Length of obtained 3' external data
African elephant γ	--	+9 bp
Asian elephant γ	+197 bp, +213 bp	+320 bp
dugong γ	+48 bp, +65 bp	+371 bp
manatee γ	--	+8 bp
elephant shrew ε	--	+70 bp
sloth ε	+93 bp	+571 bp
armadillo ε	+82 bp	n/a
armadillo ψγ	+29 bp	n/a
armadillo ψη	+118 bp	n/a