The mammalian α^D -globin gene lineage and a new model for the molecular evolution of α -globin gene clusters at the stem of the mammalian radiation

Steven J.B. Cooper^{a,b,*}, David Wheeler^{b,c}, Alison De Leo^d, Jan-Fang Cheng^e,
Robert A.B. Holland^{f,g}, Jennifer A. Marshall Graves^{d,h}, Rory M. Hope^{b,c}

^a Evolutionary Biology Unit, South Australian Museum, Adelaide, South Australia 5005, Australia

^b Centre for Evolutionary Biology and Biodiversity, The University of Adelaide, South Australia 5005, Australia

^c School of Molecular and Biomedical Science, The University of Adelaide, South Australia 5005, Australia

^d Department of Zoology, The University of Melbourne, Melbourne, Vic. 3010, Australia

^e Joint Genome Institute and Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

^f Department of Physiology and Pharmacology, School of Medical Sciences, University of New South Wales, Sydney 2052, Australia

^g Proteome Systems Ltd, North Ryde, NSW 2113, Australia

^h Comparative Genomics Research Group, Research School of Biological Sciences, Australian National University, Canberra, ACT 0200, Australia

Received 26 April 2005; revised 30 May 2005; accepted 31 May 2005

Available online 15 July 2005

Abstract

We have explored the evolution of the α -globin gene family by comparative sequence and phylogenetic analyses of mammalian α -globin genes. Our analyses reveal the existence of a new α -globin gene lineage in mammals that is related to the α^D -globin genes of birds, squamates and turtles. The gene is located in the middle of the α -globin gene cluster of a marsupial, *Sminthopsis macroura* and of humans. It exists in a wide variety of additional mammals, including pigs, cows, cats, and dogs, but is a pseudogene in American marsupials. Evolutionary analyses suggest that the gene has generally evolved under purifying selection, indicative of a functional gene. The presence of mRNA products in humans, pigs, and cows also suggest that the gene is expressed and likely to be functional. The analyses support the hypothesis that the α^D -globin gene lineage has an ancient evolutionary origin that predates the divergence of amniotes. The structural similarity of α -globin gene clusters of marsupials and humans suggest that an eight gene cluster (5'- ζ 2- ζ 1- α^D - α 3- α 2- α 1- θ - ω -3'), including seven α -like genes and one β -like globin gene (ω -globin) existed in the common ancestor of all marsupial and eutherian mammals. This basic structure has remained relatively stable in marsupials and in the lineage leading to humans, although ω -globin has been lost from the α -globin gene cluster of humans.

© 2005 Elsevier Inc. All rights reserved.

Keywords: AlphaD-globin; Alpha-globin; Globin gene evolution; Human mu-chain globin; Marsupial; PAML

1. Introduction

From the mid-1970s to the late 1980s the α -globin gene family of humans and other mammals was explored in considerable detail at the molecular level.

Following the discovery of a new pseudogene ($\varphi\alpha$ 2) (Hardison et al., 1986) and the characterization of a θ -globin gene (Hsu et al., 1988; Leung et al., 1989; Marks et al., 1986) it was commonly believed that the entire α -globin family of humans had been discovered and a relatively firm picture had emerged about how this gene family had evolved in mammals. This model of evolution revealed a complex history of tandem gene duplications,

* Corresponding author. Fax: +61 8 82077 222.

E-mail address: cooper.steve@saugov.sa.gov.au (S.J.B. Cooper).

recruitment of genes into different developmental regimes, gene inactivation and gene conversion events, providing a paradigm of how genomes evolve to produce functional complexity (Hardison, 1998). However, as will be shown in this paper, there is still much to be learnt about the evolutionary history and structure of the α -globin gene family in mammals.

The human α -globin gene cluster comprises seven tightly linked genes ($\zeta 2$ - $\psi \zeta 1$ - $\psi \alpha 2$ - $\psi \alpha 1$ - $\alpha 2$ - $\alpha 1$ - θ), with the $\zeta 2$ -globin gene expressed in early embryonic development and the α - and $\alpha 2$ -globin genes expressed during foetal and adult development (Lauer et al., 1980; Orkin, 1978; Proudfoot and Maniatis, 1980). $\psi \zeta 1$, $\psi \alpha 2$, and $\psi \alpha 1$ are reported to be pseudogenes, while the functional status of θ -globin is still uncertain, although it appears to be under strong stabilizing selection in a number of divergent mammalian lineages (Cooper et al., 2005; Shaw et al., 1987). The evolutionary history of this gene family has been explored by comparative sequence and phylogenetic studies of avian and mammalian α -globin gene clusters. There is strong support from these studies for the early duplication of an ancestral α -globin gene, prior to the divergence of the avian/mammalian lineages from a common ancestor, that produced the progenitors of ζ -globin and adult α -globin of mammals (Czelusniak et al., 1982; Proudfoot et al., 1982) and π -globin and α^A -globin of avians (Dodgson et al., 1981; Engel and Dodgson, 1980). A second duplication of an ancestral α -globin gene in the mammalian lineage, prior to the divergence of the marsupial and eutherian lineages, produced the progenitor of θ -globin (Cooper et al., 2005). Therefore, these data suggest that at least three functional α -like globin genes existed in the common ancestor of eutherian and marsupial mammals. The timing of further duplication events leading to the evolution of $\zeta 2/\psi \zeta 1$, $\psi \alpha 2$, $\psi \alpha 1/\alpha 2/\alpha 1$ -globin genes in humans, however, is unclear as much of the history of the genes has been erased by gene conversion events.

A third α -globin chain, called α^D -globin, was discovered in avians in the 1970s (Brown and Ingram, 1974; Hagopian and Ingram, 1971), but more recently has been found to exist in a wide variety of Reptilia including squamates (e.g., snakes: *Liophis miliaris*, Matsuura et al., 1989; lizards: *Varanus komodoensis*, Fushitani et al., 1996), sphenodontids (*Sphenodon punctatus*; Abbasi et al., 1988), and turtles (*Chrysemys bellii*, Rücknagel and Braunitzer, 1988; *Geochelone gigantea*; Shishikura, 2002). Phylogenetic analyses suggested that the α^D -globin gene lineage arose after an ancient duplication of an ancestral α -globin gene that pre-dated the divergence of Amniota (reptiles, avians, and mammals; Czelusniak et al., 1982; Fushitani et al., 1996), but to date, this lineage has been presumed to be lost from the genomes of mammals.

By comparative sequence analyses of α -globin gene families among distantly related mammal groups, such as marsupials and eutherians, it is possible to reconstruct the

history of events that lead to the evolution of the α -globin gene family. We have recently reported that the tammar wallaby (a marsupial) α -globin gene family consists of at least four functional genes including two ζ -globin genes, an adult α -globin and a θ -globin gene (Cooper et al., 2005; Wheeler et al., 2004). Remarkably, the adult α -globin and θ -globin genes were shown to be linked to a β -like globin gene, called ω -globin (Wheeler et al., 2001), providing evidence that the unlinked clusters of α -globin and β -globin genes in avians and mammals evolved from an ancient chromosome or genome duplication (Wheeler et al., 2004). Recently, sequence analyses confirmed the α -globin/ ω -globin linkage in another marsupial, *Sminthopsis macroura*, and showed the existence of a number of additional genes or pseudogenes (De Leo et al., 2005). A draft sequence of the α -globin gene cluster of a second marsupial, the North American opossum, *Didelphis virginiana*, has also become publicly available (via GenBank). The addition of these marsupial sequences to the extensive data available on the eutherian globin genes, now allows a much more detailed re-examination of the evolutionary history of the α -globin gene family of mammals.

In the present study, we have carried out comparative and phylogenetic analyses of globin genes and gene clusters from marsupial and eutherian mammals and identified a new mammalian α -globin gene lineage with a long-term evolutionary history that predates the divergence of amniotes. We show that this gene lineage is orthologous to the α^D -globin genes of avians/reptiles/turtles and is represented by an expressed gene of unknown function located in the human α -globin gene cluster, and either functional genes or pseudogenes in marsupials and other eutherian mammals. In addition, our analyses indicate that the basic structure of the α -globin gene cluster has remained relatively stable in a number of eutherian and marsupial species since the divergence of the marsupial and eutherian lineages from a common ancestor.

2. Methods

2.1. Sequence data

DNA sequences were obtained from GenBank for the α -globin gene clusters of four marsupial species, *Macropus eugenii* (GB# AY459590, AY459589, AY789121, and AY789121), *S. macroura* (GB# AC146781), *Monodelphis domestica* (TI# 469255534, TI# 453585430), and *Didelphis virginiana* (GB# AC148752 and AC139599).

2.2. Identification of exon sequences in α -like globin genes

The program BLAST 2 sequences (Version BLASTN 2.2.9; Tatusova and Madden, 1999) available from the Website <http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html> was used to compare two sequences to

identify exon regions of α -like globin genes. Gene structure was also derived assuming that exons in most functional vertebrate α -globin genes have a very similar sequence length and highly conserved intron/exon splicing signals. Gene structure was additionally verified using mRNA sequences, when available.

2.3. Sequence alignments and phylogenetic analyses

DNA sequences (exonic regions only) from marsupial and eutherian α -globin genes were aligned using a combination of CLUSTAL X (Thompson et al., 1997) (for aligning pseudogenes with indels) and manual alignment (for functional globin genes). A wide variety of α -like globin genes from vertebrates were used in analyses, representing exemplar sequences of known embryonic and adult-expressed genes from mammals, avians, and

amphibians (see caption of Fig. 1 for accession numbers). Fish α -globin genes (yellowtail and salmon) were used as outgroups to root the phylogeny.

Evolutionary relationships among α -like globin genes were estimated using both Maximum Parsimony (MP) and Bayesian approaches, applied to the aligned coding sequences only, to assess the robustness of various nodes in the phylogeny to variations in methods (e.g., tree search algorithms and models of DNA sequence evolution). MP analyses were carried out using PAUP*4b10 (Swofford, 2002) with a standard heuristic search option and random stepwise addition, repeated 100 times, with character state optimization using the DELTRAN option. MP bootstrap analyses (Felsenstein, 1985) were carried out using 1000 bootstrap pseudoreplicates and the default heuristic search option.

Fig. 1. A 50% majority rule consensus phylogram from a Bayesian analysis constructed using a GTR + I + G model of evolution and estimated base frequencies in an unlinked analysis using MRBAYES. Numbers above branches refer to posterior probabilities. Numbers below branches of major nodes are bootstrap proportions (>50%) of 1000 MP pseudoreplicates. GenBank accession numbers for sequences are: North American Opossum (*Didelphis virginiana*) ζ 1, ζ 2, ψ α D, α 1, α 2, θ (AC139599.2, AC148752.1); *Sminthopsis macroura* partial ζ , α D, α 1, α 2, α 3, θ (AC146781); *Monodelphis domestica* ψ α D (TI# 469255534), α (TI# 453585430); tammar (*Macropus eugenii*) θ (AY459590), α (AY459589); ζ (AY789121), ζ ' (AY789121); horse (*Equus caballus*) θ (Y00284), α 1 (M17902), ζ (X07051); pig (*Sus scrofa*) α D (AC145444); dog (*Canis familiaris*) α D (AAEX01044727); cat (*Felis catus*) α D (AC130194); cow (*Bos taurus*) α D (AC150547); goat (*Capra hircus*) α (J00043); human (*Homo sapiens*) α 1 (V00491), θ (X06482), ζ (NM 005332); μ -chain (AY698022); ψ ζ (J00184); mouse (*Mus musculus*) α A1 (NM0008218), ζ (X62302); rabbit (*Oryctolagus cuniculus*) α (X04751), ψ ζ 1 (AH001223), θ (X04751); rat (*Rattus norvegicus*) θ (X56330); *Dasyurus viverrinus* α (M14567); chicken (*Gallus gallus*) α A, π , α D (AF098919); duck (*Cairina moschata*) α D (X01831); pigeon (*Columba livia*) α D (AB001981). Turtle (*Geochelone nigra*) α D (SEG AB1165195); yellowtail (*Seriola quinquerata*) α A (AB034639); salmon (*Salmo salar*) α (X97289); salamander (*Hynobius retardatus*) larval (AB034756); salamander (*Pleurodeles waltlii*) α (M13365); *Xenopus laevis* α (X14259), tadpole α T5 (X02798).

Bayesian phylogenetic analyses were carried out using the program MRBAYES (Version 3, Huelsenbeck and Ronquist, 2001). Previous studies of the α -globin genes have indicated that a General Time Reversible model (Rodríguez et al., 1990), with a proportion of invariant sites and unequal rates among sites (Yang, 1996) modelled with a γ distribution (GTR + I + G) was the most appropriate model for these analyses (Cooper et al., 2005). Due to the lack of homogeneity of base frequencies at third positions we carried out the Bayesian analyses applying separate model parameters to codon positions using the “unlink” option in MRBAYES. The MRBAYES analysis was carried out using default uninformative priors, running four chains simultaneously for two million generations sampling trees every 100 generations. The likelihood values converged to relatively stationary values after about 25,000 generations. A conservative burnin of 500 trees (equivalent to 50,000 generations) was chosen with a >50% posterior probability consensus tree constructed from the remaining 19,500 trees. The entire procedure was repeated once to verify that analyses converged to the same overall tree topology.

2.4. ML analyses using PAML

Likelihood ratio tests of hypotheses of neutral evolution vs stabilizing selection on different branches of the α -globin phylogeny were carried out using the program

PAML (version 3.14beta3; Yang, 1997). Full details of these analyses are described in Section 3.

3. Results

3.1. The α -globin gene family of *S. macroura*

Sequence analyses comparing a BAC sequence of the α -globin gene cluster of *S. macroura* with a combination of exonic regions from ζ -globin, α -globin, θ -globin, and ω -globin genes from the tammar wallaby, *M. eugenii*, identified the presence of seven genes (Table 1). The identities of the genes were verified using phylogenetic analyses (see below and Fig. 1). The order and arrangement of these genes are shown in Fig. 2. The cluster includes a β -like globin gene, called ω -globin, located 8586 bp 3' to an α -like globin gene, θ -globin, both of which were previously identified by De Leo et al. (2005). Two genes, labelled $\psi\alpha1$ -globin and $\psi\alpha3$ -globin (Table 1) contain a number of insertions/deletions, relative to a standard α -globin gene sequence, that lead to frameshift mutations and stop codons at positions 52 and 9, respectively, in the translated globin. Therefore, it is most likely that these two genes are pseudogenes. Both of these pseudogenes form a monophyletic group with a gene, labelled $\alpha2$ -globin, supported by a high posterior probability of 99% (Bayesian trees) and bootstrap support of 94% (Fig. 1). The $\alpha2$ -globin gene

Table 1

Location of exon/intron boundaries for non-annotated GenBank sequences used in phylogenetic analyses

Species/gene	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	Stop	GenBank No.
<i>S. macroura</i>							
$\zeta1$ (= ζ')	—	—	—	—	1	85	AC146781
α^D	4028	4120	4630	4835	5184	5312	AC146781
$\psi\alpha3$	10394	10486	10775	10964	11235	11360	AC146781
$\alpha2$	17044	17139	17415	17620	17889	18017	AC146781
$\psi\alpha1$	25323	25415	25723	25899	26166	26280	AC146781
θ	36212	36307	36524	36729	36972	37094	AC146781
<i>D. virginiana</i>							
$\zeta2$	103729	103824	104753	104958	105203	105331	AC139599.2
$\zeta1$	114340	114435	115777	115982	116422	116550	AC139599.2
$\psi\alpha^D$	—	—	—	—	117581	117714	AC139599.2
$\alpha3$	123099	123194	123503	123708	124046	124174	AC139599.2
$\alpha2$	157951	158046	158347	158552	158889	158951 ^a	AC139599.2
$\alpha2$	15618	15713	16014	16219	16560	16688	AC148752
$\alpha1$	33434	33529	33838	34043	34383	34511	AC148752
θ	47020	47115	47338	47543	47646	47774	AC148752
Human α^D	26861	26953	27131	27336	27443	27571	NG_000006
Human α^D	155997	156089	156267	156472	156579	156707	AE006462
<i>M. domestica</i> α^D	—	—	—	—	201	350	Ti_469255534
Cow α^D	130991	130899	130744	130538	130427	130299	AC150547
Pig α^D	3	—	95	—	300	428	CK459716
Cat α^D	146430	146522	146682	146887	146988	147116	AC130194
Dog α^D	—	—	1	187	277	405	AAEX01044727

Numbers represent the start of the exon/intron or end of gene (stop) within the GenBank sequence specified.

^a Third exon sequence differs from this position from that given for $\alpha2$ -globin in AC148752. The sequence from AC148752 is used in phylogenetic analyses.

Fig. 2. Comparison of gene order and arrangement in the α -globin gene clusters of humans, *Sminthopsis macroura*, *Didelphis virginiana*, and tammar (*Macropus eugenii*). Note that the human α^D -globin gene replaces the pseudogene $\psi\alpha2$ -globin.

appears to have no mutations that would impair translation of the transcribed message into a functional α -globin polypeptide chain. Phylogenetic analyses show high bootstrap and posterior probabilities (100%) that $\alpha2$ -globin is orthologous to the adult α -globin gene of the Australian dasyurid marsupial *Dasyurus viverrinus*. The third exon of an α -like globin gene lies at the immediate 5' end of the BAC clone (Table 1). Phylogenetic analyses show that this gene groups most closely with the tammar ζ' -globin gene to the exclusion of a second tammar ζ -globin gene.

The coding region of an additional α -globin gene was identified by sequence alignment with tammar ζ -globin exon sequences and by the presence of conserved intron/exon splicing signals (Fig. 3). The gene consists of three exons and two introns, typical of functional vertebrate α -globin genes, but does not appear to have a conventional α -globin promoter in that it lacks any evidence of a CCAAT and TATA box (Fig. 3). Promoter analyses, using the program MATINSPECTOR V2.2 professional (http://www.genomatix.de/mat_fam) and the TRANSFAC database, reveal the possible presence of a binding site for GATA factor and SP1 (Fig. 3). A 3 bp deletion relative to other functional α -globin genes occurs in the first exon of this gene, but this does not impair an open reading frame that would lead to translation of a globin polypeptide of 141 amino acid residues, by inclusion of the N-terminal methionine. BLASTn analyses of the GenBank database revealed closest identity of the gene to a mRNA sequence of an α -like globin gene derived from humans called haemoglobin mu-chain (GB# AY698022, submitted by S.-H. Goh, Y.T. Lee, and J.L. Miller in July 2004). The predicted amino acid sequences of each gene show a sequence similarity of 63% (Fig. 3).

When the human mu-chain sequence was used in BLASTp and BLASTn searches of public databases available at NCBI (HTGS, WGS, EST; <http://www.ncbi.nlm.nih.gov/>), high scores were obtained for a

number of additional mammalian α -globin genes from cow (*Bos taurus*, GB# AC150547), cat (*Felis catus*, GB# AC130194), pig (*Sus scrofa*, cDNA: GB# CK459716, genomic DNA: GB# AC145444), dog (*Canis familiaris*, GB# AAEX01044727), and chimpanzee (*Pan troglodytes*: cDNA: GB# CR597411) (see Table 1). BLAST analyses of genomic sequences derived from mouse (*Mus domesticus*) and rat (*Rattus rattus*) genome projects failed to show any significant scores.

Alignment of each of the above eutherian and *S. macroura* α -globin genes revealed the same 3 bp deletion in the first exon of each gene. Phylogenetic analyses using both MP and Bayesian approaches provide strong evidence that the above eutherian mammal genes form a monophyletic group with the *S. macroura* unknown α -like globin gene to the exclusion of all other mammalian ζ -globin, θ -globin, and α -globin genes, supported by high posterior probabilities and bootstrap values (100%) (Fig. 1). Furthermore, there is strong evidence that this lineage of mammalian α -globin genes is orthologous to avian/turtle α^D -globin genes, supported by a high posterior probability (100%) and moderate bootstrap values (78%). Following the convention used in avians, the genes within this entire mammalian α -like globin lineage are henceforth referred to as α^D -globin.

3.2. Identification of α^D -globin pseudogenes in American marsupials

Two BAC sequences, available on GenBank, contain regions of the α -globin gene cluster of the North American opossum *D. virginiana*. Currently, these sequences are a working draft, so results using these sequences should be considered preliminary. Sequence analyses of the *D. virginiana* α -globin gene cluster, using similar methods to that described above for *S. macroura*, revealed the presence of eight genes/pseudogenes (Table 1). The identity of these genes also was assessed using phylogenetic analyses and the provisional structure of the gene cluster is shown in Fig. 2. Overall, this structure is very similar to that found in *S. macroura*, with a β -like globin gene (called ω -globin) at the 3' side of the cluster, adjacent to θ -globin and three α -globin genes related to adult α -globin genes (see Fig. 2). Within the middle of the cluster a sequence was identified that was most similar to the *S. macroura* α^D -globin third exon. However, no significant similarity to the *S. macroura* α^D -globin first and second exons could be identified in the region (990 bp) between the *D. virginiana* $\zeta1$ -globin gene (position 114,340; GB#AC139 599) and this putative α^D -globin exon sequence. In addition, the third exon sequence had two insertions, one of 1 bp in length and a second of 4 bp in length, when compared to *S. macroura* α^D -globin. These insertions lead to a frameshift and a premature stop codon in the third exon. Overall, it is most likely that *D. virginiana* α^D -globin is a pseudogene.

Fig. 3. DNA sequence and predicted protein sequence of the α^D -globin gene from *S. macroura* (A) and humans (B). Putative promoter signals (SPI and GATA) in the 5'-flanking sequence are highlighted in bold and underlined. Additional conserved signals such as intron/exon boundaries and polyadenylation signals are underlined. Full intron sequences of *S. macroura* α^D -globin are not shown. Residues highlighted in bold are in positions suggested by Knapp et al. (1999) to be involved in the formation of octomers by chicken HbD.

Searches of the *M. domestica* sequencing project also revealed the presence of an α^D -globin third exon region (TI# 469255534) in this species (Table 1). The first and second exon sequence for this gene could not be identified in the *M. domestica* DNA database (as at 1st December, 2004), in concordance with the results of the *D. virginiana* analyses. The presence of a 1 bp insertion in the third exon leads to a premature stop codon in the translated sequence, further suggesting that *M. domestica* α^D -globin is a pseudogene.

3.3. Identification of an α^D -globin genomic orthologue in the human α -globin cluster

We further explored the existence of an α^D -globin orthologue in the human α -globin cluster by sequence

comparisons of a human haemoglobin mu-chain mRNA sequence (GB# AY698022; shown above to be derived from a gene orthologous to *S. macroura* α^D -globin) with that of genomic DNA sequence of the α -globin (HBA) region on chromosome 16 (GB# NG_000006). The mRNA sequence was found to be identical to a region of the α -globin cluster previously identified to contain a pseudogene $\psi\alpha 2$ -globin (Hardison et al., 1986). Our analyses indicate that the exon/intron boundaries of $\psi\alpha 2$ -globin were mistakenly identified (see Table 1). BLAST 2-sequence analyses revealed three regions of identity of the mRNA sequence to the genomic sequence and showed that the human α^D -globin gene contains a three exon/two intron structure, conserved intron/exon splicing signals, termination and start signals typical of

known functional α -like globin genes. The promoter region showed the absence of conventional CCAAT and TATA boxes and, in concordance with *S. macroura* α^D -globin, showed the possible presence of GATA factor and SP1 binding sites.

3.4. Is mammalian α^D -globin evolving like a functional gene?

To test the possibility that α^D -globin is evolving under purifying selection (indicative of a functional gene), or at neutral rates of evolution (indicative of a pseudogene) we carried out a series of likelihood ratio tests (LRTs), applying different estimates of the nonsynonymous/synonymous divergence ratio ($\omega = d_N/d_S$) to different branches of an α -globin phylogeny using the program PAML (version 3.14beta3; Yang, 1997). To reduce computational time and potential saturation problems at synonymous sites we used a reduced number of taxa in the phylogeny to include only chicken and mammalian

adult α -globin genes and mammalian α^D -globin genes (Fig. 4). This phylogenetic tree is based on strongly supported relationships among mammalian orders (Murphy et al., 2001) and our analyses that support the orthology of marsupial and eutherian adult α -globin and α^D -globin genes, respectively. If the α^D -globin lineage was evolving at neutral rates of evolution then a ratio $\omega = 1$ would be expected, while other parts of the α -globin phylogeny would evolve with $\omega < 1$, for genes under stabilising selection. A null hypothesis model with $\omega = 1$ (log likelihood value = l_0) can, therefore, be compared with an alternative hypothesis with the ω ratio free to vary along specified branches of the α^D -globin lineage with a second ratio for the remainder of the tree (Table 2). Given that the null hypothesis is nested within the more general hypothesis with two ratios that are free to vary, twice the log likelihood difference, $2\Delta l = 2(l_1 - l_0)$, should approximate a χ^2 distribution with one degree of freedom corresponding to an increase of one parameter in the more complex model (Yang, 1998). An alternative test of a potential loss of constraint along a lineage of the phylogeny compares the null hypothesis of a single ratio in the tree with an alternative hypothesis allowing two ratios; ω_0 (background ratio) and ω_H free to vary in the mammalian α^D -globin lineage (Table 2).

When ω was free to vary along different α^D -globin lineages, maximum likelihood estimates of ω varied between 0.037 and 0.114, suggesting that the gene has been under strong purifying selection during its evolution in mammals. Furthermore, the LRT comparing a one ω ratio (null model) versus a two ω ratio model failed to reject the null model, suggesting that α^D -globin is evolving under similar selective constraints to that of the adult α -globin lineage. Likelihood ratio tests of neutrality for the mammalian, eutherian and marsupial α^D -globin lineages strongly rejected the null hypothesis of $\omega = 1$ ($p = 0.0000$) (Table 2). The LRT involving the human α^D -globin lineage failed to reject the hypothesis of neutrality at the 1% level of significance. However, it is possible that there is a lack of power in the latter test

Fig. 4. Phylogenetic tree used in the PAML (Yang, 1997) analyses showing relationships among marsupial and eutherian α -globin and α^D -globin genes. Codes and accession numbers for each of the sequences used are given in the caption of Fig. 1.

Table 2

Log likelihood values (L), parameter estimates under different models using PAML and results of likelihood ratio tests (LRTs, $2\Delta l = 2(l_1 - l_0)$)

Model	Branches tested	np	L	ω_0	ω_H	Models compared (H_0 vs H_1) and LRTs
A. One ratio: $\omega_0 = \omega_H$	—	26	−3498.83	0.1195	—	A vs B, $P = 0.3381$ (NS)
B. Two ratios: ω_0, ω_H	Mammal α^D	27	−3498.37	0.1273	0.1046	C vs B, $P = 0.0000^*$
C. Two ratios: $\omega_0, \omega_H = 1$	Mammal α^D	26	−3594.71	0.1114	1	
D. Two ratios: ω_0, ω_H	Eutherian α^D	27	−3498.79	0.1218	0.1137	E vs D, $P = 0.0000^*$
E. Two ratios: $\omega_0, \omega_H = 1$	Eutherian α^D	26	−3555.91	0.1114	1	
F. Two ratios: ω_0, ω_H	Eutherian α^D	27	−3496.83	0.1258	0.0371	G vs F, $P = 0.0000^*$
G. Two ratios: $\omega_0, \omega_H = 1$	Eutherian α^D	26	−3506.14	0.1148	1	
H. Two ratios: ω_0, ω_H	Human α^D	27	−3498.80	0.1203	0.1017	I vs H, $P = 0.0784$ (NS)
I. Two ratios: $\omega_0, \omega_H = 1$	Human α^D	26	−3500.35	0.1150	1	

The input tree used for the analysis is shown in Fig. 4.

NS, not significant.

* Significant, $p < 1\%$, under a $\chi^2(1)$ distribution.

to reject the null model due to the relative shortness of the evolutionary branch leading to human α^D -globin, where the estimated ω ratio is 0.102, indicative of a gene under purifying selection. Overall, the PAML analyses suggest strongly that the α^D -globin gene has evolved under purifying selection in each of the considered mammalian lineages and is most likely a functional gene.

4. Discussion

DNA sequence comparisons and phylogenetic analyses of marsupial and eutherian α -globin genes provide strong evidence for the existence of a new α -globin gene lineage in mammals that is orthologous to the α^D -globin genes of avians/turtles. These analyses further suggest that the α^D -globin lineage arose after an ancient duplication of an ancestral α -globin gene, supporting previous proposals that the α^D -globin gene arose prior to the divergence of all amniotes (mammals, squamates, turtles, sphenodon, crocodilians, and avians; Czelusniak et al., 1982; Fushitani et al., 1996). Remarkably, an orthologue of this gene is present in the human α -globin gene cluster where it was previously considered to be a pseudogene called $\psi\alpha 2$ -globin (Hardison et al., 1986). The recent sequencing of a mRNA called “haemoglobin μ -chain” by Goh and colleagues (GB# AY698022) has allowed this gene to be correctly annotated, revealing a structure that is highly conserved among functional α -globin genes, with the exception that the gene contains a deletion of a single codon at the start of the first exon. We, henceforth, propose that the human $\psi\alpha 2$ -globin gene and “haemoglobin μ -chain” be designated “human α^D -globin” in line with the notation used for the orthologue of this gene in avians. The location of α^D -globin, between ζ -globin and adult α -globin genes of both humans and *S. macroura*, matches the arrangement of the three genes in the chicken α -globin cluster (5'- π - α^D - α -3'; π is orthologous to mammalian ζ -globin; Dodgson et al., 1981; Engel and Dodgson, 1980), lending further support for the orthology of mammalian and avian α^D -globin.

We have also found that α^D -globin is present in the genomes of a wide variety of eutherian mammals, including cats, dogs, pigs, and cows, but to date, we have been unable to detect the presence of this gene in genome sequences from rats and mice, or rabbits. The finding of α^D -globin mRNA sequences in humans, cows, and pig confirm that the gene is expressed in these mammals. The maximum likelihood analyses using PAML suggest that in eutherian lineages α^D -globin is evolving under purifying selection and is, therefore, most likely to be functional. In marsupials a full-length α^D -globin gene exists in the dunnart *S. macroura*, which appears to have no sequence defects that would render the gene inactive. However, at present, no mRNA sequence or protein

sequence is available to verify that this gene is expressed in *S. macroura* or other marsupials. Its overall sequence conservation (63% amino acid sequence identity), when compared to human α^D -globin, and the PAML analyses suggest it is under strong selective constraints ($\omega = 0.0371$, Table 2) and is, therefore, likely to be functional. In contrast, the American marsupials (*D. virginiana* and *M. domesticus*) appear to have only a partial sequence of an α^D -globin gene, each containing insertions/deletions leading to premature stop codons, and, therefore, are likely to be pseudogenes.

The cDNA for pig α^D -globin was derived from a library made from pooled mRNA from embryos between 10 and 30 days old. Human cDNA of fetal liver was also identical to μ -chain globin which is reported to be extracted from reticulocytes in whole blood. It appears likely, therefore, that mammalian α^D -globin may be expressed in both embryonic, fetal, and adult stages of development. However, further expression analyses are necessary to confirm this possibility. Interestingly, the α^D -globin gene of avians is also expressed at similar stages of development, being found in both embryonic red blood cells of 2- to 5-day-old chickens and “definitive” cells which appear at 6–7 days (Bruns and Ingram, 1973). A number of squamates (snakes and lizards) have also been shown to have α^D -globin genes, following the isolation of α^D -globin polypeptides from adult blood (Fushitani et al., 1996; Matsuura et al., 1989). The α^D -globins of avians and turtles have 141 amino acid residues (Shishikura, 2002), despite the first exon of their encoding genes having a 3 bp deletion relative to the standard α -globin sequence length. For these species, the α^D -globin chain includes the N-terminal methionine and it is likely that α^D -globin genes of mammals, which share the 3 bp deletion, also encode a chain of 141 residues by inclusion of the N-terminal methionine. The relative amounts of α^D - to α^A -globin in adult haemoglobin of avian species vary between 15 and 45% (Eguchi and Eguchi, 2001; Oberthür et al., 1983). It is unknown what proportion of the total haemoglobin in humans contains α^D -globin, but a similar level to that found in avians might have important implications in diseases such as α thalassemia. Currently, mutations of the α^D -globin locus in humans are not assessed in molecular screens for mutations of patients with thalassemia, so it is uncertain whether mutations of this locus may be associated with disease phenotypes.

It is difficult to predict the functional significance of α^D -globin in mammals, but clues to this function may be evident from previous studies of the oxygen-binding properties of haemoglobin containing α^D (HbD) from chickens (e.g., Baumann et al., 1987; Knapp et al., 1999). These studies indicated that red cells containing sufficient quantities of HbD (e.g., ~300 mg/ml) show supercooperativity (Hill n greater than 4) in the upper part of the oxygen equilibrium curve (35–36), which results

when tetramers associate into octamers. This property may be advantageous for delivery of greater quantities of oxygen to tissues than normal tetrameric haemoglobins under standard physiological conditions. On the basis of X-ray crystallography, Knapp et al. (1999) identified 10 hydrophobic residues on the surface of the HbD tetramers, which they considered could associate with residues on the surface of another tetramer to produce an octamer. Of these residues, four were in the α^D chains (Met at residue 1, Leu at 48, Tyr at 89, and Leu at 91) and six in the β chains. Three out of four of these α^D chain residues (positions 1, 89, and 91) occur in the α^D chains of mammals (Fig. 3). However, we do not know with which β -type chains the α^D chains would combine to form a functioning molecule and so cannot know if the residues on the β chain surface could contribute to the formation of octamers, nor is it known whether α^D -globin occurs in sufficient concentrations in red blood cells to allow this process to occur. Thus, while we believe that the persistence of the α^D -globin gene in mammals indicates that it contributes to a physiologically useful molecule, we cannot yet say what its special function is.

The overall structure of the α -globin gene clusters in the two marsupial species, *S. macroura* and *D. virginiana*, are very similar (Fig. 2), although the *S. macroura* BAC clone is missing the 5'-end of the α -globin cluster. Given that the tammar wallaby has two ζ -globin genes (Cooper et al., 2005), one of which (ζ' -globin) is orthologous to the third exon of a partial ζ -globin gene at the terminal 5'-end of the BAC sequence from *S. macroura*, it is likely that *S. macroura* has an additional ζ -globin gene located 5' to the BAC sequence. The presence of similar gene clusters in both American and Australian marsupial lineages strongly indicates that the ancestral marsupial α -globin cluster consisted of eight genes including seven α -like globin genes and one β -like globin gene (ω -globin) linked in the order 5'- ζ - ζ' - α^D - $\alpha 3$ - $\alpha 2$ - $\alpha 1$ - θ - ω -3'.

The marsupial α -globin cluster is remarkably similar in both number and arrangement of genes to the human α -globin cluster (5'- $\zeta 2$ - $\psi \zeta 1$ - α^D - $\psi \alpha 1$ - $\alpha 2$ - $\alpha 1$ - θ -3'), with the exception that an ω -globin orthologue is absent in the human cluster (Flint et al., 2001; Wheeler et al., 2001, 2004). The orthology of ζ -globin, α^D -globin, α -globin, and θ -globin from marsupials and eutherians is strongly supported by phylogenetic analyses. However, the evolutionary relationships of the two ζ -globin genes and three α -globin genes have been obscured by gene conversion events (Flint et al., 1988; Proudfoot et al., 1982), making it impossible to reconstruct the actual history of duplication events that have lead to their evolution using conventional phylogenetic methods. We hypothesise that the conserved overall structure of the gene cluster with two ζ -globin genes at the 3'-end and three α -globin genes, orthologous to adult α -globin, in the middle, and the ancient origin of the β -like ω -globin gene (Wheeler

et al., 2001), suggests that an eight gene cluster, including seven α -like and one β -like globin gene, existed in the common ancestor of all marsupial and eutherian mammals (Fig. 2). Given the presence of only a three gene cluster in avians, including orthologues of ζ -globin, α^D -globin, and α -globin (Dodgson et al., 1981; Engel and Dodgson, 1980), a four gene cluster (5'- ζ - α^D - α - ω -5') was most likely the ancestral globin gene cluster of all amniotes. At least four duplication events of α -globin genes (one of ζ -globin and three of α -globin) occurred in the mammalian lineage after its divergence from a common ancestor with the avian lineage. A series of gene conversion events have also occurred in the eutherian lineage, leading to the concerted evolution of the two ζ -globin genes and three α -globin genes (Flint et al., 1988; Proudfoot et al., 1982). There is also evidence from our phylogenetic analyses for concerted evolution of the three adult α -globin genes in marsupials, as *D. virginiana* $\alpha 1$ and $\alpha 2$ -globin were found to be monophyletic to the exclusion of the three *S. macroura* α -globin genes (Fig. 1). This pattern of concerted evolution is not evident, however, for the evolution of the two ζ -globin genes in marsupials.

This new model of the evolution of α -globin gene clusters at the stem of the mammalian radiation, particularly the history of gene conversion events, requires further examination by comparative studies of α -globin gene clusters in additional divergent lineages of mammals, including monotremes. The availability of large amounts of DNA sequence data from genomic sequencing projects makes such analyses feasible in the short-term future. The comparative approach presented in this paper highlights the value of these genome projects targeting divergent lineages of mammals (e.g., platypus), in that new genes await discovery in the human genome, even within extremely well characterised genomic regions such as the human α -globin cluster.

Acknowledgments

An Australian Research Council grant to R.A.B.H., A. Gooley and R.M.H. helped fund our research on globins in marsupials. We thank A. Gooley for helpful discussion and support, for this research.

References

- Abbasi, A., Wells, R.M.G., Brittain, T., Braunitzer, G., 1988. Primary structure of the hemoglobins from *Sphenodon* (*Sphenodon punctatus*, Tuatara, Rhynchocephalia). *Biol. Chem. Hoppe-Seyler* 369, 755–764.
- Baumann, R., Fischer, J., Engelke, M., 1987. Functional properties of primitive and definitive red cells from chick embryo: oxygen-binding characteristics, pH and membrane potential, and response to hypoxia. *J. Exp. Zool. (Suppl. 1)*, S227–S238.

- Brown, J.L., Ingram, V.M., 1974. Structural studies on chick embryonic hemoglobins. *J. Biol. Chem.* 249, 3960–3972.
- Bruns, G.-A., Ingram, V.M., 1973. The erythroid cells and haemoglobins of the chick embryo. *Phil. Trans. R. Soc. London B* 266, 225–305.
- Cooper, S.J.B., Wheeler, D., Hope, R.M., Dolman, G., Saint, K.M., Gooley, A.A., Holland, R.A.B., 2005. The α -globin gene family of an Australian marsupial *Macropus eugenii*: long evolutionary history of the θ -globin gene and its functional status in mammals. *J. Mol. Evol.* (in press).
- Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M.L., Venta, P.J., Tashian, R.E., 1982. Phylogenetic origins and adaptive evolution of avian and mammalian hemoglobin genes. *Nature* 298, 297–300.
- De Leo, A.A., Wheeler, D., Lefevre, C., Cheng, J.-F., Hope, R., Kuliwaba, J., Nicholas, K.R., Westerman, M., Graves, J.A.M., 2005. Sequencing and mapping hemoglobin gene clusters in the Australian model dasyurid marsupial *Sminthopsis macroura*. *Cytogenet. Genome Res.* 108, 333–341.
- Dodgson, J.B., McCune, K.C., Rusling, D.J., Krust, A., Engel, J.D., 1981. Adult chicken α -globin genes, α^A and α^D : no anemic shock α -globin exists in domestic chickens. *Proc. Natl. Acad. Sci. USA* 78, 5998–6002.
- Eguchi, Y., Eguchi, T., 2001. Amino acid sequence of Kalinowski's Tinamou (*Nothoprocta kalinowskii*) hemoglobin and the rate of evolution of bird alpha(D)-globin. *J. Protein Chem.* 20, 543–549.
- Engel, J.D., Dodgson, J.B., 1980. Analysis of the closely linked adult chicken α -globin genes in recombinant DNAs. *Proc. Natl. Acad. Sci. USA* 77, 2596–2600.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Flint, J., Taylor, A.M., Clegg, J.B., 1988. Structure and evolution of the horse zeta globin locus. *J. Mol. Biol.* 199, 427–437.
- Flint, J., Tufarelli, C., Peden, J., et al., 2001. (11 co-authors), 2001. Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster. *Hum. Mol. Genet.* 10, 371–382.
- Fushitani, K., Higashiyama, K., Moriyama, E.N., Imai, K., Hosokawa, K., 1996. The amino acid sequences of two α chains of hemoglobins from Komodo dragon *Mu-anus komodoensis* and phylogenetic relationships of Amniotes. *Mol. Biol. Evol.* 13, 1039–1043.
- Hagopian, H.K., Ingram, V.M., 1971. Developmental changes of erythropoiesis in cultured chick blastoderms. *J. Cell Biol.* 51, 440–451.
- Hardison, R.C., Sawad, I., Cheng, J.-F., Shen, C.-K.J., Schmid, C.W., 1986. A previously undetected pseudogene in the human alpha globin gene cluster. *Nucleic Acids Res.* 14, 1903–1911.
- Hardison, R.C., 1998. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J. Exp. Biol.* 201, 1099–1117.
- Hsu, S.L., Marks, J., Shaw, J.P., Tam, M., Higgs, D.R., Shen, C.C., Shen, C.K., 1988. Structure and expression of the human theta 1 globin gene. *Nature* 331, 94–96.
- Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* 17, 754–755.
- Knapp, J.E., Oliveira, M.A., Xie, Q., Ernst, S.R., Riggs, A.F., Hackert, M.L., 1999. The structural and functional analysis of the hemoglobin D component from chicken. *J. Biol. Chem.* 274, 6411–6420.
- Lauer, J., Shen, C.-K.J., Maniatis, T., 1980. The chromosomal arrangement of human α -like globin genes: sequence homology and α -globin gene deletions. *Cell* 20, 119–130.
- Leung, S., Whitelaw, E., Proudfoot, N.J., 1989. Transcriptional and translational analysis of the human θ globin gene. *Nucleic Acids Res.* 17, 8283–8300.
- Matsuura, M.S.A., Fushitani, K., Riggs, A.F., 1989. The amino acid sequences of the α and β chains of hemoglobin from the snake, *Lio-phhis miliaris*. *J. Biol. Chem.* 264, 5515–5521.
- Marks, J., Shaw, J.-P., Shen, C.-K.J., 1986. Sequence organization and genomic complexity of primate θ 1 globin gene, a novel α -globin-like gene. *Nature* 321, 785–788.
- Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294, 2348–2351.
- Oberthür, W., Braunitzer, G., Baumann, R., Wright, P.G., 1983. Die Primärstruktur der α - und β -Ketten der Hauptkomponenten der Hämoglobine des Straußes (*Struthio camelus*) und des Nandus (*Rhea americana*) (Struthioformes). Aspekte zur Atmungsphysiologie und Systematik. *Hoppe-Seyler's Z. Physiol. Chem.* 364, 119–134.
- Orkin, S.H., 1978. The duplicated human α globin genes lie close together in cellular DNA. *Proc. Natl. Acad. Sci. USA* 75, 5950–5954.
- Proudfoot, N.J., Maniatis, T., 1980. The structure of a human α -globin pseudogene and its relationship to α -globin gene duplication. *Cell* 21, 537–544.
- Proudfoot, N.J., Gil, A., Maniatis, T., 1982. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell* 31, 553–563.
- Rodriguez, F., Oliver, J.F., Mar 'n, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitutions. *J. Theor. Biol.* 142, 485–501.
- Rücknagel, K.P., Braunitzer, G., 1988. Hemoglobins of reptiles. The primary structure of the major and minor hemoglobin component of adult Western Painted Turtle (*Chrysemys picta bellii*). *Biol. Chem. Hoppe-Seyler* 369, 123–131.
- Shaw, J.P., Marks, J., Shen, C.-K.J., 1987. Evidence that the recently discovered theta-1 globin gene is functional in higher primates. *Nature* 326, 717–720.
- Shishikura, F., 2002. The primary structure of hemoglobin D from the Aldabra giant tortoise, *Geochelone gigantea*. *Zool. Sci.* 19, 197–206.
- Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0b10. Sinauer, Sunderland, Massachusetts.
- Tatusova, T.A., Madden, T.L., 1999. Blast 2 sequences—a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* 174, 247–250.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Wheeler, D., Hope, R.M., Cooper, S.J.B., Dolman, G., Webb, G.C., Bottema, C.D., Gooley, A.A., Goodman, M., Holland, R.A.B., 2001. An orphaned mammalian β -globin gene of ancient evolutionary origin. *Proc. Natl. Acad. Sci. USA* 98, 1101–1106.
- Wheeler, D., Hope, R.M., Cooper, S.J.B., Gooley, A.A., Holland, R.A.B., 2004. Linkage of the β -like ω -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.* 58, 642–652.
- Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol.* 11, 367–372.
- Yang, Z., 1997. Phylogenetic Analysis by Maximum Likelihood (PAML) Version 1.3. University of California, Berkeley, California, USA.
- Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15, 568–573.