

Research article

Open Access

## Platypus globin genes and flanking loci suggest a new insertional model for beta-globin evolution in birds and mammals

Vidushi S Patel\*<sup>1</sup>, Steven JB Cooper<sup>2,3</sup>, Janine E Deakin<sup>1</sup>, Bob Fulton<sup>4</sup>, Tina Graves<sup>4</sup>, Wesley C Warren<sup>4</sup>, Richard K Wilson<sup>4</sup> and Jennifer AM Graves<sup>1</sup>

Address: <sup>1</sup>The ARC Centre for Kangaroo Genomics, Research School of Biological Sciences, The Australian National University, Canberra, ACT 2601, Australia, <sup>2</sup>Australian Centre for Evolutionary Biology and Biodiversity, The University of Adelaide, Adelaide, SA 5005, Australia, <sup>3</sup>Evolutionary Biology Unit, South Australian Museum, Adelaide, SA 5000, Australia and <sup>4</sup>Genome Sequencing Center, Washington University School of Medicine, St Louis, Missouri 63108, USA

Email: Vidushi S Patel\* - vidushi.patel@anu.edu.au; Steven JB Cooper - Cooper.Steve@saugov.sa.gov.au; Janine E Deakin - janine.deakin@anu.edu.au; Bob Fulton - bfulton@watson.wustl.edu; Tina Graves - tgraves@watson.wustl.edu; Wesley C Warren - wwarren@watson.wustl.edu; Richard K Wilson - rwilson@watson.wustl.edu; Jennifer AM Graves - jenny.graves@anu.edu.au

\* Corresponding author

Published: 25 July 2008

Received: 28 May 2008

BMC Biology 2008, 6:34 doi:10.1186/1741-7007-6-34

Accepted: 25 July 2008

This article is available from: <http://www.biomedcentral.com/1741-7007/6/34>

© 2008 Patel et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Vertebrate alpha ( $\alpha$ )- and beta ( $\beta$ )-globin gene families exemplify the way in which genomes evolve to produce functional complexity. From tandem duplication of a single globin locus, the  $\alpha$ - and  $\beta$ -globin clusters expanded, and then were separated onto different chromosomes. The previous finding of a fossil  $\beta$ -globin gene ( $\omega$ ) in the marsupial  $\alpha$ -cluster, however, suggested that duplication of the  $\alpha$ - $\beta$  cluster onto two chromosomes, followed by lineage-specific gene loss and duplication, produced paralogous  $\alpha$ - and  $\beta$ -globin clusters in birds and mammals. Here we analyse genomic data from an egg-laying monotreme mammal, the platypus (*Ornithorhynchus anatinus*), to explore haemoglobin evolution at the stem of the mammalian radiation.

**Results:** The platypus  $\alpha$ -globin cluster (chromosome 21) contains embryonic and adult  $\alpha$ -globin genes, a  $\beta$ -like  $\omega$ -globin gene, and the *GBY* globin gene with homology to cytoglobin, arranged as 5'- $\zeta$ - $\zeta$ '- $\alpha^D$ - $\alpha^3$ - $\alpha^2$ - $\alpha^1$ - $\omega$ -*GBY*-3'. The platypus  $\beta$ -globin cluster (chromosome 2) contains single embryonic and adult globin genes arranged as 5'- $\epsilon$ - $\beta$ -3'. Surprisingly, all of these globin genes were expressed in some adult tissues. Comparison of flanking sequences revealed that all jawed vertebrate  $\alpha$ -globin clusters are flanked by *MPG-C16orf35* and *LUC7L*, whereas all bird and mammal  $\beta$ -globin clusters are embedded in olfactory genes. Thus, the mammalian  $\alpha$ - and  $\beta$ -globin clusters are orthologous to the bird  $\alpha$ - and  $\beta$ -globin clusters respectively.

**Conclusion:** We propose that  $\alpha$ - and  $\beta$ -globin clusters evolved from an ancient *MPG-C16orf35*- $\alpha$ - $\beta$ -*GBY*-*LUC7L* arrangement 410 million years ago. A copy of the original  $\beta$  (represented by  $\omega$  in marsupials and monotremes) was inserted into an array of olfactory genes before the amniote radiation (>315 million years ago), then duplicated and diverged to form orthologous clusters of  $\beta$ -globin genes with different expression profiles in different lineages.

## Background

The evolution of the vertebrate globin superfamily has been extensively studied for many decades by comparing the structure and function of members of the gene families. These are principally haemoglobin, myoglobin, cytoglobin and neuroglobin and, more recently, globin X (in fish and amphibians [1]) and globin Y (specific to amphibians [2]).

Haemoglobin genes (alpha- and beta-globin) are of particular interest because of their critical role in oxygen transportation from the respiratory surfaces to the inner organs, and because of the dire effects of mutations in human globin genes that cause haemoglobinopathies [3]. The genes contained in the alpha ( $\alpha$ )- and beta ( $\beta$ )-globin clusters are expressed at different stages of development and in different tissues. Together, gene products from both clusters form the functional tetrameric haemoglobin molecules needed to fulfil oxygen requirements.

The evolutionary history of  $\alpha$ - and  $\beta$ -globin genes can be traced back to the common ancestors of fish, amphibians and amniotes (reptiles, birds and mammals), by comparing gene structure and composition of  $\alpha$ - and  $\beta$ -globin clusters across vertebrates. In the amphibians *Xenopus laevis* and *X. tropicalis*,  $\alpha$ - and  $\beta$ -globin genes are tightly juxtaposed as 5'- $\alpha$ - $\beta$ -3' [2,4-6]. In the Antarctic notothenioid fish (*Notothenia coriiceps*, *N. angustata*, *Trematomus hansoni*, *T. pennellii*), there is also a single 5'- $\alpha$ - $\beta$ -3' locus [7], although in pufferfish (*Fugu rubripes*) there are two globin clusters (one with  $\alpha$ -globin genes and the other with both  $\alpha$ - and  $\beta$ -globin genes), which are located on different chromosomes [8].

In amniotes,  $\alpha$ - and  $\beta$ -globin clusters are located on different chromosomes. It was proposed that the ancestral  $\alpha$ - and  $\beta$ -globin genes were located together in the common ancestor of amniotes, as they are in fish and amphibians, but became separated, either by chromosome fission or translocation between  $\alpha$ - and  $\beta$ -genes, or by chromosome/genome or *in trans* duplication and gene loss [5].

Further duplications then occurred in amniote lineages. The ancestral  $\alpha$ -globin gene is thought to have duplicated twice before the divergence of the bird-mammalian lineages, to produce progenitors of embryonic globin genes  $\pi/\zeta$ , and adult  $\alpha^D$  and  $\alpha^A$ , all of which are present in birds (for example, the chicken *Gallus gallus*) [9-11] and mammals [12,13]. The order and timing of these duplications is still debated, as is their origin: for instance,  $\alpha^D$  may have evolved by duplication either of adult  $\alpha^A$  (see [12]), or of an embryonic  $\alpha$ -like gene [14]. After the avian and mammalian lineages diverged, there were further tandem duplications of the  $\pi/\zeta$  and  $\alpha^A$  lineages to produce more complex marsupial and eutherian ('placental') mamma-

lian  $\alpha$ -globin clusters, 5'- $\zeta$ - $\psi\zeta$ '- $\alpha^D$ - $\psi\alpha^3$ - $\alpha^2$ - $\alpha^1$ - $\theta$ -3' (see [12,15-18]). The timing of these duplication events is also uncertain, because we do not know whether these seven  $\alpha$ -like globin genes all existed at the stem of the mammalian radiation.

As for many other gene families [19], comparisons of globin genes between distantly related mammals have provided unique insight into the evolution and function of the mammalian globins. Marsupials diverged from eutherian mammals about 148 million years ago (MYA), and mammalian Subclass Theria that contains these groups diverged from monotremes (Subclass Prototheria) about 166 MYA [20], so comparisons between these major mammal groups provide depth for evolutionary comparisons. Monotremes retain many anatomical and developmental features shared with birds and reptiles. Their small genome, too, and disjunct chromosome size classes are reminiscent of reptile genomes, and the 10 sex chromosomes in a karyotype of 52 chromosomes is unique among mammals [21-23]. Their importance for comparative studies is now increasingly recognised after the sequencing of the genome of a monotreme, *Ornithorhynchus anatinus* (platypus), to a depth of six to eight times by the Washington University Genome Centre, St Louis [24].

Indeed, studies of marsupial globins have clarified the timing of some of the duplications. The finding of single  $\epsilon$ - (embryonic) and  $\beta$ -globin (adult) genes together in the marsupial  $\beta$ -globin cluster indicated that a two-gene cluster ( $\epsilon$ - $\beta$ ) was present in the common therian ancestor [25-28]. Genes in the cluster were further duplicated to produce the ancestral eutherian  $\beta$ -globin cluster of 5'- $\epsilon$ - $\gamma$ - $\eta$ - $\delta$ - $\beta$ -3' (see [29-32]), which then underwent further tandem duplication events. In contrast, the bird (*G. gallus*)  $\beta$ -like globin genes ( $\epsilon$ - $\beta^H$ - $\beta^A$ - $\rho$ ) show very little homology to the mammalian  $\beta$ -like globin genes [33,34].

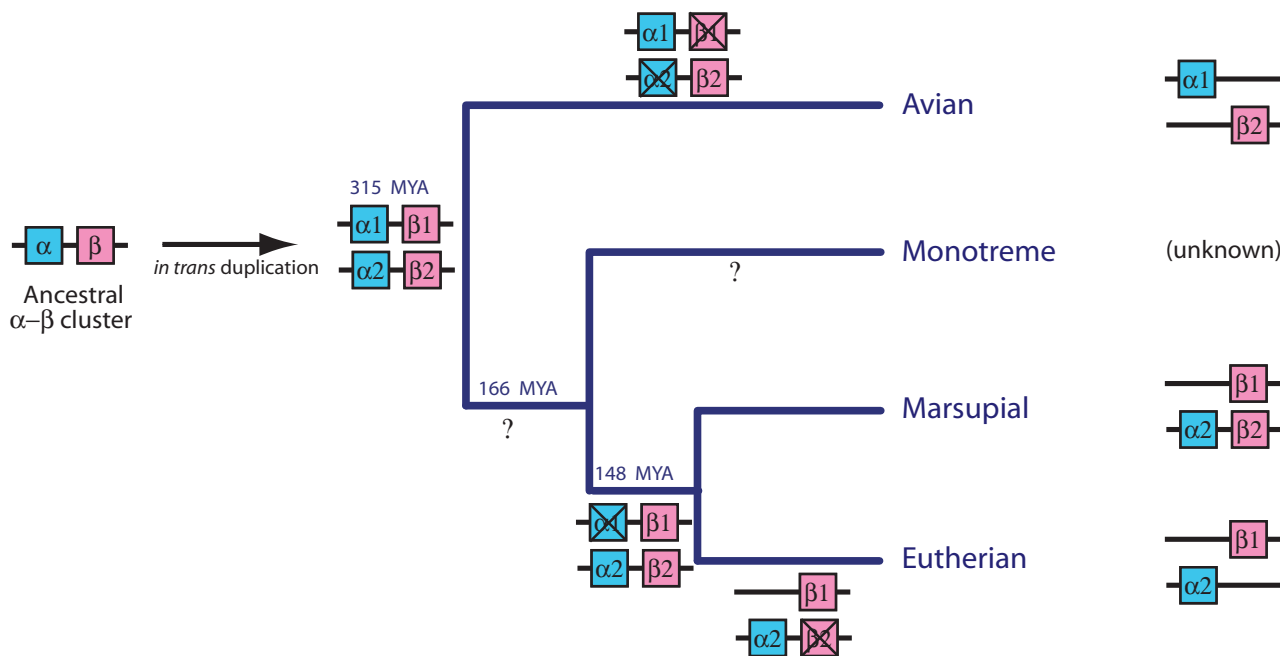
The discovery of a  $\beta$ -like globin gene ( $\omega$ -globin) adjacent (3') to the  $\alpha$ -globin cluster in marsupials led to a re-interpretation of globin evolution in birds and mammals [35,36]. Comparative sequence and phylogenetic analysis suggested that the  $\omega$ -globin gene was more closely related to bird  $\beta$ -like globin genes than to other mammalian  $\beta$ -like globin genes. The specific function of the  $\omega$ -globin gene is not yet known, but it is expressed just before birth and in the early stages of pouch young development [37]. In addition, the  $\omega$ -globin product binds to  $\alpha$ -like globin chains to form functional haemoglobin, so it is likely to be involved in oxygen transportation [35-37].

This finding of a remnant  $\beta$ -like globin gene ( $\omega$ -globin) beside the  $\alpha$ -globin cluster in marsupials [35,36] provided some support for the alternative hypothesis [5] that the  $\alpha$ -

and  $\beta$ -globin clusters in birds and mammals arose by *in trans* duplication of a chromosomal region, rather than simply by separation of the ancestral  $\alpha$ - $\beta$  globin cluster by chromosome fission or translocation. Wheeler et al. [35,36] proposed that before the divergence of birds and mammals (>315 MYA), the chromosome region bearing the ancestral  $\alpha$ - $\beta$  clusters duplicated to form two clusters ( $\alpha 1$ - $\beta 1$  and  $\alpha 2$ - $\beta 2$ ) on different chromosomes, and their contents diverged independently in mammals and birds by silencing of some genes within each cluster (Figure 1). To account for the apparent orthology of the marsupial  $\omega$ -globin gene and bird  $\beta$ -like globin genes, Wheeler et al. [35,36] suggested that the  $\alpha 1$  and  $\beta 2$  were silenced in the eutherian lineage, but  $\beta 2$  was retained in marsupials as the  $\omega$ -globin. In contrast,  $\alpha 2$  and  $\beta 1$  were silenced in the bird lineage (Figure 1). On this hypothesis, then, both the  $\alpha$  clusters and the  $\beta$  clusters of birds and mammals are paralogous (that is, evolved independently from ancient duplicates in an amniote ancestor) rather than orthologous (that is, diverged from the same ancestral cluster in an amniote ancestor).

This paralogy hypothesis (which rests on the rather weak orthology between the chicken  $\beta$  and marsupial  $\omega$ ), as well as the dates and types of other duplications, could be further tested by studying globin genes of monotreme mammals, and using comparative data to infer the ancestral globin gene arrangement of a mammal ancestor 166 MYA. The availability of platypus genomic sequences now provides an efficient way to discover all of the globin genes and regulatory signals, and to understand their function and evolution. Studies of globin genes in monotremes are also interesting because the specialized features and lifestyle of these unique mammals may have given rise to special adaptations of globin genes to fulfil unusual oxygen requirements. These features include the need for oxygen by diffusion through the egg membrane to the embryo after birth and the physiological response to hypoxic conditions during hibernation, burrowing and diving [38-40].

Little is known about monotreme  $\alpha$ - and  $\beta$ -globin families. More than 30 years ago, studies of adult blood revealed a single adult  $\alpha$  and  $\beta$  globin protein in the plat-



**Figure 1**  
**Current proposed model for the evolution of  $\alpha$ - and  $\beta$ -globin clusters from paralogous clusters in different lineages.** The unlinked  $\alpha$ - and  $\beta$ -globin clusters in birds and mammals evolved from an ancient *in trans* duplication of the ancestral linked  $\alpha$ - $\beta$  cluster, followed by differential gene silencing (marked with X). This resulted in bird  $\beta$ -like globin genes ( $\beta 2$ ) orthologous to the marsupial  $\omega$ -globin gene ( $\beta 2$  beside the  $\alpha$ -globin cluster) but paralogous to mammalian  $\beta$ -like globin genes ( $\beta 1$ ). Adapted from Wheeler et al. [36].

ypus [41,42] and echidna (*Tachyglossus aculeatus* [43,44]). Lee et al. [28] later isolated an adult  $\beta$ -globin gene in the echidna that encoded a polypeptide identical to the previously isolated echidna  $\beta$ -globin [44]. To date, there is no evidence of any monotreme embryonic  $\zeta$ - or  $\varepsilon$ -globin genes.

We used platypus genomic sequences from bacterial artificial chromosomes (BACs) to characterise the  $\alpha$ - and  $\beta$ -globin gene families of the platypus and investigate their molecular evolution. In particular, we searched for embryonic and  $\omega$ -globin genes and any novel globin genes that might fulfil the requirements for oxygen transport under hypoxic conditions. We investigated the genome context in order to infer the structure and origin of the ancestral  $\alpha$ - and  $\beta$ -globin clusters at the stem of the mammalian radiation. Our results strongly support the hypothesis that the mammalian  $\alpha$ - and  $\beta$ -globin clusters are orthologous to the avian  $\alpha$ - and  $\beta$ -globin clusters, respectively, and that the  $\beta$  cluster evolved by transposition of a copy of the beta-like  $\omega$ -globin gene in an amniote ancestor.

## Results

### Identification of BAC clones containing the $\alpha$ - and $\beta$ -globin clusters

The draft sequence assembly of platypus [24] is readily available on the University of California Santa Cruz (UCSC) Genome Browser [45]. However, currently the assembly is incomplete for the  $\alpha$ - and  $\beta$ -globin clusters, as individual globin genes appear on different contigs. There are also sequences of the platypus BAC clones available in NCBI GenBank that are not yet annotated and assembled, nor is part of the platypus genome assembly. Two of these are Oa\_Bb-2L7 [GenBank:AC195438] and Oa\_Bb-131M24 [AC203513], which were identified from the Encyclopaedia of DNA Elements Project to contain parts of the  $\alpha$ -globin cluster (see Methods). The BAC clone Oa\_Bb-484F22 [GenBank: AC192436] containing the  $\beta$ -globin cluster was obtained by screening a male platypus BAC library (Clemson University Genomic Institute, USA) and was subsequently fully sequenced and assembled by the Washington University Genome Sequencing Centre (St Louis, USA). These sequences were therefore used in this study to characterise the whole  $\alpha$ - and  $\beta$ -globin clusters in the platypus.

Genes in these sequenced BAC clones were predicted by programs GENSCAN [46] and GenomeScan [47]. Many genes were predicted, which were then used for BLAST searches of nucleotide (BlastN) and amino acid (BlastP) databases to help identify them (data not shown). Phylogenetic analyses were also conducted for the platypus  $\alpha$ - and  $\beta$ -like globin genes to further verify the identity of each gene (see below and also Figures 2, 3 and 4 below). With only one exception (platypus  $\varepsilon$ -globin, see below),

the identities of all of the genes inferred by BLAST analyses were supported by phylogenetic analyses with high posterior probabilities and bootstrap support values.

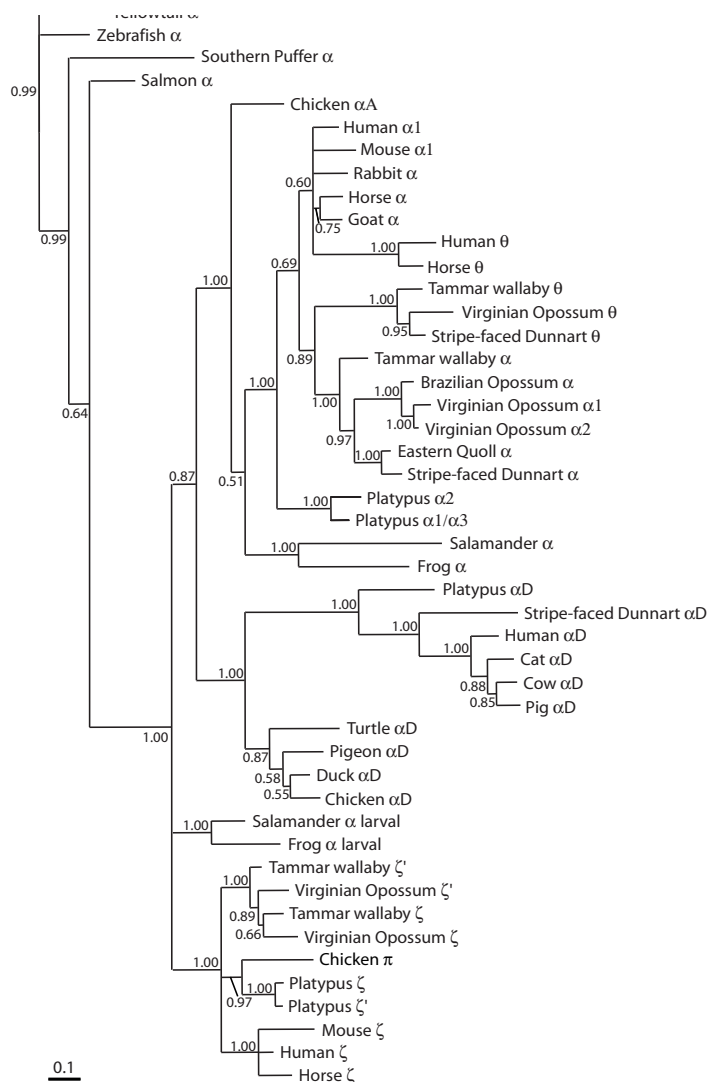
### Predictions and characterisation of genes in the platypus $\alpha$ -globin cluster

One BAC (Oa\_Bb-2L7) contained two embryonic  $\alpha$ -like globin genes, and a second BAC (Oa\_Bb-131M24) contained six  $\alpha$ -like globin genes and a  $\beta$ -like globin gene (see Additional file 1). These two BACs were found to overlap by 10,066 base pairs (bp), resulting in a contig of 330,126 bp that contained the entire platypus  $\alpha$ -globin cluster and flanking genes.

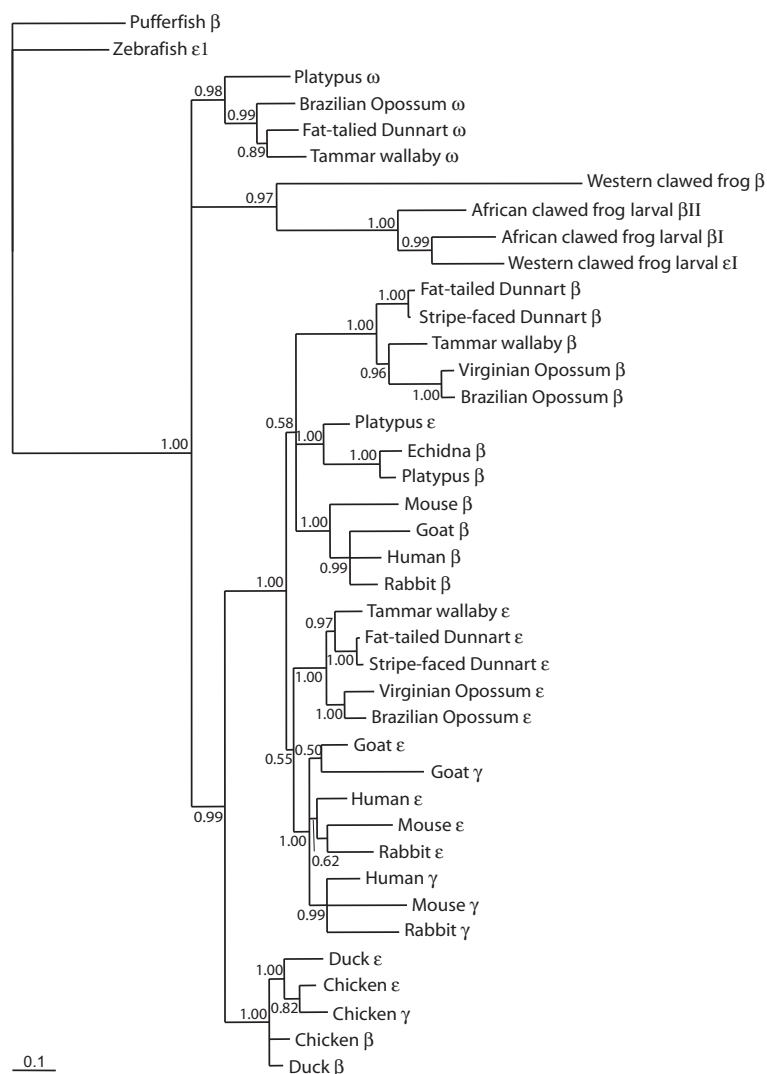
The 330,126 bp  $\alpha$ -globin contig was found to contain six  $\alpha$ -like globin genes, a  $\beta$ -like globin gene, and a gene that bore little similarity to  $\alpha$ - and  $\beta$ -like globin genes but some similarity to cytoglobins (Figure 5A). These six  $\alpha$ -like globin genes have a three-exon/two-intron structure and conserved donor/acceptor splice sites (GT/AG) typical of all vertebrate  $\alpha$ -like globin genes. They are separated from each other by 2 to 6 kilobase pairs (kb). Full details of the exon/intron lengths, location of the putative poly-A addition site (AATAAA) and the lengths of the coding domains with the predicted encoded polypeptide for each predicted gene are given in Table 1. Figure 5B shows the predictions for some of the well-characterised protein-binding sites in the 5' promoter region (about 200 bp 5' to the cap site of each gene). These include CACCC [48], CAAT [49], TATA [50], GATA 1 [51], EKLF (Erythroid Krüppel-like Factor; [52]) and have been experimentally shown to control the stage- and tissue-specific expression of  $\alpha$ - and  $\beta$ -like globin genes in other mammals [50,53-55].

Two genes at the 5' end of the  $\alpha$ -globin cluster were both identified as  $\zeta$ -like (referred to here as  $\zeta$  and  $\zeta'$ ) and predicted to encode polypeptides of 142 amino acids (aa), which are typical of known functional mammalian  $\alpha$ -like globin genes. The amino acid sequence alignment of  $\zeta$  and  $\zeta'$  shows 95% identity. In the promoter region of both genes, CACCC and CAAT consensus boxes are conserved at similar positions, and in comparable order to that of human  $\zeta$  and  $\zeta'$  (Figure 5B).

Adjoining the two  $\zeta$ -like globin genes, four other  $\alpha$ -like globin genes were identified. One was an orthologue of bird and reptilian  $\alpha^D$ , and the other three were orthologues of adult  $\alpha$  genes (here called  $\alpha^3$ ,  $\alpha^2$  and  $\alpha^1$ ). The long and uninterrupted open reading frame (ORF) of  $\alpha^D$  strongly suggests that it encodes a functional polypeptide of 141 aa, typical of known functional  $\alpha^D$  globin genes. The platypus  $\alpha^D$  globin gene contains introns of 1450 bp (intron 1) and 1610 bp (intron 2) that are very large com-

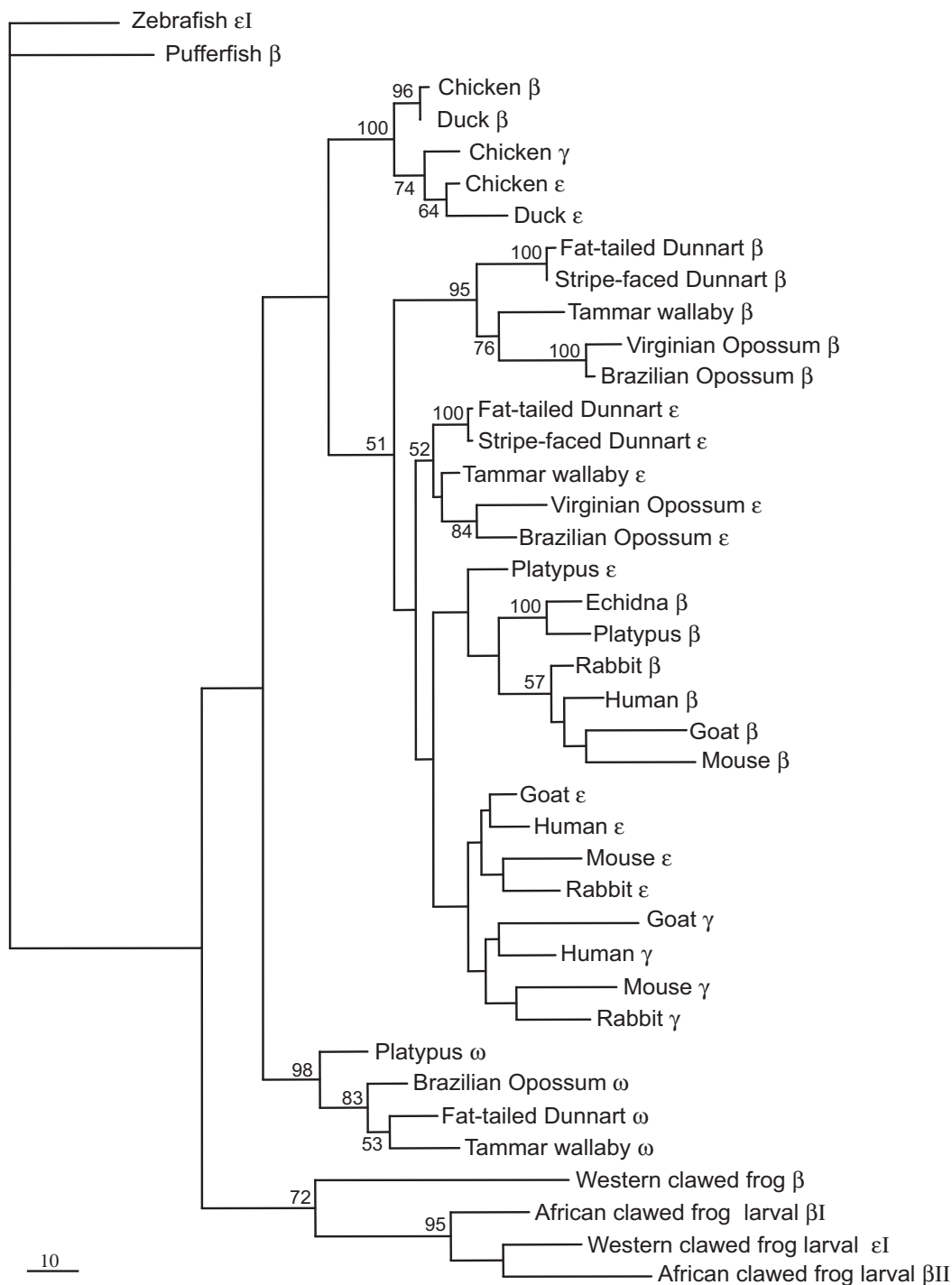


**Figure 2**  
**Evolutionary relationships among vertebrate  $\alpha$ -like globin genes using a 50% majority rule consensus phylogram from an analysis using Bayesian Inference.** The tree was constructed using mixed models of evolution for each codon position (see methods) and estimated base frequencies in an unlinked analysis using MrBayes (v. 3.1.2). Numbers adjacent to branches refer to % posterior probabilities. GenBank accession numbers for sequences are: Virginian Opossum (*Didelphis virginiana*)  $\zeta^1$ ,  $\zeta^2$ ,  $\alpha^1$ ,  $\alpha^2$ ,  $\theta$  [AC139599.2, AC148752.1]; Stripe-faced Dunnart (*Sminthopsis macroura*)  $\alpha^D$ ,  $\alpha^2$ ,  $\theta$  [AC146781]; Brazilian Opossum (*Monodelphis domestica*)  $\alpha$  [TI# 453585430]; Tammar wallaby (*Macropus eugenii*)  $\theta$  [AY459590],  $\alpha$  [AY459589];  $\zeta$  [AY789121],  $\zeta'$  [AY789122]; Horse (*Equus caballus*)  $\theta$  ( $\psi$   $\alpha$ ) [Y00284],  $\alpha^1$  [M17902],  $\zeta$  [X07051]; pig (*Sus scrofa*)  $\alpha^D$  [AC145444]; cat (*Felis catus*)  $\alpha^D$  [AC130194]; cow (*Bos taurus*)  $\alpha^D$  [AC150547]; Goat (*Capra hircus*)  $\alpha$  [J00043]; Human (*Homo sapiens*)  $\alpha^1$  [Y00491],  $\theta$  [X06482],  $\zeta$  [NM\_005332]; *mul*/ $\alpha^D$  chain [AY698022]; Mouse (*Mus musculus*)  $\alpha^1$  [NM\_008218],  $\zeta$  [X62302]; Rabbit (*Oryctolagus cuniculus*)  $\alpha$  [X04751]; Eastern Quoll (*Dasyurus viverrinus*)  $\alpha$  [M14567]; Chicken (*Gallus gallus*)  $\alpha^A$ ,  $\pi$ ,  $\alpha^D$  [AF098919]; Duck (*Cairina moschata*)  $\alpha^D$  [X01831]; Pigeon (*Columba livia*)  $\alpha^D$  [AB001981]; Turtle (*Geochelone nigra*)  $\alpha^D$  [SEG# ABI165195]; Zebrafish (*Danio rerio*)  $\alpha^1$  [NM\_131257]; Salamander (*Hynobius retardatus*) larval  $\alpha$  [AB034756]; Salamander (*Pleurodeles waltlii*)  $\alpha$  [M13365]; Frog (*Xenopus laevis*)  $\alpha^1$  [X14259], larval (tadpole)  $\alpha$  T5 [X02798]; Yellowtail (*Seriola quinqueradiata*)  $\alpha^A$  [AB034639]; Salmon (*Salmo salar*)  $\alpha$  [X97289]; Southern Puffer (*Sphoeroides nephelus*)  $\alpha^2$  [AY016023]; Platypus (*Ornithorhynchus anatinus*)  $\zeta$ ,  $\zeta'$ ,  $\alpha^D$ ,  $\alpha^3$ ,  $\alpha^2$ ,  $\alpha^1$  [AC203513].



**Figure 3**

**Evolutionary relationships among vertebrate β-like globin genes using a 50% majority rule consensus phylogram from an analysis using Bayesian inference.** The tree was constructed using mixed models of evolution for each codon position (see methods) and estimated base frequencies in an unlinked analysis using MrBayes (v. 3.1.2). Numbers adjacent to branches refer to % posterior probabilities. GenBank accession numbers for sequences are: Fat-tailed Dunnart (*Sminthopsis crassicaudata*) β [Z69592], ε [Z48632], ω [AY014770]; Stripe-faced Dunnart (*S. macroura*) β, ε [AC148754]; Virginian Opossum (*Didelphis virginiana*) β [J03643], ε [J03642]; Brazilian Opossum (*Monodelphis domestica*) β [XM\_001365299], ε [XM\_001364448], ω [XM\_001364828]; Tammar Wallaby (*Macropus eugenii*) β [AY450928], ε [AY450927], ω [AY014769]; African clawed frog (*Xenopus laevis*) larval β I [NM\_001086273], larval β II [NM\_001088028]; Western clawed frog (*X. tropicalis*) β [NM\_203528], larval ε I [NM\_001016495]; Chicken (*Gallus gallus*) β (β<sup>A</sup>) [NM\_205489], ε [NM\_001004390], γ (β<sup>A</sup>) [NM\_001031489]; Duck (*Cairina moschata*) β [J00926], ε [X15740]; Human (*Homo sapiens*) β [NM\_000518], γ [BC130459], ε [NM\_005330]; Mouse (*Mus musculus*) β (β I) [NM\_008220], γ (β h0) [NW\_001030869], ε (ε<sup>γ</sup>) [M26897]; Goat (*Capra hirus*) β (β<sup>A</sup>) [DQ350619], ε (ε I) [X01912], γ [M15388]; Rabbit (*Oryctolagus cuniculus*) β, γ, ε [M18818]; Echidna (*Tachyglossus aculeatus*) β [L23800]; Pufferfish (*Fugu rubripes*) β [AY170464]; Zebrafish (*Danio rerio*) ε I [NM\_001103130]; Platypus β, ε [AC192436], ω [AC203513].



**Figure 4**  
**Evolutionary relationships among vertebrate  $\beta$ -like globin genes analysed by maximum parsimony (MP) trees of length 926 (one of eight trees).** Third position in codons were excluded in the MP analyses, which were conducted using a heuristic search in PAUP\* v.4.0b10 [65]. The tree is rooted using pufferfish  $\beta$ -globin. Numbers adjacent to branches represent % bootstrap values (>50%) from MP heuristic analyses of 1000 pseudoreplicates. Accession numbers for sequences are given in the caption of Figure 3.

**Table 1: Gene-structure of the predicted platypus  $\alpha$ - and  $\beta$ -like globin genes and GBY**

Length/ Genes	Exon 1 (bp)	Intron 1 (bp)	Exon 2 (bp)	Intron 2 (bp)	Exon 3 (bp)	Position of Poly-A	CDS (bp)	Poly- peptide (aa)
$\zeta$	95	337	205	114	129	+119	429	142
$\zeta^1$	95	336	205	102	129	+133	429	142
$\alpha^D$	92	1450	205	1610	129	+77	426	141
$\alpha^1/\alpha^3$	92	405	205	151	129	+94	426	141
$\alpha^2$	95	720	205	155	129	+115	429	142
$\omega$	92	256	223	111	129	+69	444	147
<b>GBY</b>	98	3364	223	3053	144	+141	465	154
$\varepsilon$	92	143	223	474	129	+96	444	147
$\beta$	92	153	223	438	129	+71	444	147

For each predicted gene, the length of the exons and introns, position of poly-A addition site (AATAAA) from the stop codon, and the length of their putative coding domain (CDS) and encoded polypeptide are shown. All genes contained consensus splice sites (GT/AG) in both introns.

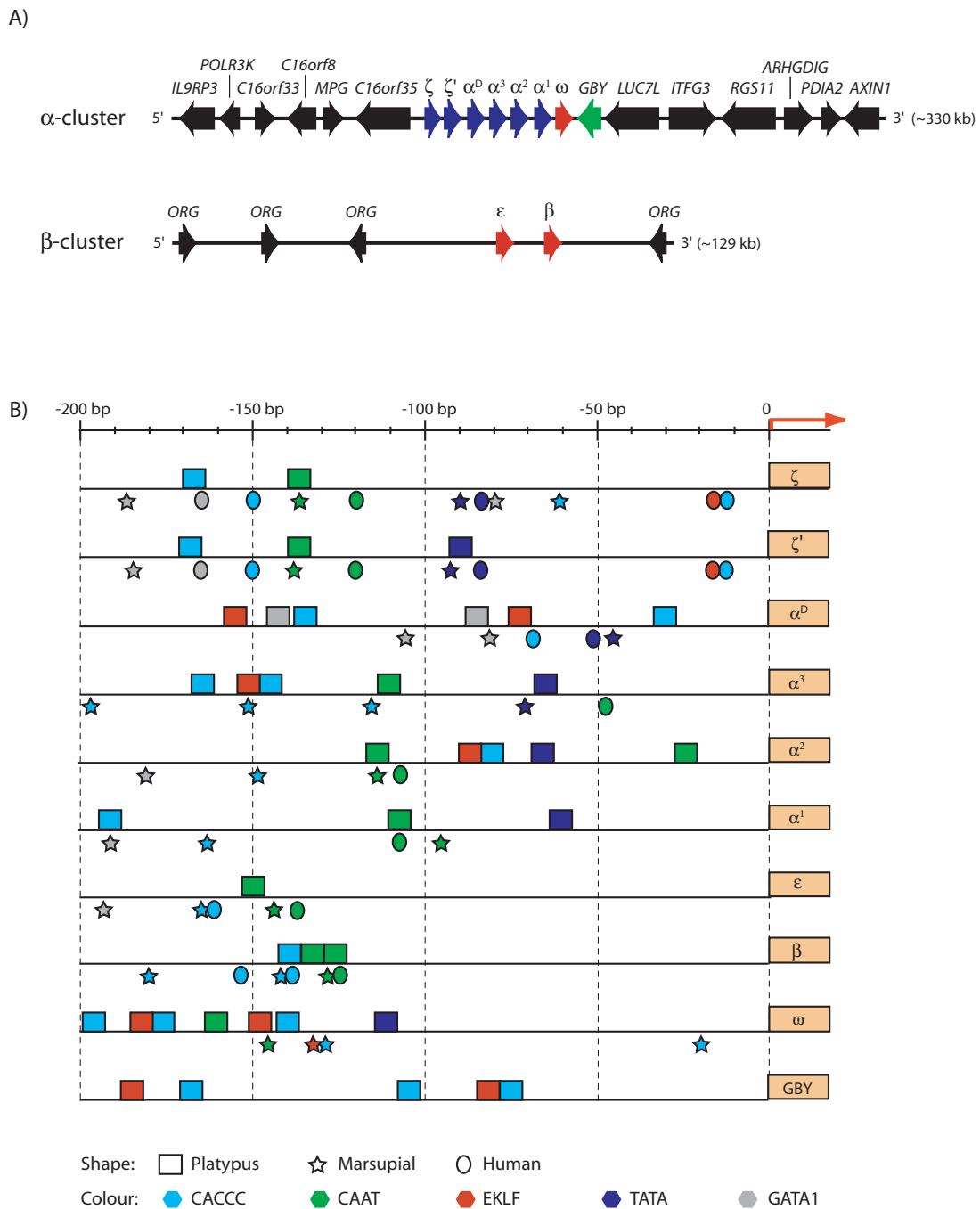
pared with those of other  $\alpha$ -like globins, which are usually less than 1000 bp.

Analyses of the platypus adult  $\alpha$ -like globin genes reveal three adult ( $\alpha^3$ ,  $\alpha^2$  and  $\alpha^1$ ) globin genes in the  $\alpha$ -globin cluster. The sequence of  $\alpha^3$  (the most 5' gene, adjacent to  $\alpha^D$ ) was found to be almost identical to  $\alpha^1$  (the most 3' gene) in their exon and intron regions, as well as in flanking regions of about 130 bp on both sides. The coding region was 100% identical, and just two sites in intron 1 were found to be different between the two genes. In order to confirm that identification of these two identical genes was not due to an error in the assembly of the original sequence data, the boundaries of the region containing the homology between  $\alpha^1$  and  $\alpha^3$  was further analysed by a BLAST search of the platypus whole-genome shotgun (WGS) database (data not shown). Two contigs were identified with homology to  $\alpha^1$  and  $\alpha^3$ ; these had identical sequences on one side of the boundary but different sequences on the other, confirming the presence of two separate genes. Further confirmation was obtained by performing a Southern blot on the  $\alpha$ -globin-containing BACs, digested with an enzyme (*EcoRV*) that does not cut within the  $\alpha^1$ ,  $\alpha^2$  and  $\alpha^3$  (data not shown). Probing with  $\alpha^1/\alpha^3$  revealed two bright bands, corresponding to  $\alpha^1$  and  $\alpha^3$ , and one fainter band between them, corresponding to  $\alpha^2$ . Probing with  $\alpha^2$  produced the same three bands, but in this case the middle one was brighter, corresponding to  $\alpha^2$ , and the outer bands were fainter, corresponding to  $\alpha^1$  and  $\alpha^3$ . These analyses confirmed the existence of separate genes  $\alpha^1$  and  $\alpha^3$  in the platypus  $\alpha$ -globin cluster. The  $\alpha^2$  gene, located between  $\alpha^1$  and  $\alpha^3$ , was distinct from both genes in the coding sequence (with 83% homology), in intron lengths (intron 1: 405 bp in  $\alpha^1/\alpha^3$  and 720 bp in  $\alpha^2$ ; intron 2: 151 bp in  $\alpha^1/\alpha^3$  and 155 bp in  $\alpha^2$ ) and in the promoter region (Figure 5B).

The amino acid sequence encoded by  $\alpha^1$  and  $\alpha^3$  was identical to the platypus adult  $\alpha$ -chain previously identified by Whittaker and Thompson [41], implying that at least one of these genes is expressed in the adult platypus. The coding domain of  $\alpha^1$  and  $\alpha^3$  is shorter (426 bp) than that of  $\alpha^2$  (429 bp), because it lacks the first three nucleotides of exon 1. The ORF of  $\alpha^2$  gives a strong indication that it is translated into a functional polypeptide of 142 aa, typical of known functional mammalian  $\alpha$ -like globin genes.

On the 3' side of the six  $\alpha$ -like globin genes, a  $\beta$ -like globin gene was predicted, which was identified as the orthologue of the marsupial  $\omega$ -globin gene. This platypus  $\omega$ -globin gene has a typical three-exon/two-intron structure, conserved donor/acceptor splice sites, and encodes a polypeptide of 146 aa, typical of all vertebrate  $\beta$ -like globin genes (Table 1). The promoter region located 5' of the  $\omega$ -globin initiation codon contains conserved sites for CAAT-EKLF-CACCC in an order identical to that of marsupial  $\omega$ -globin gene.

Unexpectedly, GenomeScan predicted a gene based on the protein similarities with the  $\alpha$ - and  $\beta$ -polypeptide chains, approximately 1.5 kb 3' of the  $\omega$ -globin gene. Like other  $\alpha$ - and  $\beta$ -globins, this gene also has a three-exon/two-intron structure and conserved donor/acceptor splice sites (Table 1). The lengths of its exons 1, 2, and 3 are 98, 223 and 144 bp, respectively, compared with 92, 223 and 129 bp in other  $\beta$ -like globin genes. However, it has much larger introns of 3364 bp (intron 1) and 3053 bp (intron 2). The long and uninterrupted ORF of this gene can be translated into a polypeptide of 154 aa, which is atypical of any known  $\alpha$ - or  $\beta$ -like globin genes. A BLAST search of the amino acid sequence of this gene obtained the best hit with Globin Y (*gby*) of the amphibian *X. laevis* (identity score of 39%), and weaker identity scores with Cytoglobins (*cygb*) of other species, such as the fish *Danio rerio* (27%), *X. tropicalis* (26%), chicken (28%) and human



**Figure 5**  
**Gene structure of the platypus  $\alpha$ - and  $\beta$ -globin clusters and flanking loci, and comparisons of their promoter regions with other mammals.** (A) The platypus  $\alpha$ -globin cluster contains six  $\alpha$ -like globin genes (red), a  $\beta$ -like ( $\omega$ ) globin gene (blue) and a distantly related globin gene, *GBY* (green), which are flanked by *IL9RP3-POLR3K-C16orf33-C16orf8-MPG-C16orf35* on the 5' end and *LUC7L-ITFG3-RGS11-ARHGDIG-PDIA2-AXIN1* on the 3' end (black). The platypus  $\beta$ -globin cluster contains only two genes,  $\epsilon$  and  $\beta$  (blue), which are flanked on both sides by *ORG* genes (black). (B) Relative positions of the putative transcription factor binding sites in the 200 bp promoter region located upstream of the predicted platypus, marsupial (*Didelphis virginiana*  $\zeta$  and  $\psi\zeta'$ , and *Sminthopsis macroura*  $\alpha^D$ ,  $\psi\alpha^3$ ,  $\alpha^2$ ,  $\alpha^1$ ,  $\omega$ ,  $\epsilon$  and  $\beta$ ) and human  $\alpha$ - and  $\beta$ -like globin genes. For the platypus *GBY* no data was available from other species, including *Xenopus tropicalis*, for comparisons.

(25%) at the protein level. We designated this gene 'GBY' based on similarities with *X. laevis gby*, and its similar position adjoining the globin cluster [2]. The predicted polypeptide of platypus GBY (154 aa) was shorter than *X. laevis gby* (156 aa), and quite different from *X. laevis cygb* (179 aa), *D. rerio cygb1* (174 aa) and *cygb2* (179 aa), and human CYGB (190 aa). Using the Expressed Sequence Tag (EST) database, a BLAST search of the platypus GBY also obtained an identity score of 38% with *X. tropicalis gby* that was expressed in both tadpoles and adults, but produced no significant matches with any other mammalian genes. The present work was the first opportunity to analyse the promoter region of any GBY gene (Figure 5B).

#### **Predictions and characterisation of genes in the platypus $\beta$ -globin cluster**

In the platypus, only two  $\beta$ -like globin genes were predicted within the 129,521 bp BAC clone (Oa\_Bb-484F22) by GENSCAN and GenomeScan (see Additional file 1). When the predicted amino acid sequences were subjected to BLAST search, the 5' gene had best hits with mammalian embryonic  $\epsilon$ -globin genes. Although the phylogenetic analyses using Bayesian inference (BI; see below) indicated that this gene was more closely related to the platypus and echidna adult  $\beta$ -globin genes than to therian  $\epsilon$ -globin genes, the position of this gene on the 5' end of the  $\beta$ -globin cluster and expression data (see below) supports its orthology with mammalian embryonic  $\epsilon$ -globin genes, and is henceforth referred to as  $\epsilon$ . The 3' gene encoded a protein identical to the previously identified platypus adult  $\beta$ -chain [42], and is henceforth referred to as  $\beta$ .

Both genes encode polypeptides of 146 aa, typical of known functional mammalian  $\beta$ -like globin genes. The promoter region of the platypus  $\beta$  has conserved sites of CACCC and CAAT in all three extant of mammals. However, the promoter region of the platypus  $\epsilon$  appears to be quite different from other mammalian  $\epsilon$ -globin genes and even from the platypus  $\beta$  (Figure 5B). The promoter of platypus  $\epsilon$  contains only one predicted motif (CAAT), whereas the promoters of other mammalian  $\epsilon$ ,  $\beta$  and the platypus  $\beta$  contain many predicted motifs.

#### **Expression studies of the platypus $\alpha$ - and $\beta$ -like globin genes**

Transcription studies were performed to gain insight into the expression and function of all of the predicted platypus globin genes. Adult liver, kidney, spleen, testis, lung and brain were obtained for this project: no embryonic samples were available (or are ever likely to be available) for this vulnerable and iconic species. Observation of the expression of any of the predicted genes would constitute a good indication that the gene is transcriptionally active and functional.

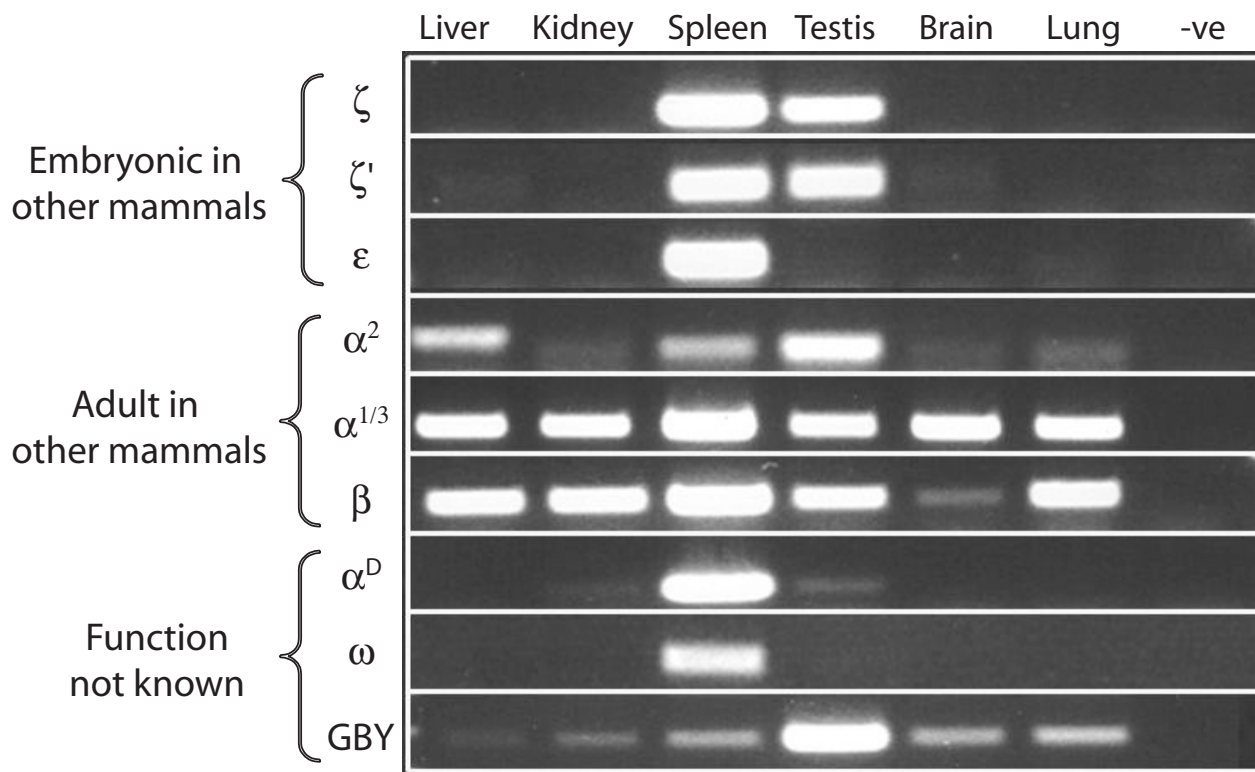
Reverse-transcriptase polymerase chain reaction (RT-PCR) of all predicted platypus genes showed that they are all expressed in at least some of these adult platypus tissues (Figure 6). Platypus genes  $\alpha^1/\alpha^3$ ,  $\alpha^2$  and  $\beta$ , whose orthologues are usually expressed in the bone marrow of an adult human, were expressed in almost all platypus tissues tested, suggesting a broader expression of these genes in the monotreme lineage. Surprisingly, the genes  $\zeta$ ,  $\zeta'$  and  $\epsilon$ , whose therian orthologues are expressed only at embryonic stages of development, were expressed in adult spleen and testis, but not in the other tissues of adult platypus. This suggests that persistent expression of these genes in some adult tissues was selected for in the platypus, perhaps in response to its aquatic lifestyle and the hypoxic conditions of a confined burrow. Also, the expression pattern of platypus  $\epsilon$  is similar to embryonic  $\alpha$ -like  $\zeta$  and  $\zeta'$  but different from that of adult globin genes ( $\alpha^1/\alpha^3$ ,  $\alpha^2$  and  $\beta$ ). The  $\omega$  and  $\alpha^D$  globin genes, whose functions are unknown, were also expressed mainly in the spleen. GBY was expressed in all adult platypus tissues, most strongly in testis.

#### **Phylogenetic analyses**

Phylogenetic analyses of the  $\alpha$ -like globin genes using BI and maximum parsimony (MP) produced several noteworthy results. The platypus adult  $\alpha$  globin genes ( $\alpha^1/\alpha^3$  and  $\alpha^2$ ) grouped closely together to the exclusion of eutherian and marsupial  $\alpha$ - and  $\theta$ -globin genes for all analyses, although posterior probability (69%) and bootstrap support (66%) for this arrangement were relatively weak (Figure 2). This finding suggests that the duplication leading to the marsupial and eutherian  $\theta$ -globin lineage occurred after the divergence of the monotreme and therian lineages. This is consistent with the absence of a  $\theta$ -globin gene from the region between platypus  $\alpha^1$ - and  $\omega$ -globin, its expected location based on its position in marsupial  $\alpha$ -globin clusters [12,56].

Both platypus  $\zeta$ -globin genes grouped closely together and formed a sister group relationship with chicken  $\pi$ , supported by a high posterior probability of 97% (Figure 2). A sister group relationship was also found in MP trees for analyses of the entire platypus coding region (bootstrap support <50%), and when third positions in the codon were excluded, was supported by 73% bootstrap pseudoreplicates (data not shown). This differs from the expectation that platypus  $\zeta$ -globin genes would group with other mammalian  $\zeta$ -globin genes to the exclusion of chicken  $\pi$ , suggesting that other factors (for example, purifying selection) operated to maintain a similar sequence in birds and monotremes.

There is still considerable uncertainty in the phylogenetic position of the  $\alpha^D$ -globin clade. It has recently been proposed that the  $\alpha^D$  globin lineage resulted from duplica-



**Figure 6**

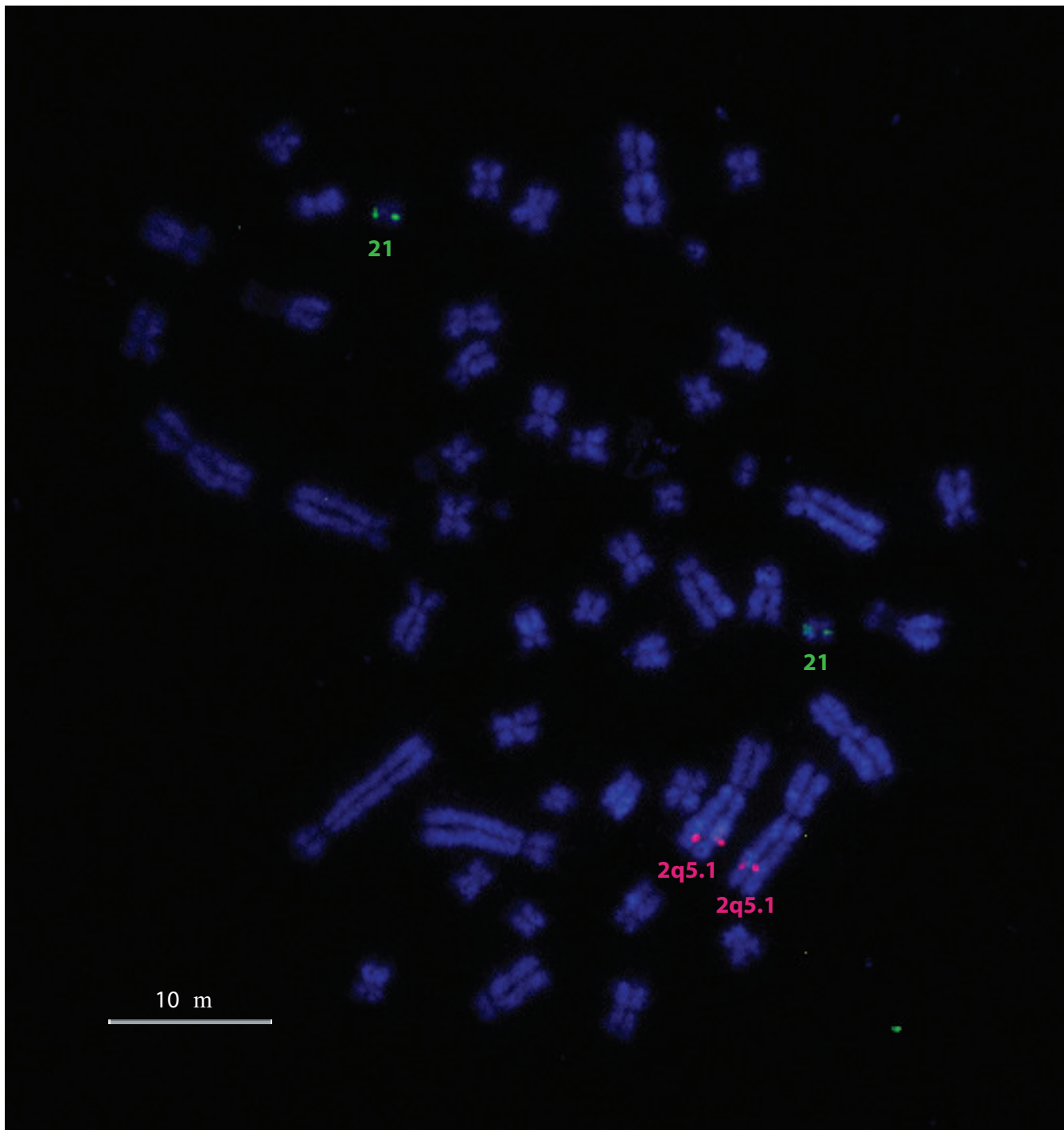
**Expression of all predicted  $\alpha$ - and  $\beta$ -like globin genes including GBY in an adult platypus.** For each of the platypus predicted genes, expression was investigated by reverse transcriptase polymerase chain reaction in adult liver, kidney, spleen, testis, brain and lung. Primers for each gene were designed between two exons so that it would result in a product distinguishable from genomic contamination of cDNA. The negative control (last lane) contained no cDNA. All genes were expressed in one or more tissues, indicating that they are transcriptionally active and might be functional.

tion of the embryonic  $\alpha$ -globin lineage, with phylogenetic analyses supporting a sister lineage relationship of these lineages to the exclusion of the adult  $\alpha$ -globin lineage [14]. However, this arrangement was not supported in BI analyses of the data set used here, and the position of the  $\alpha^D$  lineage was different in the different analyses. Analyses using BI (Figure 2) supported the sister lineage relationship of the  $\alpha^D$  and adult  $\alpha$ -globin lineages (as proposed by Cooper et al. [12]), with 87% posterior probability support. In contrast, all MP analyses supported the sister lineage status of  $\alpha^D$  and embryonic  $\alpha$ -globin genes, indicating an uncertainty in the phylogenetic position of the  $\alpha^D$ -globin clade.

Phylogenetic analyses of the  $\beta$ -globin genes provided results similar to recently reported phylogenetic analyses [35,36], with one notable exception. The BI analyses of coding sequence data (Figure 3) provided strong support (99% posterior probability) for the sister relationship of

bird and mammalian  $\beta$ -like globin genes, contradicting previously published phylogenies of mammalian  $\beta$ -globin genes showing a sister relationship of marsupial  $\omega$ -globin and bird  $\beta$ -like globin genes [35,36]. MP analyses (Figure 4), excluding third position in the codon, gave a similar tree arrangement, albeit with very low bootstrap support (<50%). In marked contrast to the BI analyses of DNA sequence data, BI protein analyses (data not shown) supported the sister relationship of bird  $\beta$ -like globin and mammal  $\omega$ -globin lineages with a high posterior probability (99%).

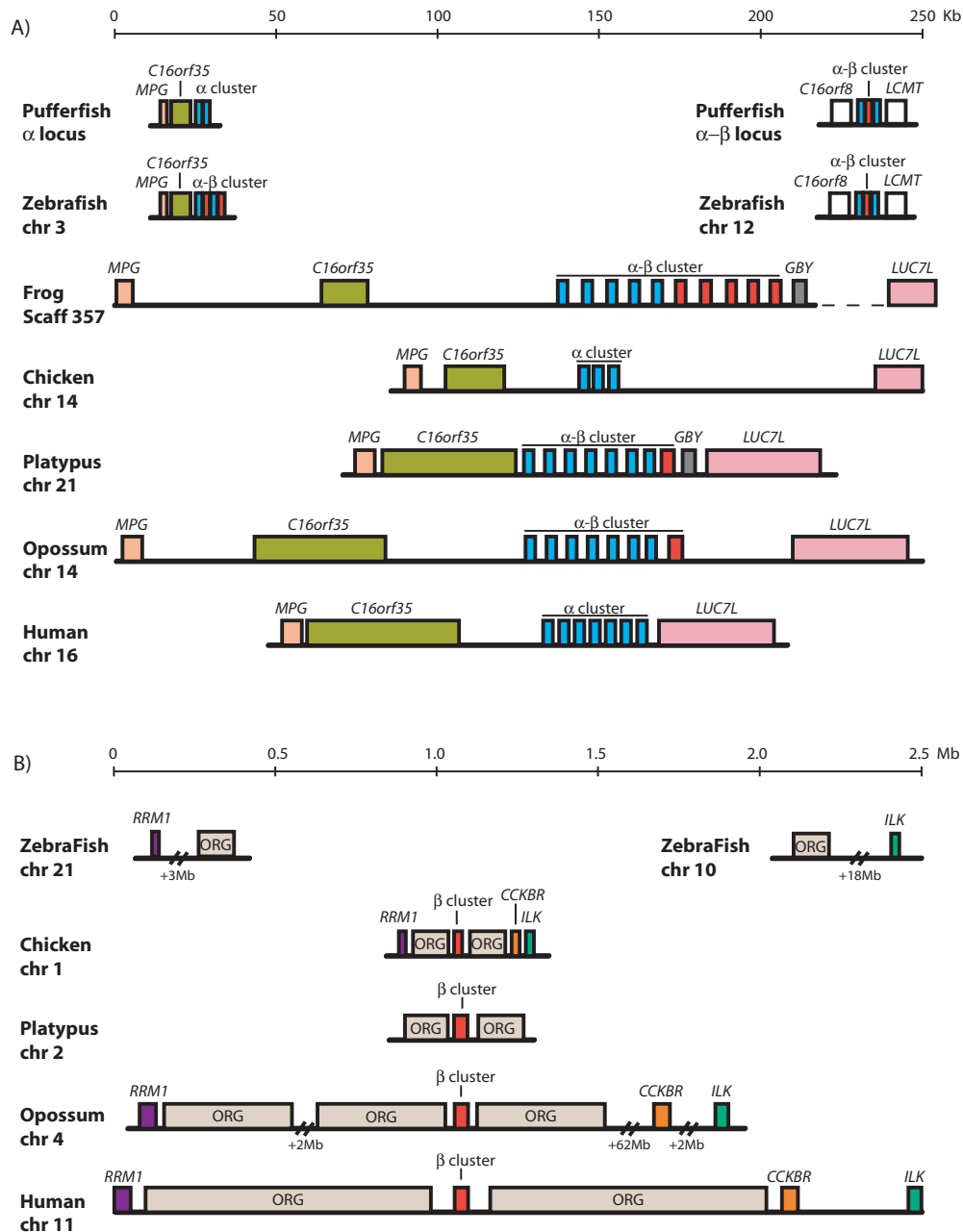
Lastly, phylogenetic analyses using BI indicated that the platypus  $\epsilon$  gene was more closely related to the platypus and echidna adult  $\beta$ -globin genes than to therian  $\epsilon$ -globin genes, suggesting it may not be orthologous to marsupial and eutherian  $\epsilon$ -globin (Figure 3). BI analyses of  $\beta$ -globin protein data and MP analyses of the coding sequence data, with third codon positions excluded, grouped the gene as

**Figure 7**

**Chromosomal location of the platypus  $\alpha$ - and  $\beta$ -globin clusters.** Two-colour fluorescence in situ hybridisation showing the location of the  $\alpha$ -globin cluster on chromosome 21 (green) and the  $\beta$ -globin cluster on chromosome 2q5.1 (red). The chromosomes are counterstained with DAPI (blue).

an ancestral lineage to eutherian and monotreme adult  $\beta$ -globin genes (see Figure 4). This ancestral position suggests that the lineage evolved following duplication of an

ancestral  $\beta$ -globin gene prior to the divergence of monotremes and therians.



**Figure 8**

**Loci flanking vertebrate α-globin (A) and β-globin (B) clusters.** The relative locations of flanking loci (A) *MPG*, *C16orf35*, *LUC7L* and *GBY* and (B) *RRM1*, *CCKBR*, *ILK* and *ORG* genes were searched for beside the α-β globin cluster in zebrafish (*Danio rerio*) and frog (*Xenopus tropicalis*), and beside the separate α-globin and β-globin clusters in chicken (*Gallus galus*), opossum (*Monodelphis domestica*) and human (*Homo sapiens*) from Ensembl [57]. The pufferfish (*Fugu rubripes*) flanking loci shown here were adapted from Gillemans et al. [8]. For the platypus, the α-globin flanking loci were characterised in this study, and *ORG* genes surrounding the platypus β-globin cluster were discovered: however, the BAC clone (484F22) was too small to cover the region containing the loci *RRM1*, *CCKBR* and *ILK*. In *X. tropicalis* *LUC7L* was found on another scaffold (466 from Ensembl) but sequence analyses by Fuchs et al. [2] suggested that *LUC7L* resides 3' to the frog α-β-*GBY* cluster. The flanking loci as well as the α- and β-globin clusters are differentiated by colour.

### Location of the $\alpha$ - and $\beta$ -globin clusters in the platypus

The location of the verified BAC clones containing the  $\alpha$ - (Oa\_Bb-2L7) and  $\beta$ -globin (Oa\_Bb-484F22) clusters in the platypus was determined by fluorescence *in situ* hybridisation (FISH) (Figure 7). The  $\beta$ -globin cluster localised to one of the largest autosomes, giving unambiguous signals on the long arm of chromosome 2 (2q5.1). The  $\alpha$ -globin cluster localised to the smallest autosome, 21, whose two arms are not distinguishable by size or DAPI banding pattern [21]. This is the first gene that has been localised on the platypus chromosome 21.

### Loci flanking the $\alpha$ - and $\beta$ -globin clusters in the platypus and other vertebrates

To explore the genome context of the  $\alpha$ - and  $\beta$ -globin clusters in the platypus and other vertebrates, the platypus BAC sequences and the genomes of other sequenced species were searched for loci residing beside the  $\alpha$ - and  $\beta$ -globin clusters.

As well as globin genes, GENSCAN predicted within the platypus  $\alpha$ -globin 330,126 bp contig many genes that flank the platypus  $\alpha$ -globin cluster (Figure 5A), which were identified by BLAST analyses. These include *IL9RP3-POLR3K-C16orf33-C16orf8-MPG-C16orf35* upstream (5') of the  $\alpha$ -globin cluster, and, *LUC7L-ITFG3-RGS11-ARH-GDIG-PDIA2-AXIN1* downstream (3') of the  $\alpha$ -globin cluster (Figure 5A).

To compare the  $\alpha$ -globin flanking loci of the platypus and other vertebrates, the genes closest to the  $\alpha$ -globin cluster, *MPG*, *C16orf35* and *LUC7L* were searched for in the human, opossum (*Monodelphis domestica*), chicken, frog (*X. tropicalis*) and zebrafish (*D. rerio*) genomes that were accessible from Ensembl [57]. Figure 8A shows that the locations of *MPG*, *C16orf35* and *LUC7L* are conserved adjacent to the  $\alpha$ -globin cluster of birds and mammals, and in the same position adjacent to the  $\alpha$ - $\beta$  cluster of amphibians, and all but *LUC7L* were also present in fish. These results are consistent with the previous analyses of Flint et al. [58] and Hughes et al. [59]. Thus the flanking loci analyses reveal that the genome context of the platypus  $\alpha$ -globin cluster is the same as the  $\alpha$ -globin clusters in therian mammals and birds, and this is the same as for the  $\alpha$ - $\beta$  cluster of fish and frogs.

GENSCAN also predicted numerous genes other than globin genes in the platypus  $\beta$ -globin BAC (484F22). These were identified by a BLAST search as members of the olfactory receptor gene (*ORG*) family that are responsible for odour detection. Three conserved *ORG* members were identified at the 5' end of the platypus  $\beta$ -globin cluster and one conserved *ORG* member at the 3' end (Figure 5A).

To compare  $\beta$ -globin flanking loci, *ORG* genes, as well as other genes that are closest to the  $\beta$ -globin cluster in other species, *RRM1*, *CCKBR* and *ILK* were searched for in the human, opossum, chicken and zebrafish genomes that were accessible from Ensembl [57]. Data from frog (*X. tropicalis*) was not useful since all of these loci lie on different contigs or scaffolds due to assembly problems. The locations of multiple *ORG* genes, *RRM1*, *CCKBR* and *ILK* were found to be conserved adjacent to  $\beta$ -globin cluster of birds and mammals [60,61], but not for the  $\alpha$ - $\beta$  cluster of fish and frogs, nor beside the second  $\alpha$ - $\beta$  cluster of zebrafish and pufferfish (Figure 8B). Thus the genome context of the platypus  $\beta$ -globin cluster is the same as in therian mammals and birds, but this is different from the  $\alpha$ - $\beta$  cluster of fish and frogs.

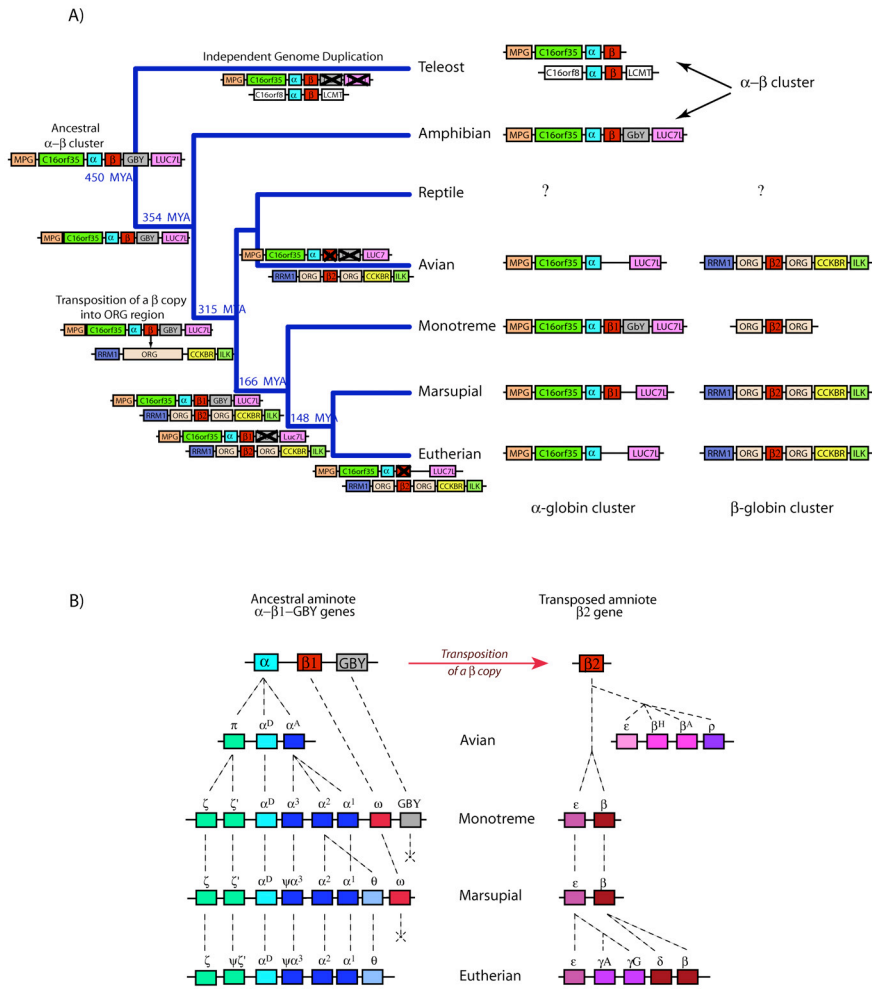
### Discussion

The phylogenetic position of monotremes makes comparisons with platypus of special value for exploring the organization, function and evolution of mammalian genes and genomes. The availability of platypus genome sequence data now makes many such studies possible, and have been used here to characterise the platypus  $\alpha$ - and  $\beta$ -globin gene clusters and explore their evolutionary history.

### The platypus $\alpha$ -globin gene cluster

The platypus  $\alpha$ -globin cluster contains at least eight genes within more than 40 kb, including six  $\alpha$ -like globin genes (including the identical  $\alpha^1$  and  $\alpha^3$ ), one  $\beta$ -like globin gene ( $\omega$ -globin) and a gene belonging to another member of the globin super-family (*GBY*) arranged in the order 5'- $\zeta$ - $\zeta^1$ - $\alpha^D$ - $\alpha^3$ - $\alpha^2$ - $\alpha^1$ - $\omega$ -*GBY*-3' (Figure 5A). The cluster maps to chromosome 21, the smallest autosome in platypus. All eight genes are likely to be functional since their expression was detected in tissues of an adult platypus.

The platypus  $\alpha$ -globin cluster is almost identical to the arrangement of  $\alpha$ -like globin genes in the ancestral therian cluster reported by Cooper et al. [12]. The one exception is the absence of a  $\theta$ -globin gene from the platypus cluster. Phylogenetic analyses support the basal position of the monotreme adult  $\alpha$ -globin lineage relative to marsupial and eutherian  $\alpha$ - and  $\theta$ -globin lineages, implying that the duplication of an adult  $\alpha$ -globin to produce  $\theta$ -globin occurred in the therian lineage after its divergence from the monotreme lineage (Figure 9B). However, although the numbers and arrangements of genes is so similar in platypus and therians, the presence of three adult  $\alpha$ -globin genes and two embryonic  $\zeta$ -globin genes in their common ancestor was not supported by phylogenetic analyses, which showed independent groupings of the three adult and embryonic genes within each separate mammalian lineage (Figure 2 and see Cooper et al. [12]). This result can be interpreted literally as resulting from



**Figure 9**

**Proposed model for the evolution of the  $\alpha$ - and  $\beta$ -globin clusters in vertebrate lineages.** (A) A region containing *MPG-C16orf35- $\alpha$ - $\beta$ -GBY-LUC7L* represented the ancient  $\alpha$ - $\beta$  globin cluster of jawed vertebrates (>450 MYA), which is seen in the amphibian lineage. This region further duplicated and underwent some gene silencing in teleost fish. In an amniote ancestor of reptiles, birds and mammals (>315 MYA), a copy of an ancestral  $\beta$ -globin gene from this region was inserted into a different chromosome within a region replete with multiple copies of *ORG* genes. The original amniote  $\beta$ -globin gene survives as the  $\omega$ -globin gene ( $\beta 1$ ) in the  $\alpha$ -globin cluster of marsupials and monotremes, whereas the transposed  $\beta$ -globin gene ( $\beta 2$ ) duplicated several times to form different clusters in the different lineages. (B) Tandem duplications of the ancestral amniote  $\alpha$ -globin gene produced a three-gene ( $\pi$ - $\alpha^D$ - $\alpha^A$ ) cluster in the avian lineage. In the mammalian lineage, further duplications gave rise to a six-gene ( $\zeta$ - $\zeta'$ - $\alpha^D$ - $\alpha^3$ - $\alpha^2$ - $\alpha^1$ ) cluster with ongoing gene conversion events homogenising the embryonic and adult genes. In monotremes, the ancestral  $\omega$  ( $\beta 1$ ) and *GBY* are retained. After the divergence of monotreme and tharian mammals, there was an additional duplication of  $\alpha^2$  to form  $\theta$ , giving rise to the seven-gene cluster ( $\zeta$ - $\zeta'$ - $\alpha^D$ - $\alpha^3$ - $\alpha^2$ - $\alpha^1$ - $\theta$ ) in marsupials and eutherians. Marsupials also retain the ancestral  $\omega$  but may have lost *GBY* gene; eutherians retain no identifiable remnant of either gene. Furthermore, the ancestral transposed  $\beta 2$ -globin gene duplicated independently in birds and mammals. Before the mammalian radiation, we propose that the ancestral  $\beta 2$  gene duplicated to form a two-gene  $\beta$ -globin cluster ( $\epsilon$ - $\beta$ ) as seen in monotremes and marsupials, except that ongoing gene conversion events homogenised platypus  $\epsilon$  to group with monotreme  $\beta$  genes. After the divergence of marsupial and eutherian mammals, there were further tandem duplications of these two genes to produce complex  $\beta$ -globin cluster ( $\epsilon$ - $\gamma$ - $\eta$ - $\delta$ - $\beta$ ) in eutherians.

independent duplications in each mammalian lineage to produce three adult and two embryonic genes in each. However, this seems unlikely to explain the convergence in gene number of the  $\alpha$ -globin cluster in these distantly related mammalian lineages. We suggest that a more parsimonious explanation is that the common ancestor of monotremes and therians contained three adult  $\alpha$ -globin genes and two  $\zeta$ -globin genes, which were homogenised by ongoing gene conversion events, leading to the gene tree that does not match the duplication history of the individual genes. The close similarity of the platypus  $\alpha^3$  and  $\alpha^1$  loci suggests a very recent gene conversion event that homogenised their sequences. Therefore, we propose that the platypus  $\alpha$ -globin cluster of eight genes ( $\zeta$ - $\zeta'$ - $\alpha^D$ - $\alpha^3$ - $\alpha^2$ - $\alpha^1$ - $\omega$ -*GBY*) represents the ancestral mammalian  $\alpha$ -globin cluster arrangement (Figure 9B), in which all genes were transcriptionally active.

Importantly, the platypus  $\alpha$ -globin cluster contains a copy of the  $\beta$ -like  $\omega$ -globin gene, also found in the marsupial  $\alpha$ -globin cluster, but absent in humans, supporting the hypothesis that  $\omega$ -globin was present in the common ancestor of all mammals. Phylogenetic analyses also confirm the ancient ancestry of the  $\omega$ -globin gene, as concluded by Wheeler et al. [35,36]. Among adult platypus tissues this gene was expressed only in the spleen. In marsupials, expression of the  $\omega$ -globin gene was detected just prior to birth and during early pouch young development [37], although the site of expression was not studied, and there was no evidence of adult expression in blood cells.

#### **Discovery of a mammalian *GBY* globin gene adjoining the $\alpha$ -globin cluster**

We discovered a globin gene *GBY* in the platypus that is adjacent (3') to  $\omega$  in the  $\alpha$ -globin cluster. It has a typical three-exon/two-intron structure like other  $\alpha/\beta$ -globin genes, contains an ORF encoding a polypeptide chain of 154 aa, and is expressed in almost all adult tissues, most strongly in testis. The amino acid sequence is unrelated to any of the other globin genes in the cluster, so it is unlikely to be derived by duplication of  $\alpha$ - or  $\omega$ -globin within the monotreme lineage. Rather, it shows sequence similarity to *gby* of *X. tropicalis* and *X. laevis*, a gene thought to be related to cytoglobins [2].

Little is known of the function of amphibian *gby*, or its relationship with other globins. Fuchs et al. [2] reported that amphibian *gby* encodes a *bona fide* globin of 156 aa, having all of the sequence features of a functional respiratory protein. *gby* was expressed in all adult tissues tested in *X. laevis*, most strongly in ovary, kidney and eye, and was present in 20 expressed sequence tag clones from different stages of *X. laevis* and *X. tropicalis* embryonic and adult development [2], suggesting that it is expressed in embryonic as well as adult stages. Phylogenetic analysis of all

vertebrate globins [2] showed that the *gby* lineage diverged at the base of two separate clades, one comprising all vertebrate cytoglobins, myoglobins, agnathan globins and bird globin E, and the other comprising the haemoglobin  $\alpha$ - and  $\beta$ -chains.

The position of platypus *GBY* adjacent to the  $\alpha$ -globin cluster and flanked by *LUC7L* mirrors its position in *X. tropicalis* between the main  $\alpha$ - $\beta$  cluster and *LUC7L* [2]. Another common feature of both was strong expression in gonads (ovary in *X. laevis* [2] and testis in platypus), so *GBY* has sex-related expression in both lineages. Thus *GBY* is not specific to amphibians, as was thought, but was a component of the cluster in an ancient tetrapod, and has been lost, or has diverged beyond recognition, in birds and therian mammals.

#### **The platypus $\beta$ -globin gene cluster**

Characterisation of the platypus  $\beta$ -globin cluster revealed two  $\beta$ -like globin genes over about 13.2 kb that are arranged in the same order as marsupials, 5'- $\epsilon$ - $\beta$ -3' (Figure 5A). This cluster is located on platypus chromosome 2q5.1. Both genes appear to be transcriptionally active and are likely to be functional.

At the time of revising this paper, an independent paper on monotreme  $\beta$ -like globin genes was published by Opazo et al. [62] in which they reported the presence of  $\omega$ ,  $\epsilon^P$  and  $\beta^P$  in the platypus. Largely on the basis of phylogenetic analyses of flanking and coding sequence data, they proposed that platypus  $\epsilon^P$  and  $\beta^P$  were not 1:1 orthologues of therian  $\epsilon$  and  $\beta$ , respectively, and arose by independent duplication of an ancestral  $\beta$ -globin gene in the monotreme lineage, with a separate duplication event, just prior to the divergence of the therians, producing the progenitors of  $\epsilon$  and  $\beta$  of therians. This hypothesis was strongly supported by our BI phylogenetic (Figure 3) analyses, but not by MP analyses of coding sequence data, with third codon sites excluded (Figure 4), or BI analyses of protein sequence data (not shown). These contradictory analyses highlight the difficulty in resolving deep relationships among globin genes, particularly when the time periods between duplication and speciation events are relatively small, the phylogenetic signal at third codon positions is potentially saturated, and non-synonymous sites may be subjected to purifying or positive selection. Despite a very high posterior probability (100%) for the grouping of platypus  $\epsilon$  with monotreme  $\beta$ , this value is a Bayesian probability and depends on the model adequately representing the evolution of the gene. Furthermore, although it was reported [62] that the 5' flanking sequences of platypus  $\epsilon$  and  $\beta$  were similar, we found no evidence for similarity of the promoter signals of these two genes (Figure 5B).

We consider that a more parsimonious explanation is that the platypus  $\epsilon$  is orthologous to the marsupial and eutherian embryonic  $\beta$ -like globin lineages ( $\epsilon$  and  $\gamma$ ), and arose by duplication of an ancestral  $\beta$ -globin gene prior to the mammalian radiation (166 MYA; Figure 9B). The sequence of platypus  $\epsilon$  may have been homogenised by some gene conversion events, leading it to group with other monotreme adult  $\beta$ -like globin genes. In addition to the MP analyses reported above, this explanation is further supported by the conserved position of  $\epsilon$  to the 5' side of the adult  $\beta$ -globin gene in the platypus cluster, which is similar to that found in other therian  $\beta$ -globin gene clusters [26]; see also [29]). Amino acid sequence analyses (BlastP) also provided additional support for the orthology of platypus  $\epsilon$  to other mammalian  $\epsilon$ -globin genes. Although we were unable to examine the expression of the genes in embryonic tissues, it was found that the expression profile of the platypus  $\epsilon$  was similar to the embryonic  $\alpha$ -like globins  $\zeta$  and  $\zeta'$  of the platypus, but not to the adult  $\beta$ -globin gene, supporting its potential role as an embryonic  $\beta$ -like globin gene.

#### **The $\omega$ -globin gene and the evolution of the $\beta$ -globin cluster**

The discovery of the marsupial  $\omega$ -globin gene in the  $\alpha$ -globin cluster [35,36] was critical in re-interpreting the relationships of the  $\alpha$ - and  $\beta$ -globin clusters in amniotes (reptiles, birds and mammals) to favour the hypothesis that these clusters in birds and mammals are paralogous, having diverged independently from different ancestral copies of the vertebrate  $\alpha$ - $\beta$ -globin locus [63].

Our observation of an  $\omega$ -globin gene in the  $\alpha$ -globin cluster in the platypus, as well as in the marsupials, confirms that the ancestral mammal  $\alpha$ -globin cluster contained a  $\beta$ -like globin gene that was lost in eutherians, as proposed by Wheeler et al. [35,36]. However, the position of monotreme and marsupial  $\omega$  in the phylogeny (Figure 3) is more consistent with the original hypothesis [5] that mammal and bird  $\beta$ -globin are orthologous, having descended from the same  $\beta$ -globin progenitor in an amniote ancestor, and this is strongly supported by flanking sequence data (see below). Our data support the proposition that the  $\omega$ -globin gene represents an ancient  $\beta$ -like globin gene lineage that is ancestral to a group containing both mammalian and bird  $\beta$ -globins with a high posterior probability (99%). This arrangement, however, was not supported by analyses of amino acid sequence data, indicating that there is uncertainty in the phylogenetic position of  $\omega$ -globin relative to bird  $\beta$ -globins, or that convergent evolution of bird  $\beta$ -globin genes and  $\omega$ -globin resulted in their similarity at the protein level. To further resolve the key question of whether bird and mammal  $\beta$ -globin gene clusters are orthologous we carried out comparative analyses of flanking loci of the  $\alpha$ - and  $\beta$ -globin clusters.

#### **Genome context of vertebrate $\alpha$ - and $\beta$ -globin clusters**

We found that the platypus  $\alpha$ -globin cluster is flanked by *MPG*, *C16orf35*, *GBY* and *LUC7L*, and that the same genes (except *GBY*) flank the  $\alpha$ -globin cluster in mammals and birds [58,59]. The same genes flank the  $\alpha$ - $\beta$  cluster of frog, and even zebrafish and the  $\alpha$ -cluster of pufferfish [8] (except *GBY* and *LUC7L*), implying that a very ancient region containing these genes (5'-*MPG-C16orf35*- $\alpha$ -*GBY-LUC7L*-3'), or perhaps an even larger region, was present in their common ancestor and has been conserved since the evolution of jawed vertebrates more than 450 MYA.

In contrast, the amniote  $\beta$ -globin clusters reside in a very different genome, sharing none of the flanking loci with the mammal and bird  $\alpha$ -globin clusters, or the  $\alpha$ - $\beta$  cluster of frogs and fish. In platypus, as well as in therian mammals [60,61], the  $\beta$ -globin clusters are flanked by numerous *ORG* genes on both sides. In birds, also, the  $\beta$ -globin cluster is embedded in *ORG* genes [60]. Even the outside loci *RRM1*, *CCKBR* and *ILK* lie in the same orientation with respect to the bird and mammalian  $\beta$ -globin clusters [60], suggesting that the 5'-*RRM1-ORG*- $\beta$  (cluster)-*ORG-CCKBR-ILK*-3' arrangement has been conserved since before the divergence of birds and mammals, more than 315 MYA. Therefore, the bird  $\beta$ -globin cluster is orthologous to the  $\beta$ -globin clusters of mammals.

#### **Conclusion**

##### **New model for the evolution of $\alpha$ - and $\beta$ -globin clusters in amniotes**

This analysis of flanking loci, in addition to the phylogenetic analyses reported above, refutes the prevailing hypothesis that mammal and bird  $\alpha$ - and  $\beta$ -globin clusters evolved from different (paralogous) copies of an ancestral  $\alpha$ - $\beta$ -globin region containing *MPG-C16orf35*- $\alpha$  (cluster)- $\beta$  (cluster)-*GBY-LUC7L*. Rather, the context of  $\beta$ -globin clusters within olfactory receptor genes in birds as well as mammals suggests that a copy of a  $\beta$ -globin locus was moved into a region replete with *ORG* genes before the divergence of birds and mammals 315 MYA. The precise mechanism for this translocation is unknown, but is likely to be either by transposition of a tandem duplicate of an ancestral  $\beta$ -globin gene, or retrotransposition of an intron-containing primary transcript. Phylogenetic analyses suggest that this ancestral  $\beta$ -globin gene within the  $\alpha$ -globin cluster is represented by the platypus and marsupial  $\omega$ -globin gene. The transposed  $\beta$ -globin gene then independently duplicated several times within the avian and mammalian lineages to form the different clusters of differentially expressed  $\beta$ -globin genes. Full details of this new model are given in Figure 9A and 9B.

This hypothesis could be further tested by investigating the gene organization of the  $\alpha$ - and  $\beta$ -globin clusters in

reptiles such as lizards and snakes, which form a sister group to birds. Our hypothesis predicts that reptiles should possess a *MPG-C16orf35- $\alpha$*  (cluster)- $\beta$  (cluster)-*GBY-LUC7L* cluster, and an unlinked *RRM1-ORG- $\beta$*  (cluster)-*ORG-CCBKR-ILK* cluster like birds and mammals. The full genome sequence of the first reptilian species, *Anolis carolinensis*, will provide an opportunity to test this hypothesis.

## Methods

### Isolation and purification of probes to screen for platypus $\beta$ -globins

At the start of this project there were no trace sequences available for any globin genes in the platypus trace archive. We therefore designed probes to screen the platypus male Oa\_Bb BAC library (Clemson University Genomic Institute, USA). The platypus  $\beta$ -globin-specific primers OaBGF (5'-tgaccagaggttcttgac-3') and OaBGR (5'-tgcaattcactcagcttgag-3') were designed from the reference tammar  $\beta$ -globin sequence [GenBank: [AY450928](#)] using Primer3 [64]. Amplification by PCR was performed in a final volume of 25  $\mu$ l, with 40 ng genomic DNA, 1 $\times$  Buffer (Roche, Australia), 0.2 mM dNTPs, 0.05 U Taq (Roche, Australia) and 1  $\mu$ M each of forward and reverse primers. PCR cycling conditions were: 94°C for 2 minutes, then 35 cycles of 94°C for 30 seconds, 50 to 60°C for 30 seconds, 72°C for 1 minute, followed by 72°C for 10 minutes. The PCR products were sub-cloned according to the TOPO TA cloning<sup>®</sup> Kit Protocol (Invitrogen, Australia) and the resulting plasmids were purified according to the centrifugation protocol of Wizard<sup>®</sup> Plus SV Minipreps DNA Purification System (Promega, Australia). The purified plasmids were confirmed to contain PCR products of a partial platypus  $\beta$ -globin gene (167 bp) by sequencing at the Australian Genome Research Facility (AGRF, Brisbane, Australia) using M13 forward (5'-gtaaacgacggccag-3') and M13 reverse (5'-caggaacagctatgac-3') primers. Once confirmed, they were used as probes to screen the platypus BAC library.

### Screening the platypus BAC library for $\beta$ -like globin genes

The platypus BAC library filters were pre-hybridised at 65°C with Church Buffer (1 mM EDTA, 0.5 M phosphate buffer, 7% (w/v) SDS) including 1% BSA for 4 hours. The platypus  $\beta$ -globin probes (25 ng) were labelled with <sup>32</sup>P-dATP using the Megaprime DNA labelling System (GE Healthcare, Australia) following the manufacturer's instructions. The probes were allowed to hybridise to the filters at stringent conditions (65°C with the above buffer) for 24 hours and then washed twice for 15 minutes each in 2  $\times$  SSC/0.1%SDS and 1  $\times$  SSC/0.1%SDS. Autoradiography was carried out for 14 days at -80°C with an intensifying cassette.

### Identification of platypus BAC clones containing $\alpha$ -like globin genes

Unlike  $\beta$ , BACs were not screened for  $\alpha$ -like globin genes. Instead they were identified directly from the Encyclopaedia Of DNA Elements Project [65], in which the  $\alpha$ -globin cluster is one of the targeted regions [12,66]. Two platypus BAC clones (Oa\_Bb-2L7 and Oa\_Bb-131M24), which were sequenced but not yet annotated, were identified by computational analysis (below) to contain parts of the  $\alpha$ -globin cluster and a  $\omega$ -globin gene.

### Isolation and purification of DNA from BAC clones

DNA from the identified BAC clones (including those that were screened) was extracted using Wizard<sup>®</sup> Plus SV Minipreps DNA Purification System (Promega, Australia). The purified BAC clones were then subjected to Dot or Southern Blot to confirm the presence of  $\alpha$ - or  $\beta$ -globin genes respectively.

### Confirmation of BACs containing globin genes

Dot blot methods were used to verify the presence of the  $\alpha$ -like globin genes. In a plate containing Luria broth agar with chloramphenicol, a Hybond N<sup>+</sup> (GE Healthcare, Australia) filter was placed and multiple 1  $\mu$ l of liquid culture BAC clones were spotted onto the filter. The plate was incubated at 37°C overnight and then the filter was soaked in Denaturation Solution (0.5 M NaOH and 1.5 M NaCl) for 5 minutes, followed by soaking twice in Neutralisation Solution (0.5 M Tris-Cl pH 7.4 and 1.5 M NaOH) for 5 minutes each. The filter was then rinsed in 2  $\times$  SSC, soaked in 0.4 M NaOH for 20 minutes and washed with 6  $\times$  SSC to remove all cellular debris. The filters were then screened with the platypus  $\alpha$ -globin probes using the standard library screening procedure (above).

Southern blotting was used to verify the presence of the  $\beta$ -like globin genes. In a 40  $\mu$ l reaction, 20 to 40 ng BAC DNA was digested with 10 U of restriction enzyme, *HIND III* (Roche, Australia). The reaction was incubated at 37°C for at least 4 hours and separated by electrophoresis on a 0.8% agarose gel overnight at 40 V. The DNA fragments were transferred onto a Hybond N<sup>+</sup> (GE Healthcare, Australia) nylon filter overnight by capillary action following the manufacturer's instructions, and cross-linked in 0.4 M NaOH for 20 minutes. These filters were then screened with the platypus  $\beta$ -globin probes using the standard library screening procedure (above).

### Fluorescence in situ hybridisation (FISH)

Male platypus metaphase spreads were prepared and *in situ* mapping was performed using two-colour FISH as described previously by McMillan et al. [21]. The verified BACs containing the  $\alpha$ -like globin genes ( $\zeta$  and  $\zeta'$ : Oa\_Bb-2L7) and  $\beta$ -like globin genes ( $\epsilon$  and  $\beta$ : Oa\_Bb-484F22) were labelled with different fluorochromes and then

hybridised to the chromosomes. The signals were detected by fluorescent microscopy, where at least twenty metaphase images were captured and analysed.

#### **Sequence data of the platypus BAC clones containing the $\alpha$ - and $\beta$ -globin clusters**

Information about the platypus BAC clones containing the  $\alpha$ -like globin genes along with the  $\omega$ -globin gene was obtained directly from the ENCODE Project [66]. Their sequence information was obtained from GenBank; accession numbers: [AC195438](#) (Oa\_Bb-2L7) and [AC203513](#) (Oa\_Bb-131M24).

The BAC clone containing the  $\beta$ -like globin genes that were found from the library screening procedure were sequenced at the Washington University Genome Sequencing Centre (St Louis, USA). The sequence information for this BAC clone was obtained from GenBank: [AC192436](#) (Oa\_Bb-484F22).

#### **Computational characterisation of the $\alpha$ - and $\beta$ -globin clusters in the platypus**

Using sequence information of [AC195438](#), [AC203513](#) and [AC192436](#), genes were predicted by GENSCAN [46] and GenomeScan [47] using default settings. All predicted gene sequences were then subjected to a BLAST search of the translated nucleotide acid (BlastX) and protein (BlastP) databases to confirm their identities.

#### **Promoter analyses**

Transcription factor binding motifs were predicted in the 200 bp promoter region located 5' to the predicted platypus  $\alpha$ - and  $\beta$ -like genes and *GBY* by rVista 2.0 [67] using user-defined consensus sequences for 'CACCC', 'CAAT', 'TATA', GATA1 ('WGATAR' [51]) and EKLF ('NGNGT-GGGN' [51]). The same criteria were used to predict the same motifs in marsupials (*Didelphis virginiana* [ $\zeta$  and  $\psi\zeta$ ': [AC139599](#)] and *Sminthopsis macroura* [ $\alpha^D$ ,  $\psi\alpha^3$ ,  $\alpha^2$ ,  $\alpha^1$ ,  $\omega$ : [AC146781](#); and  $\epsilon$ ,  $\beta$ : [AC148754](#)]) and in humans [ $\zeta$ ,  $\psi\zeta$ ',  $\alpha^D$ ,  $\psi\alpha^3$ ,  $\alpha^2$ ,  $\alpha^1$ : [NG\\_000006](#); and  $\epsilon$ ,  $\beta$ : [NG\\_000007](#)] for consistency in comparison.

#### **Confirmation of $\alpha^1$ and $\alpha^3$ by BLAST search and Southern blot**

To confirm that the presence of two almost identical genes ( $\alpha^1$  and  $\alpha^3$ ) was real rather than an assembly error, the boundaries (~300 bp) of the homologous regions were investigated by a BLAST search against the platypus WGS database. The raw sequences of best hits were extracted from NCBI GenBank, cleaned and aligned in Sequencher v4.8 (Gene Codes Corporation, Michigan) using default settings.

Southern blotting was also used to verify the presence of  $\alpha^1$  and  $\alpha^3$  genes. In a 30  $\mu$ l reaction, 100  $\mu$ g BAC DNA

(Oa\_Bb: 131M24, 130N2, 150K14 and 223I12) was digested with 10 U of restriction enzyme, *EcoRV* (Roche, Australia). The reaction was incubated at 37°C for at least 4 hours and separated by electrophoresis on a 0.8% agarose gel overnight at 40 V. The DNA fragments were transferred onto a Hybond N<sup>+</sup> (GE Healthcare, Australia) nylon filter overnight by capillary action following the manufacturer's instructions, and cross-linked in 0.4 M NaOH for 20 minutes. These filters were then screened with the platypus  $\alpha^1/\alpha^3$  (test) and  $\alpha^2$  (control) probes using the standard library screening procedure (above).

#### **RT-PCR analyses**

To remove DNA contamination, RNAs derived from adult male platypus liver, kidney, spleen, testis, brain and lungs were DNase treated using a DNA-free™ kit according to the manufacturer's instructions (Applied Biosystems, Australia). Treated RNAs were then reverse transcribed using Superscript III (Invitrogen, Australia) following the manufacturer's instructions. Primers were designed against predicted  $\alpha$ - and  $\beta$ -like and *GBY* globin gene sequences using Primer3 [64]. In each case, the region amplified spanned an intron so that the origin of the template (gDNA or cDNA) was immediately obvious. Primer sequences and the expected sizes of amplified cDNA and gDNA bands are shown in Table 2. PCR reactions and cycling conditions were the same as for screening for the platypus  $\beta$ -globin genes, above. The positive bands were directly sequenced by AGRF (Brisbane, Australia) to confirm their identities. The blood contamination in the tested samples had minimal effect on the observed expression pattern, as some tissues (for example, lung and liver) showed no amplification despite containing large quantities of blood.

#### **Phylogenetic analyses**

Phylogenetic analyses were employed to verify the identities of the platypus globin genes and study the evolutionary relationships of the different members of the  $\alpha$ - and  $\beta$ -globin gene families. This study was restricted to the coding domains of the  $\alpha$ - and  $\beta$ -globin members and the accession numbers of the sequences used are given in the legends of Figures 2 and 3. Phylogenetic analyses were conducted using MP in PAUP\* v.4.0b10 [68], and a BI approach using MrBayes v.3.1.2 [69]. Concordance of trees from each of the different methods, bootstrap proportions and posterior probability estimates were used to examine the robustness of nodes.

MP analyses were conducted for the entire coding sequence matrix and after excluding third codon positions using a heuristic search option and default options (TBR branch swapping), with the exception of using random stepwise addition repeated 100 times. Character state optimisation for MP trees used the DELTRAN option. MP

**Table 2: PCR primers used for amplification of the  $\alpha$ - and  $\beta$ -like globin genes including GBY from the platypus gDNA and cDNA**

Gene	Forward Primer	Reverse Primer	gDNA (bp)	cDNA (bp)
$\zeta$	GGCCGACAAGACCGCAGTCATCTCCC	CCCGATGGCGCTGATGACT	527	190
$\zeta^1$	TGACCAAAGGCGACAAGACCT	CCCCGATGGCACCGATGACC	534	198
$\alpha^D$	GAGGCTGTGAAGAACCTGGA	GGTGTACTCCCCTTGCAGAT	1793	153
$\alpha^2$	TGGCCCACCTCGATGACCTGG	GGGAAGGTGTCTGGCCACC	289	134
$\alpha^1/\alpha^3$	GCAAGGCCCGCCGGTCACGGC	CGCTGTCCATGTCATCGAAGTGCC	597	192
$\omega$	ATTGTGTCCATCTGGGGAAA	GCTTGGCAAAGTTGCTCTTC	488	232
<b>GBY</b>	CTGGAAACAGGTGTGCAAGA	CTATCTCCGGGTGTAGCAG	3202	149
$\varepsilon$	ATCTGAGCGCTGAGGAGAAG	GACAGTTGCCGAAGGAGTCA	285	142
$\beta$	CTGTGGGGAAAGTGAACAT	GGTCAGCACCTTAGCACCAT	321	168

bootstrap analyses [70] were carried out using 1000 bootstrap pseudoreplicates, employing a heuristic search option with random stepwise addition.

The program MODELTEST [71] and the Akaike information criterion (AIC) were used to assess the most appropriate model for BI analyses. The MODELTEST analyses were facilitated using the program MrMTgui v1.0 [72]. The MODELTEST analysis was carried out on separate codon positions for  $\alpha$ - and  $\beta$ -globin data sets. For  $\alpha$ -globin sequences, a general time reversible (GTR) model [73], with a proportion of invariant sites (I) and unequal rates among sites [74], modelled with a gamma distribution (G) was found to be the most appropriate model to use for first and second codon positions, and a GTR+G model was appropriate for third codon positions under the AIC. For  $\beta$ -globin sequences a GTR+I+G model was considered appropriate for first positions, and a GTR+G model was found to be appropriate for second and third codon positions. The MrBayes analysis was carried out applying these different models to each codon position using an unlinked analysis, with default uninformative priors. Four chains were run simultaneously for 2 million generations in two independent runs, sampling trees every 100 generations. The program TRACER (version 1.3; [75]) was used to assess tree and parameter convergence. For both the  $\alpha$ -globin and  $\beta$ -globin analyses all effective sample sizes for all parameters were >1297, indicating a sufficient sample of the parameter space had been taken. A burn-in of 2000 trees (equivalent to 200,000 generations) was chosen for each independent run of MrBayes, with a >50% posterior probability consensus tree constructed from the remaining 36,002 trees (18,001 trees each run).

A BI analysis using MrBayes (version 3.1.2) was also carried out using protein sequence data from  $\beta$ -globin genes. A mixed protein model was used, allowing the optimum model of protein evolution to be assessed from a selection of nine fixed-rate models. The optimum model was found to be the Dayhoff model with a posterior probability of 1.0. The analyses were conducted using two million gen-

erations in two independent runs, sampling trees every 100 generations. A burn-in of 2,000 trees was used for each run with a 50% consensus tree constructed from the remaining 36,002 trees.

### Authors' contributions

VSP designed and performed most of the experiments and analysed the data. VSP also drafted the main manuscript. SJBC conducted phylogenetic analyses and contributed to the writing of the manuscript. JED helped in designing the experiments and trouble-shooting experiments. BF, TG, WCW and RKW were involved in sequencing the platypus BAC clone (Oa\_Bb-484F22). JAMG conceived and supervised the research and contributed to the writing of the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Additional material

#### Additional file 1

*Annotation of the platypus  $\alpha$ - and  $\beta$ -like and GBY globin genes. This table shows the predicted positions of six  $\alpha$ -like,  $\omega$  and GBY globin genes in the platypus BAC clone Oa\_Bb-131M24 [GenBank: [AC203513](#)], two  $\alpha$ -like globin genes in BAC clone Oa\_Bb-2L7 [GenBank: [AC195438](#)], and two  $\beta$ -like globin genes in BAC clone Oa\_Bb-484F22 [GenBank: [AC192436](#) reverse direction].*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1741-7007-6-34-S1.doc>]

### Acknowledgements

We thank Dr Frank Grützner (The University of Adelaide) and Tim Hore (The Australian National University) for providing platypus tissues and RNA preparations. We also thank Dr Paul Waters for providing computational help with the identification and confirmation of two identical but different genes  $\alpha^1$  and  $\alpha^3$ .

## References

- Roesner A, Fuchs C, Hankeln T, Burmester T: **A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals.** *Mol Biol Evol* 2005, **22**:12-20.
- Fuchs C, Burmester T, Hankeln T: **The amphibian globin gene repertoire as revealed by the *Xenopus* genome.** *Cytogenet Genome Res* 2006, **112**:296-306.
- Dickerson RE, Geis I: *Hemoglobin: Structure, Function, Evolution and Pathology* Menlo Park, CA: Benjamin Cummings; 1983:65-116.
- Hentschel CC, Kay RM, Williams JG: **Analysis of *Xenopus laevis* globins during development and erythroid cell maturation and the construction of recombinant plasmids containing sequences derived from adult globin mRNA.** *Dev Biol* 1979, **72**:350-363.
- Jeffreys AJ, Wilson V, Wood D, Simons JP, Kay RM, Williams JG: **Linkage of adult alpha- and beta-globin genes in *X. laevis* and gene duplication by tetraploidization.** *Cell* 1980, **21**:555-564.
- Kay RM, Harris R, Patient RK, Williams JG: **Molecular cloning of cDNA sequences coding for the major alpha- and beta-globin polypeptides of adult *Xenopus laevis*.** *Nucleic Acids Res* 1980, **8**:2691-2707.
- Pisano E, Cocca E, Mazzei F, Ghigliotti L, di Prisco G, Detrich HW III, Ozouf-Costaz C: **Mapping of alpha- and beta-globin genes on Antarctic fish chromosomes by fluorescence *in situ* hybridization.** *Chromosome Res* 2003, **11**:633-640.
- Gillemans N, McMorrow T, Tewari R, Wai AW, Burgtorf C, Drabek D, Ventress N, Langeveld A, Higgs D, Tan-Un K, Grosveld F, Philipsen S: **Functional and comparative analysis of globin loci in pufferfish and humans.** *Blood* 2003, **101**:2842-2849.
- Dodgson JB, McCune KC, Rusling DJ, Krust A, Engel JD: **Adult chicken alpha-globin genes alpha A and alpha D: no anemic shock alpha-globin exists in domestic chickens.** *Proc Natl Acad Sci USA* 1981, **78**:5998-6002.
- Engel JD, Dodgson JB: **Analysis of the adult and embryonic chicken globin genes in chromosomal DNA.** *J Biol Chem* 1978, **253**:8239-8246.
- Engel JD, Dodgson JB: **Analysis of the closely linked adult chicken alpha-globin genes in recombinant DNAs.** *Proc Natl Acad Sci USA* 1980, **77**:2596-2600.
- Cooper SJ, Wheeler D, De Leo A, Cheng JF, Holland RA, Marshall Graves JA, Hope RM: **The mammalian alphaD-globin gene lineage and a new model for the molecular evolution of alpha-globin gene clusters at the stem of the mammalian radiation.** *Mol Phylogenet Evol* 2006, **38**:439-448.
- Goh SH, Lee YT, Bhanu NV, Cam MC, Desper R, Martin BM, Mohararam R, Gherman RB, Miller JL: **A newly discovered human alpha-globin gene.** *Blood* 2005, **106**:1466-1472.
- Hoffmann FG, Storz JF: **The alphaD-globin gene originated via duplication of an embryonic alpha-like globin gene in the ancestor of tetrapod vertebrates.** *Mol Biol Evol* 2007, **24**:1982-1990.
- Hardison RC, Sawada I, Cheng JF, Shen CK, Schmid CW: **A previously undetected pseudogene in the human alpha globin gene cluster.** *Nucleic Acids Res* 1986, **14**:1903-1911.
- Lauer J, Shen CK, Maniatis T: **The chromosomal arrangement of human alpha-like globin genes: sequence homology and alpha-globin gene deletions.** *Cell* 1980, **20**:119-130.
- Orkin SH: **The duplicated human alpha globin genes lie close together in cellular DNA.** *Proc Natl Acad Sci USA* 1978, **75**:5950-5954.
- Proudfoot NJ, Maniatis T: **The structure of a human alpha-globin pseudogene and its relationship to alpha-globin gene duplication.** *Cell* 1980, **21**:537-544.
- Wakefield MJ, Graves JA: **Marsupials and monotremes sort genome treasures from junk.** *Genome Biol* 2005, **6**:218.
- Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446**:507-512.
- McMillan D, Miethke P, Alsop AE, Rens W, O'Brien P, Trifonov V, Veyrunes F, Schatzkammer K, Kremitzki CL, Graves T, Warren W, Grutzner F, Ferguson-Smith MA, Graves JA: **Characterizing the chromosomes of the platypus (*Ornithorhynchus anatinus*).** *Chromosome Res* 2007, **15**:961-974.
- Grutzner F, Rens W, Tsend-Ayush E, El-Mogharbel N, O'Brien PC, Jones RC, Ferguson-Smith MA, Graves JA: **In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes.** *Nature* 2004, **432**:913-917.
- Rens W, Grutzner F, O'Brien PC, Fairclough H, Graves JA, Ferguson-Smith MA: **Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution.** *Proc Natl Acad Sci USA* 2004, **101**:16257-16261.
- Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F, Belov K, Miller W, Clarke L, Chinwalla AT, Yang SP, Heger A, Locke DP, Miethke P, Waters PD, Veyrunes F, Fulton L, Fulton B, Graves T, Wallis J, Puente XS, Lopez-Otin C, Ordonez GR, Eichler EE, Chen L, Cheng Z, Deakin JE, Alsop A, Thompson K, Kirby P, Papenfuss AT, Wakefield MJ, Olander T, Lancet D, Huttley GA, Smit AF, Pask A, Temple-Smith P, Batzer MA, Walker JA, Konkel MK, Harris RS, Whittington CM, Wong ES, Gemmill NJ, Buschiazio E, Vargas Jentzsch IM, Merkel A, Schmitz J, Zemmann A, Churakov G, Kriegs JO, Brosius J, Murchison EP, Sachidanandam R, Smith C, Hannon GJ, Tsend-Ayush E, McMillan D, Attenborough R, Rens W, Ferguson-Smith M, Lefevre CM, Sharp JA, Nicholas KR, Ray DA, Kube M, Reinhardt R, Pringle TH, Taylor J, Jones RC, Nixon B, Dacheux JL, Niwa H, Sekita Y, Huang X, Stark A, Kheradpour P, Kellis M, Flicek P, Chen Y, Webber C, Hardison R, Nelson J, Hallsworth-Pepin K, Delehaunty K, Markovic C, Minx P, Feng Y, Kremitzki C, Mitreva M, Glasscock J, Wylie T, Wohldmann P, Thiru P, Nhan MN, Pohl CS, Smith SM, Hou S, Renfree MB, Mardis ER, Wilson RK: **Genome analysis of the platypus reveals unique signatures of evolution.** *Nature* 2008, **453**:175-183.
- Cooper SJ, Hope RM: **Evolution and expression of a beta-like globin gene of the Australian marsupial *Sminthopsis crassicaudata*.** *Proc Natl Acad Sci USA* 1993, **90**:11777-11781.
- Cooper SJ, Murphy R, Dolman G, Hussey D, Hope RM: **A molecular and evolutionary study of the beta-globin gene family of the Australian marsupial *Sminthopsis crassicaudata*.** *Mol Biol Evol* 1996, **13**:1012-1022.
- Koop BF, Goodman M: **Evolutionary and developmental aspects of two hemoglobin beta-chain genes (epsilon M and beta M) of opossum.** *Proc Natl Acad Sci USA* 1988, **85**(11):3893-3897.
- Lee MH, Shroff R, Cooper SJ, Hope R: **Evolution and molecular characterization of a beta-globin gene from the Australian Echidna *Tachyglossus aculeatus* (Monotremata).** *Mol Phylogenet Evol* 1999, **12**:205-214.
- Goodman M, Koop BF, Czelusniak J, Weiss ML: **The eta-globin gene. Its long evolutionary history in the beta-globin gene family of mammals.** *J Mol Biol* 1984, **180**:803-823.
- Hardison R, Miller W: **Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters.** *Mol Biol Evol* 1993, **10**:73-102.
- Hardison RC: **Comparison of the beta-like globin gene families of rabbits and humans indicates that the gene cluster 5'-epsilon-gamma-delta-beta-3' predates the mammalian radiation.** *Mol Biol Evol* 1984, **1**:390-410.
- Weatherall DJ: *The Structure, Organization, and Regulation of Human Genes* 3rd edition. Oxford: Oxford University Press; 1991.
- Goodman M, Czelusniak J, Koop BF, Tagle DA, Slightom JL: **Globins: a case study in molecular phylogeny.** *Cold Spring Harb Symp Quant Biol* 1987, **52**:875-890.
- Reitman M, Grasso JA, Blumenthal R, Lewit P: **Primary sequence, evolution, and repetitive elements of the *Gallus gallus* (chicken) beta-globin cluster.** *Genomics* 1993, **18**:616-626.
- Wheeler D, Hope R, Cooper SB, Dolman G, Webb GC, Bottema CD, Gooley AA, Goodman M, Holland RA: **An orphaned mammalian beta-globin gene of ancient evolutionary origin.** *Proc Natl Acad Sci USA* 2001, **98**:1101-1106.
- Wheeler D, Hope RM, Cooper SJ, Gooley AA, Holland RA: **Linkage of the beta-like omega-globin gene to alpha-like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters.** *J Mol Evol* 2004, **58**:642-652.
- Holland RA, Gooley AA 2nd: **Characterization of the embryonic globin chains of the marsupial Tammar wallaby, *Macropus eugenii*.** *Eur J Biochem* 1997, **248**:864-871.
- Andersen NA, Mesch U, Lovell DJ, Nicol SC: **The effects of sex, season and hibernation on haematology and blood viscosity**

- of free-ranging echidnas (*Tachyglossus aculeatus*). *Can J Zool* 2000, **78**:174-181.
39. Bently PJ, Herreid CF, Schmidt-Nielsen K: **Respiration of a monotreme, the echidna, *Tachyglossus aculeatus***. *Am J Physiol* 1967, **212**:957-961.
  40. Parsons RS, Atwood J, Guiler ER, Heddl RWL: **Comparative studies on the blood of monotremes and marsupials. I. Haematology**. *Comp Biochem Physiol B* 1971, **39(2)**:203-208.
  41. Whittaker RG, Thompson EO: **Studies on monotreme proteins. V. Amino acid sequence of the alpha-chain of haemoglobin from the platypus, *Ornithorhynchus anatinus***. *Aust J Biol Sci* 1974, **27**:591-605.
  42. Whittaker RG, Thompson EO: **Studies on monotreme proteins. VI. Amino acid sequence of the beta-chain of haemoglobin from the platypus, *Ornithorhynchus anatinus***. *Aust J Biol Sci* 1975, **28**:353-365.
  43. Thompson EO, Fisher WK, Whittaker RG: **Studies on monotreme proteins. 3. Amino acid sequence of the alpha and beta-globin chains of the minor haemoglobin from the echidna, *Tachyglossus aculeatus aculeatus***. *Aust J Biol Sci* 1973, **26**:1327-1335.
  44. Whittaker RG, Fisher WK, Thompson EO: **Studies on monotreme proteins. I. Amino acid sequence of the beta-chain of haemoglobin from the echidna, *Tachyglossus aculeatus aculeatus***. *Aust J Biol Sci* 1972, **25**:989-1004.
  45. University of California Santa Cruz (UCSC) Genome Browser [<http://genome.ucsc.edu>]
  46. Burge C, Karlin S: **Prediction of complete gene structures in human genomic DNA**. *J Mol Biol* 1997, **268**:78-94.
  47. Yeh RF, Lim LP, Burge CB: **Computational inference of homologous gene structures in the human genome**. *Genome Res* 2001, **11**:803-816.
  48. Motamed K, Bastiani C, Zhang Q, Bailey A, Shen CK: **CACC box and enhancer response of the human embryonic epsilon globin promoter**. *Gene* 1993, **123**:235-240.
  49. Kim CG, Sheffery M: **Physical characterization of the purified CCAAT transcription factor, alpha-CPI**. *J Biol Chem* 1990, **265**:13362-13369.
  50. Sabath DE, Koehler KM, Yang WQ, Phan V, Wilson J: **DNA-protein interactions in the proximal zeta-globin promoter: identification of novel CCACC- and CCAAT-binding proteins**. *Blood Cells Mol Dis* 1998, **24**:183-198.
  51. Whyatt DJ, deBoer E, Grosveld F: **The two zinc finger-like domains of GATA-1 have different DNA binding specificities**. *EMBO J* 1993, **12**:4993-5005.
  52. Gillemans N, Tewari R, Lindeboom F, Rottier R, de Wit T, Wijgerde M, Grosveld F, Philippen S: **Altered DNA-binding specificity mutants of EKLf and Sp1 show that EKLf is an activator of the beta-globin locus control region in vivo**. *Genes Dev* 1998, **12**:2863-2873.
  53. Agarwal S, Arya V, Stolle CA, Pradhan M: **A novel Indian beta-thalassemia mutation in the CACCC box of the promoter region**. *Eur J Haematol* 2006, **77**:530-532.
  54. Hardison R, Chao KM, Adamkiewicz M, Price D, Jackson J, Zeigler T, Stojanovic N, Miller W: **Positive and negative regulatory elements of the rabbit embryonic epsilon-globin gene revealed by an improved multiple alignment program and functional analysis**. *DNA Seq* 1993, **4**:163-176.
  55. Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, Miller W: **Locus control regions of mammalian beta-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights**. *Gene* 1997, **205**:73-94.
  56. Cooper SJ, Wheeler D, Hope RM, Dolman G, Saint KM, Gooley AA, Holland RA: **The alpha-globin gene family of an Australian marsupial, *Macropus eugenii*: the long evolutionary history of the theta-globin gene and its functional status in mammals**. *J Mol Evol* 2005, **60**:653-664.
  57. Ensembl Genome Browser [<http://www.ensembl.org/index.html>]
  58. Flint J, Tufarelli C, Peden J, Clark K, Daniels RJ, Hardison R, Miller W, Philippen S, Tan-Un KC, McMorrow T, Frampton J, Alter BP, Frischau AM, Higgs DR: **Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the alpha globin cluster**. *Hum Mol Genet* 2001, **10**:371-382.
  59. Hughes JR, Cheng JF, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin E, Higgs DR: **Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences**. *Proc Natl Acad Sci USA* 2005, **102**:9830-9835.
  60. Taylor TD, Noguchi H, Totoki Y, Toyoda A, Kuroki Y, Dewar K, Lloyd C, Itoh T, Takeda T, Kim DW, She X, Barlow KF, Bloom T, Bruford E, Chang JL, Cuomo CA, Eichler E, FitzGerald MG, Jaffe DB, LaButti K, Nicol R, Park HS, Seaman C, Sougnez C, Yang X, Zimmer AR, Zody MC, Birren BW, Nusbaum C, Fujiiyama A, Hattori M, Rogers J, Lander ES, Sakaki Y: **Human chromosome 11 DNA sequence and analysis including novel gene identification**. *Nature* 2006, **440**:497-500.
  61. Bulger M, van Doorninck JH, Saitoh N, Telling A, Farrell C, Bender MA, Felsenfeld G, Axel R, Groudine M: **Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes**. *Proc Natl Acad Sci USA* 1999, **96**:5129-5134.
  62. Opazo JC, Hoffmann FG, Storz JF: **Genomic evidence for independent origins of beta-like globin genes in monotremes and therian mammals**. *Proc Natl Acad Sci USA* 2008, **105**:1590-1595.
  63. Hardison RC: **New views of evolution and regulation of vertebrate beta-like globin gene clusters from an orphaned gene in marsupials**. *Proc Natl Acad Sci USA* 2001, **98**:1327-1329.
  64. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers**. *Methods Mol Biol* 2000, **132**:365-386.
  65. The ENCODE Project Consortium: **The ENCODE (ENCyclopedia Of DNA Elements) Project**. *Science* 2004, **306**:636-640.
  66. NIH Intramural Sequencing Center (NISC) [<http://www.nisc.nih.gov/>]
  67. Loots GG, Ovcharenko I: **rVISTA 2.0: evolutionary analysis of transcription factor binding sites**. *Nucleic Acids Res* 2004:W217-W221.
  68. Swofford DL: **PAUP\*. Phylogenetic analysis using parsimony (\* and other methods). Version 4.0b10**. Sunderland, MA: Sinauer; 2002.
  69. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny**. *Bioinformatics* 2001, **17**:754-755.
  70. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap**. *Evolution* 1985, **39**:783-791.
  71. Posada D, Crandall K: **Modeltest: testing the model of DNA substitution**. *Bioinformatics* 1998, **14**:817-818.
  72. MrMTgui v1.0 [<http://www.genedrift.org/mrmtgui.php>]
  73. Rodriguez F, Oliver JF, Marin A, Medina JR: **The general stochastic model of nucleotide substitutions**. *J Theor Biol* 1990, **142**:485-501.
  74. Yang Z: **Among-site rate variation and its impact on phylogenetic analyses**. *Trends Ecol Evol* 1996, **11**:367-372.
  75. Rambaut A, Drummond AJ: **Tracer v1.3**. 2004 [<http://beast.bio.ed.ac.uk/Tracer>].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

