J Vector Borne Dis 48, March 2011, pp. 27–36

Comparative evolutionary analyses of beta globin gene in eutherian, dinosaurian and neopterygii taxa

Gauri Awasthi, Garima Srivastava & Aparup Das

Evolutionary Genomics and Bioinformatics Laboratory, National Institute of Malaria Research, New Delhi, India

ABSTRACT

Background & objectives: Comparative genomics and evolutionary analyses of conserved genes have enabled us to understand the complexity of genomes of closely related species. For example: β-globin gene present in human hemoglobin is one such gene that has experienced many genetic changes in many related taxa and produced more than 600 variants. One of the variant, HBS causes sickle-cell anemia in humans but offers protection against severe malaria due to *Plasmodium falciparum.* In the present study, we characterized and performed evolutionary comparative analyses of the β-globin gene in different related and unrelated taxa to have a comprehensive view of its evolution.

Methods: DNA and protein sequences of β-globin gene were downloaded from NCBI and characterized in detail in nine eutherian (*Homo sapiens*, *Pan troglodytes*, *Macaca mulatta*, *Mus musculus*, *Rattus norvegicus*, *Bos taurus*, *Canis familiaris*, *Equus caballus*, *Oryctolagus cuniculus*), a dinosaurian (*Gallus gallus*) and a neopterygii (*Danio rerio*) taxa. Three more eutherian (*Papio anubis*, *Ovis aries* and *Sus scrofa*) taxa were included for an analysis at the protein level but not included at the gene level owing to lack of genomic information. Computational and phylogenetic analyses were performed using evolutionary comparative approach.

Results: Results of comparative and phylogenetic analyses revealed less conservation of genetic architecture of β-globin compared to its protein architecture in all eutherian taxa. Both dinosaurian and neopterygii taxa served as outgroups and varied at gene and protein levels.

Interpretation & conclusion: Most remarkably, all primates from eutherian taxa including *P. anubis* showed only nine codon position differences and an absolute similarity between *H. sapiens* and *P. troglodytes*. Absolute conservation of coding region in *Equus caballus* (horse) was observed. The results were discussed with an inference on the role of evolutionary forces in maintaining such close similarities and variations across closely related taxa. Further, the need to utilize more comparative approaches in understanding the disease causing genes' evolution in closely related taxa is hoped for.

Key words β-globin gene; evolution; malaria; phylogeny; sickle-cell anemia

INTRODUCTION

Comparative genomic studies have helped to decipher within and between species variations by comparing genes conserved across species with those that have taken different functions according to need and evolved to perform specific functions[1]. From basic biology to highly complex dynamic mechanisms of genes, these studies have helped in identifying and comparing functional sequences based on high levels of evolutionary conservation. Such comparisons have proven successful not only for closely related species such as human-primate or human-mouse but also for distant evolutionary comparisons, such as human-fish and human-bird[2]. Majority of these studies have contributed to better understanding of highly important human genes related to infectious diseases that are simultaneously evolving in various taxa[3–5] and how natural selection of human genes has provided increased adaptive fitness on exposure to infectious diseases[6]. Since gene duplication with subsequent interaction divergence is one of the primary driving forces in the evolution of genetic systems and little is known about the precise mechanisms and the role of duplication divergence in evolution, these observations might prove beneficial to infer evolution of medically important genes in different taxa.

As agents of natural selection, infectious diseases like malaria have played a major role in the evolution of human species by showing association between genetic variation in beta globin gene and protection from severe malaria due to *P. falciparum*[6]. This protective effect causes a balanced polymorphism of the sickle-cell mutation in malaria endemic regions also. Considering this, the human beta globin (HBB) gene, present in hemoglobin and located on the short arm of chromosome 11 at p15.5 is a highly important gene in understanding the complexities of malaria. Hemoglobin is a major blood protein that con-

sists of four polypeptides, two alpha globins and two beta globins and belongs to the globin gene family, a group of genes involved in oxygen transport[7]. Structurally, five transcriptionally active β-like globin genes are present within a cluster of 45 Kb in the following order[8]: 5'- ε - Gγ – Aγ – δ – β - 3' (Fig. 1). Additionally, the pseudogene, $\psi\beta_1$, is located between the Aγ and δ genes[8] (Fig. 1). The HBB gene provides instructions for making a protein called beta-globin. The ε- and γ-globin genes diverged from the ancient ε/γ ancestral gene about 100 Myrs ago followed by duplication of the γ-globin gene to form the Gγ and Aγ genes, an event which occurred before the divergence of the Old World and New World monkeys[9]. The human alpha globin gene (HBA), located on chromosome 16, is also a member of this gene family. Under normal conditions, the alpha and beta genes in hemoglobin are transmitted separately as they are located on separate chromosomes[7]. Individuals inherit a copy of alpha gene and a beta gene from each of their parents and then recombine to determine individual genotypes.

Altered forms of genes (otherwise called as alleles), such as sickle-cell hemoglobin HbS, originate only by mutation or a change in the nucleotide sequence of the DNA[10]. Sickle-cell hemoglobin is a single point mutation occurring when there is a substitution of valine (V) for glutamic acid (E) at the sixth amino acid position in the beta globin gene[10]. HbS mutation forms three genotypes; for example, when the A hemoglobin and the S hemoglobin genes (replacing the normal hemoglobin gene) combine they form: AA homozygous dominant (normal genotype and phenotype), AS heterozygote (sickle-cell genotype but not phenotype: carrier) and the SS homozygous recessive[7] (abnormal genotype and phenotype: leads to sickle-cell anemia). Different types of variations of beta globin gene are associated with human disease such as Hemoglobin SC, Sickle/beta-thalassemia, Hemoglobin E/beta-thalassemia and Alpha thalassemia/Hemoglobin constant spring. An advantage of inheriting HBB, if heterozygous, is increased immunity to malaria while the disadvantage is that homozygous recessive alleles (SS) develop sickle-cell anemia[7]. The HbS allele has been identified in four genetic backgrounds in different African populations, suggesting that the same mutation arose independently several times through convergent evolution[11]. Since the discovery of HbS in 1949, the number of hemoglobin variants is increasing and more than 600 are known today, mainly, the HbC and HbE alleles, which arose and spread in Africa and in southeast Asia, respectively[12]. Different uncommon beta globin variants have also been reported from different parts of India, like HbD Iran (β22 Glu→Gln), Hb Hofu (β126 Val→Glu) besides unspecified ones such as HbJ, HbK and HbM[8]. Heterozygosity for two different globin gene mutations were also recorded from different parts of India and globe, e.g. HbSD, HbSE, HbDK and HbSC[8].

Indeed, the β-globin gene cluster in humans represents a good model for investigating mechanisms and processes of genome evolution, because it is one of the most intensively studied multigene families from the standpoint of molecular genetics and phylogenetic history[13]. Apart from humans, β-globin gene or its variants (present in the cluster; Fig. 1) have been reported in various other mammals such as chimps, rhesus monkeys, baboons, cows, sheeps, dogs, rabbits and horses. Chimpanzees are the closest extant relatives of humans, having shared a common ancestry of 4–6 million years[14]. However, recent studies

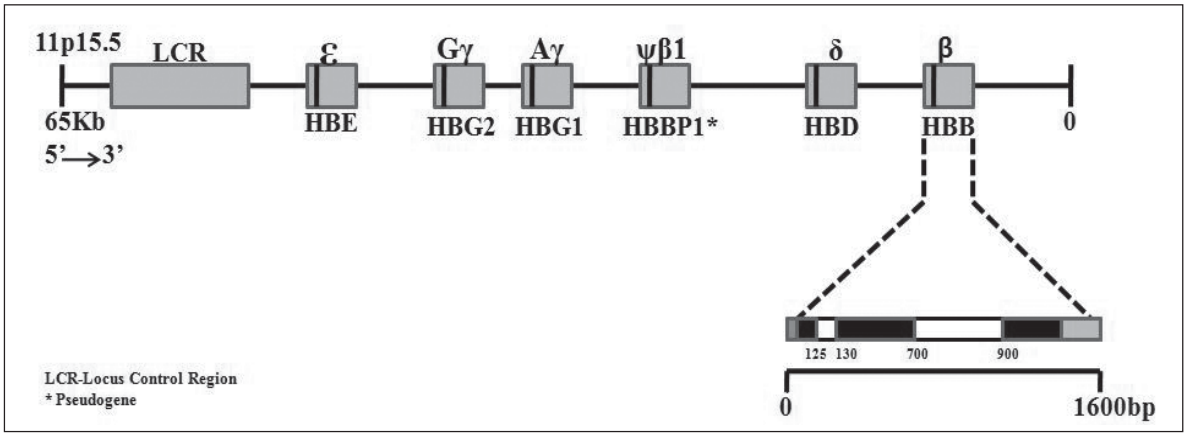

Fig. 1: Schematic representation of the human β-like globin gene cluster and the detailed structure of human β-globin gene. In the enlarged view of the β-globin gene, the dark colour boxes are the exons, the light are UTR's and the white colour boxes are introns. The first and second introns are 125–130 bp and 700–900 bp long, respectively. The other variants of β-globin gene present in the cluster are also represented.

on β-globin gene in chimpanzees have neither provided any evidence of mutations that confer resistance to malaria nor of long-term balancing selection at the genetic loci[15] unlike humans[16,17], although in rodents a separate study has reported presence of complex signatures of selection and gene conversion in the duplicated globin genes[18].

Herein, we report the results of a comparative genomic analyses of the β-globin gene in 11 different taxa; nine eutherian [*Homo sapiens* (human), *Macaca mulatta* (rhesus monkey), *Pan troglodytes* (chimpanzee), *Rattus norvegicus* (mouse), *Mus musculus* (rat), *Canis familiaris* (dog), *Bos taurus* (cow), *Equus caballus* (horse), *Oryctolagus cuniculus* (rabbit)], one dinosaurian (avian) *Gallus gallus* (chicken) and a neopterygii (marine) *Danio rerio* (fish), describing fundamental similarities and differences among taxa to enable better evolutionary understanding of functional β-globin gene (Fig. 2). Since the presence of shared conserved insertion or deletions (indels) in protein sequences is a special type of signature that shows considerable promise for phylogenetic inference[19] and also a species place on an evolutionary tree is a valuable predictor of the structure and function[20], we inferred the evolutionary relationships based on β-globin gene among all the studied 11 taxa through phylogenetic analysis.

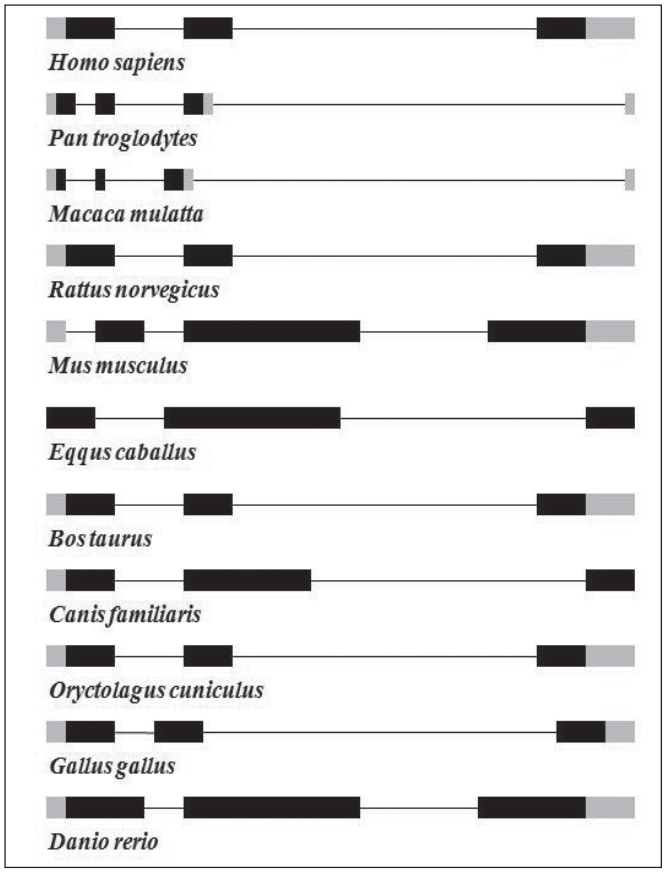

Fig 2: β-globin gene characterization in nine eutherian, one dinosaurian and one neopterygii taxa (not in scale). The light boxes are the untranslated regions (UTR's) and dark boxes are exons. Introns are displayed as black continuous line between two exons.

METHODS

Nucleotide and protein sequences of β-globin gene in nine eutherian (taxonomic classification given in Table 1a), *Homo sapiens* (GenBank Accession No. NC_000011.9), *Macaca mulatta* (GenBank Accession No. NC_007871.1), *Pan troglodytes* (GenBank Accession No. NC_006478.2), *Mus musculus* (GenBank Accession No. NC_095534.1), *Rattus norvegicus* (GenBank Accession No. NC_005100.2), *Bos taurus* (GenBank Accession No. NC_007313.3), *Equus caballus* (GenBank Accession No. NC_009150.2), *Canis familiaris* (GenBank Accession No. NC_006603.2), *Oryctolagus cuniculus* (GenBank Accession number NC_013669.1), a dinosaurian—*Gallus gallus* (GenBank Accession No. NC_006088.2) and a neopterygii—*Danio rerio* (GenBank Accession No. NC_007114.4) taxa were downloaded from National Centre for Biotechnology Information (NCBI) database (*http://www.ncbi.nlm.nih.gov/*) during February/March 2010. These sequences were subjected to different computational and statistical analyses for different variables; such as total gene, exon and intron lengths. Coding/non-coding ratios and mean coding percentages were computed in all 11 taxa. The computer software DNASTAR (DNASTAR Inc. Madison, USA, www.dnastar.com) was used to align the nucleotide sequences of β-globin gene followed by Clustal W algorithm. Phylogenetic tree based on neighbour-joining (NJ) method was constructed to infer the evolutionary relationships among various taxa at the β-globin gene level using phylogeny option in DNASTAR and validated with MEGA[21] computer program version 4.0 (*http://www.megasoftware.net/*). Length of each branch and bootstrapped values for each internal node were also estimated. To further establish the relationship of altered β-globin gene (variants) in 11 taxa, HomoloGene option of NCBI (*http://www.ncbi.nlm.nih.gov/sites/homologene*) was used and nucleotide sequences of these variants were downloaded. Three more taxa [*Sus scrofa* (pig) GeneID: 407066, *Ovis aries* (sheep) GeneID: 100049064 and *Papio anubis* (Baboon) GeneID: 100137310] were also included for phylogenetic analysis of variants, as the β-globin gene was found orthologous in these taxa. However, these taxa were not included in genetic characterization study owing to non-availability of positional information in their re-

Table 1 (a). Taxonomic positions of Eutherian, Neopterygii and Dinosauria organisms

Taxa/ Classification	*Homo sapiens*	*Pan troglodytes*	*Macaca mulatta*	*Rattus norvegicus*	*Mus musculus*	*Equus caballus*	*Bos taurus*	*Canis familiaris*	*Oryctolagus cuniculus*	*Gallus gallus*	*Danio rerio*
Domain	Eukaryota	Eukaryota	Eukaryota	Eukaryota	Eukaryota	Eukaryota	Eukaryota	Eukaryota	Eukaryota	Eukaryota	Eukaryota
Kingdom	Metazoa	Metazoa	Metazoa	Metazoa	Metazoa	Metazoa	Metazoa	Metazoa	Metazoa	Metazoa	Metazoa
Phyllum	Chordata	Chordata	Chordata	Chordata	Chordata	Chordata	Chordata	Chordata	Chordata	Chordata	Chordata
Clade	Craniata	Craniata	Craniata	Craniata	Craniata	Craniata	Craniata	Craniata	Craniata	Craniata	Craniata
Sub phylum	Euteleostomi	Euteleostomi	Euteleostomi	Euteleostomi	Euteleostomi	Euteleostomi	Euteleostomi	Euteleostomi	Euteleostomi	Euteleostomi	Euteleostomi
Class	Mammalia	Mammalia	Mammalia	Mammalia	Mammalia	Mammalia	Mammalia	Mammalia	Mammalia	Archosauria	Actinopterygii
Infra class	Eutheria	Eutheria	Eutheria	Eutheria	Eutheria	Eutheria	Eutheria	Eutheria	Eutheria	Dinosauria	Neopterygii
Super order	Euarchontoglires	Euarchontoglires	Euarchontoglires	Euarchontoglires	Euarchontoglires	Laurasiatheria	Laurasiatheria	Laurasiatheria	Euarchontoglires	Saurischia	Teleostei
Order	Primates	Primates	Primates	Glires	Glires	Perissodactyla	Cetartiodactyla	Carnivora	Glires	Theropoda	Ostariophysi
Suborder	Haplorrhini	Haplorrhini	Haplorrhini	Rodentia	Rodentia	Ruminantia	Caniformia	Lagomorpha	Coelurosauria	Cypriniformes	
Family	Hominidae	Hominidae	Hominidae	Murinae	Murinae	Equidae	Bovidae	Canidae	Leporidae	Phasianidae	Cyprinidae
Genus	Homo	Pan	Macaca	Rattus	Mus	Equus	Bos	Canis	Oryctolagus	Gallus	Danio
Species	sapiens	troglodytes	mulatta	norvegicus	musculus	caballus	taurus	familiaris	cuniculus	gallus	rerio

Table 1 (b). Properties of â-globin gene in 11 taxa

Taxa	NCBI Accession No.	Chr. No.	Exon1	Intron1	Exon 2	Intron 2	Exon 3	Intron 3	Exon 4	TGS (bp)	Ratio (E/I)	Coding region (bp)	Coding (%) age	NCR + UTRs
H. sapiens	NC_000011.9	11	142	130	223	850	261	–	–	1606	0.638	444	70.92	1162
P. troglodytes	NC_006478.2	11	267	130	223	850	265	13353	10	15098	0.053	456	59.60	14654
M. mulatta	NC_007871.1	14	267	130	223	859	265	21270	16	23030	0.034	462	59.92	22586
M. musculus	NT_095534.1	7	79	125	103	116	223	654	264	1564	0.142	444	66.36	4928
R. norvegicus	NC_005100.2	1	139	109	223	683	258	–	–	1412	0.780	445	71.77	968
B. taurus	NC_007313.3	15	138	128	223	900	254	–	–	1643	0.598	438	71.21	1205
E. caballus	NC_009150.2	7	92	128	223	625	129	–	–	1197	0.465	444	100.0	953
C. familiaris	NC_006603.2	21	185	128	223	597	129	–	–	1262	0.610	444	65.00	818
O. cuniculus	NC_013669.1	1	145	126	223	573	221	–	–	1288	0.842	444	75.38	844
G. gallus	NC_006088.2	1	137	108	223	987	177	–	–	1632	0.490	434	80.80	1188
D. rerio	NC_007114.4	3	135	100	223	99	245	–	–	802	3.030	447	74.12	355

spective genomes in the NCBI. Since amino acid sequences were available for these three taxa, these sequences were used for phylogenetic tree reconstruction (total 14 taxa were considered for a phylogenetic analysis using the amino acid sequences and only 11 taxa using the nucleotide sequences).

RESULTS & DISCUSSION

Results of the comparative genetic analyses revealed that the β-globin structure (5'- ε - Gγ – Aγ – δ – β - 3') has been subjected to many evolutionary changes and thus is unique in all taxa except human and chimp β-globin genes. Variations in the β-globin gene structure of many primates have been reported[13]. Moreover, evolutionary history of the beta globin gene in eutherian taxa is quite intriguing and perplexing. For example, it was reported that in higher primates, the γ-globin gene was duplicated before the divergence of Old World and New World monkeys, probably by an unequal homologous crossing over event mediated by LINE elements[9] and after duplication the γ-globin gene became fetally expressed as a direct result of the accumulation of sequence changes in the 5' flanking region such as Gγ and Aγ[9]. Studies between mammalian and marsupial's beta-like globin genes confirmed the hypothesis that a two-gene cluster, containing an embryonic- and an adult-expressed beta-like globin gene, existed in the most recent common ancestor of marsupials and eutherians[22]. This was evident from the analysis of RNA from embryos and neo-natals that indicated a switch from embryonic to adult gene expression occurring at the time of birth, coinciding with the transfer of the marsupial from a uterus to a pouch environment[22]. Moreover, this cluster of genes was proposed to have arisen by tandem duplication of ancestral beta-globin genes, with the first duplication occurring 200 to 155 MYBP just prior to a period in mammalian evolution when eutherians and marsupials diverged from a common ancestor[22].

The present study reveals several interesting evolutionary observations on the HBB gene among the eutherian, dinosaurian and neopterygii taxa. The β-globin gene in primate taxa (human, chimp and monkey) was found to have differed at the nucleotide sequence level but closely resembled at the amino acid level. Remarkably, *P. troglodytes* and *M. mulatta* shared same genetic compositions (similar size of exon 1, 2 and intron 1 in both taxa) with only 9 bp and 15 bp differences in intron 2 and in intron 3, respectively (Table 1b). Even the first exon, first and second introns were found to be similar in both human and chimp, suggesting closer genetic resemblances among two taxa (Table 1b). The presence of long extra third intron (14333 bp in *P. troglodytes*, 22259 bp in *M. mulatta*) and its absence in *H. sapiens* at the 3' end suggests that the β-globin gene in *P. troglodytes* and *M. mulatta* is able to accommodate more genetic changes compared to human β-globin gene. Presence of a long intron also contributes to a larger gene size (15098 bp in *P. troglodytes* and 23030 bp in *M. mulatta*) and eventually a larger genome size. It is also probable that an unequal homologous crossing over and duplication of γ–globin gene could have resulted in this long intron of *P. troglodytes* and *M. mulatta*. The exon-intron ratio computed in all the three primate taxa differed in *H. sapiens* (0.638), *P. troglodytes* (0.053) and *M. mulatta* (0.034) signifying that introns are generally longer compared to exons (Table 1b). Furthermore, observation of a significant positive correlation between the intron size and total gene size (data not shown) suggests that the size of gene has increased across taxa due to accumulation of non-coding nucleotides[11]. Like humans, in rodents also, new genes originated via multiple recombinational pathways[23]. In the present study, it was evident that *M. musculus* (rat) exhibits an extra 264 bp long exon and 654 bp long intron compared to *R. norvegicus* (mouse) gene composition (Table 1b). Organisms belonging to Laurasiatheria super order such as *B. taurus, C. familiaris* and *E. caballus* (Table 1a) showed similarity with conservation in the first intron and second exon (Table 1b). Interestingly, in all 10 taxa, exon 2 was found to be conserved except *M. musculus* (almost a difference of more than 100 bp). The size of the UTR also varied among taxa (1301 bp in chimp, 1316 bp in monkey, 182 bp in human, 4033 bp in rat, 176 bp in mouse, 177 bp in cow, 200 bp in horse, 93 bp in dog, 145 bp in rabbit, 93 bp in chicken and 156 bp in fish). It is widely known that UTRs help in a better gene regulation and are known to play crucial roles in the post-transcriptional regulation of gene expression, including modulation of the transport of mRNAs out of the nucleus and of translation efficiency, subcellular localization and stability[24]. Thus, the observed variation in the size of UTRs in evolutionarily close taxa signifies that the mechanism of globin gene regulation might be different in different taxa. However, if size of the UTR matters to the efficiency of gene regulation, it might be true that efficient gene regulation mechanism possibly exists in rat in comparison to other taxa. The observation on an almost-uniform exon-intron ratio over all taxa except the primates signifies evolutionary conservation of nucleotides belonging to exon and intron in these taxa. The results otherwise indicate that introns are major determinants on the size of the gene. However, in *D. rerio*, the observed highest exon/intron ratio (3.030) suggests the major contribution of coding nucleotides to gene size in marine or-

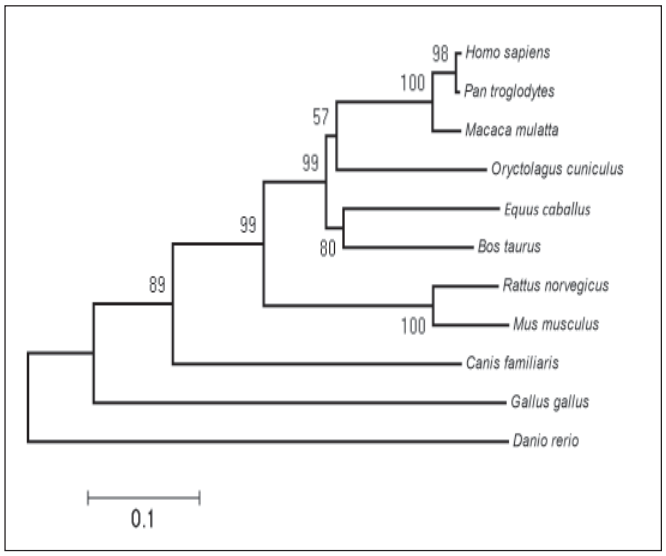

Fig. 3(a): Phylogenetic tree constructed on the basis of β-globin gene nucleotide sequences of nine eutherian, one dinosaurian and one neopterygii clade. The figures indicate the strength of each internal node as determined from 1000 bootstrapped analysis.

ganisms. Further, percentage of coding nucleotides in the β-globin gene varied across taxa; the highest was detected in *E. caballus* (100%) and lowest in *P. troglodytes* (59.6%) testifying the fact that the coding region of β-globin gene in *E. caballus* is absolutely conserved.

With a view to understand phylogenetic inter-relationship among all the 11 taxa at genetic level, an un-rooted neighbour-joining (NJ) tree was constructed (Fig. 3a). As expected, all primate taxa fall in one clade and rodents in another. Human and chimp were closer to each other. Moreover, *B. taurus* and *E. caballus* belonged to one clade with *O. cuniculus* slightly diverging from these two taxa. Surprisingly, *C. familiaris* belonging to the super order Laurasiatheria (to which *B. taurus* and *O. cuniculus* also belong) fell separately to Euarchontoglires and Laurasiatheria super order organisms. Being outgroups, *G. gallus* and *D. rerio* fell in separate branches. It is suggested that a number of events including gene conversion have occurred during evolution which has altered the clusters of β-globin gene in specific ways in different eutherian (mammalian) orders. These include insertion of repeat sequences, change in expression profile, gene duplication, gene fusion and gene loss or inactivation[9]. In order to verify this hypothesis, such altered genes showing homologies to β-globin gene in different taxa were selected (Table 2a) and a neighbour-joining phylogenetic tree using nucleotide sequences was constructed (Fig. 3b). Even the original β-globin gene nucleotide sequences present in all taxa were included for this phylogeny construction. It was apparent that all altered forms of β-globin gene in *B. taurus* formed a separate clade, but the actual β-globin gene in *B. taurus* was closer to β-globin gene variant (HBBA) in *O. aries*. This might be due to the fact that both *B. taurus* and *O. aries* belong to the bovine class. This result also signifies that *H. sapiens* and *P. troglodytes* are very closely related to each other at the HBG1 gene variants. It is already known that functionally both these primate taxa are closer to each other than to any other apes[25]; hence, sequence similarities in both these taxa at the gene and protein level is no surprise. It has already been reported that the δ-globin gene (HBD) of eutherian organism exhibits a propensity for recombinational exchange with the closely linked β-globin gene and has been independently converted by the β-globin gene in many lineages[26]. For example, in African elephant (*Loxodonta africana*), presence of a chimeric β/δ fusion gene was created by unequal crossing-over between misaligned HBD and HBB paralogs[26]. Hence, the knowledge of β-globin variants is certainly very important to have a clear picture

Table 2(a). Altered forms or variants of â-globin gene in 14 taxa

Taxa	Altered forms of β-globin gene						
H. sapiens	HBG 1	HBE 1	—	—	—	—	—
P. troglodytes	HBG 1	—	—	—	—	—	—
M. mulatta	—	—	—	—	—	—	—
R. norvegicus	HBG 1	HBE 1	LOC689064	MGC72973	—	—	—
M. musculus	HBB–B 1	HBB–BH 1	HBB–Y				
B. taurus	HBG 1	HBE 1	LOC781088	LOC788610	LOC781651	LOC783672	LOC781674
C. familiaris	Beta–A–globin	HBE 1	LOC480784	LOC609402	—	—	—
E. caballus	—	—	—	—	—	—	—
O. cuniculus	—	—	HBB2	—	—	—	—
S. scrofa	—	—	—	LOC100127142			
P. anubis	—	—	—	—	—	—	—
O. aries	—	—	—	LOC100134870	—	—	—
G. gallus	HBG 1	—	—	—	—	—	—
D. rerio	HBBE 3	LOC100000780	ba1 l	ba1	HBBE 2	Si_busm 1-118j2.5	b Y 119C3.2

Fig. 3(b): Phylogenetic tree constructed on the basis of β-globin gene variants nucleotide sequences of 12 eutherian, one dinosaurian and one neopterygii clade. The figures indicate the strength of each internal node as determined from 1000 bootstrapped analysis.

of the evolution of this gene in all taxa. Both *G. gallus* and *D. rerio* being outgroups, formed separate clades even with their variants. It has been suggested that major gene duplications have given rise to the paralogous beta globin genes or variants that are associated with significant evolutionary rate variation among gene lineages[27]. Further, genes arising from more recent gene duplications (e.g. tandem duplications within lineages) do not appear to differ greatly in rate[27]. Such a pattern also reflects a complex interplay of evolutionary forces where natural selection for diversifying paralogous functions and lineage-specific effects contribute to rate variation on a long-term basis, while gene conversion tends to increase sequence similarity[27]. Moreover, gene conversion effects appear to be stronger on recent gene duplicates, as their sequences are highly similar[27].

It is known that in Old World monkeys, apes and humans, both the γ1 and γ2-globin genes are functional but the expression of γ1 gene is three-fold higher than that of the γ2 gene; whereas in New World monkeys only one-globin gene is functional, usually γ2[9]. This has been explained by different models of gene family evolution that

explain the mechanism whereby gene copies created by gene duplications are maintained and diverge in functions[28]. Two models have been proposed to explain this phenomena; one that the nonsynonymous substitutions increase following gene duplication and preserve the duplicates through positive selection[28]. Another alternative model, the duplication–degeneration–complementation (DDC) model, does not explicitly require the action of positive Darwinian selection for the maintenance of duplicated gene copies, although purifying selection is assumed to continue to act on both copies[27]. Since gene duplication and divergence play significant role in eukaryotic protein network evolution, pointing to either discontinuous likely to be adaptive or continuous unlikely to be adaptive taxonomical differences, inferences at the protein level of the beta locus are as vital as inferences at the gene level[29]. Therefore, apart from characterization at the gene level, we also considered the protein sequences in all the 14 taxa (including sheep, baboon and pig) to ascertain functional level changes across taxa. Results of the β-globin protein alignment (147 codons) in all 14 taxa were found absolutely similar in human and chimp. Monkey and baboon sequences were found to be similar with only three substitutions at codon position 10, 14 and 53. All the four primate taxa were dissimilar at only nine codon positions (Table 2b), signifying that plenty of conservation exists at the β-globin protein level in closely related species, unlike the genetic level. Interestingly, contrary to the traditional belief[10], we found that neither glutamic acid (E) nor valine (V) was present at the sixth amino acid position in human β-globin protein sequence. Instead, proline (P) was observed at the sixth codon position. Similar to human β-globin protein, proline was also observed at the sixth position for chimp, monkey, baboon, dog and pig (Table 2c). Absolute protein homology was also seen for the rodent taxa. Interestingly, in bovine group (cow and sheep) the first and second codons were found missing (Table 2c). To further assess the evolutionary relationship among all the 14 taxa, an unrooted phylogenetic tree was constructed based on the protein sequences and inferences were drawn. Human and chimp fell in a single clade neighboured by monkey and baboon with chicken and fish as outgroups (Fig. 3c).

Table 2(b). Variations at nine codon positions in primates

Taxa	10	14	44	51	53	77	88	105	126
H. sapiens	S	A	E	T	D	A	T	R	P
P. troglodytes	S	A	E	T	D	A	T	R	P
M. mulatta	T	T	D	S	D	N	Q	K	Q
P. anubis	N	A	D	S	A	N	Q	K	Q

Table 2(c). Depiction of the sixth codon in all the 14 taxa

Taxa	1	2	3	4	5	6	7	8	9	10
H. sapiens	M	V	H	L	T	**P**	E	E	K	S
P. troglodytes	M	V	H	L	T	**P**	E	E	K	S
M. mulatta	M	V	H	L	T	**P**	E	E	K	T
M. musculus	M	V	H	L	T	**D**	A	E	K	A
R. norvegicus	M	V	H	L	T	**D**	A	E	K	A
B. taurus	–	–	M	L	T	**A**	E	E	K	A
C. familiaris	M	V	H	L	T	**P**	E	E	K	S
E. caballus	M	V	Q	L	S	**G**	E	E	K	A
S. scrofa	M	V	H	L	T	**P**	E	E	K	N
P. anubis	M	V	H	L	T	**P**	E	E	K	N
O. aries	–	–	M	L	T	**A**	E	E	K	A
O. cuniculus	M	V	H	L	S	**S**	E	E	K	S
G. gallus	M	V	H	W	S	**A**	E	E	K	Q
D. rerio	M	V	E	W	T	**D**	A	E	R	T

The study therefore not only provides a comprehensive overview of the β-globin gene in evolutionarily closely related taxa and establishes the relationship among them based on the beta globin gene but also among the unrelated taxa. Absolute conservedness between human and chimp β-globin protein is certainly a highlight of the study. However, absence of evidence for a genomic signature of malaria at β-globin gene in chimpanzee[15] challenges the correlation of β-globin gene with mild and severe malaria in human, since both proteins are absolutely similar. Either, the polymorphisms in β-globin gene in humans are no longer under the effect of balancing selection to impart protectiveness against malaria or that the β-globin gene is not so adaptive in chimpanzees since malaria is less detri-

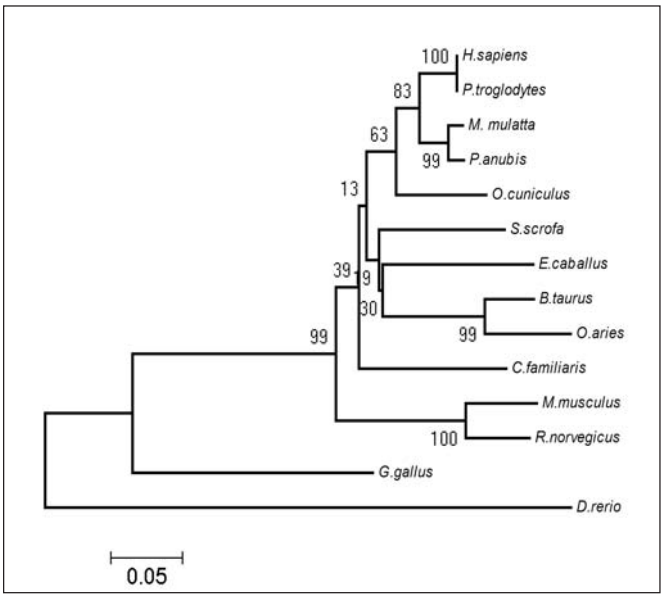

Fig. 3(c): Phylogenetic tree constructed on the basis of β-globin gene protein sequences of twelve eutherian, one dinosaurian and one neopterygii clade. The figures indicate the strength of each internal node as determined from 1000 bootstrapped analysis.

mental in them[15]. Alternatively, it could be possible that chimpanzees might utilize mechanisms different from humans for protection against malaria[15]. Not all variants in humans are under the effect of balancing selection, as a recent study suggests that the beta globin recombinational hotspot around HbC reduces the effects of strong selection[30]. Evidences of natural selection suggest that gene duplications in paralogous copies of beta-globin genes evolve under a non-episodic process of functional divergence[28]. Considering the fact that β-globin gene variant (HBS) in humans is responsible for sickle-cell anemia and highly associated with severe malaria; more in-depth evolutionary and genetic diversity studies from field samples are needed for a better understanding of intricate relationship between infectious diseases and medically important genes.

In conclusion, the present study clearly indicates that genetic changes have occurred during evolution of different mammalian and closely related taxa in the globin gene. Although controversial[31] this gene has been shown to be correlated with *P. falciparum* malaria in humans[32]. The evolutionary genetic studies performed here could thus not only be useful in testing the hypothesis that whether this particular gene has any role in *P. falciparum* malaria infection, but also could ascertain, with further studies in a malaria endemic region, if this gene is under evolutionary pressure in humans differentially in *P. falciparum* endemic and non-endemic areas.

ACKNOWLEDGEMENTS

Gauri Awasthi is a Senior Research Fellow of the Indian Council of Medical Research (ICMR) and Garima Srivastava spent summer internship in the EGB Lab. The authors thank the Director Incharge of NIMR for providing facilities. Dr Aparup Das thanks the ICMR for intramural funding.

REFERENCES

1. Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet* 2005; *39:* 309–38.
2. Nobrega MA, Pennacchio LA. Comparative genomic analysis as a tool for biological discovery. *J Physiol* 2004; *554:* 31–9.
3. Awasthi G, Dash AP, Das A. Characterization and evolutionary analysis of human CD36 gene. *Indian J Med Res* 2008; *129*: 534–41.
4. Awasthi G, Singh S, Dash AP, Das A. Genetic characterization and evolutionary inference of TNF-α with computational analyses. *Braz J Infect Dis* 2008; *12*(5): 374–9.
5. Awasthi G, Dash AP, Das A. Evolutionary insights into Duffy gene in mammalian taxa with comparative genetic analysis. *J Vector Borne Dis* 2009; *46*: 230–6.
6. Pasvol G. Protective hemoglobinopathies and *Plasmodium falciparum* transmission. *Nat Genet* 2010; *42*: 284–5.
7. Bindon J. Natural selection and adaptation 2003: sickle-cell. Available from: *http://www.as.ua.edu/ant/bindon/ant475/Sicklecell/Sicklecell.pdf.* [accessed on February 11, 2004].
8. Das SK, Talukder G. Beta globin gene and related diseases: A review. *Int J Hum Genet* 2002; *2*: 139–52.
9. Chiu CH, Gregoire L, Gumucio DL, Muniz JAPC, Lancaster WD, Goodman M. Model for the fetal recruitment of simian γ-globin genes based on findings from two new world monkeys *Cebus apella* and *Callithrix jacchus* (Platyrrhini, Primates). *J Exp Zool* Part B: *Mol Dev Evol* 1999; *285:* 27– 40.
10. Koch AA, Olney RS, Yang Q. Sickle hemoglobin allele and sickle cell disease. *Am J Epi* 2002; *9:* 839–45.
11. Sironi M, Menozzi G, Comi GP, Bresolin N, Cogliani R, Pozzoli U. Fixation of conserved sequences shapes human intron size and influences transposon-insertion dynamics. *Trends Genet* 2005; *21:* 484–8.
12. Sabeti P. Natural selection: uncovering mechanisms of evolutionary adaptation to infectious disease. *Nat Edu* 2008; *1*(1).
13. Opazo JC, Hoffmann HG, Storz JF. Differential loss of embryonic globin genes during the radiation of placental mammals. *Proc Natl Acad Sci USA* 2008; *105:* 12950–5.
14. Chen FC, Li WH. Genomic divergence between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* 2001; *68*: 444–56.
15. MacFie TS, Nerrienet E, Bontrop RE, Mundy NI. The action of *falciparum* malaria on the human and chimpanzee genomes compared: Absence of evidence for a genomic signature of malaria at HBB and G6PD in three subspecies of chimpanzee. *Infect Genet Evol* 2009; *9:* 1248–52.
16. May J, Evans JA, Timmann C, Ehmen C, Busch W, Thye T, Agbenyega T, Horstmann RD. Hemoglobin variants and disease manifestations in severe falciparum malaria. *JAMA* 2007, *297:* 2220–6.
17. Williams TN, Mwangi TW, Roberts DJ, Alexander ND, Weatherall DJ, Wambua S, Kortok M, Snow RW, Marsh K. An immune basis for malaria protection by the sickle-cell trait. *PLOS Med* 2005; *2:* e128.
18. Storz JF, Baze M, Waite JL, Hoffmann FG, Opazo JC, Hayes JP. Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. *Genet* 2007; *177:* 481–500.
19. Gupta RS. Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev* 1998; *62:* 1092–2172.
20. Zhaurova, K. Genomes of other organisms: DNA barcoding and metagenomics. *Nat Edu* 2008; *1*(1).
21. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) Software version 4.0. *Mol Biol Evo* 2007; *24:* 1596–9.
22. 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:* 1012–22.
23. Hoffmann FG, Opazo JC, Storz JF. New genes originated via multiple recombinational pathways in the β-globin gene family of rodents. *Mol Biol Evol* 2008; *25:* 2589–600.
24. Bashirullah A, Cooperstock RL, Lipshitz HD. Spatial and temporal control of RNA stability. *Proc Natl Acad Sci USA* 2001;

98: 7025–8.
25. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H *et al*. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 2005; *309:* 1850–4.
26. Opazo JC, Sloan AM, Campbell KL, Storz JF. Origin and ascendancy of a chimeric fusion gene: the β/δ-globin gene of paenungulate mammals. *Mol Biol Evol* 2009; *26*: 1469–78.
27. Aguileta G, Bielawski JP, Yang Z. Evolutionary rate variation among vertebrate beta globin genes: implications for dating gene family duplication events. *Genet* 2006; *380:* 21–9.
28. Ratmann O, Jørgens, Hinkley T, Stumpf M, Richardson S, Wiuf C. Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*. *PLoS Comput Biol* 2007; *3*(11): e230.
29. Aguileta G, Bielawski JP, Yang Z. Gene conversion and functional divergence in the beta-globin gene family. *J Mol Evol* 2004; *59:* 177–89.
30. Wood ET, Stover DA, Slatkin M, Nachman MW, Hammer MF. The β-globin recombinational hotspot reduces the effects of strong selection around HbC, a recently arisen mutation providing resistance to malaria. *Am J Hum Genet* 2005; *77:* 637–42.
31. Naka I, Ohashi J, Nuchnoi P, Hananantachai H, Looareesuwan S, Tokunaga K, Patarapotikul J. Lack of association of the HbE variant with protection from cerebral malaria in Thailand. *Biochem Genet* 2008; *46:* 708–11.
32. Naka I, Ohashi J, Patarapotikul J, Hananantachai H, Looareesuwan S, Tokunaga K. Genetic variants of beta-globin gene in Thai malaria patients. *Southeast Asian J Trop Med Public Health* 2003; *34:* 29–31.

Correspondence to: Dr Aparup Das, Evolutionary Genomics and Bioinformatics Laboratory, National Institute of Malaria Research, Sector 8, Dwarka, New Delhi–10 077, India.
E-mail: aparup@mrcindia.org

Received: 1 September 2010 *Accepted in revised form:* 20 January 2011