# Evolution of Duplicated $\beta$-Globin Genes and the Structural Basis of Hemoglobin Isoform Differentiation in *Mus*

*Amy M. Runck, Hideaki Moriyama, and Jay F. Storz*

School of Biological Sciences, University of Nebraska

The functional diversification of multigene families may be strongly influenced by mechanisms of concerted evolution such as interparalog gene conversion. The $\beta$-globin gene family of house mice (genus *Mus*) represents an especially promising system for evaluating the effects of gene conversion on the functional divergence of duplicated genes. Whereas the majority of mammalian species possess tandemly duplicated copies of the adult $\beta$-globin gene that are identical in sequence, natural populations of house mice are often polymorphic for distinct two-locus haplotypes that differ in levels of functional divergence between duplicated $\beta$-globin genes, HBB-T1 and HBB-T2. Here, we use a phylogenetic approach to unravel the complex evolutionary history of the HBB-T1 and HBB-T2 paralogs in a taxonomically diverse set of species in the genus *Mus*. The main objectives of this study were 1) to reconstruct the evolutionary history of the different HBB haplotypes of house mice, 2) to assess the role of recombinational exchange between HBB-T1 and HBB-T2 in promoting concerted evolution, 3) to assess the role of recombinational exchange between HBB-T1 and HBB-T2 in creating chimeric genes, and 4) to assess the structural basis of hemoglobin isoform differentiation in species that possess distinct HBB paralogs. Results of our phylogenetic survey revealed that the HBB-T1 and HBB-T2 genes in different species of *Mus* exhibit the full range of evolutionary outcomes with respect to levels of interparalog divergence. At one end of the spectrum, the two identical HBB paralogs on the $Hbb^s$ haplotype (shared by *Mus domesticus*, *Mus musculus*, and *Mus spretus*) represent a classic example of concerted evolution. At the other end of the spectrum, the two distinct HBB paralogs on the $Hbb^d$, $Hbb^p$, $Hbb^{w1}$, and $Hbb^{w2}$ haplotypes (shared by multiple species in the subgenus *Mus*) show no trace of gene conversion and are distinguished by a number of functionally important amino acid substitutions. Because the possession of distinct HBB paralogs expands the repertoire of functionally distinct hemoglobin isoforms that can be synthesized during fetal development and postnatal life, variation in the level of functional divergence between HBB-T1 and HBB-T2 may underlie important physiological variation within and among species.

## Introduction

An important goal of evolutionary genomics is to identify mechanisms responsible for the initial retention and subsequent functional divergence of duplicated genes. In some cases, identical gene duplicates may be retained in the genome because selection favors the production of increased quantities of the encoded RNA or protein (Sugino and Innan 2006). In other cases, duplicated genes may acquire novel functions or partition ancestral functions of the single-copy progenitor gene (Ohno 1970; Lynch et al. 2001; Zhang 2003; Lynch and Katju 2004). Each of these evolutionary outcomes may be strongly influenced by mechanisms of concerted evolution such as interparalog gene conversion. In some cases, concerted evolution may facilitate the spread of an adaptive mutation to multiple members of a multigene family (Mano and Innan 2008). By contrast, in cases where selection favors some type of division of labor between the products of functionally distinct paralogs, the homogenizing effects of gene conversion may counteract adaptive sequence divergence (Innan 2003; Teshima and Innan 2004, 2008). Finally, gene conversion between distinct paralogs can also create chimeric genes with novel functions or expression patterns.

The $\beta$-globin gene family of house mice (genus *Mus*) holds much promise as a model system for understanding the evolutionary dynamics of duplicated genes. First, the $\beta$-globin gene cluster of *Mus musculus* is an intensively studied system from the standpoints of molecular genetics (Leder et al. 1980) and functional genomics (Hardies et al. 1984; Hill et al. 1984; Shehee et al. 1989; Moon and Ley 1990; Hardison and Miller 1993; Hoffmann et al. 2008a). Secondly, natural populations of house mice are often polymorphic for distinct two-locus $\beta$-globin haplotypes that differ in levels of amino acid divergence between the two paralogs. The two tandemly duplicated $\beta$-globin genes of house mice, HBB-T1 and HBB-T2, encode the $\beta$-chain subunit of adult hemoglobin (Hb) and are separated by ~12–15 kb on Chromosome 7 (Hoffmann et al. 2008a; Sato et al. 2008).

Five main classes of HBB haplotypes have been characterized in house mice: $Hbb^d$, $Hbb^p$, $Hbb^s$, $Hbb^{w1}$, and $Hbb^{w2}$. Figure 1*A* provides a graphical summary of amino acid variation within and between the HBB genes on each of these different haplotype backgrounds. As illustrated in figure 1*A*, the $Hbb^d$ haplotype harbors two distinct HBB paralogs that are distinguished from one another by nine amino acid substitutions. The more highly expressed HBB-T1 gene encodes the $\beta$-chains of the major Hb isoform (isoHb), $d_{maj}$, whereas HBB-T2 encodes the $\beta$-chains of the minor isoHb, $d_{min}$ (Hutton et al. 1962; Gilman 1974; Whitney 1977). The $Hbb^{w1}$ haplotype also harbors two distinct HBB paralogs, and the two sequences are distinguished from one another by 12 amino acid substitutions. Similar to the case with the $Hbb^d$ haplotype, the $\beta$-chain subunits of the major and minor isoHbs are encoded by HBB-T1 and HBB-T2, respectively. The $Hbb^d$ and $Hbb^{w1}$ haplotypes are distinguished by three amino acid substitutions at HBB-T1 and two amino acid substitutions at HBB-T2. As illustrated in figure 1*B*, intergenic recombination in $Hbb^d/Hbb^{w1}$ heterozygotes has produced one recombinant chromosome ($Hbb^p$) that carries an HBB-T1 allele derived from $Hbb^d$ and an HBB-T2 allele derived from $Hbb^{w1}$, and another recombinant chromosome ($Hbb^{w2}$) that carries an

Key words: Concerted evolution, gene duplication, gene family evolution, gene conversion, hemoglobin, *Mus*.

E-mail: jstorz2@unl.edu.

*Mol. Biol. Evol.* 26(11):2521–2532. 2009
doi:10.1093/molbev/msp165
Advance Access publication August 12, 2009

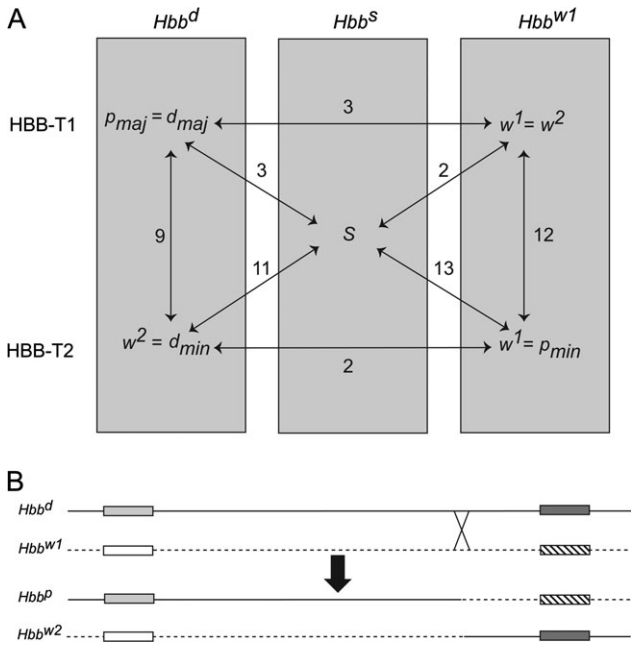


FIG. 1.—(*A*) Amino acid variation within and between HBB-T1 and HBB-T2 paralogs on the $Hbb^d$, $Hbb^s$, $Hbb^{w1}$, $Hbb^p$, and $Hbb^{w2}$ haplotypes of house mice. Numbers refer to pairwise amino acid differences between alleles of the same gene or between paralogous HBB-T1 and HBB-T2 sequences. (*B*) Diagram showing the inferred intergenic recombination event that gave rise to the $Hbb^p$ and $Hbb^{w2}$ haplotypes in house mice. The HBB-T1 and HBB-T2 paralogs are depicted as boxes separated by ~14.9 kb of noncoding DNA on *Mus* Chromosome 7. The inferred recombination break point, indicated by the X, is located ~2.6 kb upstream of the initiation codon of HBB-T2 (Sato et al. 2008).

HBB-T1 allele derived from $Hbb^{w1}$ and an HBB-T2 allele derived from $Hbb^d$ (Ueda et al. 1999; Sato et al. 2006, 2008). In contrast to the $Hbb^d$, $Hbb^p$, $Hbb^{w1}$, and $Hbb^{w2}$ haplotypes, the $Hbb^s$ haplotype harbors two HBB paralogs that are identical in sequence due to a history of HBB-T1 → HBB-T2 gene conversion (Erhart et al. 1985; Storz, Baze, et al. 2007). Consequently, mice that carry two copies of $Hbb^s$ synthesize a single $\beta$-chain isoHb during postnatal life.

Electrophoretic surveys of $\beta$-globin polymorphism in natural populations of *Mus domesticus* and *M. musculus* have revealed that the $Hbb^d$ and $Hbb^s$ haplotypes are nearly always present at intermediate frequencies in mice sampled from disparate geographic localities across Europe and the Americas, and it has been suggested that the polymorphism is maintained by overdominance of fitness or some other form of balancing selection (Selander and Yang 1969; Selander et al. 1969; Berry and Murphy 1970; Wheeler and Selander 1972; Myers 1974; Berry and Peters 1975, 1977; Berry 1978; Berry et al. 1978; Gilman 1979; Petras and Topping 1983). Consistent with this hypothesis, levels of nucleotide variation and linkage disequilibrium in wild mice indicate that the $Hbb^d$ and $Hbb^s$ haplotypes have been maintained as a long-term balanced polymorphism (Storz, Baze, et al. 2007). Sequence data from additional species of *Mus* are needed to elucidate the evolutionary origins and antiquity of these different $\beta$-globin haplotypes.

Although there does not appear to be any segregating variation in HBB copy number in house mice, there is extensive variation in levels of amino acid divergence between the two HBB paralogs on each of the different haplotype backgrounds (fig. 1*A*). As a result of gene conversion between HBB-T1 and HBB-T2, the $Hbb^s$ haplotype has essentially reverted to an unduplicated state. This pattern of concerted evolution is typical of the globin gene families in mammals (Hardison and Gelinas 1986; Hardison and Miller 1993; Hoffmann et al. 2008a, 2008b; Opazo et al. 2008a, 2008b; Storz et al. 2008, 2009; Opazo et al. 2009). The majority of mammals possess two or more tandemly duplicated HBB genes, and the paralogous copies are typically identical in sequence (Opazo et al. 2008a, 2008b). Thus, the $Hbb^s$ haplotype, with its two identical HBB paralogs, is typical of the situation observed in most mammals, whereas the $Hbb^d$ and $Hbb^p$ haplotypes, with their two highly divergent HBB paralogs, are quite unusual. Sequence data from the HBB paralogs of additional *Mus* species are needed to determine which pattern is the norm in this particular group.

In cases where tandem gene duplicates have escaped from concerted evolution, as in the case of the two distinct HBB paralogs on all haplotypes other than $Hbb^s$, recombinational exchanges between the two paralogs can produce novel chimeric sequences (Zangenberg et al. 1995; Storz, Sabatino, et al. 2007; von Salome et al. 2007; Hoffmann et al. 2008b; Storz and Kelly 2008; Opazo et al. 2009). For example, Gilman (1972, 1974) reported that *Mus caroli* possesses a single, chimeric HBB gene that is characterized by T2-like sequence at the 5′ end and T1-like sequence at the 3′ end. It was hypothesized that the unusual HBB of *M. caroli* is a chimeric fusion gene that was produced by unequal crossing-over between distinct HBB-T1 and HBB-T2 parent genes. In the genus *Mus*, it thus appears that recombinational exchanges between tandemly duplicated HBB genes have produced a variety of different evolutionary outcomes, in some cases, promoting concerted evolution, as in the case of the $Hbb^s$ haplotype, and in other cases, creating novel, chimeric genes, as in the case of *M. caroli*.

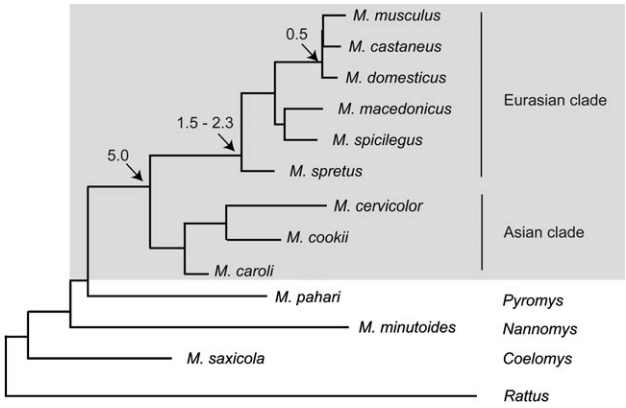


FIG. 2.—Inferred phylogenetic relationships of 12 species in the genus *Mus*. The clade shown in the gray box contains members of the subgenus *Mus*. Numbers above nodes are estimated divergence dates in millions of years before present; based on data from Lundrigan et al. (2002).

Here, we use a phylogenetic approach to unravel the complex evolutionary history of the HBB-T1 and HBB-T2 paralogs in a taxonomically diverse set of mouse species in the genus *Mus*. This set of species includes house mice of the Eurasian *musculus* group that carry the $Hbb^d$, $Hbb^p$, $Hbb^{w1}$, $Hbb^{w2}$, and $Hbb^s$ haplotypes as well as representatives of three other subgenera of *Mus* (*Coelomys*, *Nannomys*, and *Pyromys*). The main objectives of this study were 1) to reconstruct the evolutionary history of the different HBB haplotypes of house mice, 2) to assess the role of recombinational exchange between HBB-T1 and HBB-T2 in promoting concerted evolution, 3) to assess the role of recombinational exchange between HBB-T1 and HBB-T2 in creating chimeric genes; and 4) to identify the structural basis of isoHb differentiation in species that possess distinct HBB paralogs.

## Materials and Methods

### Sampling

Our phylogenetic survey of nucleotide variation in the HBB-T1 and HBB-T2 genes included 12 species in the genus *Mus*. This set of species included nine members of the subgenus *Mus* (*M. caroli*, *M. castaneus*, *M. cookii*, *Mus cervicolor*, *M. domesticus*, *M. macedonicus*, *M. musculus*, *M. spicilegus*, and *M. spretus*) and single representatives of three other subgenera: *Coelomys* (*M. pahari*), *Nannomys* (*M. minutoides*), and *Pyromys* (*M. saxicola*; fig. 2). We cloned and sequenced the HBB-T1 and HBB-T2 genes in each of the 12 species listed above, and we retrieved additional sequences from public databases. HBB sequences from the $Hbb^d$ and $Hbb^s$ haplotypes of *M. domesticus* were taken from the study of Storz, Baze, et al. (2007; GenBank accession numbers EF605358, EF605359, EF605487, and EF605488). We also retrieved publicly available sequences for the following haplotypes of *M. musculus*: $Hbb^s$ from the C57BL/6J strain (NC_000073), $Hbb^d$ from the BALB/cByJ strain (NT_095534), $Hbb^p$ from the MSM/Ms strain (AB020015, AB020016, and AB189411–AB189418), $Hbb^{w1}$ from the BALB/c-$Hbb^{w1}$ congenic strain (AB020013, AB020014, AB189420-189427), and $Hbb^{w2}$ from the BALB/c-$HBB^{w2}$ congenic strain (AB364474 and AB364475). We used the HBB-T1 and HBB-T4 genes of *Rattus* as outgroup sequences (NC_005100), as these genes are 1:1 orthologs of the HBB-T1 and HBB-T2 genes in *Mus*, respectively (Hoffmann et al. 2008a). Tissue samples from all *Mus* species other than *M. domesticus* were kindly provided by P. Tucker (University of Michigan).

### Molecular Cloning and Sequencing

We designed paralog-specific primer sets for HBB-T1 and HBB-T2 by using a multispecies sequence alignment of orthologous genes from *Rattus*, *Mus*, and deer mouse (*Peromyscus maniculatus*). Each of the two locus-specific primer combinations (HBB-T1F 5′-CAATTCAGTAGTTGATTGAGC and HBB-T1R 5′-CAAGCTATGTTATTGGTGCAA) and (HBB-T2F 5′-GTG GCT TAC TGC TTG CTG TCC and HBB-T2R 5′-CTC TTT GGT ATT TTA TT CTT G) amplified a ~1.8-kb DNA fragment that spanned the complete coding region of each HBB paralog in addition to 338 bp of 5′-flanking sequence and 290 bp of 3′-flanking sequence. Amplification of the two paralogs was conducted using the Roche High Fidelity PCR System (Roche Diagnostics, Indianapolis, IN). We used the following thermal cycling protocol: 94 °C (120 s) initial denaturing (94 °C [30 s], 44 °C–53 °C [30 s], 72 °C [105 s]) 35 cycles and a final extension of 72 °C (7 min). PCR products were cloned into pCR4-TOPO vector following the manufacturer's protocols (Invitrogen, Carlsbad, CA). For each species, we sequenced a total of 8–10 colonies per gene using the vector primers T3 and T7 (54 °C annealing). In several species, we recovered distinct alleles at one or both HBB paralogs. In such cases, the cloning of diploid PCR products allowed us to determine the exact haplotype phase for all heterozygous sites. Sequences were run on an ABI 3730 capillary sequencer using Big Dye chemistry (Applied Biosystems, Foster City, CA). Sequences were deposited in GenBank under the accession numbers GQ250367–GQ250397.

### Alignment

Sequences were assembled into contigs using Sequencher (Gene Codes, Ann Arbor, MI) and were aligned using ClustalX (Thompson et al. 1997) with manual adjustment. Intron 2 of HBB-T2 was manually aligned because this gene region is characterized by an extremely high density of insertions and deletions (Erhart et al. 1985; Sato et al. 2006, 2008; Storz, Baze, et al. 2007).

### Phylogenetic Reconstruction of HBB Gene Trees and Detection of Interparalog Gene Conversion

Because gene conversion is primarily restricted to the coding regions of mammalian globin genes, reliable inferences about orthologous relationships require an examination of flanking sequence or intronic sequence (Hardison and Gelinas 1986; Hardison and Miller 1993; Hoffmann

et al. 2008a, 2008b; Opazo et al. 2008a, 2008b; Storz et al. 2008; Opazo et al. 2009). We therefore conducted phylogenetic reconstructions that were based on four different partitions of the alignment: 5′-flanking sequence (338 bp), coding sequence (441 bp), intron 2 sequence (773 bp), and 3′-flanking sequence (290 bp). We inferred phylogenetic relationships among HBB-T1 and HBB-T2 sequences in a maximum likelihood framework using Treefinder, version April 2008 (Jobb et al. 2004), and assessed support for the nodes with 1,000 bootstrap pseudoreplicates. The Bayesian Information Criterion in Treefinder was used to select the best fitting model of nucleotide substitution for each data partition. Phylogenetic reconstructions of the flanking and coding regions were conducted using the HKY model of nucleotide substitution (Hasegawa et al. 1985) in which rate variation conformed to a discrete gamma distribution (HKY + $\gamma$). Phylogenetic reconstructions of intron 2 sequences were conducted using the TN93 model (Tamura and Nei 1993) with a gamma distribution (TN93 + $\gamma$). Both global and simple tree searches were conducted. Global searches for each data partition were conducted using seven different starting trees.

### Ancestral Sequence Reconstruction and Analysis of Selective Constraints

To reconstruct amino acid sequences of HBB-T1 and HBB-T2 in the common ancestor of *Mus*, we used the maximum likelihood approach of Yang et al. (1995) and Koshi and Goldstein (1996). Specifically, we reconstructed ancestral sequences using the 3 × 4 codon model in PAML 4 (Yang 2007). Ancestral reconstructions were conducted separately for each paralog and sequences that harbored ectopic conversion tracts were excluded from the analysis. Marginal posterior probabilities were calculated for each reconstructed residue position. We also applied a codon substitution model to the same alignment of unconverted HBB-T1 and HBB-T2 sequences to estimate relative rates of synonymous and nonsynonymous substitution. This allowed us to evaluate possible differences in selective constraint between the two paralogs.

We used an alignment of adult $\beta$-globin sequences from 51 species of mammals to characterize site-specific variation in structural constraint across the $\beta$-globin polypeptide. Conservation scores were calculated at each amino acid residue using the method of Valdar (2002) with a modified PET91 distance matrix. For visualization purposes, we used the Pymol program (DeLano, http://www.pymol.org) to project color-coded conservation scores onto the 3D structure of the Hb molecule.

### Structural Modeling

To characterize physicochemical differences between the $\beta$-chain products of HBB-T1 and HBB-T2 for each species, we used an in silico approach (Gasteiger et al. 2003) to compute the isoelectric point (*pI*), the inhibition constant, $K_i$ (a measure of the free energy of oxygen binding), and the grand average of hydropathicity (a measure of hydrophobicity; Kyte and Doolittle 1982). For each pair of $\beta$-chain isoHbs, we calculated a normalized 3D distance based on calculated values of *pI*, $K_i$, and hydrophobicity. To characterize the structural basis of isoHb differentiation, we used SWISS-Model (Arnold et al. 2006) to map observed amino acid substitutions onto a 3D homology-based model of *Mus* Hb. The D chain of 1JEB (Kidd et al. 2001) was used as a template for all models.

## Results

### Phylogenetic Relationships of HBB-T1 and HBB-T2 Sequences

We successfully cloned two adult $\beta$-globin genes from each of the 12 species of *Mus*, including *M. caroli*, which was previously thought to have only one HBB gene copy (Gilman 1972, 1974). Remarkably, the $Hbb^s$ haplotype (previously characterized in the C57BL/6J inbred strain) was shared between *M. domesticus*, *M. musculus*, and *M. spretus*, the $Hbb^d$ haplotype (previously characterized in the BALB/cByJ inbred strain) was shared between *M. castaneus*, *M. domesticus*, *M. macedonicus*, *M. musculus*, and *M. spicilegus*, and the $Hbb^p$ haplotype (previously characterized in the AU/SsJ inbred strain) was shared between *M. castaneus* and *M. musculus*.

Phylogenetic reconstructions of 5′- and 3′-flanking sequences grouped HBB-T1 and HBB-T2 into two reciprocally monophyletic groups (fig. 3). The sole exception was the 5′-flanking sequence of the HBB-T2 gene in *M. pahari*. This sequence was not nested within the clade of HBB-T2 sequences from the other species, although it was more closely allied with the HBB-T2 clade than with the HBB-T1 clade. Closer inspection revealed perfect sequence identity between the HBB-T1 and HBB-T2 genes of *M. pahari* in the 104 bp immediately upstream of the start codon. In *M. pahari*, it appears that the 5′-flanking sequence of HBB-T2 has been partially converted by HBB-T1. Aside from the HBB-T2 gene of *M. pahari*, we found no evidence of gene conversion in the flanking regions of HBB-T1 or HBB-T2 in any of the other species of *Mus*.

In general, phylogenetic reconstructions based on intron 2 also recovered the same two clades of orthologous HBB-T1 and HBB-T2 sequences (fig. 3). There were five cases of paraphyly: HBB-T1 sequences of *M. saxicola* and *M. cookii* were both nested within the HBB-T2 clade (indicating a T2 → T1 conversion of intron 2), and HBB-T2 sequences from the $Hbb^s$ haplotype of *M. domesticus*, *M. musculus*, and *M. spretus* were nested within the HBB-T1 clade (indicating a T1 → T2 conversion of intron 2).

Despite these few cases of paraphyly, phylogenies based on flanking regions and intron 2 clearly group the HBB-T1 and HBB-T2 sequences into two distinct clades. Although phylogenetic signal was relatively weak due to the restricted number of informative sites within each partition of the multiple alignment, the tree topologies were largely consistent with species phylogenies inferred from independent data (Lundrigan et al. 2002; Tucker et al. 2005; Tucker 2007). In contrast to the generally well-defined HBB-T1 and HBB-T2 clades in the phylogenies of flanking and intronic sequences, the phylogeny of coding

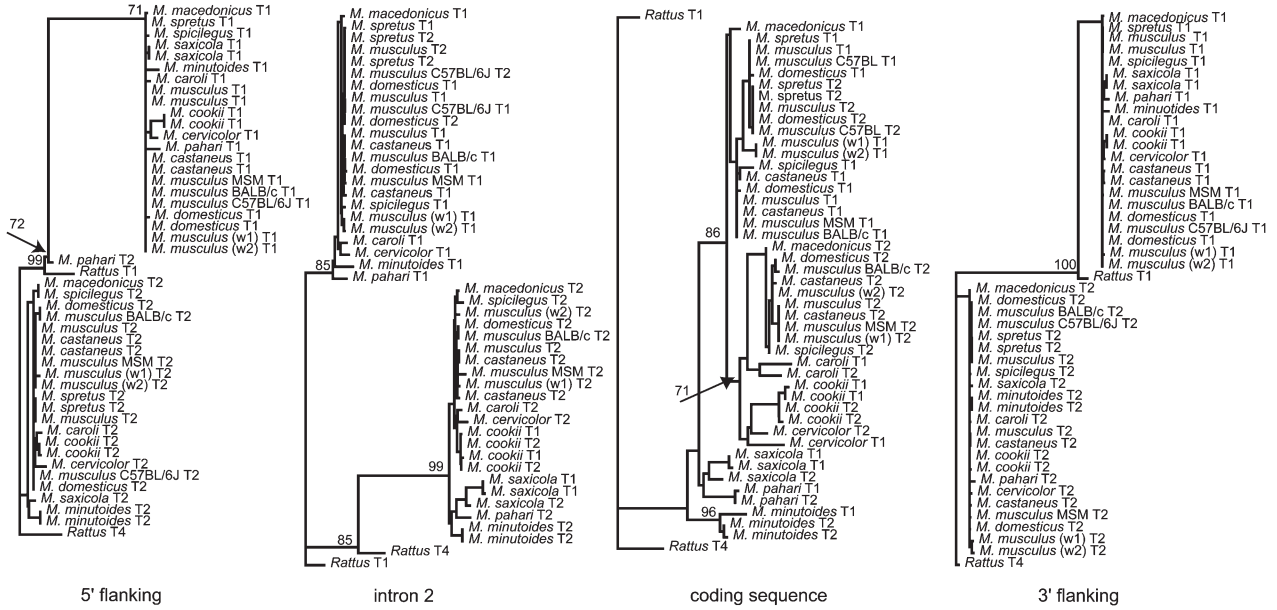


FIG. 3.—Maximum likelihood phylogenies depicting evolutionary relationships among adult $\beta$-globin genes in 12 species of *Mus*. From left to right, phylogenies were reconstructed from the 5′-flanking sequence (338 bp), intron 2 (773 bp), coding sequence (441 bp), and 3′-flanking sequence (290 bp). Trees were constructed under a maximum likelihood framework using the HKY + $\gamma$ model of nucleotide substitution for the flanking and coding sequences and TN93 + $\gamma$ model of sequence evolution for intron 2. For each of the four data partitions, the gamma-distributed rate heterogeneity parameter, $\alpha$, was 5.56, 1.21, 0.75, and 1.79, respectively. For each of the four reconstructed trees, the likelihood values were −1610.03, −4462.94, −1698.98, and −1486.62, respectively. Numbers above the nodes are bootstrap support values >65.

sequences was characterized by extensive paraphyly as HBB-T1 and HBB-T2 sequences were intermingled throughout the tree. In the phylogeny of coding sequences, *M. cervicolor*, *M. cookii*, *M. minutoides*, *M. pahari*, and *M. saxicola*, each exhibited the hallmarks of concerted evolution, as paralogs from the same species grouped together to the exclusion of their presumed orthologs in other species (fig. 3). The hypothesized fusion gene of *M. caroli* did not show clear affinities with the HBB-T1 or HBB-T2 genes of other species, but the HBB-T2 of *M. caroli* formed a clade with HBB-T1/$d_{maj}$ and HBB-T2/$d_{min}$ sequences from Eurasian members of the subgenus *Mus*.

## Is the $\beta$-Globin of *M. caroli* the Product of a Chimeric Fusion Gene?

Gilman (1972, 1974) reported that the $\beta$-chain subunit of *M. caroli* Hb is a hybrid polypeptide characterized by a C-terminal portion that is nearly identical to $d_{min}$ (the HBB-T2 allele on the *Hbb*$^d$ haplotype) and an N-terminal portion that is nearly identical to $d_{maj}$ (the HBB-T1 allele on the *Hbb*$^d$ haplotype). Gilman hypothesized that this chimeric $d_{min}/d_{maj}$ fusion gene was created by unequal crossing-over between misaligned copies of HBB-T1 ($d_{maj}$) and HBB-T2 ($d_{min}$). The product of this $d_{min}/d_{maj}$ fusion gene would be structurally similar to the $\beta$-chains of "Hb Lepore," a human Hb mutant that incorporates the products of a chimeric $\delta/\beta$-globin fusion gene (Forget 2001). Based on a comparison of amino acid sequences between the *M. caroli* $\beta$-chain and the $d_{maj}$ and $d_{min}$ $\beta$-chains, Gilman (1972, 1974) hypothesized that the crossover break point was located in the interval of exon 2 that encodes amino acid residues 58–73. To test this unequal cross-over hypothesis, we conducted a phylogenetic analysis of the HBB-T1 and HBB-T2 genes of *M. caroli* and the corresponding genes on the *Hbb*$^d$ haplotype of *M. castaneus* and *M. musculus*. We reconstructed separate phylogenies for four partitions of the multiple sequence alignment: fragment 1 (5′-flanking sequence), fragment 2 (exon 1 + intron 1 + exon 2), fragment 3 (intron 2), and fragment 4 (exon 3 + 3′-flanking sequence). According to Gilman's unequal cross-over hypothesis, sequence from the 5′ end of *M. caroli* HBB-T1 (fragments 1 and 2) should group with HBB-T2/$d_{min}$ sequences of *M. castaneus* and *M. musculus*, whereas sequence from the 3′ end of *M. caroli* HBB-T1 (fragments 3 and 4) should group with HBB-T1/$d_{maj}$ sequences of the other species.

Phylogenetic reconstructions showed that the 5′-flanking sequence of *M. caroli* HBB-T1 (fragment 1) grouped with HBB-T1/$d_{maj}$ sequences of *M. castaneus* and *M. musculus*, and likewise, the HBB-T2 sequence of *M. caroli* grouped with HBB-T2/$d_{min}$ sequences of the other species (fig. 4). By contrast, in the case of fragment 2, the *M. caroli* HBB-T1 and HBB-T2 sequences grouped together to the exclusion of $d_{maj}$ and $d_{min}$ sequences in the other species. Further downstream, phylogenies of fragments 3 and 4 reverted to the same pattern of reciprocal monophyly between HBB-T1 and HBB-T2 sequences that was observed for fragment 1. Thus, contrary to Gilman's (1972, 1974) hypothesis, the chimeric sequence of the *M. caroli* HBB-T1 gene is not attributable to unequal crossing-over. Rather, it is attributable to a HBB-T2 → HBB-T1 gene conversion event that was restricted to exon 1, intron 1, and exon 2.

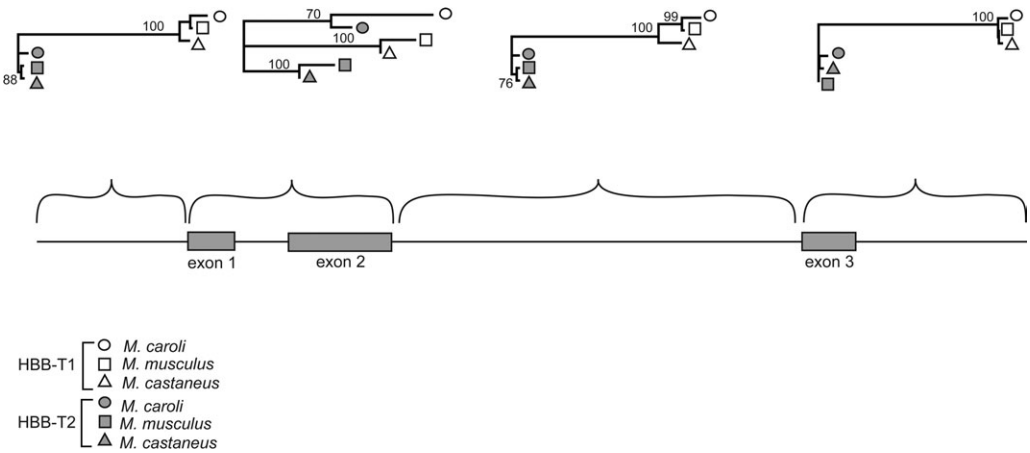


FIG. 4.—Maximum likelihood phylogenies depicting relationships of HBB-T1 and HBB-T2 genes of *Mus caroli* with orthologous and paralogous sequences from the $Hbb^d$ haplotype of *Mus castaneus* and *Mus musculus*. Separate phylogenies were reconstructed from four separate partitions of the multiple alignment (from left to right): fragment 1 (5′-flanking sequence), fragment 2 (exon 1 + intron 1 + and exon 2), fragment 3 (intron 2), and fragment 4 (exon 3 and 3′-flanking sequence). Numbers above the nodes are bootstrap support values >65.

## Patterns of Interparalog Gene Conversion

Gene conversion between HBB-T1 and HBB-T2 was pervasive in all species other than those that carried the $Hbb^d$, $Hbb^p$, $Hbb^{w1}$, and $Hbb^{w2}$ haplotypes (*M. castaneus*, *M. macedonicus*, *M. musculus*, and *M. spicilegus*; table 1). Our analysis of gene conversion between the two HBB paralogs revealed four noteworthy patterns. First, conversion tracts were almost exclusively restricted to coding regions. Second, conversion tracts spanned the entire coding region in some cases (HBB-T1 of *M. cookii* and *M. saxicola*, and HBB-T2 on the $Hbb^s$ haplotype of *M. domesticus*, *M. musculus*, and *M. spretus*), and in the remaining cases, the conversion tracts generally spanned just the 5′ portion of the gene (exon 1, intron 1, and exon 2; table 1, fig. 5). Third, gene conversion was bidirectional as roughly equal numbers of identified conversion events occurred in the 5′ → 3′ direction (T1 → T2) and in the 3′ → 5′ direction (T2 → T1; table 1 and fig. 5). And fourth, despite the pervasiveness of interparalog gene conversion, the HBB genes of most species have at least partially escaped from concerted evolution. Gene conversion has completely homogenized amino acid sequence variation between the HBB paralogs of *M. cookii* and those species carrying the $Hbb^s$ haplotype, but all other species carry HBB paralogs that are distinguished by 1–12 amino acid substitutions (table 1 and fig. 5).

## Structural Differentiation of β-chain IsoHbs

We inferred that the common ancestor of *Mus* possessed two distinct HBB-T1 and HBB-T2 paralogs that were distinguished by seven amino acid substitutions at residues 20, 58, 76, 80, 121, 125, and 135 (fig. 6). The three substitutions at residues 58, 76, and 80 also distinguish the HBB paralogs on the $Hbb^d$, $Hbb^p$, $Hbb^{w1}$, and $Hbb^{w2}$

**Table 1**
**Differences between HBB-T1 and HBB-T2 Paralogs as Measured by Average Number of Nucleotide Substitutions per Site ($Dxy_{total}$), Average Number of Nucleotide Substitutions per Silent Site ($Dxy_{silent}$), and the Number of Amino Acid Differences**

| Species | $Dxy_{total}$ | $Dxy_{silent}$ | Number of Amino Acid Differences | Physicochemical Distance | Direction of Gene Conversion and Length of Identified Conversion Tract |
|---|---|---|---|---|---|
| *Mus castaneus* ($Hbb^d$) | 0.24 | 0.29 | 9 | 0.52 | — |
| *M. castaneus* ($Hbb^p$) | 0.24 | 0.28 | 10 | 0.63 | — |
| *Mus caroli* | 0.25 | 0.33 | 8 | 0.08 | T1 ← T2 (exon 1–exon 2) |
| *Mus cervicolor* | 0.24 | 0.31 | 10 | 0.55 | T1 ← T2 (exon 1–220 bp into intron 2) |
| *M. cooki* | 0.14 | 0.18 | 0 | 0 | T1 ← T2 (exon 1–exon 3) |
| *Mus domesticus* ($Hbb^s$) | 0.05 | 0.07 | 0 | 0 | T1 → T2 (exon 1–exon 3) |
| *M. domesticus* ($Hbb^d$) | 0.18 | 0.24 | 9 | 0.52 | — |
| *Mus macedonicus* ($Hbb^d$) | 0.25 | 0.32 | 9 | 0.53 | — |
| *Mus minutoides* | 0.24 | 0.31 | 4 | 0.46 | T1 ← T2 (exon 1–220 bp into intron 2) |
| *Mus musculus* ($Hbb^s$) | 0.14 | 0.17 | 0 | 0 | T1 → T2 (exon 1–exon 3) |
| *M. musculus* ($Hbb^p$) | 0.25 | 0.30 | 11 | 0.76 | — |
| *M. musculus* ($Hbb^{w1}$) | 0.19 | 0.24 | 12 | 0.74 | — |
| *M. musculus* ($Hbb^{w2}$) | 0.19 | 0.25 | 10 | 1 | — |
| *Mus pahari* | 0.24 | 0.32 | 1 | 0.01 | T1 → T2 (104 bp 5′ flanking–220 bp into intron 2) |
| *Mus saxicola* | 0.19 | 0.24 | 6 | 0.52 | T1 ←T2 (exon 1–exon 3) |
| *Mus spicilegus* ($Hbb^d$) | 0.24 | 0.31 | 9 | 0.34 | — |
| *Mus spretus* | 0.14 | 0.17 | 0 | 0 | T1 → T2 (exon 1–exon 3) |

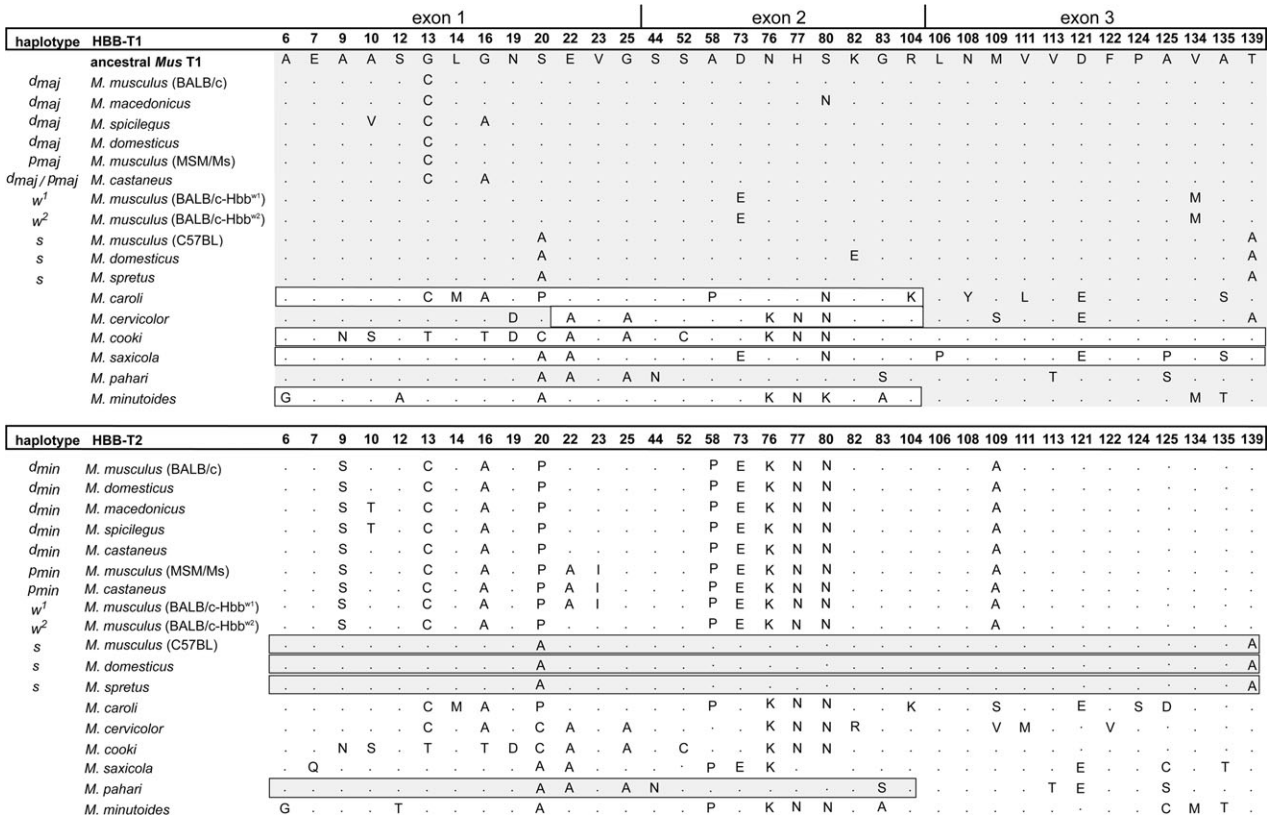


FIG. 5.—Alignment of HBB-T1 and HBB-T2 amino acid sequences from 12 *Mus* species showing sites that are variable in one or both paralogs. The $Hbb^s$ and $Hbb^p$ haplotypes recovered from *Mus musculus* are not shown as they are identical to the sequences found in the C57BL ($Hbb^s$) and MSM/Ms ($Hbb^p$) strains. Representative sequences are displayed for the HBB genes of *M. castaneus*, *M. cookii*, *M. domesticus*, and *M. saxicola*. HBB-T1 sequences are shown in gray and HBB-T2 sequences are shown in white. Identified gene conversion tracts are shown in boxes.

haplotypes. The remaining differences between HBB-T1 and HBB-T2 on each of these haplotypes are attributable to substitutions that accumulated in the HBB-T2 paralog. In the case of the five main haplotypes that are found in the Eurasian members of the subgenus *Mus*, the HBB-T2 sequences have accumulated a preponderance of amino acid changes, whereas the HBB-T1 sequences are more highly conserved (fig. 6). The difference in rates of amino acid substitution between the two paralogs on the $Hbb^d$, $Hbb^p$, $Hbb^{w1}$, and $Hbb^{w2}$ haplotypes is mirrored by differences in the ratio of nonsynonymous to synonymous substitution rates for the full set of *Mus* species ($d_N/d_S = 0.31$ for HBB-T1 and 0.51 for HBB-T2). There is no way of knowing whether the HBB paralogs on the $Hbb^s$ haplotype have experienced a similar disparity in rates of amino acid substitution as any changes that accumulated in the

| HBB-T1 | 9 | 13 | 16 | 20 | 22 | 23 | 58 | 73 | 76 | 77 | 80 | 109 | 121 | 125 | 134 | 135 | 139 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ancestral *Mus* T1 | Ala | Gly | Gly | Ser | Glu | Val | Ala | Asp | Asn | His | Ser | Met | Asp | Ala | Val | Ala | Thr |
| $Hbb^s$ | Ala | Gly | Gly | Ala | Glu | Val | Ala | Asp | Asn | His | Ser | Met | Asp | Ala | Val | Ala | Ala |
| $Hbb^{dmaj}$ | Ala | Cys | Gly | Ser | Glu | Val | Ala | Asp | Asn | His | Ser | Met | Asp | Ala | Val | Ala | Thr |
| $Hbb^{pmaj}$ | Ala | Cys | Gly | Ser | Glu | Val | Ala | Asp | Asn | His | Ser | Met | Asp | Ala | Val | Ala | Thr |
| $Hbb^{w1}$ | Ala | Gly | Gly | Ser | Glu | Val | Ala | Glu | Asn | His | Ser | Met | Asp | Ala | Met | Ala | Thr |
| $Hbb^{w2}$ | Ala | Gly | Gly | Ser | Glu | Val | Ala | Glu | Asn | His | Ser | Met | Asp | Ala | Met | Ala | Thr |

| HBB-T1 | 9 | 13 | 16 | 20 | 22 | 23 | 58 | 73 | 76 | 77 | 80 | 109 | 121 | 125 | 134 | 135 | 139 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ancestral *Mus* T2 | Ala | Gly | Gly | Ala | Glu | Val | Pro | Asp | Lys | His | Asn | Met | Glu | Cys | Val | Thr | Thr |
| $Hbb^s$ | Ala | Gly | Gly | Ala | Glu | Val | Ala | Asp | Asn | His | Ser | Met | Asp | Ala | Val | Ala | Ala |
| $Hbb^{dmin}$ | Ser | Cys | Ala | Pro | Glu | Val | Pro | Glu | Lys | Asn | Asn | Ala | Asp | Ala | Val | Ala | Thr |
| $Hbb^{pmin}$ | Ser | Cys | Ala | Pro | Ala | Ile | Pro | Glu | Lys | Asn | Asn | Ala | Asp | Ala | Val | Ala | Thr |
| $Hbb^{w1}$ | Ser | Cys | Ala | Pro | Ala | Ile | Pro | Glu | Lys | Asn | Asn | Ala | Asp | Ala | Val | Ala | Thr |
| $Hbb^{w2}$ | Ser | Cys | Ala | Pro | Glu | Val | Pro | Glu | Lys | Asn | Asn | Ala | Asp | Ala | Val | Ala | Thr |

FIG. 6.—Alignment of variable amino acid sites in the HBB-T1 and HBB-T2 genes from five different haplotype backgrounds. HBB-T1 sequences are shown in gray and HBB-T2 sequences are shown in white. The $Hbb^s$ allele at HBB-T2 is shown in gray, which reflects its ectopic origin via T1 → T2 gene conversion. Substitutions that distinguish the inferred ancestral sequences of *Mus* HBB-T1 and HBB-T2 are denoted by boxes. Derived amino acids in HBB-T1 and HBB-T2 are shown in blue and green, respectively.

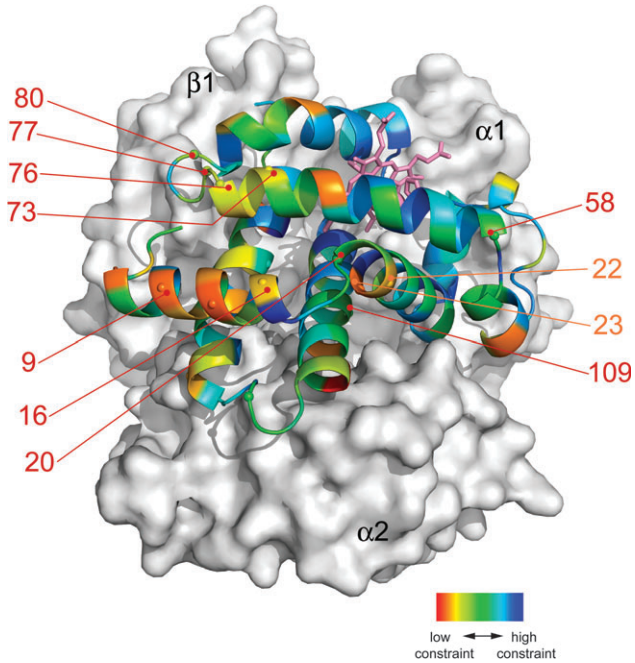


FIG. 7.—Homology-based structural model of the *Mus* β-globin polypeptide showing the location of nine amino acid substitutions (red numbers) that distinguish HBB-T1 and HBB-T2 on the $Hbb^d$ haplotype ($d_{maj}$ and $d_{min}$, respectively) and two additional substitutions (orange numbers) that distinguish HBB-T1 and HBB-T2 on the $Hbb^p$ haplotype ($p_{maj}$ and $p_{min}$, respectively). Because $d_{maj}$ and $p_{maj}$ sequences are identical, the $Hbb^d$ and $Hbb^p$ haplotypes only differ at HBB-T2 sites 22 and 23. Color coding of the ribbon structure shows site-specific variation in conservation scores (see text for details). Residue positions in blue are invariant or nearly invariant (and are therefore presumably subject to stringent functional constraints), whereas residue positions that are further toward the red end of the spectrum are more variable (and are therefore presumably subject to less stringent functional constraints).

HBB-T2 sequence have since been overwritten by gene conversion from HBB-T1.

The $Hbb^d$, $Hbb^p$, $Hbb^{w1}$, and $Hbb^{w2}$ haplotypes are characterized by the highest level of physicochemical differentiation between the products of HBB-T1 and HBB-T2 (table 1). In species that possess distinct HBB-T1 and HBB-T2 genes, most of the amino acid differences between products of the two paralogs involve exterior, solvent-exposed residues. Most of the amino acid substitutions that distinguish the two paralogs on the $Hbb^d$ and $Hbb^p$ haplotypes ($d_{maj}$ vs $d_{min}$, and $p_{maj}$ vs $p_{min}$) are located in positions that appear to be subject to relatively low levels of functional constraint, with the exception of sites 20, 58, and 109, each of which had conservation scores $\geq$0.75 (fig. 7). In the case of $Hbb^d$ and $Hbb^p$, the especially high levels of physiochemical differentiation between the two coexpressed isoHbs is largely attributable to the β109(Ala → Met) substitution in the internal, water-filled cavity of the Hb tetramer (fig. 8*A*). Whereas the β-chain Hbs of almost all mammals studied to date contain Val at position 109, the β-chain product of HBB-T1/$d_{maj}$ contains Met. In human Hb, the rare β109Met mutant (Hb San Diego) is characterized by unusually high $O_2$-binding affinity and impaired cooperativity and is associated with pathological erythrocytosis (Anderson 1974; Nute et al. 1974). Residue position 109 is located immediately adjacent to an $\alpha_1\beta_1$ intersubunit contact, and substitution of Met at this highly conserved site disrupts an H-bond between β35Tyr (the N-terminal residue of the β-chain C helix) and α122His on the α-chain H helix (Anderson 1974; fig. 8). This loss of intradimer contact between α- and β-chain subunits destabilizes the low-affinity deoxyHb structure, thereby shifting the allosteric equilibrium in favor of the high-affinity oxyHb (Anderson 1974). Thus, the red blood cells of mice that carry the $Hbb^d$ and $Hbb^p$ haplotypes contain a mixture of distinct β-chain isoHbs that may differ in allosteric equilibria between the deoxy-Hb and oxy-Hb conformations.

## Discussion

Among species in the genus *Mus*, the HBB-T1 and HBB-T2 genes exhibit the full range of evolutionary outcomes with respect to levels of interparalog divergence. At one end of the spectrum, the two identical HBB paralogs on the $Hbb^s$ haplotype (shared by *M. domesticus*, *M. musculus*, and *M. spretus*) represent a textbook example of concerted evolution. At the other end of the spectrum, the two distinct HBB paralogs on the $Hbb^d$, $Hbb^p$, $Hbb^{w1}$, and $Hbb^{w2}$ haplotypes (shared by multiple species in the subgenus *Mus*) show no trace of gene conversion and are distinguished by a number of amino acid substitutions that alter important biochemical properties of the Hb protein. Moreover, the ancestral sequence reconstructions indicate that the species of *Mus* included in our analysis descend from a common ancestor that possessed two HBB paralogs that were distinguished by seven amino acid substitutions. Thus, with the exception of individuals of *M. cookii* that are homozygous for the same HBB haplotype and individuals of *M. musculus*, *M. domesticus*, and *M. spretus* that are

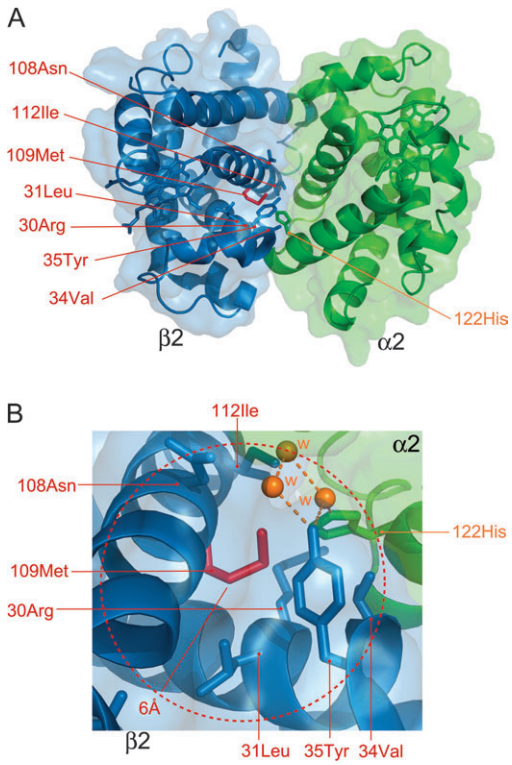


FIG. 8.—(*A*) Homology-based structural model of the $\alpha_2\beta_2$ dimer of *Mus* Hb, representing one-half molecule of the Hb tetramer. Amino acid side chains are highlighted for six β-chain residues (30, 31, 34, 35, 108, and 112) and one α-chain residue (122) that are located within a six Å radius of the sulfur atom of β109Met where the α- and β-chain subunits come into contact. (*B*) Enlarged view of the $\alpha_2\beta_2$ intersubunit contact surface showing how, in the presence of the derived β109Met residue, the phenol side chain of β35Tyr coordinates two water molecules (oxygen atoms denoted by w), one of which is jointly coordinated with the imidazole side chain of α122His. By contrast, in the presence of the ancestral β109Val residue, the free rotation of the β35Tyr phenol side chain permits the formation of an H-bond with α122His. This H-bond between β35Tyr and α122His helps stabilize the intradimer contact, thereby shifting the allosteric equilibrium in favor of the low-affinity deoxyHb conformation (Anderson 1974).

homozygous for the $Hbb^s$ haplotype, mice in the genus *Mus* are unusual among mammals in that they are capable of synthesizing two distinct β-chain isoHbs during fetal development and postnatal life.

## Patterns of Gene Conversion

Phylogenetic analysis of the coding sequences showed that the HBB paralogs from *M. caroli*, *M. cervicolor*, *M. minutoides*, *M. musculus*, *M. pahari*, *M. saxicola*, and *M. spretus* were more similar to each other than to their orthologs in other species. In principle, this pattern could be attributable to the effects of gene conversion or it may reflect recent ancestry between the products of de novo gene duplication events that occurred independently in multiple lineages. In the α-globin gene family of primates, sequence similarity between paralogous genes in the same species was sometimes attributable to gene conversion, but in a surprising number of cases, it was attributable to recent ancestry between the products of lineage-specific gene duplications (Hoffmann et al. 2008b). In the case of the HBB-T1 and HBB-T2 paralogs of *Mus*, the phylogenies of flanking sequence and intronic sequence provided no evidence of lineage-specific gene duplications. It is clear that each of the 12 mouse species included in our analysis inherited the same pair of HBB-T1 and HBB-T2 genes from a common ancestor, but the antiquity of the two paralogs has been obscured by recurrent gene conversion that has occurred independently in each descendant lineage. It appears that the original duplication event that gave rise to the β-globin genes of *Mus* predated the diversification of muroid rodents as 1:1 orthologs of the HBB-T1 and HBB-T2 genes have been identified in *Rattus* and two species of *Peromyscus* (Hoffmann et al. 2008a).

Estimates of gene conversion tract lengths in the human β-globin gene cluster range from 113 to 2266 bp (Papadakis and Patrinos 1999). The conversion events that we detected in the present study all fall well within this range. The largest conversion tracts that we detected, such as the HBB-T1 → HBB-T2 conversion event on the $Hbb^s$ haplotype, did not extend much beyond the initiation and termination codons and were therefore less than 1.4 kb in length. In contrast to other eukaryotic gene families in which interparalog gene conversion has been documented (Chen et al. 2007), we observed no consistent bias in the directionality of conversion events. In the human β-globin gene cluster, the directionality of gene conversion is associated with the relative expression levels of the two genes involved in the exchange as the gene that is expressed at a higher level is more likely to convert the gene that is expressed at a lower level (Papadakis and Patrinos 1999). In *M. musculus*, the expression level of HBB-T1 is roughly 4-fold higher than that of HBB-T2 (Hutton et al. 1962; Gilman 1974; Whitney 1977). If this discrepancy in relative expression levels between the two HBB paralogs is consistent among other species of *Mus*, then it would appear that the association between expression level and directionality of gene conversion does not hold in mice.

## Implications for Functional Differentiation of IsoHbs

Because the possession of distinct HBB paralogs expands the repertoire of functionally distinct isoHbs that can be synthesized during fetal development and postnatal life, variation in functional divergence between HBB-T1 and HBB-T2 may underlie important physiological variation within and among species. For example, coexpression of multiple isoHbs may permit higher intraerythrocytic Hb concentrations by increasing solubility and inhibiting protein aggregation (Weber 1990; Storz and Moriyama 2008). It is interesting that the alternative 2-locus haplotypes that represent opposite ends of the spectrum with respect to interparalog divergence—$Hbb^s$ and $Hbb^d$—are maintained at intermediate frequencies in natural populations of *M. domesticus* and *M. musculus*. Two of the most commonly used inbred strains of laboratory mice, C57BL and BALB/c, are homozygous for the $Hbb^s$ and $Hbb^d$ haplotypes, respectively. Like C57BL, humans effectively express a single major isoHb during postnatal life as the

minor $HbA_2$ isoHb (which incorporates $\beta$-type chains that are encoded by the δ-globin gene) typically accounts for $<2\%$ of Hb in circulating red blood cells. Thus, strains of mice like BALB/c that coexpress multiple isoHbs may not be ideal models for research on pathologies of the cardiopulmonary system.

### Sharing of *Hbb*$^s$, *Hbb*$^d$, *and Hbb*$^p$ Haplotypes Among Species

The $\beta$-globin haplotypes *Hbb*$^s$, *Hbb*$^d$, and *Hbb*$^p$ are shared among multiple species in the subgenus *Mus*. *Mus castaneus* is known to segregate the *Hbb*$^d$ and *Hbb*$^p$ haplotypes (Gilman 1976; Bonhomme et al. 1984; Miyashita et al. 1985), whereas *M. domesticus* and *M. musculus* both segregate the *Hbb*$^d$ and *Hbb*$^s$ haplotypes (Selander and Yang 1969; Selander et al. 1969; Selander 1970; Storz, Baze, et al. 2007). It will be necessary to collect polymorphism data for *M. macedonicus* and *M. spicilegus* to determine whether the *Hbb*$^d$ haplotype is fixed in these two species or whether they are also polymorphic for two or more haplotypes. In principle, the sharing of identical 2-locus $\beta$-globin haplotypes among species could be attributable to introgressive hybridization or the retention of ancestral polymorphism. At face value, introgressive hybridization seems like a plausible explanation for the sharing of identical HBB haplotypes among some of the species that were included in our study as admixture has been documented between natural populations of *M. castaneus* and *M. domesticus*, between *M. domesticus* and *M. musculus* and between *M. domesticus* and *M. spretus* (Moriwaki et al. 1979; Ferris et al. 1983; Yonekawa et al. 1988; Bonhomme et al. 1989; Boursot et al. 1989; Orth et al. 2002; Payseur et al. 2004; Geraldes et al. 2008; Teeter et al. 2008). Even in the absence of introgressive hybridization, the sharing of identical haplotypes among *M. castaneus*, *M. domesticus*, and *M. musculus* can also be plausibly explained by the retention of ancestral polymorphism. These three species are thought to have diverged from one another ~500,000 yrs ago, and gene genealogies at many unlinked autosomal and X-linked loci exhibit paraphyletic and polyphyletic patterns of relationship (Salcedo et al. 2007; Geraldes et al. 2008). In other words, it is not uncommon for alleles at a given gene in *M. castaneus* to be more closely related to alleles in *M. domesticus* than to other alleles in *M. castaneus* (and vice versa). In the case of more distantly related sets of species in the subgenus *Mus* that share the *Hbb*$^s$ haplotype (*M. musculus* and *M. spretus*) and the set of five species that all share the *Hbb*$^d$ haplotype (*M. castaneus*, *M. domesticus*, *M. macedonicus*, *M. musculus*, and *M. spicilegus*), each of the alternative haplotypes would have to be maintained for several million years in order to explain transspecific polymorphism without invoking introgression (Lundrigan et al. 2002; Salcedo et al. 2007). The retention of alternative alleles for especially longtime spans becomes more plausible if polymorphism is actively maintained by some form of balancing selection, as has been suggested in the case of the *Hbb*$^s$ and *Hbb*$^d$ haplotypes (Storz, Baze, et al. 2007). Surveys of $\beta$-globin polymorphism in additional species in the subgenus *Mus* would be useful to assess whether balancing selection needs to be invoked to explain the observed patterns of transspecific polymorphism.

## Acknowledgments

We thank P. Tucker for providing tissue samples for 11 of the species included in this study, and we thank two anonymous reviewers for helpful comments on the manuscript. This work was supported by a National Science Foundation Fellowship in Bioinformatics to A.M.R. (0630779), grants to J.F.S. and H.M. from the National Institutes of Health/NHLBI (R01 HL087216), the Nebraska Research Council, and a grant to J.F.S. from the National Science Foundation (DEB-0614342).

## Literature Cited

Anderson NL. 1974. Hemoglobin San Diego ($\beta$109 (G11) Val→Met) crystal structure of the deoxy form. J Clin Invest. 53:329–333.

Arnold K, Bordoli L, Koop J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. Bioinformatics. 22:195–201.

Berry R. 1978. Genetic variation in wild house mice: where natural selection and history meet. Am Sci. 66:52–60.

Berry R, Murphy H. 1970. The biochemical genetics of an island population of the house mouse. Proc R Soc B Biol Sci. 176:87–103.

Berry R, Peters J. 1975. Macquarie Island house mice: a genetical isolate on a sub-Antarctic island. J Zool. 176:375–389.

Berry R, Peters J. 1977. Heterogeneous heterozygosities in *Mus musculus* populations. Proc Roy Soc B-Biol Sci. 197:485–503.

Berry R, Peters J, Van Aarde R. 1978. Sub-Antarctic house mice: colonization, survival, and selection. J Zool. 184:127–141.

Bonhomme F, Catalan J, Britton-Davidian J, Chapman V, Moriwaki K, Nevo E. 1989. Thaler. 1984. Biochemical diversity and evolution in the genus *Mus*. Biochem Genet. 22:275–303.

Bonhomme F, Miyashita N, Boursot P, Catalan J, Moriwaki K. 1989. Genetical variation and polyphyletic origin in Japanese *Mus musculus*. Heredity. 63:299–308.

Boursot P, Bonhomme F, Catalan J. 1989. Variation of a Y chromosome repeated sequence across subspecies of *Mus musculus*. Heredity. 63:289–297.

Chen J-M, Cooper DM, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution, and human disease. Nat Rev Genet. 8:762–775.

Erhart MA, Simons KS, Weaver S. 1985. Evolution of the mouse $\beta$-globin gene: a recent gene conversion in the Hbb$^s$ haplotypes. Mol Biol Evol. 2:304–320.

Ferris S, Sage R, Huang C-M, Nielsen J, Ritte U, Wilson A. 1983. Flow of mitochondrial DNA across a species boundary. Proc Natl Acad Sci USA. 80:2290–2294.

Forget BG. 2001. Molecular mechanisms of $\beta$-thalassemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, editors. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge: Cambridge University Press. p. 252–276.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31: 3784–3788.

Geraldes A, Basset P, Gibson B, Smith K, Harr B, Yu H-T, Bulatova N, Ziv Y, Nachman MW. 2008. Inferring the history

of speciation in house mice from autosomal, X-linked, Y-linked, and mitochondrial genes. Mol Ecol. 17:5349–5363.
Gilman JG. 1972. Hemoglobin β chain structural variation in mice: evolutionary and functional implications. Science. 178: 873–874.
Gilman JG. 1974. Rodent hemoglobin structure: a comparison of several species of mice. Ann N Y Acad Sci. 241:416–433.
Gilman JG. 1979. Evolutionary potential: a mathematical hypothesis of mouse hemoglobin beta chain evolution. J Mol Evol. 13:1–14.
Hardies SC, Edgell MH, Hutchison CA. 1984. Evolution of the mammalian β-globin gene cluster. J Biol Chem. 259: 3748–3756.
Hardison R, Gelinas R. 1986. Assignment of orthologous relationships among mammalian α-globin genes by examining flanking regions reveals a rapid rate of evolution. Mol Biol Evol. 3:243–261.
Hardison R, Miller W. 1993. Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. Mol Biol Evol. 10:73–102.
Hasegawa M, Kishino K, Yano T. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol. 22:160–174.
Hill A, Hardies S, Phillips S, Davis M, Hutchison C, Edgell M. 1984. Two mouse early embryonic β-globin gene sequences. Evolution of the nonadult β-globins. J Biol Chem. 259: 3739–3747.
Hoffmann FG, Opazo JC, Storz JF. 2008a. New genes originated via multiple recombinational pathways in the β-globin gene family of rodents. Mol Biol Evol. 25:2589–2600.
Hoffmann FG, Opazo JC, Storz JF. 2008b. Rapid rates of lineage-specific gene duplication and deletion in the α-globin gene family. Mol Biol Evol. 25:591–602.
Hutton J, Bishop J, Schweet R, Russell E. 1962. Hemoglobin inheritance in inbred mouse strains, II. Genetic studies. Proc Natl Acad Sci USA. 48:1718–1724.
Innan H. 2003. A two-locus gene conversion model with selection and its application to the human RHCE and RHD genes. Proc Natl Acad Sci USA. 100:8793–8798.
Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol. 4:18. doi: 10.1186/1471-2148-4-18.
Kidd RD, Russell JE, Watmough NJ, Baker EN, Brittain T. 2001. The role of beta chains in the control of hemoglobin oxygen binding function: chimeric human/mouse protein, structure, and function. Biochemistry. 40:15669–15675.
Koshi J, Goldstein R. 1996. Probabilistic reconstruction of ancestral protein sequences. J Mol Evol. 42:313–320.
Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 157:105–132.
Leder P, Hansen J, Konkel D, Leder A, Nishioka Y, Talkington C. 1980. Mouse globin system: a functional and evolutionary analysis. Science. 77:1336–1342.
Lundrigan BL, Jansa SA, Tucker PK. 2002. Phylogenetic relationship in the genus *Mus*, based on paternally, maternally, and biparentally inherited characters. Syst Biol. 51:410–431.
Lynch M, Katju V. 2004. The altered evolutionary trajectories of gene duplicates. Trends Genet. 20:544–549.
Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. Genetics. 159:1789–1804.
Mano S, Innan H. 2008. The evolutionary rate of duplicated genes under concerted evolution. Genetics. 180:493–505.
Miyashita N, Moriwaki K, Minezawa M, Yonekawa H, Bonhomme F, Migita S, Yu ZC, Lu DY, Cho WS, Tjohari M. 1985. Allelic consitution of the hemoglobin β chain in wild poputionals of the house mouse, *Mus musculus*. Biochem Genet. 23:975–986.
Moon AM, Ley TJ. 1990. Conservation of the primary structure, organization, and function of the human and mouse β-globin locus-activating regions. Proc Natl Acad Sci USA. 87: 7693–7697.
Moriwaki K, Shiroishi T, Minezawa M, Aotsuka T. 1979. Frequency distribution of hitocompatibility-2 antigenic specificities in the Japanese wild mouse genetically remote from the European subspecies. J Immunogenet. 6:99–113.
Myers J. 1974. Genetic and social structure of feral house mouse populations on Grizzly Island, California. Ecology. 55: 747–759.
Nei M, Rooney AP. 2005. Concerted birth-and-death evolution of multigene families. Annu Rev Genet. 39:121–152.
Nute PE, Stamatoyannopoulos G, Hermodson MA, Roth D. 1974. Hemoglobinopathic erythrocytosis due to a new electrophoretically silent variant, hemoglobin san Diego (β109(G11) Val→ Met). J Clin Invest. 53:320–328.
Ohno S. 1970. Evolution by gene duplication. Springer-Verlag: New York.
Opazo JC, Hoffmann FG, Storz JF. 2008a. Differential loss of embryonic globin genes during the radiation of placental mammals. Proc Natl Acad Sci USA. 105:12950–12955.
Opazo JC, Hoffmann FG, Storz JF. 2008b. Genomic evidence for independent origins of β-like globin genes in monotremes and therian mammals. Proc Natl Acad Sci USA. 105: 1590–1595.
Opazo JC, Sloan AM, Campbell KL, Storz JF. 2009. Origin and ascendancy of a chimeric fusion gene: the β/δ-globin gene of paenungulate mammals. Mol Biol Evol. 26:1469–1478.
Orth A, Belkhir K, Britton-Davidian J, Boursot P, Benazzou T, Bonhomme F. 2002. Natural hybridisation between two sympatric species of mice *Mus musculus domesticus* L. and *Mus spretus* Lataste. C R Biol. 325:89–97.
Papadakis MN, Patrinos GP. 1999. Contribution of gene conversion in the evolution of the human β-like globin gene family. Hum Genet. 104:117–125.
Payseur BA, Krenz JG, Nachman MW. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evolution. 58: 2064–2078.
Petras M, Topping J. 1983. The maintenance of polymorphisms at two loci in house mouse (*Mus musculus*) populations. Can J Genet Cytol. 25:190–201.
Salcedo T, Geraldes A, Nachman MW. 2007. Nucleotide variation in wild and inbred mice. Genetics. 177:2277–2291.
Sato JJ, Shinohara A, Miyashita N, Koshimoto C, Tsuchiya K, Nakahara I, Morita T, Yonekawa H, Moriwaki K, Yamaguchi Y. 2008. Discovery of a new HBB haplotype w2 in a wild-derived house mouse, *Mus musculus*. Mamm Genome. 19:155–162.
Sato JJ, Tsuru Y, Hirai K, Yamaguchi Y, Mekada K, Takahata N, Moriwaki K. 2006. Further evidence for recombination between mouse hemoglobin beta b1 and b2 genes based on the nucleotide sequences of intron, UTR, and intergenic spacer regions. Genes Genet Syst. 81:201–209.
Selander R. 1970. Behavior and genetic variation in natural populations. Am. Zool. 10:53–66.
Selander R, Yang S. 1969. Protein polymorphism and genic heterozygosity in a wild population of the house mouse. Genetics. 63:653–667.
Selander R, Yang S, Hunt W. 1969. Polymorphism in esterases and hemoglobin in wild populations of the house mouse (*Mus musculus*). Univ Tex Publ. 6918:271–338.
Shehee W, Loeb D, Adey N, Burton F, Casavant N, Cole P, Davies C, McGraw R, Schichman S, Severynse D. 1989.

Nucleotide sequence of the BALB/c mouse $\beta$-globin complex. J Mol Biol. 205:41–62.
Storz JF, Baze M, Waite J, Hoffmann FG, Opazo JC, Hayes JP. 2007. Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. Genetics. 177:481–500.
Storz JF, Hoffmann FG, Opazo JC, Moriyama H. 2008. Adaptive functional divergence among triplicated $\alpha$-globin genes in rodents. Genetics. 178:1623–1638.
Storz JF, Kelly JK. 2008. Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: insights from deer mouse globin genes. Genetics. 180:367–379.
Storz JF, Moriyama H. 2008. Mechanisms of hemoglobin adaptation to high-altitude hypoxia. High Alt Med Biol. 9:148–157.
Storz JF, Runck AM, Sabatino SJ, Kelly JK, Ferrand N, Moriyama H, Weber RE, Fago A. 2009. Evolutionary and functional insights into the mechanism underlying high altitude adaptation of deer mouse hemoglobin. Proc Natl Acad Sci USA. 105:12950–12955.
Storz JF, Sabatino SJ, Hoffmann FG, Gering EJ, Moriyama H, Ferrand N, Monteiro B, Nachman MW. 2007. The molecular basis of high-altitude adaptation in deer mice. PLoS Genet. 3(e5):448–459.
Sugino R, Innan H. 2006. Selection for more of the same product as a force to enhance concerted evolution of duplicated genes. Trends Genet. 22:642–644.
Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10:512–526.
Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, O'Brien JE, Krenz JG, Sans-Fuentes MA, Nachman MW, Tucker PK. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. 18:67–76.
Teshima KM, Innan H. 2004. The effects of gene conversion on the divergence between duplicated genes. Genetics. 166: 1553–1560.
Teshima KM, Innan H. 2008. Neofunctionalization of duplicated genes under the pressure of gene conversion. Genetics. 178:1385–1398.
Thompson J, Gibso T, Plewniak F, Jeanmougin F, Higgins D. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.
Tucker PK. 2007. Systematics of the genus *Mus*. In: Fox J, Newcomer C, Smith A, Barthold S, Quimby F, Davisson M, editors. The mouse in biochemical research. Boston: Elsevier Press. p. 13–23.
Tucker PK, Sandstedt SA, Lundgrigan BL. 2005. Phylogenetic relationships in the subgenus *Mus* (genus *Mus*, family Muridae, subfamily Murinae): examining gene trees and species trees. Biol J Linn Soc. 84:653–662.
Ueda Y, Miyashita N, Imai K, Yamaguchi Y, Takamura K, Notohara M, Shiroishi T, Kawashima T, Ning L, Wang CY, Wu X, Moriwaki K. 1999. Nucleotide sequences of the mouse globin $\beta$ gene cDNAs in a wild derived new haplotype $Hbb^{w1}$. Mamm Genome. 10:879–882.
Valdar WS. 2002. Scoring residue conservation. Proteins. 28:227–241.
von Salome J, Gyllensten U, Bergstrom TF. 2007. Full-length sequence analysis of the HLA-DRB1 locus suggests a recent origin of alleles. Immunogenetics. 59:261–271.
Weber RE. 1990. Functional significance and structural basis of multiple hemoglobins with special reference to ectothermic vertebrates. In: Truchot JP, Lahlou B, editors. Animal nutrition and transport processes 2: transport, respiration, and excretion: comparative and environmental aspects. Comparative physiology. Vol. 6. SwitzerlandL: S. Karger Publishers. p. 58–45.
Wheeler L, Selander R. 1972. Genetic variation in populations of the house mouse, *Mus musculus*, in the Hawaiian Islands. Univ Tex Publ. 7213:269–296.
Whitney JB. 1977. Differential control of the synthesis of two hemoglobin beta chains in normal mice. Cell. 12:863–871.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yang Z, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics. 141:1641–1650.
Yonekawa H, Moriwaki K, Gotoh O, Miyashita N, Matsushima Y, Shi LM, Cho WS, Zhen XL, Tagashira Y. 1988. Hybrid origin of Japanese mice "*Mus musculus molossinus*": evidence from restriction analysis of mitochondrial DNA. Mol Biol Evol. 5:63–78.
Zangenberg G, Huang MM, Arnheim N, Erlich H. 1995. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nat Genet. 10:407–414.
Zhang J. 2003. Evolution by gene duplication. Trends Ecol Evol. 18:292–298.

Michael Nachman, Associate Editor

Accepted July 22, 2009