Am. J. Hum. Genet. 64:31–39, 1999

MOLECULAR EVOLUTION '99
The Genomic Record of Humankind's Evolutionary Roots

Morris Goodman

Department of Anatomy and Cell Biology, Wayne State University School of Medicine, Detroit

In response to the call for a human genome–evolution project (McConkey and Goodman 1997), the view has been expressed that what makes us human resides in the 1.5% difference in genomic DNA that separates us from chimpanzees (Gibbons 1998). This view is far too narrow. Features that we associate with being human did not just arise de novo in the past 6 million years since the lineage to humans separated from that to chimpanzees. Rather, some of the most striking human features, such as greatly enlarged brains and prolonged childhoods in social nurturing societies, have deep roots in our evolutionary history. Forty to 30 million years ago (Ma) neocortical portions of the brain increased in the two emerging branches of anthropoid primates—the platyrrhines (or New World monkeys) and the catarrhines. Within the catarrhine branch, additional marked enlargements occurred by 18–6 Ma in the lineage to the ancestors of modern hominids, and the largest neocortical increases occurred in the past 3 million years in the lineage to modern humans.

A parallel evolutionary trend prolonged fetal life and the periods of postnatal life needed to reach full maturity. We may surmise that the genetic program for our enlarged neocortex has both ancient conserved features and more-recently derived features—in particular, the anthropoid-specific features shared with New and Old World monkeys and apes, the hominid-specific features shared with apes, and some human-specific features. Although many mutations in the past 40 million years have shaped the neurogenetic program for an enlarged neocortex, it is possible that just a small number of regulatory mutations in the past 6 million years have brought about the final enlargement of our neocortex compared with that of chimpanzees.

Behaviorally, the separation between chimpanzees and humans is much smaller than once thought. Chimpanzees are emotionally complex and intelligent. They use tools and have material cultures (McGrew 1992), are ecological generalists, are highly social (De Waal 1995; McGrew et al. 1997), and apparently can learn and use rudimentary forms of language (Savage-Rumbaugh and Lewin 1994; Fouts 1997). In agreement with the newer information on the social lives and intelligence of chimpanzees and other apes (McGrew et al. 1997), the results of molecular studies of primate phylogeny (Goodman et al. 1998, and in press) challenge the traditional anthropocentric view that humans are very different from all other animals. Rather, the molecular results reveal that genetically we humans are only slightly remodeled apes. We share with our most distant living ape relatives (the gibbons and siamangs) >95% identity in genomic DNA, and with our closest relatives (the chimpanzees and bonobos, or pygmy chimpanzees) >98.3% identity in typical noncoding DNA and probably ~99.5% identity in the active coding sequences of functional nuclear genes.

Traditional primate classifications, still favored by many anthropologists, use the nebulous concept of grades of evolutionary advancement to place both extinct and living small-brained primates—those of the Paleocene and Eocene epochs of 65–35 Ma—along with the living lemurs, lorises, bush babies, pottos, and tarsiers, in the suborder Prosimii, the primitive grade. In turn, these traditional primate classifications place the larger-brained primates in the suborder **Anthropoidea**, the advanced grade. Moreover, within Anthropoidea, a gradistic grouping places the African great apes (chimpanzees, bonobos, and gorillas) with the Asian great apes (orangutans) in subfamily Ponginae of family Pongidae, whereas humans, viewed as the most advanced primates, are the sole living members of family Hominidae. This traditional anthropocentric view of our place in the order **Primates** ignores (1) the overwhelming evidence that the African great apes share their more recent common ancestry with humans rather than with orangutans and (2) the mounting evidence that the clade of chimpanzees and bonobos is the sister group of humans—that is, shares a more recent common ancestry with humans than with gorillas (Goodman et al. 1998). In contrast, the cladistic view of how to classify organisms calls for classifying our species, *Homo sapiens,* in a radically new

Received October 19, 1998; accepted for publication November 13, 1998; electronically published January 13, 1999.

Address for correspondence and reprints: Dr. Morris Goodman, Department of Anatomy and Cell Biology, Wayne Sate University, 540 E. Canfield Ave., Detroit, MI 48201. E-mail: mg@tree.roc.wayne.edu

This article represents the opinion of the author and has not been peer reviewed.

0002-9297/99/6401-0007$02.00

but strictly objective way, one without arbitrary anthropocentric biases. In this new way, the cladistic evidence from molecules and morphology, as well as the fossil and molecular evidence on branch times in primate phylogeny, favors a phylogenetic classification (Goodman et al. 1998, and in press) in which humans are very close to apes, especially to chimpanzees and bonobos.

In the derivation of this classification, three principles proposed by Hennig (1966) were followed. The first principle is that each taxon should represent a monophyletic group or clade—that is, it should represent all species descended from a common ancestor. The second principle is that the hierarchical groupings of lower-ranked taxa into higher-ranked taxa should describe the phylogenetic relationships of the clades. The third principle is that, ideally, taxa at the same hierarchical level or rank should represent clades that are equally old—that is, at an equivalent evolutionary age. When these principles are followed, not only do all the apes group with humans within the family Hominidae, but also chimpanzees and bonobos as one subgenus and humans as the other subgenus group together into the same genus, *Homo*.

Below I briefly discuss this classification of primates, in which humans share their genus with chimpanzees and bonobos. Phylogenetic relationships depicted in the classification are supported by DNA hybridization data and DNA sequence data from the mitochondrial genome and from a growing number of unlinked nuclear genomic loci (reviewed in Goodman et al. 1998, and in press). Some of the best data come from noncoding sequences of the nuclear genomic region called the "β-globin gene cluster." Phylogenetic analyses of these noncoding sequences have provided evidence not only on primate phylogeny (Goodman et al. 1998) but also on functionally important *cis*-regulatory elements that control the developmental expression of the cluster's functional β-type globin genes, including ϵ, γ, and β (Gumucio et al. 1996). In particular, mutations in *cis*-regulatory elements during the emergence and evolution of the anthropoid primates changed the pattern of γ-globin gene expression, from being strictly embryonic to being primarily fetal. These results, too, will be reviewed, since they illustrate how molecular evolution has shaped functionally important components in the human genome. The emergence in anthropoid primates of a fetal hemoglobin that unloads oxygen in tissues more readily than does adult hemoglobin correlates with the evolutionary trend that prolonged fetal life.

The Primates: Their Phylogeny and Classification

A recent tabulation of living mammals lists for the order Primates 60 genera and 233 species (Groves 1993). The DNA sequences from the β-globin gene cluster that provide evidence on primate phylogeny represent 61 primate species belonging to 41 of the 60 recognized genera and almost all primate clades with ages older than that of genera (Goodman et al., in press). More than 85% of the β-globin gene cluster consists of flanking and intergenic noncoding sequences that separate the cluster's β-type globin genes (ϵ, γ, $\psi\eta$, δ, and β) from one another, and these β-type globin genes have twice as much sequence in their two introns (all noncoding) as in their three exons. Because noncoding sequences evolve at a relatively rapid rate and because a majority of the β-globin gene–cluster sequence data are noncoding, maximum-parsimony trees constructed for each series of aligned orthologous sequences in these data have provided a well-resolved picture of phylogenetic relationships among primate clades. Maximum-likelihood trees, constructed for these aligned sequences, and so-called neighbor-joining trees, constructed from matrices of pairwise distances among the aligned sequences, have depicted the same phylogenetic relationships among primate clades as are depicted in the maximum-parsimony trees.

The percentages of sequence change on the branches of the globin phylogenetic trees were used in conjunction with fossil evidence (reviewed in Goodman et al. 1998) to estimate lineage divergence dates, by means of the model of local molecular clocks (Bailey et al. 1992). On the basis of fossil evidence, the lineage divergence date or last common ancestor (LCA) of Old World monkeys (family Cercopithecidae) and humans and apes (family Hominidae) was placed at 25 Ma, the LCA of platyrrhines and catarrhines at 40 Ma, and the LCA of strepsirhines and haplorhines (i.e., of all living primates) at 63 Ma. The paleontologically based age of 25 Ma for the LCA of cercopithecids and hominids served as the starting reference date for estimating, from relevant branch lengths of the globin phylogenetic trees, the divergence dates for lineages within the hominid clade and separately within the cercopithecid clade. The age of 40 Ma for the LCA of platyrrhines and catarrhines served as the starting date for estimating the divergence dates for lineages within the platyrrhine clade. The age of 63 Ma for the LCA of strepsirhines and haplorhines served as the starting date for estimating the divergence dates for lineages within the strepsirhine clade and also for the haplorhine lineage to tarsiers. The local-molecular-clock model took account of lineage variations in rates of base substitution by having each base substitution in a more slowly evolving lineage occur over a longer period of time than that for each base substitution in a more rapidly evolving lineage.

The results on phylogenetic relationships among primate clades and on the ages of the clades were converted into the taxonomic classification presented by Goodman et al. (1998, and in press) and are shown here, in ab-

breviated form, in the Appendix. This classification portrays a series of phylogenetic branchings during the course of primate evolution from the Paleocene epoch to the present day. The division of a higher-ranked taxon into subordinate lower-ranked taxa denotes a phylogenetic branching. The age (in Ma) placed after the name of a taxon is the estimated age of that taxon treated as a *crown group* but also of that taxon's closest (at a step below in rank) subordinate taxa treated as *total groups*. A crown group includes both the LCA of the extant species in a clade and all descendant species (extinct and extant) of the LCA but does not include the stem of the LCA (Jeffries 1979). The total group includes, in addition to all members of the crown group, the stem of the LCA and all extinct offshoots of the stem. Thus the age of 63 Ma for the LCA of all living primates—that is, the age for Primates as a crown group—is the age for Strepsirhini as a total group and for Haplorhini as a total group. In turn, the ages of 50 Ma and 58 Ma listed alongside Strepsirhini and Haplorhini, respectively, are the ages for these two taxa treated as crown groups.

After this first major branching, in the early Paleocene epoch, into semiordinal clades, subordinal clades emerged. The late Paleocene haplorhines divided into Tarsiiformes and Anthropoidea. The anthropoideans of the middle Eocene epoch (at ~40 Ma) divided into the infraorders Platyrrhini and Catarrhini. Families (as total groups) originated from superfamilial clades within infraorders in the middle to late Oligocene epoch (~28–25 Ma), subfamilies in the early Miocene epoch (~23–22 Ma), tribes in the early to middle Miocene (~20–15 Ma), subtribes in the middle to late Miocene (~14–10 Ma), and genera in the late Miocene (~10–7 Ma). Estimated branch times at the infrageneric level placed the ages of genera as crown groups in the late Miocene to early Pliocene epoch (~6–4 Ma) (Goodman et al., in press). For example, 6 Ma is the estimated divergence time of the tarsier species *Tarsius syrichta* from *Tarsius bancanus* and of the platyrrhine species *Callicebus moloch* from *Callicebus torquatus*.

In this primate classification, in which the taxa represent clades and the ages of the clades determine the ranks of the taxa, many of the names for taxa are the same as those commonly used in other primate classifications. This is possible because, in traditional primate classifications, despite the use of the grade concept to name some of the taxa, most taxa do represent monophyletic groups. For example, in the traditional primate classification used by Martin (1990), there are extant members in 5 infraorders, 6 superfamilies, 12 families, and 13 subfamilies.The molecular evidence shows that, of all these extant infraorders and superfamilies, 9 of the 12 extant families and 10 or, possibly, 11 of the 13 extant subfamilies are monophyletic taxa. However, sister-group relationships are not well depicted, nor are taxa at the same rank necessarily at an equivalent age in traditional primate classifications. However, a crude correlation does exist between age of origin of a taxon and its rank. As Romer (1962, p. 32) observed, the rise of modern orders and suborders of mammals occurred in the Eocene epoch, the rise of modern families in the Oligocene epoch, and the rise of modern subfamilies in the Miocene epoch. In correlation, the strictly phylogenetic classification of primate taxa, with its age equivalence among taxa at the same rank (Appendix), places suborders, families, and subfamilies, when treated as total groups, in the Eocene, Oligocene, and Miocene geologic epochs, respectively. The names for genera used in tabulations of the living primates (e.g., see Groves 1993) are also used in the age-related phylogenetic classification (Appendix), in most cases as full generic names but in a few cases as subgeneric names. An exception is that Groves (1993) treats gibbons and siamangs as members of the same genus, *Hylobates*. However, the estimated LCA age for gibbons and siamangs is 8 Ma. Thus, in this case, the phylogenetic classification places these two apes in separate genera but groups them together in the same subtribe (Appendix).

In contrast with the traditional family Hominidae, which has *Homo sapiens* as its only living species, the age-equals-rank system places all living apes and humans in subfamily Homininae. A phylogenetic branching (at ~18 Ma) divided this subfamily into tribes Hylobatini and Hominini. Within Hylobatini, the phylogenetic branching (at ~8 Ma) in the subtribe Hylobatina separated *Symphalangus* (siamangs) from *Hylobates* (gibbons). Within Hominini, a phylogenetic branching (at ~14 Ma) separated the monogeneric subtribe Pongina (orangutans in the genus *Pongo*) from the subtribe Hominina. Within Hominina, a phylogenetic branching (at ~7 Ma) separated *Gorilla* from *Homo*. Within *Homo,* a phylogenetic branching (at ~6 Ma) separated the subgenus for common chimpanzees and bonobos—that is, *H.* (*Pan*)—from the subgenus for humans—that is, *H.* (*Homo*). Thus, the principle of rank equivalence with other primate clades of the same age requires this grouping—of the chimpanzee clade with the human clade—within the same genus.

Molecular Evolution of Primate β-Type Globin Genes

Reconstruction of the evolutionary history of the mammalian β-type globin genes reveals that a tandem duplication of the prototypic mammalian β gene occurred at ~180 Ma and that by 135 Ma, in the common ancestor of metatherian (marsupial) and eutherian (placental) mammals, the 5′ gene (ϵ) had control elements associated with embryonic expression, whereas the 3′ gene (β) had regulatory features associated with postembryonic expression (Koop and Goodman 1988). This

two-gene cluster has persisted in the lineages to present-day marsupials (Koop and Goodman 1988; Cooper et al. 1996), but ϵ duplications in a common ancestor of primates and other eutherian groups—such as lagomorphs, rodents, and artiodactyls—produced three embryonic genes (ϵ, γ, and η), and a β duplication produced two postembryonic genes (δ and β) (Goodman et al. 1984). The main features of this ancestral eutherian five-gene cluster with its upstream locus control region (LCR)—that is, 5′-LCR-ϵ(embryonic)-γ(embryonic)-η(embryonic)-δ(fetal and postnatal)-β(fetal and postnatal)-3′—are discernible from the β-globin clusters of present-day eutherians (Hardison et al. 1997). The β-globin gene cluster in the early (stem) primate lineage ancestral to strepsirhine and haplorhine primates retained the same features, except that the η gene was silenced—that is, became a pseudogene (fig. 1, stage 1). In the bush baby or galago (suborder Loriformes), ϵ and γ are embryonically expressed genes that are repressed at the onset of fetal life, whereas δ and β are not expressed during embryonic life but are expressed genes from fetal life on (Tagle et al. 1988).

In contrast with the persistence of the ancestral pattern of globin gene switches in the lineage to galago, important new expression patterns evolved in the catarrhines and platyrrhines, the two branches of anthropoid primates. The anthropoids have two linked γ loci rather than the single locus found in tarsiers and strepsirhines (fig. 1, stage 2). The catarrhine γ genes are up-regulated in fetal life rather than repressed (fig. 1, stage 3), with the fetal expression of the γ^1 (human $^{G}\gamma$) locus being three times that of the γ^2 (human $^{A}\gamma$) locus, and δ and β genes are not up-regulated (fully switched on) until postnatal life, when the two γ genes are down-regulated (Bunn and Forget 1986). Platyrrhine γ genes are also up-regulated in fetal life. However, γ^2 rather than γ^1 is the primary fetally expressed gene (Johnson et al. 1996; Chiu et al., in press). Most of the γ^1 locus is deleted in all members of the platyrrhine family Atelidae (Meireles et al. 1995), and disabling promoter mutations at the γ^1

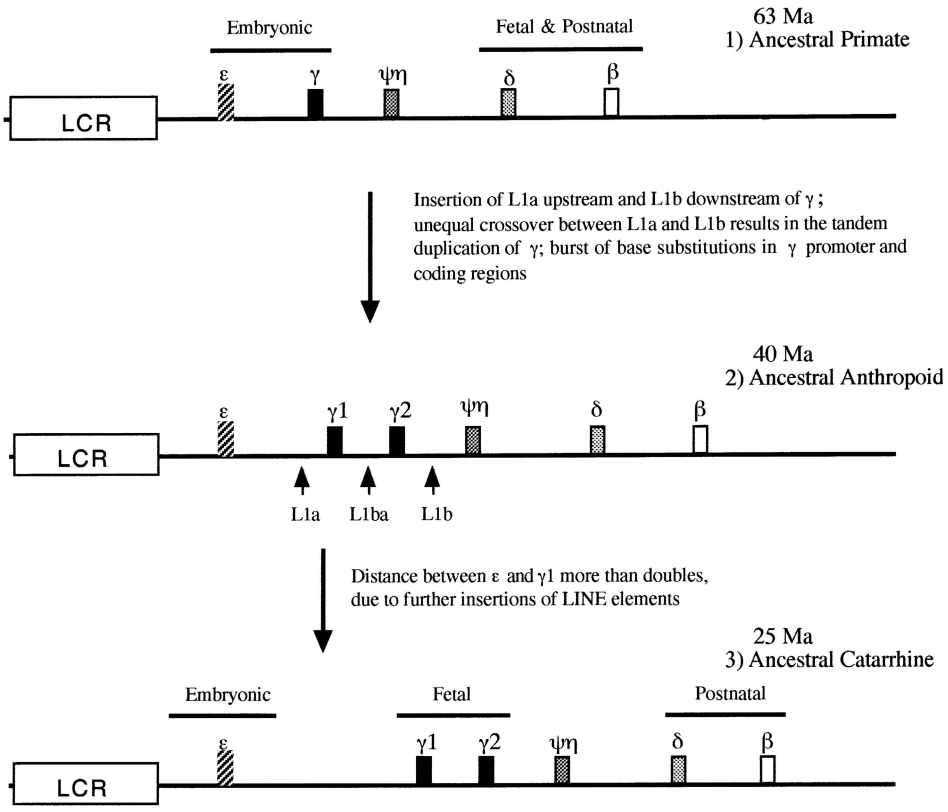

Figure 1 Descent of the β-globin gene cluster in the lineage leading to the LCA of the present-day catarrhine primates. The ancestral form of the β-globin cluster underwent a series of changes in the evolutionary lineage that led to our species. Functionally important changes include both the duplication of the γ-globin gene to form γ^1- and γ^2-globin and the gradual accumulation of noncoding DNA separating the ϵ-globin gene from the rest of the cluster. This latter change may account for species-specific developmental patterns of globin gene expression that distinguish our branch of the anthropoid primates—the catarrhines—from the more-distantly related platyrrhines—for example, the *Cebus*, or capuchin monkeys.

locus in *Cebus* (a member of the platyrrhine family Cebidae) apparently account for the near silence of the γ^1 locus in *Cebus* fetuses (Chiu et al. 1997, and in press). Distinct promoter mutations with similar functional consequences at the γ^1 locus are also present in members of the platyrrhine family Pitheciidae. Aside from the tendency of platyrrhines to have only one functional γ-globin gene, they differ from catarrhines by up-regulating β expression in midfetal life rather than at birth (Johnson et al. 1996). Thus, for their γ and β genes, platyrrhines have expression patterns that are intermediate between those of strepsirhines and those of catarrhines.

A possible factor contributing to the differences, in globin-expression patterns, between catarrhines and platyrrhines is that the intergenic ϵ-to-γ^1 distance is short (~6 kb) in platyrrhines (Chiu et al., in press; C. M. M. Meireles, M. Goodman, unpublished data) and long (13.4 kb) in catarrhines (Barrie et al. 1981; Collins and Weisman 1984) (fig. 1, stage 3). The short platyrrhine distance is the same length as the intergenic ϵ-to-γ distance in tarsier and galago. However, the distance between ϵ and γ^2 in platyrrhines is ~11.5 kb and is comparable to the distance between ϵ and γ^1 in catarrhines. Conceivably, a short distance between ϵ and γ genes, as in galago, allows the LCR to activate both genes in embryonic life, whereas a longer distance between ϵ and γ genes, as between ϵ and γ^2 in platyrrhines or between ϵ and γ^1 in catarrhines, reduces the chances of activating that γ gene during embryonic life but increases the chances of activating it during fetal life.

The first steps in the transition from an embryonic γ gene to the fetally expressed γ genes of present-day anthropoids clearly began in a common ancestor of platyrrhines and catarrhines. A likely triggering event was the tandem duplication of a 5.5-kb DNA fragment containing the γ-globin gene. The tandem duplication took place in the stem-anthropoid lineage after its separation from the tarsiiform lineage but before its divergence into platyrrhines and catarrhines. In sequences flanking the stem-anthropoid γ gene, insertions of two truncated but homologous LINE elements occurred, one (L1a) upstream and the other (L1b) downstream of the γ gene; a crossover, in misaligned chromatids, between the two homologous L1 elements then produced the 11-kb-long tandem duplicate 5′-L1a-γ^1-L1ba-γ^2-L1b-3′ (Fitch et al. 1991). Also, a burst of promoter and coding-region base substitutions occurred in the evolving stem-anthropoid γ genes, and most of these base substitutions were subsequently retained in the further evolution of platyrrhine and catarrhine primates (Goodman et al. 1996). A consequence at the protein level was loss of 2,3-diphosphoglycerate binding ability, resulting in a fetal hemoglobin molecule that binds oxygen with increased affinity, facilitating the transfer of oxygen from mother to fetus. Such a fetal hemoglobin could have helped make possible the prolonged intrauterine fetal life and extensive prenatal brain development of anthropoid primates.

Physiological Consequences of Globin Gene Evolution

Functional studies have provided evidence that base substitutions in *cis*-regulatory elements caused or permitted the anthropoid γ genes to be fetal genes rather than exclusively embryonic. An analysis of galago and human γ genes in transgenic mice has demonstrated that the *cis* differences between galago and human sequences in a 4.0-kb region surrounding the γ gene resulted in distinctly different expression patterns: the galago γ transgene expression was embryonic and was silenced in the mouse fetal liver, whereas the activity of the human γ transgene peaked in fetal life (TomHon et al. 1997). Using a strategy called differential phylogenetic footprinting, Gumucio et al. (1994) identified some *cis* changes that are good candidates for further investigation in transgenic studies; in particular, they found evidence of a two-step change in the nucleotide sequence of the γ promoter–proximal CCAAT box region. In the first step, anthropoid-specific base substitutions occurred in the stem of the anthropoids (i.e., in the common ancestor of platyrrhines and catarrhines). In the second step, catarrhine-specific base substitutions occurred in the stem of the catarrhines (i.e., in the common ancestor of cercopithecids and hominids). Each step resulted in an alteration in the binding affinity of a set of putative fetal repressor proteins: these proteins bind galago and lemur sequences with high affinity, platyrrhine sequences with moderate affinity, and human sequences with very low affinity (Gumucio et al. 1994). Anthropoid-specific base substitutions also occurred in the γ promoter -50 region; these substitutions permitted the binding of the fetal activator, which is called "stage-specific protein" and favors the competitive expression of γ over β during fetal life (Jane et al. 1992).

Just as the LINE insertions in the 5′ and 3′ flanking sequences of the early stem-anthropoid γ gene may have initiated the train of events that led from embryonic to fetal expression, additional LINE-sequence insertions (Smit et al. 1995), which more than doubled the intergenic ϵ-to-γ^1 distance in the stem-catarrhine lineage (fig. 1, stage 3), may have been responsible for the difference between catarrhines and platyrrhines, as to which of the two linked γ genes is most expressed during fetal life: γ^1 in catarrhines but γ^2 in platyrrhines. In addition to *cis*-regulatory mutations, changes in the distance between linked genes could, under some circumstances, alter the levels of expression of these genes. These circumstances would revolve around whether the change in distance increased or decreased the chances for pro-

moter/LCR interaction at that developmental stage. Ideas and experimental evidence on the central role of the LCR in the enhancement of transcription implicate both the linkage order of the different β-type globin genes and the distances of these genes from the LCR as factors in the timing of developmental expression (Hanscombe et al. 1991; Peterson and Stamatoyannopoulos 1993; Wijgerde et al. 1995). Building upon the concept that the distances of ϵ and γ genes from each other and from the LCR are a factor in the timing of developmental expression, Chiu et al. (1997, and in press) have presented a model for the emergence and subsequent evolution of fetal γ expression patterns in the anthropoid primates. The model attempts to make sense of the various findings, reviewed above, on LINE insertions, gene duplications, *cis*-regulatory mutations, distances between ϵ and γ genes, and temporal expression patterns. This model may be briefly described as follows.

The early stem-anthropoid γ-globin gene (at a distance of ~6 kb from the ϵ-globin gene) functioned, like ϵ, as an embryonic gene—that is, the promoter/LCR interaction could "flip flop" readily between the two genes, permitting both γ and ϵ to be highly expressed during embryonic life. Immediately after the stem-anthropoid γ gene tandemly duplicated, the γ^1 gene (at a distance of 6 kb from ϵ) was still induced, by the LCR, to function as an embryonic gene, whereas the γ^2 gene (at a distance of 11–12 kb from ϵ) was poorly expressed during embryonic life (being outcompeted by ϵ and γ^1, for promoter/LCR interaction) and was actively repressed during all of fetal and postnatal life. Being relatively silent—and thus not scrutinized by purifying selection—the nascent γ^2 locus freely accumulated base changes, including those that specified fetal expression (by disrupting binding of fetal repressors and, conversely, by permitting binding of fetal activators). These *cis*-regulatory changes, as well as coding-sequence changes that favored the transport of oxygen from maternal hemoglobin to the new fetal hemoglobin, were positively selected. Such γ^2 changes were transferred to γ^1 by gene conversion (see Schimenti 1999 [in this issue]). However, the short distance between ϵ and γ^1, in the context of the longer distance between ϵ and γ^2, might not have been optimal for either embryonic or fetal expression, interfering with the LCR's interaction with both ϵ, during embryonic life, and γ^2, during fetal life. In most platyrrhine clades, γ^1 accumulated mutations that silenced or drastically reduced its expression. The tendency in platyrrhines to have only one functional γ gene may have been positively selected. In contrast, the increase of the intergenic ϵ-to-γ^1 distance, from 6 to 13 kb, in the stem catarrhines placed the γ^1 gene at an optimal distance for LCR interaction during fetal life, and γ^1 is the primary fetally expressed γ gene both in hominids (Bunn and Forget 1986) and in cercopithecids (R. M. Johnson, C.-H. Chiu, M. Goodman, unpublished data). Also, the delay in the γ-to-β switch, to a later developmental stage in catarrhines (birth) than in platyrrhines (midgestation), correlates with the distance from the primary fetal γ gene to the β gene being longer in catarrhines than in platyrrhines. To explore a possibly related aspect of this puzzle, it is important to find out if there are catarrhine-specific mutations in *cis* regulators of the β-locus of cercopithecids and hominids. If there are, then these mutations could account, in part, for the difference, in β-expression patterns, between catarrhines and platyrrhines—for example, such *cis* mutations might disrupt binding of a fetal activator protein or, conversely, permit binding of a fetal suppressor protein.

Classification Revisited: What Counts in Evolution?

The goal of completely sequencing a human genome is likely to be reached within the next 3–6 years. This will facilitate the sequencing of chimpanzee and bonobo genomes and, ultimately, the genomes of a series of other primates, a long-range goal envisioned in the call for a human-genome-evolution project. By application of techniques such as phylogenetic and differential phylogenetic footprinting (Gumucio et al. 1996) to the aligned genome sequences, it should be possible to identify not only the course of nonsynonymous substitutions that have shaped the proteins encoded by each of the 100,000 or so human genes but also each gene's *cis*-regulatory elements.

This analysis will highlight both conserved elements, common to primates and other eutherians, and those evolutionarily later elements—for example, those common only to anthropoids or those common only to humans, chimpanzees, and bonobos. Tracing the course of functionally significant sequence changes from the reconstructed ancestral primate genome to each sequenced extant-primate genome will then provide the data needed to test objectively the assumption made by advocates of traditional primate classifications: that chimpanzees, bonobos, and gorillas are evolutionarily closer to orangutans than to humans. My guess is that this assumption will not withstand scrutiny. Rather, I believe that, if divisions are based on molecular features that lead to distinctive physiological and developmental patterns, then not only will the African great apes be found to be closer to humans than to orangutans but chimpanzees and bonobos will prove to be closer to humans than to gorillas. If such comparisons of gene regulation are indeed consistent with our present molecular data on noncoding DNA sequences, we will have a powerful additional rationale for a phylogenetic classification that places chimpanzees and bonobos along with our species in the genus *Homo*.

Acknowledgments

This work has been supported by grants from the National Science Foundation (INT-9602913) and the National Institutes of Health (HL33940). The assistance of Elaine Krobock and Scott L. Page in formatting the Appendix and figure 1 is much appreciated. Also, discussions with Deborah Gumucio and Chi-Hua Chiu helped frame the model of β-globin gene–cluster regulatory evolution presented here.

Appendix

Age-Related Molecular Phylogenetic Classification of Living Primates

Taxa above the genus level that are referred to in the text are indicated here in boldface.

Order **Primates** (63 Ma)
- Semiorder **Strepsirhini** (50 Ma)
  - Suborder Lemuriformes (45 Ma)
    - Infraorder Chiromyiformes
      - Family Daubentoniidae, *Daubentonia:* aye-ayes
    - Infraorder Eulemurides
      - Superfamily Lemuroidea (28 Ma)
        - Family Cheirogaleidae (22 Ma)
          - Subfamily Microcebinae, *Microcebus:* mouse lemurs
          - Subfamily Cheirogaleinae, *Cheirogaleus:* dwarf lemurs
        - Family Indridae, *Propithecus:* sifakas
        - Family Lemuridae, *Eulemur:* brown lemurs
  - Suborder **Loriformes**
    - Family Loridae (23 Ma)
      - Subfamily Galagoninae, *Otolemur:* bush babies
      - Subfamily Perodicticinae, *Perodicticus:* pottos
      - Subfamily Lorinae, *Nycticebus:* slow lorises
- Semiorder **Haplorhini** (58 Ma)
  - Suborder **Tarsiiformes**
    - Family Tarsiidae, *Tarsius:* tarsiers
  - Suborder **Anthropoidea** (40 Ma)
    - Infraorder **Platyrrhini**
      - Superfamily Ceboidea (26 Ma)
        - Family **Cebidae** (22 Ma)
          - Subfamily Cebinae (20 Ma)
            - Tribe Cebini, *Cebus:* capuchin monkeys
            - Tribe Saimiriini, *Saimiri:* squirrel monkeys
          - Subfamily Aotinae, *Aotus:* night monkeys
          - Subfamily Callitrichinae
            - Tribe Callitrichini (13 Ma)
              - Subtribe Saguinina, *Saguinus:* tamarins
              - Subtribe Leontopithecina, *Leontopithecus:* golden lion tamarins
              - Subtribe Callimiconina, *Callimico:* goeldi's monkeys
              - Subtribe Callitrichina
                - *Callithrix* (*Callithrix*)*:* marmosets (*jacchus* group)
                - *Callithrix* (*Cebuella*)*:* pygmy marmosets
                - *Callithrix* (*Mico*)*:* marmosets (*argentata* group)
        - Family **Pitheciidae**
          - Subfamily Pitheciinae (19 Ma)
            - Tribe Callicebini, *Callicebus:* titi monkeys
            - Tribe Pitheciini
              - Subtribe Pitheciina (11 Ma)
                - *Pithecia:* saki monkeys
                - *Chiropotes* (*Chiropotes*)*:* bearded saki monkeys

Chiropotes (*Cacajao*): uacari monkeys
Family **Atelidae**
Subfamily Atelinae (17 Ma)
Tribe Alouattini, *Alouatta:* howler monkeys
Tribe Atelini (12 Ma)
Subtribe Atelina, *Ateles:* spider monkeys
Subtribe Brachytelina (10 Ma)
Lagothrix: woolly monkeys
Brachyteles: woolly spider monkeys
Infraorder **Catarrhini**
Superfamily Cercopithecoidea (25 Ma)
Family **Cercopithecidae**
Subfamily Cercopithecinae (15 Ma)
Tribe Colobini (10 Ma)
Subtribe Colobina, *Colobus:* colobus monkeys
Subtribe Presbytina (7 Ma)
Trachypithecus: langurs
Nasalis: proboscis monkeys
Tribe Cercopithecini (10 Ma)
Subtribe Cercopithecina (9 Ma)
Cercopithecus: guenons
Erythrocebus: patas monkeys
Chlorocebus: green monkeys
Subtribe Papionina (9 Ma)
Macaca: macaques
Cercocebus (*Cercocebus*): terrestrial mangabeys
Cercocebus (*Mandrillus*): mandrills, drills
Papio (*Papio*): baboons (hamadryas group)
Papio (*Theropithecus*): gelada baboons
Papio (*Lophocebus*): arboreal mangabeys
Family **Hominidae**
Subfamily **Homininae** (18 Ma)
Tribe **Hylobatini**
Subtribe **Hylobatina** (8 Ma)
Symphalangus: siamangs
Hylobates: gibbons
Tribe **Hominini** (14 Ma)
Subtribe **Pongina**, *Pongo:* orangutans
Subtribe **Hominina** (7 Ma)
Gorilla: gorillas
Homo (*Homo*): humans
Homo (*Pan*): chimpanzees, bonobos

References

Additional relevant references are included in a Supplemental Reading List, which appears, in the electronic version of this article, immediately after the References.

Bailey WJ, Hayasaka K, Skinner CG, Kehoe S, Sieu LC, Slightomn JL, Goodman M (1992) Reexamination of the African hominoid trichotomy with additional sequences from the primate β-globin gene cluster. Mol Phylogenet Evol 1: 97–135

Barrie PA, Jeffreys AJ, Scott AF (1981) Evolution of the β-globin gene cluster in man and the primates. J Mol Biol 149: 319–336

Bunn HF, Forget BG (1986) Hemoglobin: molecular genetic and clinical aspects. WB Saunders, Philadelphia

Chiu C-H, 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 (in press)

Chiu C-H, Schneider H, Slightom JL, Gumucio DL, Goodman M (1997) Dynamics of regulatory evolution in primate β-globin gene clusters: *cis*-mediated acquisition of simian γ fetal expression patterns. Gene 205:47–57

Collins FS, Weissman SM (1984) The molecular genetics of human hemoglobin. Prog Nucleic Acid Res Mol Biol 31: 315–462

Cooper STB, Murphy R, Dohman G, Hussey D, Hope RM (1996) A molecular and evolutionary study of the β-globin gene family of the Australian marsupial *Sminthopis crassicaudata*. Mol Biol Evol 13:1012–1022

De Waal FBM (1995) Bonobo sex and society. Sci Am 272: 82–88

Fitch DHA, Bailey WJ, Tagle DA, Goodman M, Sieu L, Slightom JL (1991) Duplication of the γ-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates. Proc Natl Acad Sci USA 88: 7396–7400

Fouts R (1997) Next of kin. William Morrow, New York

Gibbons A (1998) Which of our genes makes us human? Science 281:1432–1434

Goodman M, Koop BF, Czelusniak J, Weiss ML, Slightom JL (1984) The η-globin gene: its long evolutionary history in the β-globin gene family of mammals. J Mol Biol 180: 803–823

Goodman M, Page SL, Meireles CM, Czelusniak J. Primate phylogeny and classification elucidated at the molecular level. In: Wasser SP (ed) Evolutionary theory and processes: modern perspectives. Kluwer Academic, Dordrecht, The Netherlands (in press)

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, et al (1998) Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9:585–598

Goodman M, Slightom JL, Gumucio DL (1996) Molecular evolution in the β-globin gene family of mammals, emergence of redundant genes, important new genes, and new expression patterns. In: Holmes RS, Lim HA (eds) Gene families: structure, function, genetics, and evolution. World Scientific, Singapore, pp 43–52

Groves CP (1993) Order primates. In: Wilson DE, Reader DM (eds) Mammalian species of the world: a taxonomic and geographic reference, 2d ed. Smithsonian Institution Press, Washington, DC, pp 243–277

Gumucio DL, Shelton DA, Blanchard-McQuate K, Gray T, Tarle S, Heilstedt-Williamson H, Slightom JL, et al (1994) Differential phylogenetic footprinting as a means to identify base changes responsible for recruitment of the anthropoid γ gene to a fetal expression pattern. J Biol Chem 269: 15371–15380

Gumucio DL, Shelton DA, Zhu W, Millinoff D, Gray T, Bock JH, Slightom JL, et al (1996) Evolutionary strategies for the elucidation of *cis* and *trans* factors that regulate the developmental switching programs of the β-like globin genes. Mol Phylogenet Evol 5:18–32

Hanscombe O, Whyatt D, Fraser P, Yannoutsos N, Greaves D, Dillon N, Grosveld F (1991) Importance of globin gene order for correct developmental expression. Genes Dev 5: 1387–1394

Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, Miller W (1997) Locus control regions of mammalian β-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. Gene 205:73–94

Hennig W (1966 [reissued 1979]) Phylogenetic systematics. University of Illinois Press, Urbana

Jane SM, Ney PA, Vanin EF, Gumucio DL, Nienhuis AW (1992) Identification of a stage selector element in the human γ-globin promoter that fosters preferential interaction with the 5′ HS2 enhancer when in competition with the β-promoter. EMBO J 11:2961–2969

Jeffries RSP (1979) The origin of the chordates—a methodological essay. In: House MR (ed) The origin of major invertebrate groups. Academic Press, London, pp 443–477

Johnson RM, Buck S, Chiu C-H, Schneider H, Sampaio I, Gage DA, et al (1996) Fetal globin expression in New World monkeys. J Biol Chem 271:14684–14691

Koop BF, Goodman M (1988) Evolutionary and developmental aspects of two hemoglobin β chain genes (ϵ^M and β^M) of opossum. Proc Natl Acad Sci USA 85:3893–3897

Martin RD (1990) Primate origins and evolution: a phylogenetic reconstruction. Chapman & Hall, London

McConkey EH, Goodman M (1997) A human evolution project is needed. Trends Genet 13:350–351

McGrew WC (1992) Chimpanzee material culture—implications for human evolution. Cambridge University Press, Cambridge

McGrew WC, Marchant LF, Nishida T (1997) Great ape societies. Cambridge University Press, Cambridge

Meireles CMM, Schneider MPC, Sampaio MIC, Schneider H, Slightom JL, Chiu C-H, Neiswanger K, et al (1995) Fate of a redundant γ-globin gene in the atelid clade of New World monkeys: implications concerning fetal globin gene expression. Proc Natl Acad Sci USA 92:2607–2611

Peterson KR, Stamatoyannopoulos G (1993) Role of gene order in developmental control of human γ- and β-globin gene expression. Mol Cell Biol 13:4836–4843

Romer AS (1962) The vertebrate body. WB Saunders, Philadelphia

Savage-Rumbaugh S, Lewin R (1994) Kanzi—the ape at the brink of the human mind. John Wiley, New York

Schimenti JC (1999) Mice and the role of unequal recombination in gene-family evolution. Am J Hum Genet 64:40-45 (in this issue)

Smit AFA, Toth C, Riggs AD, Jurka J (1995) Ancestral mammalian-wide subfamilies LINE-1 repetitive sequences. J Mol Biol 246:401–417

Tagle DA, Koop BF, Goodman M, Slightom JL, Hess D, Jones RT (1988) Embryonicε and γ globin genes of a prosimian primate (*Galago crassicaudatus*): nucleotide and amino acid sequences, developmental regulation, and phylogenetic footprints. J Mol Biol 203:439–455

TomHon C, Zhu W, Millinoff D, Hayasaka K, Slightom JL, Goodman M, Gumucio DL (1997) Evolution of a fetal expression pattern via *cis* changes near the γ globin gene. J Biol Chem 272:14062–14066

Wijgerde M, Grosveld F, Fraser P (1995) Transcription complex stability and chromatin dynamics *in vivo*. Nature 377: 209–213