Functionally important glycosyltransferase gain and loss during catarrhine primate emergence

Chihiro Koike[†‡], Monica Uddin[§], Derek E. Wildman[§¶], Edward A. Gray[†‡], Massimo Trucco[††], Thomas E. Starzl[†‡], and Morris Goodman[§‡‡§§]

[†]Department of Surgery and [‡]Thomas E. Starzl Transplantation Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213; [§]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201; [††]Division of Immunogenetics, Department of Pediatrics, Children's Hospital of Pittsburgh, University of Pittsburgh, Pittsburgh, PA 15213; Departments of [‡‡]Anatomy and Cell Biology and [¶]Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, MI 48201; and [∥]Perinatology Research Branch, National Institute of Child Health and Human Development, National Institutes of Health/Department of Health and Human Services, Bethesda, MD 20892

Contributed by Morris Goodman, November 10, 2006 (sent for review October 18, 2006)

A glycosyltransferase, α1,3galactosyltransferase, catalyzes the terminal step in biosynthesis of Galα1,3Galβ1–4GlcNAc-R (αGal), an oligosaccharide cell surface epitope. This epitope or antigenically similar epitopes are widely distributed among the different forms of life. Although abundant in most mammals, αGal is not normally found in catarrhine primates (Old World monkeys and apes, including humans), all of which produce anti-αGal antibodies from infancy onward. Natural selection favoring enhanced resistance to αGal-positive pathogens has been the primary reason offered to account for the loss of αGal in catarrhines. Here, we question the primacy of this immune defense hypothesis with results that elucidate the evolutionary history of *GGTA1* gene and pseudogene loci. One such locus, *GGTA1P*, a processed (intronless) pseudogene (PPG), is present in platyrrhines, i.e., New World monkeys, and catarrhines but not in prosimians. PPG arose in an early ancestor of anthropoids (catarrhines and platyrrhines), and *GGTA1* itself became an unprocessed pseudogene in the late catarrhine stem lineage. Strong purifying selection, denoted by low nonsynonymous substitutions per nonsynonymous site/synonymous substitutions per synonymous site values, preserved *GGTA1* in noncatarrhine mammals, indicating that the functional gene product is subjected to considerable physiological constraint. Thus, we propose that a pattern of alternative and/or more beneficial glycosyltransferase activity had to first evolve in the stem catarrhines before *GGTA1* inactivation could occur. Enhanced defense against αGal-positive pathogens could then have accelerated the replacement of αGal-positive catarrhines by αGal-negative catarrhines. However, we emphasize that positively selected regulatory changes in sugar chain metabolism might well have contributed in a major way to catarrhine origins.

adaptive evolution | glycobiology | pseudogene

The catarrhine primates (Old World monkeys and apes, including humans) are distinguished from other primates by a suite of biological characteristics among which are relatively long life spans, extended periods of growth and development, long generation times, large body sizes, atrophied vomeronasal organs but trichromatic vision, and increased encephalization (1–4). Unearthing the genetic, epigenetic, environmental, and behavioral changes associated with the emergence of these features is a major goal of evolutionary primatologists. It is evident from the fossil record and molecular divergence estimates that many of the genetic changes associated with the emergence of catarrhines occurred between 40 and 25 million years ago (mya) (5). Some of the genetic changes during this period of stem-catarrhine evolution drastically altered the prevalence of different cell surface sugar chains in tissues. In this article, we focus on elucidating the evolutionary history of the genetic locus, *GGTA1*, that was involved in an especially enigmatic alteration, the loss of the cell-surface sugar chain epitope termed αGal (Galα1,3Gal β1–4GlcNAc-R). The *GGTA1* locus encodes a particular glycosyltransferase, the Golgi transmembrane-bound

α1,3 galactosyltransferase (α1,3GT, EC 2.4.1.87) that catalyzes the terminal step in biosynthesis of the αGal epitope (6).

With the sole exception of catarrhines (7), the αGal epitope is expressed on the surface of cells in all mammalian species examined to date, including artiodactyls, carnivores, rodents, prosimians, and platyrrhines (New World monkeys) (7, 8). The epitope is abundant on such different cells as those of the vascular system (8), the digestive mucosa (9,) and the vomeronasal sensory epithelium (10). The αGal-negative catarrhines produce high titers of anti-αGal antibodies from infancy onward (11). These antibodies are similar to the anti-A and anti-B isoagglutinins of humans who lack the A or B antigens of the ABH histo-blood type oligosachharides. Because of the high titer of catarrhine anti-αGal antibodies, immediate (hyperacute) rejection occurs when organs, tissues, or cells are transplanted from αGal-positive species to αGal-negative catarrhines (12).

The molecular basis for αGal expression was established by the cloning of bovine (13), mouse (14), and porcine (15) *GGTA1* cDNA, by examination of the mouse (16) and porcine (17–20) *GGTA1* genomic organization, and by production of *GGTA1* knockout mice (21) and pigs (22). Insight into why catarrhines lack a functional α1,3GT enzyme was forestalled by the shortage of information on the catarrhine *GGTA1* coding region. The initial information was limited to partial sequences and chromosomal locations of an unprocessed human pseudogene and a processed (i.e., intronless) human pseudogene (23, 24), i.e., an orthologue and a paralogue of *GGTA1*. Based on the finding that the unprocessed pseudogene sequence shared the same in-frame termination codon with the processed pseudogene sequence, it was assumed that the processed pseudogene was generated from an already inactivated source gene (8, 23). It was further assumed from partial catarrhine sequences (25) that the *GGTA1* locus became an unprocessed pseudogene twice, once in apes and separately in Old World monkeys. Because the αGal epitope or antigenically similar epitopes occur widely among the different forms of life, including

© 2006 by The National Academy of Sciences of the USA

EVOLUTION

pathogens, it has been suggested that the primary reason for why the catarrhines would lose αGal is that there would be an immunological advantage afforded by anti-αGal antibodies (8, 23, 25–27). There might also be enhanced defense against those pathogens that have binding sites for αGal epitopes, because these pathogens might then not be able to attach themselves to the αGal-negative cells of catarrhines.

The molecular basis for the absence of αGal epitope expression in catarrhine primates was definitively elucidated by delineation of the full coding region and exon–intron structure of the unprocessed pseudogene (termed UPG) in rhesus, orangutan, and human genomes (28). In contrast to the hypothesis of independent inactivation in apes and Old World monkeys, the presence of two shared derived substitutions in the UPG of the three catarrhine species indicated that *GGTA1* inactivation occurred in the stem lineage of catarrhines (28). Sequence comparisons also indicated that the origin of the processed pseudogene (*GGTA1P*, termed PPG) preceded the origin of UPG, in contradiction to the previous view that the PPG arose from the UPG (8, 23).

The uncertainty as to PPG's origin has been resolved by the finding in this study that the PPG locus is present in both platyrrhines and catarrhines but not in the prosimians that are strepsirrhine primates. We identified the expressed *GGTA1* locus and sequenced cDNA representing this locus in two strepsirrhines (lemur and loris) and partially sequenced such cDNA in the haplorhine prosimian (tarsier). With a data set of *GGTA1*, UPG, and PPG orthologues and paralogues, we reconstructed the evolutionary history of these sequences, dating not only when the PPG arose in the stem anthropoids but also when the active gene became a pseudogene, i.e., UPG, in the stem catarrhines. Our data contradict the view that the active gene was relatively neutral to natural selection. Instead, our data suggest that purifying selection pressure favored retention of the *GGTA1* gene in noncatarrhine mammals even though they then lacked the possible immunological advantage that would be afforded by producing anti-αGal antibodies. Furthermore, we infer that pseudogenization of the functional *GGTA1* gene in catarrhines became possible only after alternative and/or more beneficial glycosyltransferase activity evolved in the stem lineage of the catarrhine primates.

Results

Alignment File. In this study, full coding region *GGTA1* sequences were generated from lemur, loris, and howler monkey samples, and PPG sequences were generated from capuchin, marmoset, rhesus, orangutan, and chimpanzee samples. Also, a partial coding region sequence from a putative *GGTA1* locus was generated from a tarsier sample. These sequences and previously reported sequences for the active *GGTA1* gene and the PPG and UPG homologues [see supporting information (SI) Table 2] were aligned for subsequent phylogenetic analyses. An alignment file containing the full set of *GGTA1*, UPG, and PPG nucleotide sequences referred to in this study is available in Nexus format as SI Data Set 1. The species with the active α1,3GT enzyme show many conserved amino acid residues, among which are the 16 that have been previously described as required for the enzyme to function properly (29–32). These 16 residues are identical in all species with an active *GGTA1* coding region except in the howler monkey, which differed at three of the critical residues: S207C, R210S, and H288Y (numbered according to the marmoset amino acid sequence). The species with an inactivated *GGTA1* coding region, i.e., all catarrhine UPG sequences, show in the alignment file a single nucleotide position gap located where the marmoset sequence codes for amino acid residue 81. This gap designates a frame shift deletion just 5′ of the coding sequence region specifying the enzyme's catalytic domain (28). Thus, if any transcribed message with this frame shift deletion had been translated in the stem catarrhines, the resulting polypeptide would not have been functional.

Nonsynonymous to Synonymous Rates (ω). Phylogenetic Analysis by Maximum Likelihood (PAML) (33) rejected the hypothesis ($\chi 2$ = 116.44, 37 df, $P < 0.0001$) of a molecular clock under the "one-ratio" M0 model ($-\ln L = 7014.74$) with a single ω [i.e., nonsynonymous substitutions per nonsynonymous site (dN)/synonymous substitutions per synonymous site (dS)] value on all branches in favor of a "free ratio" M1 model ($-\ln L = 6956.52$) that allowed the ω value to vary on every branch. The M1 tree is depicted in Fig. 1. In general, the ω values are well below 1 for all lineages in which *GGTA1* was active. This pattern of purifying selection is consistent with the hypothesis that *GGTA1* is a functionally important gene. The highest ω values for active lineages are on the lineage leading to the howler monkey and on the stem primate lineage. The ω values for the UPG and PPG are much more compatible with neutrally evolving lineages, with ω values that tend toward a value of 1. Branch test results comparing the M0 ($-\ln L = 4766.92$) and M2 ($-\ln l = 4764.96$) models for the active gene data set were significantly different ($\chi 2$ = 3.93, 1 df, $P < 0.05$), with the howler lineage showing a dN/dS ratio more than two times higher than the ratio estimated for the rest of the tree (0.64 vs. 0.28, respectively). Maximum parsimony (MP) and maximum likelihood (ML)-reconstructed ancestral sequences yielded dN/dS ratios that were similar to those obtained by PAML analysis except that the primate stem showed a lower ratio (0.36 and 0.20 for MP and ML, respectively) (data not shown), indicative of an active gene evolving under strong purifying selection.

Phylogenetic Analyses. The optimal MP, ML, and Bayesian phylogenetic results are in close agreement (Fig. 1). Parsimony searches recovered three most parsimonious trees of length 1,121, and the only source of incongruence among these topologies is the branching order of the *GGTA1* sequences within the New World monkey clade. The optimal ML topology agrees with both the MP and Bayesian topologies. Branch support for all clades (SI Table 3) is generally quite high.

Time of PPG and UPG Origin. The PPG lineage is estimated to have originated 57.6 mya, indicating an appearance on the stem anthropoid lineage at a date close to that of the tarsier-anthropoid last common ancestor (LCA): i.e., the crown haplorhine LCA. The UPG lineage is estimated to have originated 28.1 mya, supporting the inference that the conversion of the *GGTA1* to the UPG locus preceded the crown catarrhine LCA by only a few million years.

Tarsier Sequence. This study also provides sequence information from the genus *Tarsius* for its putative *GGTA1* locus. It is not certain whether the sequences obtained represent PPG or *GGTA1* sequences. In favor of the former, phylogenetic analyses group the tarsier sequences with sequences from the PPG clade, albeit with low branch support (SI Table 4). This finding is consistent with the calculated appearance of the PPG at a time close to the haplorhine LCA (see above). Conversely, there are no disruptive mutations in the sequence that would cause either reading frameshifts or premature termination codons. Also, in the phylogenetic tree that included the partial tarsier sequence, the tarsier lineage has a low ω value (0.21, SI Fig. 3), which indicates a functional gene under purifying selection. It will be necessary to obtain either intronic or full-length cDNA sequence data to fully determine whether the tarsier possesses an active *GGTA1* locus.

Evolutionary Rates. The "Hominoid Slowdown" hypothesis was originally proposed on the basis of data that revealed very small degrees of serum albumin antigenic divergence among hominoids (apes, including humans) and also on the basis of ideas concerning such hominoid organismal features as lengthened gestations and increased generation times (34–36). That a slowdown in rates of molecular evolution occurred has been affirmed in numerous studies since then (37–42). A comparison of rates calculated from
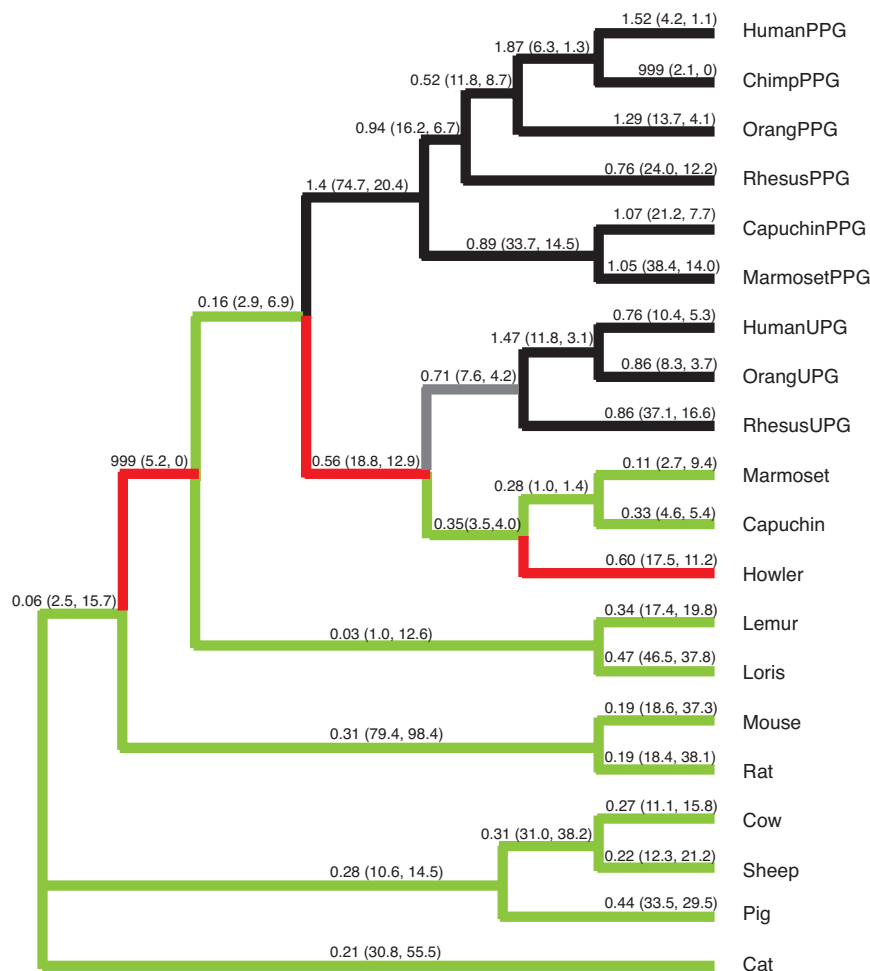
Fig. 1. PAML-estimated ω (i.e., dN/dS) and (N*dN, S*dS) values under the M1 ''free ratio'' model, which allows the ω value to vary on each branch. Estimated number of nonsynonymous and synonymous sites under this model are 827.4 and 318.6, respectively. Branches leading to unprocessed (UPG) and processed (PPG) loci are depicted in black. Branches leading to *GGTA1* genes are depicted in color, with green indicating those branches with dN/dS values <0.5 (i.e., purifying selection) and red indicating those branches with dN/dS values >0.5 (i.e., relaxed purifying selection). Depicted in gray is the branch on which the UPG locus originated. Tree topology also reflects the optimal maximum likelihood and Bayesian phylogenetic results (−lnL = 7125.71 and 7140.76, respectively); maximum parsimony bootstrap results differed only in showing a trichotomy among the active New World monkey genes.

the PPG clade supports the concept of a rate slowdown in anthropoid lineages, especially in the catarrhines (Table 1). The UPG rates also show the hominoid rate being slower than the Old World monkey rate.

Discussion

From our results, we can infer that, in noncatarrhine mammals, the *GGTA1* gene functions under strong physiological constraints. The reconstructed evolutionary history of this gene (Fig. 2) reveals that, in the stem lineage to anthropoid primates, near the time of the tarsier–anthropoid LCA (≈57–58 mya), *GGTA1* gave rise to a processed, retrotranscribed pseudogene. The two paralogously related loci (*GGTA1* and PPG) then coexisted in the same anthropoid lineages but evolved independently of each other. The retrotranscribed PPG locus freely accumulated mutations that escaped the scrutiny of natural selection, whereas the parent *GGTA1* gene did not escape such scrutiny in the anthropoid stem and platyrrhine lineages but, nevertheless, did become a UPG in the late catarrhine stem lineage (≈28 mya). Thereafter, the UPG sequences accumulated mutations freely in catarrhine lineages.

Despite being freed of physiological constraints, both the PPG and UPG sequences have nucleotide substitution rates that slowed in hominoid lineages. This slowdown indicates that rates of occur-

rence of new mutations decreased during the descent of hominoids. In turn, the decreased *de novo* mutation rates probably resulted directly from lengthened generation times and indirectly from the whole suite of adaptive changes that shaped catarrhine origins and evolution. We suspect that among the adaptive changes were those that brought about a new pattern of sugar chain expression that then permitted inactivation of the *GGTA1* gene.

If the production of anti-αGal antibodies was the crucial survival advantage worth loss of the α1,3GT enzyme, inactivation of the *GGTA1* gene should have occurred in many mammalian species. Yet over the past 100 or so million years, no examples of αGal-negative lineages are known to have arisen among noncatarrhine mammals (13–20, 28, 43, and this study). Indeed, strong purifying selection with preservation of the functional *GGTA1* locus in noncatarrhine mammals is evident from our finding that the αGal-positive lineages accumulated nonsynonymous substitutions at this locus much more slowly than synonymous substitutions (Fig. 1).

That loss of the *GGTA1* gene can be deleterious has become evident by the production of *GGTA1* KO mice and pigs for xenotransplantation research (21, 22, 44). Animals with *GGTA1* KO have both subtle and conspicuous health abnormalities. For example, the KO mice develop early-onset cataracts (21). The

Table 1. Evolutionary rates in selectively neutral DNA

Time period/lineage	Total changes*	Nucleotide substitution rate†
Rates from PPG data		
57.6–40 mya	95.1	4.7
40 mya, marmoset	100.6	2.2
40 mya, capuchin	77.1	1.7
40 mya, rhesus	59.1	1.3
40 mya, orangutan	61.2	1.3
40 mya, chimpanzee	53.1	1.2
40 mya, human	56.3	1.2
25 mya, rhesus	36.2	1.3
25 mya, orangutan	38.3	1.3
25 mya, chimpanzee	30.2	1.1
25 mya, human	33.4	1.2
14 mya, orangutan	17.8	1.1
14 mya, chimpanzee	9.7	0.6
14 mya, human	12.9	0.8
6 mya, chimpanzee	2.1	0.3
6 mya, human	5.3	0.8
Rates from UPG data‡		
To rhesus UPG	53.7	1.9
To orangutan UPG	26.9	0.9
To human UPG	30.6	1.1

*Nonsynonymous and synonymous changes.

†Reported as nucleotide substitutions per million years per 1,000 nt positions.

‡Twenty-five million years ago to present.

Fig. 2. Phylogeny of the *GGTA1* locus. Shown in black are the lineages in which *GGTA1* remained active. Shown in red are the lineages in which *GGTA1* was inactivated. Blue and red arrows indicate, respectively, the origin of the processed pseudogene (PPG) and the origin of the unprocessed pseudogene (UPG). The latter event corresponds to the time of *GGTA1* inactivation that occurs in the stem-catarrhine lineage and precedes the last common ancestor of living catarrhines by only a few million years. The vertical line at the left indicates time (millions of years ago).

founder KO pigs were only a few months old and appeared to be healthy at the time they were reported (22). However, they proved to be small, frail, and difficult to breed (unpublished observations of C.K., M.T., and T.E.S., who also are coauthors of ref. 22). Such findings are consistent with the view that the *GGTA1* gene plays an important role in the endogenous metabolic homeostasis of αGal-positive species.

However, the fact that the KO is not lethal suggests that the activities of the α1,3GT enzyme can be substituted for (albeit incompletely) by another glycosyltransferase or by multiple glycosyltransferases. We propose that adaptive coevolution of other glycosyltransferase genes made such a substitution possible in the late stem-catarrhines. If the functional consequences of the replacement glycosyltransferase activity were more beneficial than the original α1,3GT activity, that could have provided the major impetus for *GGTA1* gene inactivation. However, the inactivation would still leave open the possibility that pathogen resistance played a substantial role in *GGTA1* evolution.

No doubt, the evolution of glycosyltransferase gene families (26, 45–48) involved not only changes in enzyme encoding sequences but also in promoter and other regulatory sequences. Such changes would be expected to result in species-specific patterns of developmental and tissue expression. For example, the ABH epitopes are expressed in the digestive mucosa of all mammalian species but are not expressed on the vascular endothelial cells (VEC) of noncatarrhine mammals (49). All catarrhines express ABH epitopes on VEC. Apes also express ABH epitopes on RBC but Old World monkeys do not (49). Another example of species differences in oligosaccharide tissue expression patterns involves the sialic acid moiety. This moiety is found at higher concentrations in the brains of apes than in other mammals and, among apes, is more abundant in human than in chimpanzee brains (50). A striking finding is the loss of the sialic acid moiety *N*-glycolylneuraminic acid (Neu5Gc) in humans and its replacement by a greater abundance
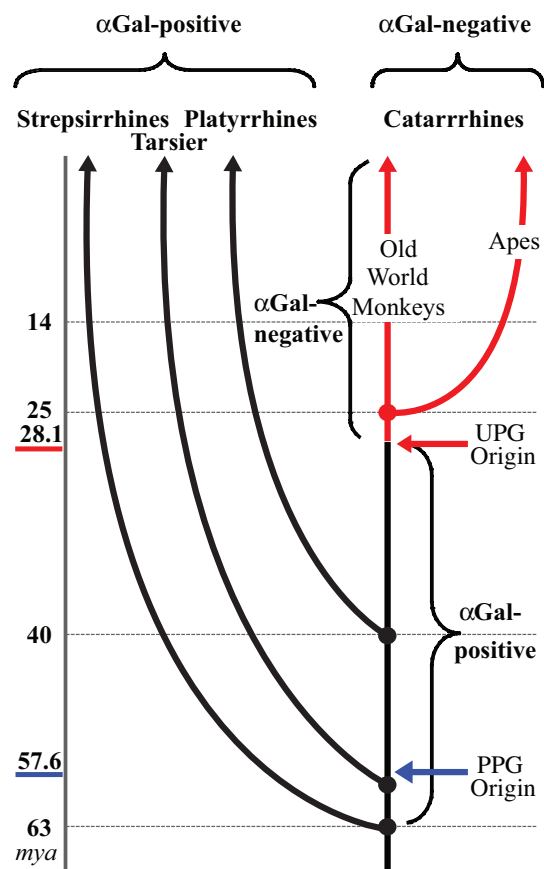
of its precursor, *N*-acetylneuraminic acid (Neu5Ac) (51). This human-specific loss resulted from an inactivating mutation in the gene encoding cytidine-monophosphate *N*-acetylneuraminic acid hydroxylase (CMAH) (51). Siglecs, i.e., receptors that can interact with the cell-surface sialic acid oligosaccharides, also show striking evolutionary changes in descent of catarrhine lineages (51, 52).

Clearly, the changes associated first with catarrhine primate emergence and then with later ape lineages involved striking shifts in the expression of sugar chains. The evidence from our study indicates that maintenance of the *GGTA1* gene was an obligatory condition for survival of noncatarrhine mammals in the wild. In our metabolic hypothesis, the αGal-negative state could not be established until the activity of the α1,3GT enzyme was replaced by alternative and/or more beneficial glycosyltransferase activity, i.e., by a major adaptive change in sugar chain metabolism. The resulting production of anti-αGal antibodies may well have enhanced protection against αGal-positive pathogens (26, 27), but this added benefit from the loss of αGal would have required first the alterations in sugar-chain metabolism.

Loss of αGal in the late stem-catarrhines may have been implicated in the decreased reliance on olfaction that accompanied the greater reliance on vision. Observations made on the sensory epithelium of the vomeronasal organ (VNO) of rats (10) show that αGal epitopes are abundantly expressed in this sensory epithelium and thus presumably engaged in the VNOs important olfactory

function of detecting pheromones. Catarrhines have atrophic VNO tissue, loss of olfactory receptor genes, and decreased pheromone detection abilities (3, 53, 54). It may be significant that the only noncatarrhine lineage in our *GGTA1* analysis showing an increased dN/dS ratio is the platyrrhine lineage to the howler monkey, suggesting that purifying selection for maintenance of α1,3GT function acted less strongly on this lineage than on all other noncatarrhine lineages. Possibly, the howler monkey relies less on olfaction than on vision as suggested by Gilad *et al.* (54). The howler monkey is the only primate other than catarrhines that is known to have full trichromatic vision (4, 55–58). Examination of many other primates, especially other members of the Atelidae (the family to which howler monkeys belong) is needed to determine whether, among noncatarrhine primates, the decreased purifying selection acting on the *GGTA1* gene is specific to the howler monkey.

Finally, however, we emphasize that studies of the *GGTA1* gene cannot be interpreted in isolation. Since discovery of the ABO (H) histo-blood groups, such systems have been seen mainly from the perspective of immunological reactions to cell-surface antigens (26, 45, 46). In place of the hypothesis that loss of αGal was primarily driven by an immunologic advantage against pathogens, we propose that *GGTA1* inactivation could not be established until alternative glycosyltransferase activity replaced α1,3GT activity. This replacement involved striking shifts in sugar-chain expression patterns. In subjecting the many glycosyltransferase genes to detailed phylogenetic studies, the species chosen should reflect the great adaptive diversity that exists among primates and other mammals. The results of such studies should help test our hypothesis that positively selected changes in sugar chain metabolism coincided with and helped bring about emergence and evolution of the catarrhine primates.

Materials and Methods

Materials. Whole blood from the ring-tailed lemur (*Lemur catta*), slender loris (*Loris tardigradus*), Philippine tarsier (*Tarsier syrichta*), common marmoset (*Callithrix jacchus*), white-throated capuchin (*Cebus capucinus*), black-and-gold howler (*Alouatta caraya*), rhesus macaque (*Macaca mulatta*), orangutan (*Pongo pygmaeus*), and chimpanzee (*Pan troglodytes*) was kindly provided by the Pittsburgh Zoo (Pittsburgh, PA), University of Wisconsin (Madison, WI), University of Pittsburgh, Wayne State University, Yerkes National Primate Research Center Emory University (Atlanta, GA), or the Duke University Primate Research Center (Durham, NC).

Methods. Standard methods were used to isolate high-molecular-weight genomic DNA from the various samples. Total RNA was extracted from the samples with TRIzol reagent (GIBCO, Carlsbad, CA). Poly(A)$^+$ RNA was separated from total RNA by using the Dynabeads mRNA Purification Kit (Dynal, Oslo, Norway) according to the manufacturer's instructions.

DNA Amplification and Sequencing. GenomeWalker libraries for the respective species were constructed by using the Universal GenomeWalker Library kit (Clontech, Palo Alto, CA). Human PPG amplification was obtained with GenomeWalker-PCR. Species- or clade-specific PPG primers were designed from sequences initially obtained by using the human PPG primer set. For the lemur *GGTA1* gene, primers were designed based on sequence obtained by using the previously published (28) human UPG primer set; these newly designed primers were also used to obtain loris and tarsier gene sequences. Primer sequences used to identify the various genes are as follows: Lemur *GGTA1* gene: L9A, 5′-CATCATGCTGGACGACATCTCCAAGATGC-3′; L9B, 5′-CAAGCCTGAGAAGAGGTGGCAGGACATC-3′; L9P, 5′-GTATGCTGACTTTACGCCTCTCATAGG-3′; L9Q, 5′-GTAGCTGAGCCACTGACTGGCCCAG-3′. New World monkey PPG: M4a, 5′-AGGAGAAAATAATGAATGTCAAAGG-AAACGT-3′; M6a, 5′-ACACCCAGAAGTTGTTGACAGCG

GCAC-3′; M8p, 5′-TCTTTCCACAGCAAACCCATCAA-CCCCA-3′; M8q, 5′-GCCTTCCCACATAACCGGC-ACATTCCA-3′. Rhesus PPG: Rpa, 5′-GCTGAGTGGATGGA-TGATGGGGAGGAG-3′;Rpq,5′-CAAGCTGATCTCGAACTC-CTGACCTCACGTG-3′. Orangutan PPG: Upa, 5′-GTCAA-AGCCGATACGTTTTCCCGGCAG-3′; Upq, 5′-ACCA-TAGATTCATTCTCTCATATTACAGTGCTC-3′.

TaKaRa LA *Taq* (Takara Shuzo, Shiga, Japan) and Titanium *Taq* (Clontech) enzyme were used for all PCR experiments. The PCR thermal cycling conditions, recommended by the manufacturer, were performed on a Gene Amp System 9600 or 9700 thermocycler (PerkinElmer, Wellesley, MA).

To identify the 5′ and 3′ ends of the *GGTA1* gene or UPG transcripts of the lemur, loris, tarsier, marmoset, capuchin, rhesus, orangutan, and chimpanzee, the Marathon RACE (rapid amplification of cDNA end) libraries (Clontech) were constructed from total RNA of the different species in accordance with the manufacturer's specified protocol.

PCR products amplified by the GW-PCR, RACE-PCR, and RT-PCR were subcloned into the pCR II vector provided with the Original TA Cloning kit (Invitrogen, Carlsbad, CA). Automated fluorescent sequencing of cloned inserts was performed by using an ABI 377 DNA Sequence Analyzer (Applied Biosystems, Foster City, CA).

Sequence Analyses. Newly obtained and previously reported DNA sequences (SI Table 2) were manually aligned in MacClade 4.08 (59). Deduced amino acid sequences of *GGTA1* genes were checked for accuracy against protein sequences associated with the nucleotide GenBank accession numbers. Gene trees were inferred by using MP and ML methods as implemented in PAUP* 4.0b10 (60) and by using Bayesian methods as implemented by MrBayes 3.1 (61). For the latter two methods, the optimal model of sequence evolution was determined by using ModelTest 3.7 (62) and MrModeltest2.2 (63), respectively, before inferring optimal gene trees. ModelTest 3.7 selected TVM+Γ as the best-fit model (−lnL = 7125.71) according to the Akaike Information Criterion (AIC) (64), with the following associated parameters: Lset Base = (0.2876 0.2117 0.2446), Nst = 6, Rmat = (1.7559 5.6656 0.7355 1.7448 5.6656), Rates = γ, Shape = 1.1091, Pinvar = 0. MrModeltest2.2 selected a GTR+ Γ as the best-fit model according to the AIC, with the following model form: Prset statefreqpr = dirichlet (1, 1, 1, 1); Lset nst = 6 rates = γ.

Patterns of selection across the inferred tree for the entire data set were estimated by using PAML 3.15 (33). In addition, ancestral sequences were reconstructed by using MP (with the DELTRAN algorithm) and ML (by using the TVM+Γ model) methods for each node in the tree depicted in Fig. 1; dN/dS for each node and its descendent branch was then calculated by using MEGA 3.1 (65) with the Pamilo–Bianchi–Li method (66, 67). The existence of a molecular clock was tested by comparing the M0 model, which estimates a single ω (dN/dS) ratio for all links contained in the tree, with the M1 model, which is allowed to estimate a different ω ratio for each link contained in the tree. In addition, to test whether the appearance of color vision correlates with a pattern of selection that suggests a reduction of functional constraints, a data set containing only the active genes was analyzed by comparing the M0 model with an M2 model, which allowed one rate for the howler lineage and another for the rest of the tree. In all cases, the existence of multiple local optima was evaluated by running each model three times, with three different staring ω values (0.5, 1, and 2). Where differences existed, the value with the highest likelihood score was used to compare models by using the likelihood ratio test.

Estimating the Time of PPG and UPG Origin. The origin of the PPG and UPG sequences was estimated by using both the topology

EVOLUTION

determined by the gene tree analyses and the nonsynonymous (N) and synonymous (S) values depicted in Fig. 1. As in Goodman *et al.* (5), Goodman (68), and Wildman and Goodman (69), 63 mya was taken as the date for the LCA of the living primates, 40 mya for the LCA of living anthropoids (platyrrhines and catarrhines), 26 mya for the LCA of living platyrrhines, 25 mya for the LCA of living catarrhines, and 14 mya for the orangutan–human LCA.

PPG Origin. By using the N and S changes determined for each lineage depicted in Fig. 1, the distance (number of N plus number of S) between the primate LCA and the PPG-stem anthropoid LCA added to the distance between PPG-stem anthropoid LCA and crown anthropoid LCA was equated to 23 million years (63–40 mya); this value totaled 41.5. Then, by proportionality, the number of million years (X) was determined between: (*i*) primate LCA and PPG-stem anthropoid LCA; 23 million years/41.5 = X/9.8, where X = 5.4 million years, indicating a time of PPG origin at (63–5.4) or 57.6 mya. (*ii*) PPG-stem anthropoid LCA and crown anthropoid LCA; 23 million years/41.5 = X/31.7, where X = 17.6 million years, indicating a time of UPG origin at (40 + 17.6) or 57.6 million years.

UPG Origin. Next, the conversion of the ACT to UPG on the catarrhine stem was determined by assuming that, at the time of pseudogenization, a slow rate for N functional substitutions became a fast rate for N pseudogene substitutions. By using the N changes depicted in Fig. 1, a functional N rate was estimated from the platyrrhine stem (e.g., 3.5/14 = 0.25 N per million years) and a pseudogene N rate was estimated from the UPG branch to the rhesus monkey (e.g., 37.1/25 = 1.48 N per million years). With these two rates and the total number of N changes on the 15 million years represented by the catarrhine stem (i.e., 7.6), the number of millions of years that the stem catarrhine *GGTA1* locus existed before it became the UPG locus (i.e., X) was calculated in the following manner: $0.25X + 1.48 (15 - X) = 7.6$. Thus, $X = 11.87$ million years, and the UPG locus originated ≈ 28.1 mya.

Finally, overall evolutionary rates were calculated for selectively neutral DNA as represented by the PPG clade and for PPG and UPG rates from the crown catarrhine LCA until the present day.

1. Martin RD (1990) *Primate Origins and Evolution: A Phylogenetic Reconstruction.* (Princeton Univ Press, Princeton).
2. Harvey PH, Martin RD, Clutton-Brock TH (1987) in *Primate Societies*, eds Smuts BB, Cheyney DL, Seyfarth RM, Wrangham RW, Struhsaker TT (Univ of Chicago Press, Chicago), pp 181–196.
3. Zhang D, Webb DM (2003) *Proc Natl Acad Sci USA* 100:8337–8341.
4. Dominy NJ, Lucas PW (2001) *Nature* 410:363–366.
5. Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) *Mol Phylogenet Evol* 9:585–598.
6. Blanken WM, Van den Eijnden DH (1985) *J Biol Chem* 260:12927–12934.
7. Galili U, Shohet SB, Kobrin E, Stults CLM, Macher BA (1988) *J Biol Chem* 263:17755–17762.
8. Joziasse DH, Oriol R (1999) *Biochim Biophy Acta* 1455:403–418.
9. McKenzie IF, Xing PX, Vaughan HA, Prenzoska J, Dabkowski PL, Sandrin MS (1994) *Transplant Immunol* 2:81–86.
10. Takami S, Getchell ML, Getchell TV (1995) *Cell Tissue Res* 280:211–216.
11. Galili U, Clark MR, Shohet SB, Buehler J, Macher BA (1987) *Proc Natl Acad Sci USA* 84:1369–1373.
12. Good AH, Cooper DKC, Malcolm AJ, Ippolito RM, Koren E, Neethling FA, Ye Y, Zuhdi N, Lamontagne LR (1992) *Transplant Proc* 24:559–562.
13. Joziasse DH, Shaper JH, Van den Eijnden DH, Van Tunen AJ, Shaper NL (1989) *J Biol Chem* 264:14290–14297.
14. Larsen RD, Rajan VP, Ruff MM, Kukowska-Latallo J, Cummings RD, Lowe JB (1989) *Proc Natl Acad Sci USA* 86:8227–8231.
15. Sandrin MS, Dabkowski PL, Henning MM, Mouhtouris E, McKenzie IFC (1994) *Xenotransplant* 1:81–85.
16. Joziasse DH, Shaper NL, Kim D, Van den Eijnden DH, Shaper JH (1992) *J Biol Chem* 267:5534–5541.
17. Strahan KM, Gu F, Preece AF, Gustavsson I, Andersson L, Gustafsson K (1995) *Immunogenetics* 41:101–105.
18. Vanhove B, Goret F, Soulillou JP, Pourcel C (1997) *Biochim Biophys Acta* 1356:1–11.
19. Katayama A, Ogawa H, Kadomatsu K, Kurosawa N, Kobayashi T, Kaneda N, Uchimura K, Yokoyama I, Muramatsu T, Takagi H (1998) *Glycoconj J* 6:583–589.
20. Koike C, Friday RP, Nakashima I, Luppi P, Fung JJ, Rao AS, Starzl TE, Trucco M (2000) *Transplant* 70:1275–1283.
21. Tearle RG, Tange MJ, Zannettino ZL, Katerelos M, Shinkel TA, Van Denderen BJW, Lonie AJ, Lyons I, Nottle MB, Cox T, *et al.* (1996) *Transplant* 61:13–19.
22. Phelps CJ, Koike C, Vaught TD, Boone J, Wells KD, Chen SH, Ball S, Specht SM, Polejaeva IA, Monahan JA, *et al.* (2003) *Science* 299:411–414.
23. Joziasse DH, Shaper JH, Jabs EW, Shaper NL (1991) *J Biol Chem* 266:6991–6998.
24. Shaper NL, Lin SP, Joziasse DH, Kim DY, Yang-Feng TL (1992) *Genomics* 12:613–615.
25. Galili U, Swanson K (1991) *Proc Natl Acad Sci USA* 88:7401–7404.
26. Gagneux P, Varki A (1999) *Glycobiology* 9:747–755.
27. Varki A (2006) *Cell* 126:841–845.
28. Koike C, Fung JJ, Geller DA, Kannagi R, Libert T, Luppi P, Nakashima I, Profozich J, Rudert W, Sharma SB, *et al.* (2002) *J Biol Chem* 277:10114–10120.
29. Gastinel LN, Bignon C, Misra AK, Hindsgaul O, Shaper JH, Joziasse DH (2001) *EMBO J* 20:638–649.
30. Boix E, Swaminathan GJ, Zhang Y, Natesh R, Brew K, Acharya KR (2001) *J Biol Chem* 276:48608–48614.
31. Shetterly S, Tom I, Yen TY, Joshi R, Lee L, Wang PG, Macher BA (2001) *J Glycobiol* 11:645–653.
32. Lazarus BD, Milland J, Ramsland PA, Mouhtouris E, Sandrin MS (2002) *Glycobiology* 12:793–802.
33. Yang Z (1997) *Comput Appl Biosci* 13:555–556.
34. Goodman M (1961) *Hum Biol* 33:131–162.
35. Goodman M (1962) *Hum Biol* 34:104–150.
36. Goodman M (1963) in *Classification and Human Evolution*, ed Washburn SL (Aldine, Chicago), pp 204–234.
37. Kim S-H, Elango N, Warden C, Vigoda E, Yi SV (2006) *PLoS Genet* 2:e163.
38. Steiper ME, Young NM, Sukarna TY (2004) *Proc Natl Acad Sci USA* 101:17021–17026.
39. Bailey WJ, Fitch DH, Czelusniak J, Slightom JL, Goodman M (1991) *Mol Biol Evol* 8:155–184.
40. Seino S, Bell GI, Li WH (1992) *Mol Biol Evol* 9:193–203.
41. Ellsworth DL, Hewett-Emmett D, Li WH (1993) *Mol Phylogenet Evol* 2:315–321.
42. Elango N, Thomas JW, NISC Comparative Sequencing Program, Yi SV (2006) *Proc Natl Acad Sci USA* 103:1370–1375.
43. Henion TR, Galili U (1999) *Subcell Biochem* 32:49–77.
44. Thall AD, Malys P, Lowe JB (1995) *J Biol Chem* 270:21437–21440.
45. Hennet T (2002) *Cell Mol Life Sci* 59:1081–1095.
46. Costache M, Apoil P-A, Cailleau A, Elmgren A, Larson G, Henry S, Blancher A, Iordachescu D, Oriol R, Mollicone R (1997) *J Biol Chem* 272:29721–29728.
47. Harduin-Lepers A, Vallejo-Ruiz V, Krzewinski-Recchi MA, Samyn-Petit B, Julien S, Delannoy P (2001) *Biochimie* 83:727–737.
48. Saito N, Yamamoto F (1997) *Mol Biol Evol* 14:399–411.
49. Oriol R, Le Pendu J, Mollicone R (1986) *Vox Sang* 51:161–171.
50. Wang B, Brand-Miller J (2003) *Eur J Clin Nutr* 57:1351–1369.
51. Varki A (2001) *Am J Phys Anthropol* Suppl 33:54–69.
52. Angata T, Margulies EH, Green ED, Varki A (2004) *Proc Natl Acad Sci USA* 101:13251–13256.
53. Liman ER, Innan H (2003) *Proc Natl Acad Sci USA* 100:3328–3332.
54. Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S (2004) *PLOS Biol* 2:120–125.
55. Jacobs GH, Neitz M, Deegan JF, Neitz J (1996) *Nature* 382:156–158.
56. Boissinot S, Zhou Y-H, Qui L, Dulai KS, Neiswanger K, Schneider H, Sampaio I, Hunat DM, Hewett-Emmett D, Li WH (1997) *Zool Stud* 36:360–369.
57. Hunt DM, Dulai KS, Cowing JA, Julliot C, Mollon JD, Bowmaker JK, Li WH, Hewett-Emmet D (1998) *Vision Res* 38:3299–3306.
58. Kainz PM, Neitz J, Neitz M (1998) *Vision Res* 38:3315–3320.
59. Maddison DR, Maddison WP (2000) *MacClade 4: Version 4.0* (Sinauer, Sunderland, MA).
60. Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4* (Sinauer, Sunderland, MA).
61. Ronquist F, Huelsenbeck JP (2003) *Bioinformatics* 19:1572–1574.
62. Posada D, Crandall KA (1998) *Bioinformatics* 14:817–818.
63. Nylander JAA (2004) *MrModeltest v2* (Evolutionary Biology Centre, Uppsala Univ, Uppsala, Sweden).
64. Akaike H (1974) *IEE Trans Aut Cont* 19:716–722.
65. Kumar S, Tamura K, Nei M (2004) *Brief Bioinform* 5:150–163.
66. Pamilo P, Bianchi NO (1993) *Mol Biol Evol* 10:271–281.
67. Li W-H (1993) *J Mol Evol* 36:96–99.
68. Goodman M (1999) *Am J Hum Genet* 64:31–39.
69. Wildman DE, Goodman M (2004) in *Evolutionary Theory and Processes: Modern Horizons*, ed Wasser SP (Kluwer, Dordrecht, The Netherlands), pp 293–311.