Evolution of the *Phosphoglycerate mutase* Processed Gene in Human and Chimpanzee Revealing the Origin of a New Primate Gene

Esther Betrán, Wen Wang,* Li Jin†, and Manyuan Long**

*Department of Ecology and Evolution, The University of Chicago; and †Department of Environmental Health, University of Cincinnati

Processed genes are created by retroposition from messenger RNA of expressed genes. The estimated amount of processed copies of genes in the human genome is 10,000–14,000. Some of these might be pseudogenes with the expected pattern for nonfunctional sequences, but some others might be an important source of new genes. We have studied the evolution of a *Phosphoglycerate mutase* processed gene (*PGAM3*) described in humans and believed to be a pseudogene. We sequenced *PGAM3* in chimpanzee and macaque and obtained polymorphism data for human coding region. We found evidence that *PGAM3* likely produces a functional protein, as an example of addressing functionality for human processed pseudogenes. First, the open reading frame was intact despite many deletions that occurred in the 3′ untranslated region. Second, it appears that the gene is expressed. Finally, interspecies and intraspecies variation for *PGAM3* was not consistent with the neutral model proposed for pseudogenes, suggesting that a new functional primate gene has originated. Amino acid divergence was significantly higher than synonymous divergence in *PGAM3* lineage, supporting positive selection acting in this gene. This role of selection was further supported by the excess of rare alleles in a population genetic analysis. *PGAM3* is located in a region of very low recombination; therefore, it is conceivable that the rapid fixation events in this newly arising gene may have contributed to a selective sweep of variation in the region.

Introduction

Processed copies of genes, initially found in mammals because of their abundance (Wagner 1986), are present in many organisms (Mighell et al. 2000). The amount of processed genes in mammalian genomes is remarkable. For example, 545 genes have been described, and 134 pseudogenes have been claimed in human chromosome 22. Eighty-two percent of these pseudogenes are intronless processed copies that originate via retroposition from known genes (Dunham et al. 1999). In human chromosome 21, 225 genes have been described, and 59 pseudogenes have been found with many of them being processed copies (Hattori et al. 2000). Given that the human genome contains 30,000–40,000 genes (Lander et al. 2001; Venter et al. 2001), the amount of processed genes in the human genome (Gonçalves, Duret, and Mouchiroud 2000) is estimated to be between 10,000 and 14,000.

Processed genes have been experimentally generated in human cells by Maestre et al. (1995). In the experimental work of Maestre and co-workers, mRNAs are reverse transcribed and integrated in the human genome. The processed gene made by the integrated sequence has the length of the mRNA of the original gene, possesses a poly-A tail and, often, direct flanking repeats. Many of these processed genes found in organisms have traits that preclude their functionality and thus are pseudogenes; however, more and more cases of functional processed copies of genes are accumulating (see Brosius 1999 for a review). Two well-studied instances in humans are *Pyruvate dehydrogenase 2* (*Pdha2*) and *Phosphoglycerate kinase 2* (*Pgk-2*). Both *Pdha2* and *Pgk-2* are intronless autosomal copies of *Pdha1* and *Pgk-1*, respectively. *Pdha1* and *Pgk-1* are intron-containing genes located in the X chromosome. The original parental copies, in both cases, have a constitutive function, but their presence at the X chromosome prevents them from being expressed in the testis because of X-inactivation. Unlike *Pdha1* and *Pgk-1*, *Pdha2* and *Pgk-2* are expressed only in the testis. *Pdha2* and *Pgk-2* have 86% and 87% identity at the protein level, respectively, with the parental gene and maintain similar function (McCarrey and Tomas 1987; Dahl et al. 1990; Fitzgerald et al. 1996; McCarrey et al. 1996). The *Pdha2*-promoter region is derived from a recent retroposon insertion from *Pdha1* gene (Datta et al. 1999), whereas the *Pgk-2*–promoter region arose originally from a rare aberrant transcript that included the promoter region of *Pgk-1* (McCarrey 1987).

Hence, processed genes are common in humans, and many of them are likely to be functional. How do we know whether or not a processed gene is functional from available sequence data? The neutral theory of molecular evolution (Kimura 1983, p. 178) predicts that pseudogenes should evolve as a neutral sequence (Li, Gojobori, and Nei 1981). They should change with time at an even substitution rate across the whole sequence. Flanking and coding sequences should show the same evolutionary patterns. In addition, and as a consequence of this, stop codons are likely to appear (Li 1997, p. 347). Synonymous substitutions per synonymous site (K_S) and nonsynonymous substitutions per nonsynonymous site (K_A) should be equivalent. Thus, the K_A/K_S ratio of pseudogenes is expected to be equal to one (Miyata and Hayashida 1981; Kimura 1983). Similarly, if we look at the first, second, and third codon positions since the pseudogene formation, rates of divergence for every one of these positions should also be equal (Li, Gojobori, and Nei 1981). Furthermore, these regions

Key words: *Phosphoglycerate mutase,* functional processed gene, primates, retroposition, positive selection.

Address for correspondence and reprints: Manyuan Long, Department of Ecology and Evolution, The University of Chicago, 1101 East 57th Street, Chicago, Illinois 60637.
E-mail: mlong@midway.uchicago.edu.

Mol. Biol. Evol. 19(5):654–663. 2002

will show a high level of polymorphism and divergence along the whole region because levels of diversity inversely correlate with level of biological constraint. The analysis of polymorphism and divergence for a processed gene can give important insights complementary to functional studies (e.g., expression).

Here, we report that a *Phosphoglycerate mutase* processed gene (*PGAM3*) in primates evolved as a new functional gene. *PGAM3* had so far been only found in humans, and it was described as a pseudogene: *Phosphoglycerate mutase 1* processed pseudogene (*ψ-PGAM1*; Dierick, Mercer, and Glover 1997). However, our data reveal its functionality, and we suggest naming this newly found functional processed gene as *Phosphoglycerate mutase 3* (*PGAM3*). *PGAM3* originated by retrotransposition, as suggested by its several molecular features. The parental gene, phosphoglycerate mutase brain isoform gene (*PGAM1*), is an intron-containing gene with a coding region of 762 bp, a 5′ untranslated region (UTR) of 12 bp, and a 3′UTR of 912 bp (Dierick, Mercer, and Glover 1997; Lander et al. 2001), and it is located on chromosome 10, 10q25.3 (Dierick, Mercer, and Glover 1997; Lander et al. 2001). However, *PGAM3* is intronless, and it is located in the first intron of the Menkes disease gene (*MNK*) in the X chromosome region Xq13.3 (Dierick, Mercer, and Glover 1997). This position is different from the *PGAM1* location, and note that it is inserted within the coding region of another gene. *PGAM3* homology with *PGAM1* includes 3′ and 5′UTR (Dierick, Mercer, and Glover 1997). *PGAM3* has a poly-A tail 16 bp after the polyadenylation signal at the end of the 3′UTR and is flanked by 10 bp direct repeats (Dierick, Mercer, and Glover 1997), as expected from a processed gene (Vanin 1985; Maestre et al. 1995; Mighell et al. 2000). All these features were revealed in an earlier analysis by Dierick, Mercer, and Glover (1997). We have found *PGAM3* in chimpanzee and macaque. We have investigated polymorphism in human populations and examined the expression of this gene. Our data support the functionality of *PGAM3*. These results are discussed considering published information about the genomic region and how pseudogenes evolve.

Materials and Methods

DNA Samples and Sequencing

The *PGAM3* gene in human (*Homo sapiens*) and chimpanzee (*Pan troglodytes*) was amplified by PCR from genomic DNA. Primers were designed to amplify this gene sequence specifically. A 5′ primer was located in the second exon of the Menkes disease gene (*MNK* or *ATP7A* gene), 5′-CACCATTCACTTTTCCAATC-3′, and a 3′ primer was located in the 3′UTR of *PGAM3*, maximizing mismatches with the *PGAM1* copy, 5′-ACATCACCATGCAGATTACATTCA-3′ (fig. 1). For some individuals, nested primers (5′-CAATCTGCTGCTCAATGGTC-3′ and 5′-CTAGAGCCCCCAGGCAGTGG-3′) designed in the same regions were used to reamplify (fig. 1). The *PGAM3* coding regions of 15 human males, representing all continents, and a chimpanzee male sample were sequenced (see table 1 for details on the origin of the samples). Partial coding sequence (324 bp) was obtained in a macaque male (*Macaca mulata*) using primers flanking *PGAM3* coding region (5′-AGGGAATAAGGGTGGGAAAAAGAAT-3′ and 5′-ACATCACCATGCAGATTACATTCA-3′). Haplotypes were determined directly by sequencing both strands from every individual, except for macaque in which only primers in one direction produced high quality sequence.

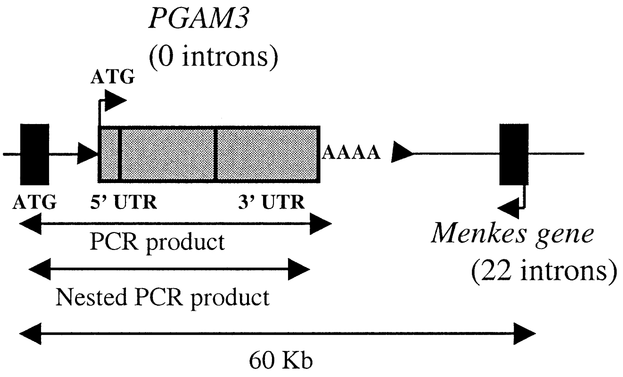

FIG. 1.—Genomic region showing *PGAM3* insertion in the first intron of the Menkes gene (*MNK*). Note that both genes are encoded in opposite strands. Primers and nested primers used in this study are depicted in the figure.

Sequence Analysis

Sequences were aligned by means of Clustal W (Thompson, Higgins, and Gibson 1994) and manually adjusted. Corrected number of synonymous and nonsynonymous substitutions per site (K_S and K_A) were computed following the method described by Kumar (Nei and Kumar 2000, p. 64) and using MEGA2 software (Kumar et al. 2000). This method corrects for multiple hits, treating arginine and isoleucine codons accurately (Kumar et al. 2000). Standard errors (SEs) were computed analytically, according to Nei and Kumar (2000). In a rough scale, the K_A/K_S ratio is assumed to be equal to one under the neutral model, smaller than one under purifying selection, and greater than one under positive selection (Kimura 1983). The difference between K_A and K_S was tested using the normal deviate or Z test (Nei and Kumar 2000). This test relies on substitutions being normally distributed. $D = K_A - K_S$ and its SE (SE[D]) to compute Z were obtained analytically using MEGA2. MEGA2 was also used to obtain the neighbor-joining trees for K_A and K_S. Because the assumption of normal distribution for substitutions may not be met, we will further investigate the results of this comparison with the following different approaches.

Tajima's relative rate test (Tajima 1993) was performed for the first, second, and third codon position substitutions using MEGA2 (Kumar et al. 2000). After pseudogene formation (see fig. 2), the sequence is supposed to evolve neutrally. The pseudogene lineage will show an accelerated rate of evolution because of the loss of constraint along the entire sequence. First, the rates of evolution among all the codon positions for a pseudogene should be similar (Li, Gojobori, and Nei 1981).

Table 1
DNA Sequence Variation in the Coding Region of *PGAM3* Genes Compared with *PGAM1*

Codon Position	1	1	3	1	1	2	2	2	3	2	2	2	2	2	2	2	1	3	3	2	1	3
Coding Position	43	70	87	118	166	188	269	290	390	443	458	467	524	563	572	581	589	633	648	713	718	729
Human *PGAM1*	G	G	C	C	T	C	G	G	T	A	G	T	T	A	G	T	C	G	T	C	C	G
Macaque *PGAM3*	—	—	—	—	—	—	—	—	—	**G**	**A**	.	.	**T**	**A**	**C**	**G**	A	C	**G**	.	**A**
Chimpanzee *PGAM3*	**A**	**T**	.	.	**C**	**T**	.	**C**	C	.	**A**	**C**	.	.	**A**	**C**	**G**	.	.	.	**T**	.
Human *PGAM3*																						
Berg 36	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
P92	**A**	**T**	T	.	**C**	**T**	**A**	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
NE16-208	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
Ul	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
JL	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
FD1 01	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
FD1 10	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
FD1 38	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
FD1 58	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	**C**	.	**A**	**C**	**G**	.	C	.	**T**	**A**
JK2667	**A**	**T**	T	**T**	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
JK2668	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
JK2669	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
JK2670	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
JK2673	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
BI57	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**
Consensus	**A**	**T**	T	.	**C**	**T**	.	.	.	.	**A**	**C**	.	.	**A**	**C**	**G**	.	C	.	**T**	**A**

NOTE.—Samples are: P92 from an African Pygmy, Berg36 and NE16-208 from Northern Europeans, U1 from a Southern European, JL, FD1 01, FD1 10, FD1 38, and FD1 58 from Han Chinese, JK2667, JK2668, JK2669, JK2670, and JK2673 from New Guinean Highlanders, and BI57 from an American Karitiana. Replacement changes with respect to *PGAM1* sequence are in bold. Coding region of *PGAM3* is 762 bp. Only 324 bp of coding region were sequenced from Macaque *PGAM3*.

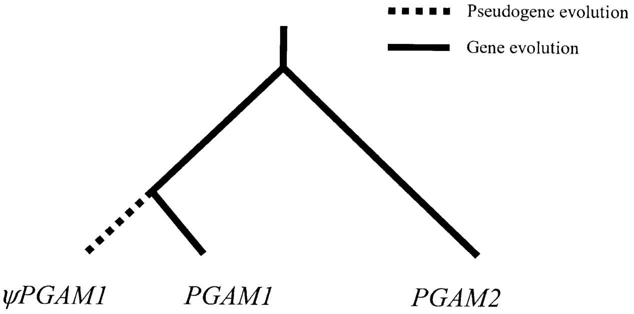

FIG. 2.—Phylogenetic tree for *PGAM3* (*ψPGAM1;* Dierick, Mercer, and Glover 1997), *PGAM1,* and *PGAM2.* After *PGAM3* formation, its sequence is expected to evolve neutrally if it were a pseudogene. This *PGAM3* lineage should show accelerated rate of evolution because of the loss of constraint. Comparison with an outgroup sequence, *PGAM2* in this case, provides the necessary information to determine changes in each lineage (Li, Gojobori, and Nei 1981).

Comparison with an outgroup sequence, *PGAM2* in this case (phosphoglycerate mutase muscle isoform gene; Dierick et al. 1995), provides the necessary information to determine changes in each lineage using a parsimony criterion.

Using this information, we have designed a probability test for two hypotheses: (1) if *PGAM3* is evolving as a pseudogene, i.e., all three codon positions evolve at the same rate, and (2) if *PGAM3* is evolving as a constrained protein. We use the likelihood ratio test (Sokal and Rohlf 1995) on the trinomial probability of the number of nucleotide substitutions over codon positions under the two models and under observed sequence evolution. Trinomial distribution probability is computed: $P(n_1,n_2,n_3) = n!/(n_1!n_2!n_3!)(f_1^{n_1}f_2^{n_2}f_3^{n_3})$, where f_1, f_2, and f_3 are the frequencies and n_1, n_2, and n_3 are the numbers of the observed changes. Likelihood ratio test is

$$G = 2\ \ln\frac{\dfrac{n!}{n_1!n_2!n_3!}f_1^{n_1}f_2^{n_2}f_3^{n_3}}{\dfrac{n!}{n_1!n_2!n_3!}f_1'^{n_1}f_2'^{n_2}f_3'^{n_3}} = 2\ \ln\frac{f_1^{n_1}f_2^{n_2}f_3^{n_3}}{f_1'^{n_1}f_2'^{n_2}f_3'^{n_3}}$$

where f_1, f_2, and f_3 are the observed frequencies and f_1', f_2', and f_3' are the expected frequencies under the hypothesis (Sokal and Rohlf 1995). G can be approximated by the chi-square distribution with as many degrees of freedom as cells minus one (Sokal and Rohlf 1995).

Nucleotide diversity, π, defined as the average number of nucleotide differences per site between two random sequences (Tajima 1989), and θ_W, Watterson estimate of 3Neμ from the number of segregating sites (Watterson 1975), were calculated. Both values estimate the neutral parameter $\theta = 3\text{Ne}\mu$ for X-linked loci, where Ne is the effective population size and μ is the neutral mutation rate under equilibrium conditions. Deviations from equality of these values reveal nonequilibrium conditions in the history of the sample. Tajima's D (Tajima 1989) measures those deviations. $D_T = (\pi - \theta_W)/(V(\pi - \theta_W))^{1/2}$. Fu and Li (1993) propose another test to measure departure from neutral expectations. $D_{F-L} = (\eta_e - \eta_i/(a - 1))/[V(\eta_e - \eta_i/(a - 1))]^{1/2}$. They demonstrated that η_e, the total number of mutations in the external branches of a genealogy of n sequences, and η_i, the total number of mutations in the internal branches, can be used to estimate θ. External branches are known to be more affected by selection because recent mutations are close to the tips. Tajima's test (Tajima 1989) and Fu and Li's test with outgroup (1993) were carried out using the program DNAsp, Version 3.52 (Rozas and Rozas 1999). Significance was tested by 10,000 coalescent simulations.

Expression Analysis

PCR from four human cDNA libraries (Human leukocyte 5′-stretch cDNA library from Clontech, Human testis 5′-strech plus cDNA library from Clontech, Human Tcell lambda cDNA library from Stratagene, and λGEX5 Hela cDNA library of Fukunaga and Hunter 1997) was carried out with primers specific for *PGAM3* (5′-CAGAAGATCAGCTACCCTCCT**A**-3′ and 5′-ACATCACCA**T**GCAG**A**TTACATTC**A**-3′) and nested primers specific for *PGAM3* (5′-CTACCCTCCT**A**TGAGAGTC**C**-3′ and 5′-GGGCAGAGGGACAAGAC**CA**-3′). Primers contain from 1 to 3 mismatches to the *PGAM1* gene are shown in bold type. Specificity was achieved at 59°C using Expand High Fidelity *Taq* polymerase (Roche). Products were digested with BstXI to reveal that the amplified copy was *PGAM3* (Dierick, Mercer and Glover 1997). The BstXI restriction site is present in *PGAM3* but not in *PGAM1.* In addition, products from these PCRs were sequenced to confirm the specificity of the amplification.

Results

Evolution of *PGAM3* Coding Region

Complete sequence of the coding region was determined for human and chimpanzee samples. A summary of the variation between the sequences obtained in this work is shown in table 1. Only partial sequence was obtained for macaque (324 bp). Fixed differences between *PGAM1* (original copy) and *PGAM3* genes (positions 458, 572, 581, 589, 648, and 729) suggest that *PGAM3* originated before cercopithecoidea-hominoidea split (table 1). We include this sequence to illustrate the presence of *PGAM3* in macaque but exclude it from the analysis because the sequence is partial. Human *PGAM3* consensus sequence differs in two positions from the sequence reported by Dierick, Mercer, and Glover (1997). Only partial sequence is available from human sequence draft (Lander et al. 2001), and it is identical to the consensus in table 1. The difference in Dierick, Mercer, and Glover (1997) might be a rare polymorphism or a sequencing error.

Synonymous and nonsynonynous substitutions per site (K_S and K_A) are shown in table 2. K_S in a pairwise comparison between *PGAM1* and *PGAM3* was on an average 0.006. This value does not reveal high levels of divergence for this type of site. The average sequence divergence was 0.0124 ± 0.0007 for the human and chimpanzee pair in a recent study of 24 kb of autosomal intergenic DNA segments (Chen and Li 2001) and ranged from 0.0026 to 0.0138 for noncoding regions of

Table 2
K_A and K_S Values of Divergence and SE Computed Following the Method of Kumar (Nei and Kumar 2000)

	Human *PGAM1*	Human *PGAM3*	Chimpanzee *PGAM3*
Human *PGAM1*		0.008 ± 0.005	**0.004 ± 0.004**
Human *PGAM3*	0.021 ± 0.007		0.012 ± 0.007
Chimpanzee *PGAM3*	**0.020 ± 0.006**	0.003 ± 0.003	

NOTE.—Consensus sequence for Human *PGAM3* was used in these comparisons. K_A (below diagonal) and K_S (above diagonal) values are bold whenever there is a significant difference from one of the K_A/K_S ratio ($P < 0.05$).

the X chromosome (see Przeworski, Hudson, and Di Rienzo 2000 for a review). The low K_S values produce an unresolved tree (fig. 3). Given that *PGAM3* is older than human-chimp divergence, 0.006 is not a high value as it would be expected if *PGAM3* is evolving without constraint. K_A values in the comparison of *PGAM1* with *PGAM3* are on an average 0.0205. This value is high compared with synonymous substitutions (see subsequently).

A way to test for deviation of the neutral pattern is to compare synonymous and nonsynonymous changes per site that have accumulated since two sequences diverged from their common ancestor. Another way is to study whether all codon positions evolve at the same rate. We address both these aspects.

Under the neutral model, it is expected that K_A equals K_S (Kimura 1983). However, comparison of *PGAM1* with human *PGAM3* and chimpanzee *PGAM3* showed K_A to be greater than K_S. K_A/K_S values were 2.625 and 5.000, respectively, with probabilities of 0.052 and 0.008. The ratio does reveal high levels of amino acid divergence but only in the *PGAM3* lineage (see subsequently).

Comparison with an outgroup sequence, *PGAM2*, provides the necessary information to determine changes in each lineage using a parsimony criterion. Most of the nonsynonymous changes between *PGAM1* and *PGAM3* occurred in the *PGAM3* lineage (see fig. 3 and table 3). The number of amino acid changes in this lineage was significantly higher than that in the lineage leading to *PGAM1* (table 3): 1 versus 9 ($P = 0.011$) in human and chimpanzee *PGAM3* lineages. Changes occurred preferentially not only in the second codon position (see significant values in table 3) but also in the first codon position. These two codon positions together explain the significant difference in amino acid rate.

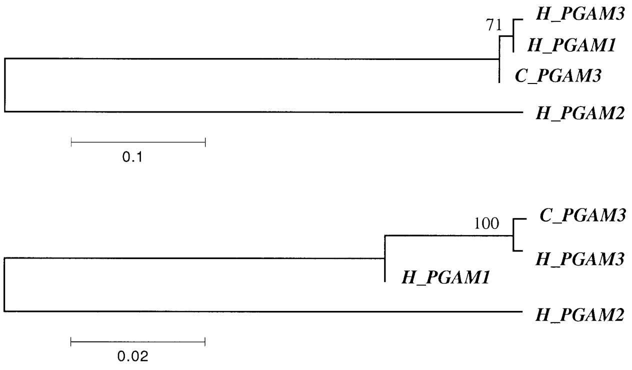

FIG. 3.—Synonymous and nonsynonymous neighbor-joining trees (Saitou and Nei 1987) using Kumar correction (Nei and Kumar 2000). MEGA 2.00 (Kumar et al. 2000) was used to compute and to display the phylogeny. Human *PGAM2* was used as an outgroup in all the comparisons because it is believed to come from an earlier duplication than the one from which *PGAM3* arises (Dierick, Mercer, and Glover 1997; Sakoda et al. 1988). Synonymous and nonsynonymous trees (in this order) are shown. 254 codons and consensus human *PGAM3* were used in the comparisons.

We tested whether or not the substitutions at the three codon positions of *PGAM3* are consistent with the substitution pattern for a pseudogene. Do substitutions occur more often at the first and second codon positions than expected in a pseudogene? In a pseudogene, we expect $f_1 + f_2 = 2f_3$. The probability of our observation in chimpanzee under this expectation is 0.0172. First and second positions appear to be evolving faster than expected in a pseudogene (unconstrained sequence), supporting the view that positive selection is acting in this gene changing amino acid composition.

We also tested whether or not the substitutions at the three codon positions are consistent with a general substitution pattern for common genes. A general substitution rate pattern for the three codon positions is $f_3 > f_1 > f_2$ (Li 1997). In the particular case of globins, $f_1' = 0.24$, $f_2' = 0.20$, and $f_3' = 0.56$ (Kimura 1983). However, *PGAM3* shows $f_2(0.600) > f_1(0.400) > f_3(0)$ in chimpanzee and $f_2(0.455) > f_1(0.365) > f_3(0.18)$ in human. The likelihood ratio tests for these two comparisons yield probabilities of 0.0002 and 0.0297, respectively. First and second positions are evolving faster than expected in a coding gene, again supporting the view that positive selection is acting on this gene.

Pseudogenes often have deletions and insertions (indel) in the coding region (Vanin 1985; Mighell et al. 2000). However, this is not the case for *PGAM3* in human and chimpanzee in which we sequenced the complete coding sequence (762 bp, table 1). There are no deletions in the partial sequence obtained for macaque either.

Why do we observe no indels in the coding region of *PGAM3*? A possibility is that the gene is too young to accumulate any indels. However, Dierick, Mercer, and Glover (1997) found nine indels and 18 bp changes in the 3′UTR of *PGAM3* and 0 indels in the coding region when they sequenced the whole *PGAM3* and compared this sequence with *PGAM1* (Dierick, Mercer, and Glover 1997). We have statistically compared coding versus UTR in the *PGAM3* lineage. In the coding region, there are 11 mutations (table 3). For the 3′UTR, there is no outgroup sequence that allows us to infer in what lineage the mutations occurred. We think that the 3′UTR is

Table 3
Tajima's Relative Rate Test with *PGAM2* as Outgroup

	Amino acid changes	1st Codon Position	2nd Codon Position	3rd Codon Position
H_PGAM1 versus *H_PGAM3*	1 versus 9 **(*P* = 0.011)**	1 versus 4 (*P* = 0.180)	0 versus 5 **(*P* = 0.025)**	0 versus 2 (*P* = 0.157)
H_PGAM1 versus *C_PGAM3*	1 versus 9 **(*P* = 0.011)**	1 versus 4 (*P* = 0.180)	0 versus 6 **(*P* = 0.014)**	1 versus 0 (*P* = 0.317)

not as constrained as the coding region and that both *PGAM3* and *PGAM1* can accumulate changes in that region. Under this assumption that half of the changes in the 3′UTR occurred in the *PGAM3* lineage, we can construct a two (indels-substitutions) by two (coding region-3′UTR) contingency table for *PGAM3*. Chi-square test probability under the observed pattern is 0.0267. This implies that both regions behaved differently. We would expect a similar distribution of the variation if *PGAM3* were a pseudogene.

These analyses uncover the functionality of *PGAM3* (1) functional constraint revealed by the absence of indels in the protein coding region, and (2) higher protein substitution rate than in a pseudogene and higher protein substitution rate than in a normal functional gene like globin genes, revealing the action of positive selection. Thus, we suggest that primate *PGAM3* is a newly evolved functional gene encoding a rapidly evolving protein.

Levels of Polymorphism

PGAM3 polymorphism data in human are shown in table 1. Coding region of *PGAM3* was sequenced for 15 human males from around the world. Only three segregating sites were observed in the human *PGAM3*. The three segregating sites are singletons, transitions, and produce replacement change. π is 0.052% (SE = 0.024%) and θ_W is 0.121% (SE = 0.078%). These values are low, but they do not differ from the diversity reported in normal human X-linked genes: π of 0.000%–0.178% and θ_W of 0.000%–0.148% (Li and Sadler 1991; Przeworski, Hudson, and Di Rienzo 2000).

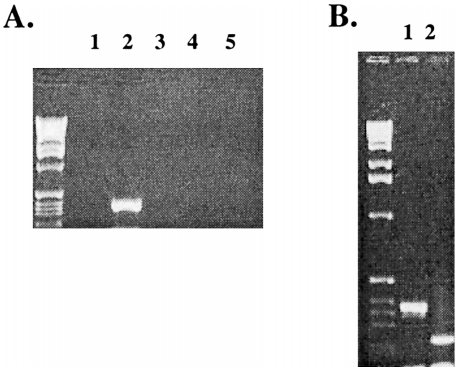

FIG. 4.—PCR results for *PGAM3* from cDNA libraries (*A*) Nested RT-PCR from four cDNA libraries: (1) Hela cells, (2) Human leukocyte, (3) Human testis, (4) Human T cells, and (5) negative control. (*B*) (1) Undigested, and (2) digested product of human leukocytes.

Tajima's *D* value for this data was negative: −1.68501 ($P < 0.05$). A negative value is caused by segregating sites at low frequency because rare alleles contribute less to π than to θ_W. Negative values are expected under exponential growth (Slatkin and Hudson 1991) or after a selective sweep, rapid fixation of a new allele (Braverman et al. 1995). Fu and Li's *D* value using *PGAM1* or chimpanzee *PGAM3* as an outgroup was −2.39127 ($P < 0.05$). This negative value is again a consequence of the presence of segregating sites at low frequency. This is expected under rapid fixation of a new allele or exponential growth (Fu and Li 1993). The high rate of amino acid substitution we observe in *PGAM3* indicates that *PGAM3* may have been under positive selection in the human lineage and could explain the observed pattern for the gene variation. Rapid fixation of new mutations could explain the observed polymorphism pattern (see *Discussion*).

Expression Analysis and Functionality

Given the data of the population genetics and other evolutionary studies that reveal functionality of *PGAM3*, we tested a few cDNA libraries for the expression of this gene. Figure 4*A* shows nested amplifications from these cDNA libraries. A band is produced from leukocyte cDNA library. This product was digested with BstXI restriction enzyme to reveal an amplified copy (Dierick, Mercer, and Glover 1997). BstXI restriction site is present in *PGAM3* but not in *PGAM1*. Undigested and digested products of this band are shown in figure 4*B*. This band was also sequenced reconfirming that it corresponds to *PGAM3*. The detected tissue-specific expression pattern suggests that the expression may be authentic.

Although no transcript was observed for this gene by Dierick, Mercer, and Glover (1997), we observed the transcript when we used a different strategy. These authors tried to amplify this gene with primers that amplify both copies and after that digest with BstXI to reveal the presence of the new copy. This procedure could mask the low expression of the new copy because *PGAM1* is very highly expressed in all the essayed tissues (data not shown). In our case we used two rounds of specific primers (see *Materials and Methods*). Interestingly, our observation of expression is consistent with a possible promoter (putative TATA box and CAAT box) already described by Dierick, Mercer, and Glover (1997).

To infer the possible function of PGAM3, sequences of the putative human and chimpanzee PGAM3 were

```
C_PGAM3    MAAYKLVLIRHGESTWNLENRFSCWYDADLSPAGHEEAKRGGQALRDAGYEFDICLTSVQKRVIRTLWTVLDAIDQMWLPVVRTW
H_PGAM3    MAAYKLVLIRHGESTWNLENRFSCWYDADLSPAGHEEAKRGGQALRDAGYEFDICLTSVQKRVIRTLWTVLDAIDQMWLPVVRTW
H_PGAM1    MAAYKLVLIRHGESAWNLENRFSGWYDADLSPAGHEEAKRGGQALRDAGYEFDICFTSVQKRAIRTLWTVLDAIDQMWLPVVRTW
H_PGAM2    MATHRLVMVRHGESTWNQENRFCGWFDAELSEKGTEEAKRGAKAIKDAKMEFDICYTSVLKRAIRTLWAILDGTDQMWLPVVRTW
H_DPGAM    MSKYKLIMLRHGEGAWNKENRFCSWVDQKLNSEGMEEARNCGKQLKALNFEFDLVFTSVLNRSIHTAWLILEELGQEWVPVESSW
R_DPGAM    MSKYKLIMLRHGEGAWNKENRFCSWVDQKLNSEGMEEARNCGKQLKALNFEFDLVFTSVLNRSIHTAWLILEELGQEWVPVESSW
           *    *   ****  ** ****  * *  *   * ***            ***   ***  * * * *  *    * * **   *

C_PGAM3    RLNERHYGGLTALNKAETAAKHGEAQVKIWRRSYDVPPPPMEPDHPFYSNISKDRRY--ADLTEDQLPSYESPKDTIARALPFWN
H_PGAM3    RLNERHYGGLTGLNKAETAAKHGEAQVKIWRRSYDVPPPPMEPDHPFYSNISKDRRY--ADLTEDQLPSYESPKDTIARALPFWN
H_PGAM1    RLNERHYGGLTGLNKAETAAKHGEAQVKIWRRSYDVPPPPMEPDHPFYSNISKDRRY--ADLTEDQLPSCESLKDTIARALPFWN
H_PGAM2    RLNERHYGGLTGLNKAETAAKHGEEQVKIWRRSFDIPPPPMDEKHPYYNSISKERRY--AGLKPGELPTCESLKDTIARALPFWN
H_DPGAM    RLNERHYGALIGLNREQMALNHGEEQVRLWRRSYNVTPPPIEESHPYYQEIYNDRRYKVCDVPLDQLPRSESLKDVLERLLPYWN
R_DPGAM    RLNERHYGALIGLNREKMALNHGEEQVRIWRRSYNVTPPPIEESHPYYHEIYSDRRYRVCDVPLDQLPRSESLKDVLERLLPYWN
           ******** *  **    *  *** **  ****    ***    ** *  *   ***         **  ** **   * ** **

C_PGAM3    EEIVPQIKEGKRVLIAAHGNSLQGIAKHVEGLSEEAIMELNLPTGIPIVYELDKNLKPIKPMQFLGDEETVCKAMEAVAAQGKAK
H_PGAM3    EEIVPQIKEGKRVLIAAHGNSLQGIAKHVEGLSEEAIMELNLPTGIPIVYELDKNLKPIKPMQFLGDEETVCKAIEAVAAQGKAK
H_PGAM1    EEIVPQIKEGKRVLIAAHGNSLRGIVKHLEGLSEEAIMELNLPTGIPIVYELDKNLKPIKPMQFLGDEETVRKAMEAVAAQGKAK
H_PGAM2    EEIVPQIKAGKRVLIAAHGNSLRGIVKHLEGMSDQAIMELNLPTGIPIVYELNKELKPTKPMQFLGDEETVRKAMEAVAAQGKAK
H_DPGAM    ERIAPEVLRGKTILISAHGNSSRALLKHLEGISDEDIINITLPTGVPILLELDENLRAVGPHQFLGDQEAIQAAIKKVEDQGKVK
R_DPGAM    ERIAPEVLRGKTVLISAHGNSSRALLKHLEGISDED11NITLPTGVPILLELDENLRAVGPHQFLGDQEAIQAAIKKVEDQGKVK
           * * *    **  ** *****     ** ** *   *     **** **  **   *    * ***** *    *   *  *** *

C_PGAM3    K---
H_PGAM3    K---
H_PGAM1    K---
H_PGAM2    ----
H_DPGAM    QAKK
R_DPGAM    RAEK
```

FIG. 5.—Sequences of the putative human and chimpanzee PGAM3. Alignment with other related proteins is shown for comparison. See text for details.

aligned (fig. 5). Human PGAM3 consensus sequence is shown. PGAM proteins are dimeric glycolytic enzymes that catalyze the reaction from 2-phosphoglycerate to 3-phosphoglycerate (Grisolia and Joyce 1959; Grisolia and Carreras 1975). Two isoforms have been described in mammals: PGAM-M (muscle specific form; PGAM2 in human) and PGAM-B (brain form; PGAM1 in human). Erythrocytes contain another related enzyme that catalyzes the conversion of 1,3-biphosphoglycerate in 2,3-biphosphoglycerate: the diphosphoglycerate mutase (DPGAM; Sakoda et al. 1988). Figure 5 shows an alignment with these other related proteins. Amino acids known to be at the enzyme active site are outlined in bold type: His at position 11, Arg at position 62, and His at position 186 (Grisolia and Joyce 1959). All of them remain intact in human, chimpanzee, and macaque PGAM3, revealing that PGAM3 can still keep PGAM function.

Discussion

Pseudogenes are defined as regions of DNA that are similar to functional genes, but various mutations do not allow the sequence to generate a functional product (Proundfoot and Maniatis 1980). As we reviewed earlier in the article (see *Introduction*), processed genes will often lack promoter and thus become silent functionless pseudogenes. Even so, some processed copies of genes can attain functionality through the acquisition of regulatory regions (Brosius 1999). As of now, there are up to ~20 cases in which the functionality of the processed genes in humans has been demonstrated (Betrán and Long 2001). Our data revealed a new primate processed gene, i.e., *PGAM3*, that originated by retroposition and evolved under positive Darwinian selection.

The Age of *PGAM3*

PGAM3 was first described to be a pseudogene present only in human beings. We have found *PGAM3* in chimpanzee and macaque. This means that *PGAM3* is older than originally proposed. According to its distribution, it is no less than 25 Myr old (Goodman et al. 1998).

PGAM3 Encoding a Functional Protein

Many features of *PGAM3* evolutionary pattern support the finding that the gene is encoding for a functional protein. The coding region does not show deletions or insertions (or both) or nonsense mutations in human, chimpanzee, or macaque sequences reported in this work or previous work (Dierick, Mercer, and Glover 1997), whereas many deletions took place in the 3′UTR, implying that the coding region is under certain constraint. Pseudogene evolution has been studied since the late-1970s. They often accumulate mutations at an accelerated rate compared with parental functional genes, indels, and nonsense mutations that destroy their function (Vanin 1985; Mighell et al. 2000). For example, *HLA-H* is a pseudogene that arose by duplication in the human leukocyte antigen complex. It has a single base pair deletion in exon 4 that makes the gene nonfunctional (Grimsley, Mather, and Ober 1998). This is not the case for *PGAM3*.

In addition, *PGAM3* coding region evolution is not consistent with the neutral model proposed for unconstrained sequences. First, levels of amino acid divergence are higher than synonymous divergence when compared with that of the original copy *PGAM1* in human and in chimpanzee. Amino acid changes are more rapid than synonymous changes in the *PGAM3* lineage. Evolutionary changes occurred preferentially in the first and second codon positions. This supports the view that *PGAM3* may have been under positive selection, rapid fixation of some amino acid changes. One usually expects the effect of positive selection at the amino acid level only if the gene is functional: i.e., translated.

Low Levels of Variation

Polymorphism within *PGAM3* in humans is very low and is biased toward rare alleles; there are significant negative Tajima's *D* and negative Fu and Li's *D*. The neutral model under which one expects high levels of polymorphism and no frequency spectrum bias for pseudogenes, again, cannot interpret these results.

Negative values of Tajima's *D* and Fu and Li's *D* are expected under exponential growth (Slatkin and Hudson 1991) or after a selective sweep, rapid fixation of a new allele (Braverman et al. 1995). Under exponential growth, one expects that every single locus of the genome will show negative Tajima's *D*. Three studies of 10 kb noncoding sequence in human have been carried out: Xq13.3 region (Kaessmann et al. 1999), chromosome 22 (Zhao et al. 2000), and chromosome 1 (Yu et al. 2001). Tajima's *D* values in these studies were -1.62 ($P > 0.05$), -1.03 ($P > 0.10$), and -1.22 ($0.05 < P < 0.10$), respectively (Kaessmann et al. 1999; Zhao et al. 2000; Yu et al. 2001). All values were negative but not significant. The study of polymorphism of Xq13.3 was carried out in a region of very low recombination (1.3 cM/Mb) and has recently been reviewed (Zhao et al. 2000). It showed a significant Tajima's *D* and Fu and Li's *D* test: -1.57 ($P < 0.03$) and -3.29 ($P < 0.05$; Kaessmann et al. 1999; see Zhao et al. 2000 for new computation of Tajima's *D*). Zhao et al. (2000) concluded that the rejection of neutrality in this region might indicate linkage of this noncoding region to a gene under selection. Xq13.3 10 kb region is 1 Mb apart from *PGK-1*, the closest gene, and 1.3 Mb apart from *PGAM3*, the gene studied here. The high rate of amino acid substitution we present exists in *PGAM3*, indicating that *PGAM3* has been under positive selection in the human lineage. Because *PGAM3* is located in the Xq13.3 region that shows very low recombination (Kaessmann et al. 1999), it is conceivable that rapid fixation events in this gene contributed to sweep variation away in that part of the genome (as observed; Kaessmann et al. 1999; Zhao et al. 2000 for new computations).

PGAM3 Function

Although no transcript was observed for this gene by Dierick, Mercer, and Glover (1997), they already pointed out that the region upstream of the gene shows some features of possible promoter region (TATA box and CAAT box). Here, we report preliminary data on the expression of *PGAM3* in white blood cells, suggesting that mRNA is produced and that the protein can be produced in some tissues. We know that the three amino acids in the active center of the PGAM proteins are intact in PGAM3 in human and chimpanzee. This reveals that it is likely that the putative PGAM3 retains PGAM activity.

We have studied the evolution of *Phosphoglycerate mutase 3* processed gene (*PGAM3*) that was previously believed to be a pseudogene (Dierick, Mercer, and Glover 1997). However, polymorphism and divergence data as well as expression analysis support its functionality. Interestingly, many amino acid substitutions took place in *PGAM3* in a short period of time. This suggests a scenario in which this gene could be gaining a new function in the genomes of human and chimpanzee.

The estimated amount of processed copies of genes in the human genome is large. Different pseudogenes might be in different stages: recent acquisitions (Mighell et al. 2000), old acquisitions (Vanin 1985), recent loss of function (Winter et al. 2001), recent regain of function (Mighell et al. 2000), or functional throughout its evolution (Brosius 1999). Some of them might be pseudogenes following the expected pattern for nonfunctional sequences, but some others might be an important source of new genes. The evolutionary and functional analyses as presented in this investigation may be an efficient approach to revealing the evolutionary processes of these processed genes.

Supplementary Material

The *PGAM1* and *PGAM2* (phosphoglycerate mutase muscle isoform gene; Dierick et al. 1995) sequences used in the analysis were retrieved from GenBank (accession numbers NM_002629 and XM_011580, respectively). New sequences have been submitted to GenBank under accession numbers AF465731–46.

Acknowledgments

We are indebted to Dr. Y. Wan for human cDNA libraries. We thank Josep Comeron and Ana Llopart for comments on the data analysis, Kateryna Makova and Janice Spofford for critically reading the manuscript and three anonymous reviewers for their suggestions. M.L. was supported by grants from NSF and Packard Foundation.

LITERATURE CITED

Betrán, E., and M. Long. Expansion of genome coding regions by acquisition of new genes. Genetica (in press).

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140**(2):783–796.

Brosius, J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene **238**:115–134.

Chen, F.-C., and W.-H. Li. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am. J. Hum. Genet. **68**:444–456.

Dahl, H.-H. M., R. M. Brown, W. M. Hutchison, C. Maragos, and G. K. Brown. 1990. A testis-specific form of the human pyruvate Dehydrogenase E1_subunit is coded for by and introless gene on chromosome 4. Genomics **8**:225–232.

Datta, U., I. D. Wexler, D. S. Kerr, I. Raz, and M. S. Patel. 1999. Characterization of the regulatory region of the human testis-specific form of the pyruvate dehydrogenase _subunit (PDHA-2) gene. Biochim. Biophys. Acta **1447**:236–243.

Dierick, H. A., L. Ambrosini, J. Spencer, T. W. Glover, and J. F. B. Mercer. 1995. Molecular structure of Menkes disease gene (*ATP7A*). Genomics **28**:462–469.

Dierick, H. A., J. F. B. Mercer, and T. W. Glover. 1997. A phosphoglycerate mutase brain isoform (*PGAM1*) pseudogene is localized within the human Menkes disease gene (*ATP7A*). Gene **198**:37–41.

Dunham, I., N. Shimizu, B. A. Roe et al. (214 co-authors). 1999. The DNA sequence of human chromosome 22. Nature **402**:489–495.

Fitzgerald, J., H.-H. M. Dahl, I. B. Jakobsen, and S. Easteal. 1996. Evolution of mammalian X-linked and autosomal *Pgk* and *Pdh E1_* subunit genes. Mol. Biol. Evol. **13**(7):1023–1031.

Fu, Y. X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693–709.

Fukunaga, R., and T. Hunter. 1997. MNK1, a new MAP kinase–activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. EMBO J. **16**(8):1921–1933.

Gonçalves, I., L. Duret, and D. Mouchiroud. 2000. Nature and structure of human genes that generate retropseudogenes. Genome Res. **10**:672–678.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider, J. Shoshani, G. Gunnell, and C. P. Groves 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **9**(3):585–598.

Grimsley, C., K. A. Mather, and C. Ober. 1998. HLA-H: a pseudogene with increased variation due to balancing selection at neighboring loci. Mol. Biol. Evol. **15**(12):1581–1588.

Grisolia, S., and J. Carreras. 1975. Phosphoglycerate mutase from yeast, chicken, breast muscle and kidney (2,3-PGA-dependent). Methods Enzymol. **42**:435–450.

Grisolia, S., and B. K. Joyce. 1959. Distribution of two types of phosphoglyceric acid mutase, diphosphoglycerate mutase and D-2, 3-Dipphosphoglyceric acid. J. Biol. Chem. **234**(6): 1335–1337.

Hattori, M., A. Fujiyama, T. D. Taylor et al. (63 co-authors). 2000. The DNA sequence of human Chromosome 21. Nature **405**:311–319.

Kaessmann, H., F. Heißig, A. von Haeseler, and S. Pääbo. 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat. Genet. **22**:78–81.

Kimura, M. 1983. The Neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2000. MEGA: molecular evolutionary genetics analysis. Version 2.0. Pennsylvania State University, University Park, and Arizona State University, Tempe.

Lander, E. S., L. M. Linton, B. Birren et al. (248 co-authors). 2001. Initial sequencing and analysis of the human genome. Nature **409**:860–921.

Li, W.-H. 1997. Molecular Evolution. Sinauer Associates, Sunderland, Mass.

Li, W.-H., T. Gojobori, and M. Nei. 1981. Pseudogenes as a paradigm of neutral evolution. Nature **292**:237–239.

Li, W.-H., and L. A. Sadler. 1991. Low nucleotide diversity in man. Genetics **129**:513–523.

Maestre, J., T. Tchénio, O. Dhellin, and T. Heidmann. 1995. mRNA retroposition in human cells: processed pseudogene formation. EMBO J. **14**(24):6333–6338.

McCarrey, J. R. 1987. Nucleotide sequence of the promoter region of a tissue-specific human retroposon: comparison with its housekeeping progenitor. Gene **61**:291–298.

McCarrey, J. R., M. Kumari, M. J. Aivaliotis, Z. Wang, P. Zhang, F. Marshall, and J. L. Vandeberg. 1996. Analysis of the cDNA and encoded protein of the human testis-specific *PGK-2* gene. Dev. Gen. **19**:321–332.

McCarrey, J. R., and K. Tomas. 1987. Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. Nature **326**:501–505.

Mighell, A. J., N. R. Smith, P. A. Robinson, and A. F. Markham. 2000. Vertebrate pseudogenes. FEBS Lett. **468**:109–114.

Miyata, T., and H. Hayashida. 1981. Extraordinarily high evolutionary rate of pseudogenes: evidence for the presence of selective pressure against changes between synonymous codons. Proc. Natl. Acad. Sci. USA **78**(9):5739–5743.

Nei, M., and S. Kumar. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, UK.

Proundfoot, N. J., and T. Maniatis. 1980. The structure of a human α-globin pseudogene and its relationship to a α-globin gene duplication. Cell **21**:537–544.

Przeworski, M., R. R. Hudson, and A. Di Rienzo. 2000. Adjusting the focus on human variation. TIG **16**(7):296–302.

Rozas, J., and R. Rozas. 1999. DnaSP version 3.52: an integrated program for molecular population genetics and molecular evolution. Bioinformatics **15**:174–175.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

Sakoda, S., S. Shanske, S. DiMauro, and E. A. Schon. 1988. Isolation of cDNA encoding the B isozyme of human phosphoglycerate mutase (PGAM) and characterization of the PGAM gene family. J. Biol. Chem. **263**(32):16899–16905.

Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics **129**(2):555–562.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. 3rd edition. Freeman, New York.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

———. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135**:599–601.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

Vanin, E. F. 1985. Processed pseudogenes: characteristics and evolution. Annu. Rev. Genet. **19**:253–272.

Venter, J. C., M. D. Adams, E. W. Myers et al. (274 co-authors). 2001. The sequence of the human genome. Science **291**:1304–1351.

Wagner, M. 1986. A consideration of the origin of processed pseudogenes. TIG 134–137.

Watterson, G. A. 1975. On the number of segregation sites in genetical models without recombination. Theor. Popul. Biol. 256–276.

WINTER, H., L. LANGBEIN, M. KRAWCZAK, D. N. COOPER, L. F. JAVE-SUAREZ, M. A. ROGERS, S. PRAETZEL, P. J. HEIDT, and J. SCHWEIZER. 2001. Human type I hair keratin pseudogene phihHaA has functional orthologs in the chimpanzee and gorilla: evidence for recent inactivation of the human gene after the Pan-Homo divergence. Hum. Genet. **108**(1):37–42.

YU, N., Z. ZHAO, Y.-X. FU et al. (11 co-authors). 2001. Global patterns of human DNA sequence variation in a 10 kb region on chromosome 1. Mol. Biol. Evol. **18**(2):214–222.

ZHAO, Z., L. JIN, Y.-X. FU et al. (13 co-authors). 2000. World-wide DNA sequence variation in a 10 kb noncoding region on human chromosome 22. PNAS **97**(21):11354–11358.

DAVID IRWIN, reviewing editor

Accepted December 17, 2001