Evolution of Mammalian X-linked and Autosomal $Pgk$ and $Pdh\ E1\alpha$ Subunit Genes

Jamie Fitzgerald, *1 Hans-Henrik M. Dahl,*1 Ingrid B. Jakobsen,† and Simon Easteal†

*The Murdoch Institute, Royal Children’s Hospital, Melbourne; and †The John Curtin School of Medical Research, The Australian National University

The phylogeny and substitution rates of the mammalian X chromosome-located and autosomal phosphoglycerate kinase and pyruvate dehydrogenase genes were investigated. Compatibility analysis was used to show reticulate evolution in these genes. Analysis of the marsupial, mouse, and human phosphoglycerate kinase genes suggests that at least two recombination events have taken place, one occurring around the time of the placental-marsupial split involving exons 1-5 and the other before the primate-rat split involving exons 9-10. Similar analysis of the pyruvate dehydrogenase genes indicates a recombination event involving exons 2-3 at a time before the primate-rat split and a gene conversion between exons 3-4 in the human somatic and testis-specific pyruvate dehydrogenase genes after the primate-rat split. This demonstrates that genetic exchange can occur between paralogous genes at widely separated chromosomal locations. Estimation of nucleotide substitution rates in these genes confirmed a higher substitution rate in the primate-rat dehydrogenase genes. In the phosphoglycerate kinase genes, there is no difference between the substitution rates in mice and humans and between the X chromosome- and autosomelocated genes. A greater substitution rate was noted in the mouse autosomal pyruvate dehydrogenase gene when compared with the other mouse and human genes. This relaxation of functional constraint at this specific gene.

Introduction

The genes that code for phosphoglycerate kinase ($Pgk$, E.C.2.7.2.3) and pyruvate dehydrogenase Ela subunit ($Pdh$, E.C. 1.2.4.1) share similarities in organization and expression pattern (Brown, Dahl, and Brown 1990; McCarrer 1990). In placental mammals, both enzymes have two variants. The somatic isoforms of both proteins ($Pgk-1$ and $Pdh-1$) are encoded by X chromosome loci, and these variants are expressed exclusively in somatic tissues. The testis-specific isoforms ($Pgk-2$ and $Pdh-2$) are expressed only in spermatogenic cells and are encoded by autosomal-located, intronless genes (McCarrer and Thomas 1987; Dahl et al. 1990).

In both cases, it has been suggested that the testis-specific genes arose as retroposons. Both $Pgk$ and $Pdh$ are coded for by single autosomal genes in monotremes (Graves 1987; Wrigley and Graves 1987; J. Graves, personal communication). In marsupials, there are autosomal and X-linked functional $Pgk$ genes, although the presence of a large family of $Pgk$ pseudogenes has complicated their molecular isolation and characterization (Cooper et al. 1994). DNA sequence has been obtained from one marsupial $Pgk$ pseudogene (van Daal, Cooper, and Molloy 1989), but it is not yet available for the marsupial testis-specific $Pgk-2$ gene. The presence of a testis-specific $Pgk$ isoform in some species (Graves 1990) suggests a $Pgk-1/Pgk-2$ gene duplication after the monotreme-therian divergence and before the placental-marsupial divergence.

In contrast, marsupials have only one $Pdh$ gene, which is autosomal and contains introns (Fitzgerald et al. 1993). It is situated in a chromosomal region that appears to have been translocated to the X chromosome in placents (Graves 1987). The absence of an intronless testis-specific $Pdh$ variant in marsupials and the autosomal location of the marsupial gene suggest that the $Pdh-1/Pdh-2$ duplication happened after the marsupial-placental divergence.

The $Pgk$ and $Pdh$ genes have been isolated and sequenced in a number of species, including humans (Homo sapiens), mouse (Mus musculus), two marsupials—the Striped-faced Dunnart (Sminthopsis macroura) and the Tammar Wallaby (Macropus eugenii), Drosophila melanogaster, and Ascaris suum. This allows a number of issues about the evolution of these genes to be addressed through comparative sequence analysis.

First, the phylogeny of the genes, indicated by their chromosomal locations and patterns of expression in monotremes, marsupials, and placentals, can be tested.

Second, the sequence data can be used to investigate the possible occurrence of reticulate evolution (i.e., the exchange of information through recombination and gene conversion) between the autosomal and X-linked gene copies. Reticulate evolution between closely related paralogues is well documented (e.g., Fitch et al. 1990). These genes are usually tandemly arranged on the same chromosome, although chromosomal proximity does not appear to be a necessary condition for this kind of genetic exchange because reticulate evolution has been demonstrated between unlinked tRNA genes in yeast (Amstutz et al. 1985) and it is indicated between X- and Y-linked zinc-finger-containing genes (Hayashida, Kuma, and Miyata 1992). The generality of reticulate evolution of dispersed genes can be tested by analysis of the $Pgk$ and $Pdh$ paralogues because they are situated at widely separated chromosomal locations.

Third, substitution rates can be compared between orthologues in different mammalian lineages and between X-linked and autosomal paralogues. Sequence...
Comparisons in mammals have been interpreted as showing a higher rate of nucleotide substitution in rodents than in primates (Li, Tanimura, and Sharp 1987; Gu and Li 1992; Li 1993a), although this interpretation is disputed (Eastal and Collet 1994; Eastal, Collet, and Betty 1995). They have also been interpreted as showing a higher rate of substitution on Y chromosomes than on X chromosomes (Shimmin, Chang, and Li 1993; Chang et al. 1994). In both types of comparison, differences in cell generation time have been proposed as the explanation for differences in substitution rate. The results of these comparisons thus have important implications with respect to the mechanisms by which mutations arise and are substituted during evolution.

We present here a comparative analysis of the Pgk and Pdha gene sequences to address these issues.

Material and Methods

Sequences were aligned using Clustal V (Higgins, Bleasby, and Fuchs 1992), with minor adjustments made by eye using Genetic Data Environment (GDE) version 2 (Smith et al. 1994). Few indels were indicated by the alignments, which were largely unambiguous and which are available on the World Wide Web at http://jcsmr.anu.edu.au/dmm/humgen.html and at the EMBL anonymous ftp site: ftp.ebi.ac.uk. The GenBank accession numbers for the compared sequences are as follows: human Pgk-I-V00572; human Pgk-2—X05246; mouse Pgk-Z-X53509; mouse Pgk-2-X55310; marsupial Pgk-I-X64296; marsupial Pdgk pseudogene—J03076; Drosophila melanogaster Pgk-Z14029; human Pdha-1—L13318; human Pdha-2-M76808; mouse Pdha-1-M76727; mouse Pdha-2-M76728; pig Pdha-1-X52990; marsupial Pdha-L20774; Ascaris suum Pdha-M76555.

Compatibility analyses of aligned sets of Pgk and Pdha sequences were used to detect evidence of reticulate evolution (Sneath, Sackin, and Ambler 1975; Jakobsen and Eastal, in press). Each site in the aligned sequences was compared for phylogenetic compatibility with all other sites in a pairwise fashion. Sites were identified as phylogenetically compatible if at the different sites a tree with the same topology could be constructed that implied the unique occurrence of all substitutions. Conversely, sites were identified as incompatible when all possible tree topologies applied to the different sites indicated the multiple occurrence of at least one substitution (Le Quesne 1969). Some sites, e.g., those that are invariant or at which only one sequence differs from all others, are necessarily compatible with all other sites and are thus not informative. These were removed from the analysis.

Matrices were constructed showing the compatibility or incompatibility of all pairs of informative sites arranged in their linear order in the sequence (fig. 1) using the program “reticulate,” available at http://jcsmr.anu.edu.au/dmm/humgen.html. Incompatibility between sites can arise from the multiple occurrence, within sites, of the same base changes in different sequences. It can also arise from the occurrence of recombination or gene conversion between some of the aligned sequences. These two causes can be distinguished because incompatibility resulting from recombination will be confined to the recombining regions of the aligned sequences and they will thus tend to be spatially clustered. In contrast, incompatibility resulting from multiple substitutions will not be clustered, except in the unusual case that there is variation in substitution rate among sites and clustering of sites with high substitution rates. The compatibility matrices were examined for clustering of incompatible comparisons as evidence of reticulate evolution. Regions that had relatively high and low proportions of incompatible comparisons were identified by visual inspection.
Results and Discussion

Evolutionary Pattern

The compatibility matrix for the human Pgk-1 and -2, mouse Pgk-1 and -2, and marsupial Pgk gene sequences (fig. 1A) indicates a region of relatively high compatibility extending from the start of exon 6 to the end of exon 8. Seventy-seven percent of the comparisons within this region are compatible. This is significantly greater than the mean for the comparable region of 1,000 alignments with randomly shuffled sites (65.2 ± 4.2%; P < 0.05). In the relatively “incompatible” region encompassing exons 1-5, 57.8% of comparisons are compatible, which is significantly less than the mean number for shuffled alignments (65.1 ± 3.6%; P < 0.05). In exons 9-10, 67.4% of comparisons are compatible, which is not significantly different from the value observed for the shuffled alignments (65.2 ± 4.9%). This pattern might be explained by a relatively faster rate of substitution, and hence multiple substitution, in these regions compared with exons 6-8.

Comparison of the numbers of substitutions in the three regions does not, however, indicate any such difference in rate. The complete distance matrices for all three regions are not presented here but can be obtained from http://jcsmr.anu.edu.au/dmm/humgen.html. The mean numbers of substitutions per nonsynonymous site for all the mammalian sequences compared with the Drosophila Pgk sequence are 0.248 in exons 1-5, 0.322 in exons 6-8, and 0.220 in exons 9-10. This indicates that, if anything, the substitution rate in exons 6-8 is the highest, not the lowest, of the three regions. These comparison are too distant to allow reliable estimates of substitution rates at synonymous sites. In comparisons between less divergent sequences there is also no indication of a difference in substitution rate of a kind that would explain the differences in the proportion of compatible comparisons in the three regions. Thus, for example, the numbers of substitutions per site between the human and mouse Pgk-1 genes are 0.347 in exons 1-5, 0.260 in exons 6-8, and 0.385 in exons 9-10 at synonymous sites and 0.013 in exons 1-5, 0.025 in exons 6-8, and 0.004 in exons 9-10 at nonsynonymous sites. Between the human and mouse Pgk-2 genes, the numbers of substitutions per site are 0.581 in exons 1-5, 0.579 in exons 6-8, and 0.329 in exons 9-10 at synonymous sites and 0.082 in exons 1-5, 0.058 in exons 6-8, and 0.078 in exons 9-10 at nonsynonymous sites.

Thus, the variation among regions evident in figure 1A is not readily explained by differences in substitution rate. It is, however, also not immediately apparent that it can be explained by recombination between regions in some of the sequences. Recombination between regions will normally result in greater incompatibility between the recombined regions than within them. This is not the pattern observed in figure 1A. The proportion of compatible comparisons between sites in exons 1-5 and sites in exons 6-8 is 63.9%. This is intermediate between the proportion of compatible comparisons between sites within exons 1-5 (57.8%) and sites within exons 6-8 (77.5%). Similarly, 70.5% of comparisons between exons 6-8 sites and exons 9-10 sites are compatible, which is also intermediate between the 77.5% of compatible comparisons within exons 6-8 and the 67.4% of compatible comparisons within exons 9-10.

Phylogenetic analyses of the five sequences indicate the following pattern of relationship for all three regions: (human Pgk-1, mouse Pgk-Z), (marsupial Pgk), (human Pgk-2, mouse Pgk-2), with clustering of the two Pgk-1 genes with each other and the two Pgk-2 genes with each other in a high proportion of bootstrap samples (figs. 2A-C). This phylogeny indicates that no reticulate evolution occurred in either the primate or the rodent lineage after the two lineages split. If reticulate evolution had occurred in one of these lineages, the phylogeny of the converted region would depend on the lineage converted and the direction of conversion. It would be one of the following: (1) (human Pgk-1, human Pgk-2), (mouse Pgk-1, mouse Pgk-Z), (2) (human Pgk-1, human Pgk-2), (mouse Pgk-2), (mouse Pgk-Z, marsupial Pgk), (3) (human Pgk-1, marsupial Pgk), (human Pgk-2, mouse Pgk-Z), (4) (marsupial Pgk, human Pgk-2), (human Pgk-1, mouse Pgk-1, mouse Pgk-Z). Conversion in both lineages would give the following phylogeny: (human Pgk-1, human Pgk-2), (marsupial Pgk), (mouse Pgk-Z, mouse Pgk-1, mouse Pgk-2). The sites compatible with each of these five topologies are as follows: (1) (34, 40, 45, 51, 52, 78, 117, 192, 210, 248, 300, 316, 337, 338, 346, 372, 452, 561, 565, 600, 618, 684, 726, 951, 987, 1050, 1083, 1089, 1125, 1161); (2) (12, 42, 45, 66, 69, 78, 132, 174, 192, 248, 294, 300, 378, 600, 612, 634, 669, 684, 705, 1249); (3) (462, 468, 507, 630, 639, 741, 804, 918, 985, 990, 1005, 1048, 1200); (4) (15, 225, 262, 286, 342, 462, 468, 477, 918, 939, 1023); (5) (45, 78, 1023).
FIG. 2-A-C, Phylogeny of the marsupial $Pgk\text{-}1$, human $Pgk\text{-}1$ and -2, and mouse $Pgk\text{-}1$ and -2 sequences estimated using the neighbor-joining method, with the Drosophila $Pgk$ sequence as an outgroup for: A: exons 1–5; B: exons 6–8; C: exons 9–10. The percentage of bootstrap samples supporting the internal branches of the trees are shown. D. Overall phylogeny of the $Pgk$ genes with proposed recombination events involving exons 1-5 and 9-10 indicated by arrows.

192, 248, 300, 462, 468, 600, 684, 918). In no case are these clustered, and most are synonymous sites at which base differences result from transition-type mutations. This suggests that at these sites incompatibility with the expected phylogeny results from the multiple occurrence of the same substitution in different sequences rather than from reticulate evolution.

The use of the Drosophila sequence as an outgroup, however, shows that the topology of the exons 9-10 phylogeny (fig. 2C) differs from that of the phylogeny of the other regions (figs. 2A and 2B). The phylogenies of exons 1-5 and of exons 6–8 indicate a separation of the X-linked and autosomal genes before the placental-marsupial divergence. As discussed in the Introduction, this is the phylogeny expected from the chromosomal locations and expression patterns of the genes in different mammalian taxa. The exons 9-10 phylogeny indicates a separation of the placental X-linked and autosomal genes after the placental-marsupial divergence. A clustering of the placental X-linked and autosomal genes is indicated in 97.8% of bootstrap samples, providing a high degree of confidence in this aspect of the tree topology. The different phylogeny for exons 9-10 can be explained as resulting from recombination between the X-linked and autosomal genes occurring before the primate-rodent split, but after the placental-marsupial split.

The phylogeny estimated for exons 1-5 is the same as that estimated for exons 68. However, the branching position of the marsupial lineage appears much closer to separation of the X-linked and autosomal genes in the placental species and is supported by only 44% of bootstrap samples. The possibility that in this region of the gene the marsupial sequence diverged before the two placental genes cannot be discounted. In any case, the two divergences appear to be very close, in contrast to their clear separation in exons 6-8. This may indicate recombination occurring between the two copies of the gene in placentals, close to the time of the placental-marsupial divergence.

From the preceding analysis two possible recombination events are indicated, for which there is not strong evidence, at about the time of the placental-marsupial split and involving exons 1-5, and another, for which there is strong evidence, occurring later, but before the primate-rodent split, and involving exons 9 and 10 (fig. 2D). In the five-sequence tree ([human $Pgk\text{-}2$, mouse $Pgk\text{-}2$], [marsupial $Pgk$], [human $Pgk\text{-}1$, mouse $Pgk\text{-}1$]), these events would have occurred on the internal branches on either side of the node that separates the marsupial branch. In the four sequence tree, in which the marsupial sequence is omitted ([human $Pgk\text{-}2$, mouse $Pgk\text{-}2$], [human $Pgk\text{-}1$, mouse $Pgk\text{-}1$]), the two branches on either side of the marsupial node are fused into a single internal branch. In a compatibility matrix for these four sequences, any incompatibilities arising from recombination between these two internal branches would no longer be evident.

In the four-sequence compatibility matrix (fig. 1B), there are fewer informative sites (98) than in the five-sequence matrix (157). This is expected with a reduction in the number of aligned sequences. There is also a sub-
A substantial increase in the overall proportion of compatible comparisons (77.6% compared with 65.2% for the five-sequence matrix). This is not seen in four-sequence compatibility matrices involving the removal of individual placental sequences. In all of these, the proportion of compatible comparisons is substantially reduced, ranging from 38.3% to 50%. This further indicates the role of recombination, occurring between the placental-marsupial and the primate-rodent splits, in generating the clustered pattern of incompatibility seen in the five-sequence matrix.

**Substitution Rates**

Rates of substitution can be compared between some pairs of sequence lineages despite the reticulate nature of the evolution of these genes indicated by the above analysis. Estimates of K(4) and K(O) for all sequences, including the marsupial pseudogene, are shown in table 1. The relative rate test involves a comparison of compatible comparisons is substantially reduced, ranging from 38.3% to 50%. This further indicates the role of recombination, occurring between the placental-marsupial and the primate-rodent splits, in generating the clustered pattern of incompatibility seen in the five-sequence matrix.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Numbers of Nucleotide Substitutions per 100 Fourfold-degenerate Sites (Above the Diagonal) and per 100 Nondegenerate Sites (Below the Diagonal) Between Mammalian Pgk Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Human 1</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Human 1</td>
<td>31.8 ± 4.8</td>
</tr>
<tr>
<td>Mouse 1</td>
<td>67.0 ± 9.6</td>
</tr>
<tr>
<td>Human 2</td>
<td>43.9 ± 6.3</td>
</tr>
<tr>
<td>Mouse 2</td>
<td>7.2 ± 1.0</td>
</tr>
<tr>
<td>Marsupial 1</td>
<td>8.9 ± 1.1</td>
</tr>
<tr>
<td>Marsupial 2</td>
<td>13.2 ± 1.4</td>
</tr>
<tr>
<td>Drosophila</td>
<td>24.2 ± 2.0</td>
</tr>
</tbody>
</table>

Substitution rates in terminal branches leading to genes i and j, with reference gene k, can be estimated from $K_{ij} = (K_{ij} + K_{ik} - K_{jk})/2$ and equivalently for $K_{0j}$, where $O$ is the branch point of genes i and j. The standard errors of $(K_{ik} - K_{jk})$ can be used to estimate the 95% confidence interval of $K_{ij}/K_{0j}$. Thus, for example, in comparing fourfold degenerate sites between human 1 (gene 1), mouse 1 (gene 2), and marsupial (gene 3), $K_{12} = 31.8$ (from table 1). $K_{12}/K_{02} = 15.9$. The standard error of $K_{12} - K_{23} = 6.2$ (from table 2), giving $K_{01}/K_{02} = 15.9$. The standard error of $K_{12} - K_{23}$ is 6.2 (from table 2), giving a 95% confidence interval of 0.44 to 2.3 for $K_{01}/K_{02}$.

The results of the present analysis can be discussed in the context of previous work on relative substitution rates between rodents and primates and between sex chromosomes and autosomes. However, because the analysis is of individual genes, it is not sufficient to allow general conclusions to be drawn about relative substitution rates. Fourfold degenerate sites are likely to be largely free from the effects of natural selection. Differences in substitution rate at these sites are expected to result from general lineage effects such as variation in mutation rate. Easteel and Collet (1994) found no evidence of a rate difference at fourfold degenerate sites between rodents and primates, and the present results are consistent with this.

At nonsynonymous and nondegenerate sites, rate variation can result from both general lineage effects and

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Differences in the Number of Nucleotide Substitutions per 100 Sites (K) Between Pgk Genes in Different Taxa or on Different Chromosomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
<td>Human 1</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Human 1</td>
<td>31.8 ± 4.8</td>
</tr>
<tr>
<td>Mouse 1</td>
<td>67.0 ± 9.6</td>
</tr>
<tr>
<td>Human 2</td>
<td>43.9 ± 6.3</td>
</tr>
<tr>
<td>Mouse 2</td>
<td>7.2 ± 1.0</td>
</tr>
<tr>
<td>Marsupial 1</td>
<td>8.9 ± 1.1</td>
</tr>
<tr>
<td>Marsupial 2</td>
<td>13.2 ± 1.4</td>
</tr>
<tr>
<td>Drosophila</td>
<td>24.2 ± 2.0</td>
</tr>
</tbody>
</table>

$K_{12} = 31.8$ (from table 1). $K_{12}/K_{02} = 15.9$. The standard error of $K_{12} - K_{23} = 6.2$ (from table 2), giving a 95% confidence interval of 0.44 to 2.3 for $K_{01}/K_{02}$.

<table>
<thead>
<tr>
<th>COMPARISON</th>
<th>GENE</th>
<th>NONDEGENERATE SITES</th>
<th>FOURFOLD-DEGENERATE SITES</th>
</tr>
</thead>
<tbody>
<tr>
<td>P r i m a t e - R o d e n t</td>
<td>Human 1</td>
<td>Mouse 1</td>
<td>Marsupial 1</td>
</tr>
<tr>
<td>Human 1</td>
<td>-0.5 ± 0.7</td>
<td>0 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>Human 2</td>
<td>-0.7 ± 0.9</td>
<td>-1.3 ± 0.9</td>
<td></td>
</tr>
<tr>
<td>X chromosome-autosome</td>
<td>Human 1</td>
<td>Human 2</td>
<td>Drosophila</td>
</tr>
<tr>
<td>X chromosome-autosome</td>
<td>Mouse 1</td>
<td>Mouse 2</td>
<td>Drosophila</td>
</tr>
<tr>
<td>P l a c e n t a l - M a r s u p i a l</td>
<td>Human 2</td>
<td>Marsupial 1</td>
<td>Drosophila</td>
</tr>
<tr>
<td>P l a c e n t a l - M a r s u p i a l</td>
<td>Mouse 2</td>
<td>Marsupial 1</td>
<td>Drosophila</td>
</tr>
</tbody>
</table>

NOTE.—$L =$ no. of sites.
gene-specific effects of natural selection. Although not significant, the direction of the difference in rate in the two primate‐rodent comparisons is consistent with the general finding of a 40% greater rate of substitution in rodents compared with primates (Gu and Li 1992, Easteal and Collet 1994). Gu and Li (1992) argue that this difference is the result of a greater mutation rate in the rodent lineage. This, however, is inconsistent with the lack of difference in rate at fourfold degenerate sites reported by Easteal and Collet (1994). Eastal and Collet (1994) concluded that the difference reflects the differential effects of natural selection.

Similarly, although not significant, the direction of the differences between the X-chromosomal and autosomal genes is consistent with a slightly lower rate of substitution on the X chromosome (Miyata et al. 1990). This prediction is based on the assumption that mutation rate depends on cell generation time and the fact that there are more cell generations in male than in female germ lines. Miyata et al.’s (1990) predictions (including that the Y chromosome will evolve faster than autosomes) are borne out by a number of studies (Shimmin, Chang, and Li 1993, Chang et al. 1994). However, cell generation time is not the only difference between males and females that could explain the difference in mutation rate. Other explanations include differential methylation between male and female germ cells, differences in effective population size, and differences in the intensity of purifying selection (Charlesworth 1993, Kettering et al. 1993, Eastal, Collet, and Betty 1995, pp. 139-142). These factors may also account for differences in substitution rate between genes on the different kinds of chromosome.

There is no evidence of rate difference between marsupial and placental genes. However, many more genes will need to be analyzed before any firm conclusions are drawn about the relative substitution rates between these mammalian subclasses.

Pdha
Evolutionary Pattern

The compatibility matrix for the marsupial, the two human and the two mouse Pdha genes (fig. 3A) shows an overall level of 51.7% compatibility, much less than for the Pgk genes. This probably reflects, in part, the greater rate of substitution in this gene (Eastal and Collet 1994) and, hence, a greater rate of multiple substitutions at individual sites. There is also less indication of presence of compatibility/compatibility. Inspection of the matrix indicates a region with a relatively higher proportion of compatible comparisons encompassing exons 7-10. Within this region, 65.4% of comparisons are compatible, which is significantly more than the mean for the comparable region of 1,000 alignments with randomly shuffled sites (51.7 ± 3.7; P < 0.01). There is also an indication of low compatibility (46.2%) in comparisons between sites in exons 1-3 and sites in exons 4-6.

These patterns were investigated further by constructing compatibility matrices with each of the five sequences removed in turn. As in the case of the Pgk genes, the greatest reduction in overall compatibility is achieved by removing the marsupial sequence (fig. 3B). Of the resulting 61 informative sites, 39 indicate the expected phylogeny: ([human Pdha-1, mouse Pdha-1] and -2, and mouse Pdha-2] and -2 genes; and [B, All pairs of the 61 informative sites in the aligned sequences of the human Pdha-1 and -2 and mouse Pdha-1 and -2 genes. Exons are numbered and their boundaries are indicated by arrows. The informative sites in A are 7, 21, 24, 25, 33, 39, 64, 66, 75, 78, 81, 88, 89, 93, 105, 108, 110, 114, 123, 135, 153, 159, 165, 171, 183, 190, 192, 195, 211, 231, 234, 237, 245, 252, 261, 264, 267, 270, 273, 274, 279, 283, 288, 291, 303, 304, 305, 308, 309, 318, 321, 322, 324, 327, 331, 332, 333, 348, 355, 359, 378, 381, 402, 417, 420, 432, 447, 456, 468, 480, 483, 488, 499, 504, 513, 516, 534, 540, 549, 589, 611, 612, 631, 657, 666, 672, 681, 687, 690, 696, 709, 711, 717, 720, 723, 735, 746, 750, 753, 792, 801, 813, 822, 837, 843, 864, 879, 885, 895, 897, 901, 914, 921, 928, 930, 953, 954, 959, 993, 997, 1005, 1008, 1024, 1042, 1043, 1047, 1074, 1091, 1098 in the alignment (http://jesmr.anu.edu.au/dmm/hungen.html). The informative sites in B are indicated in bold. Fig. 3-Matrices of
phylogenetic compatibility (white squares) and incompatibility (black squares) between: A, All pairs of the 130 informative sites in the aligned sequences of the marsupial Pdha, human Pdha-1 and -2, and mouse Pdha-1 and -2 genes; and B, All pairs of the 61 informative sites in the aligned sequences of the human
Evolution of *Pgk* and *Pdha* Genes

**Rodent Split**

The clustering in this region of the human *Pdha-1* and *Pdha-2* genes with mouse *Pdha-1* suggests that the direction of this event was from *Pdha-1* to *Pdha-2* (fig. 40).

In neither of the remaining two regions of the gene is there strong support for the expected phylogeny in which the *Pdha-1* and *Pdha-2* genes diverged after the marsupial-placental split. However, alternative phylogenies are not strongly supported, and the data appear inadequate to confidently resolve the gene phylogeny.

**Substitution Rates**

*K(4)* and *K(0)* values between all pairs of *Pdha* sequences (including the pig *Pdha-1* sequence) are shown in table 3. The uncertainty about the phylogeny and the possibility of recombination events make analysis of evolutionary rates between sequence lineages difficult. The marsupial sequence can probably be used as a legitimate reference in comparing human and mouse X-linked genes. In this comparison, no difference in *K(4)* (not shown) or *K(0)* is apparent between the species, as previously reported by Easteal and Collet (1994). For all other comparisons, the only reference that can be reliably used is the *Ascaris* sequence. Because this sequence is very divergent, reliable *K(4)* values between it and the mammalian sequences cannot be estimated. The relative rate comparisons using the *Ascaris* sequence as a reference (table 4) show that *K(0)* in the mouse autosomal gene is significantly greater than in the mouse X-linked gene and the marsupial gene. No significant differences were apparent in any of the other comparisons.

The greater rate of substitution in mouse *Pdha-2* might be a result of a relaxation of functional constraint.

---

### Table 3

Numbers of Nucleotide Substitutions per 100 Fourfold-degenerate Sites (Above the Diagonal) and per 100 Nondegenerate Sites (Below the Diagonal) Between Mammalian *Pdha* Genes

<table>
<thead>
<tr>
<th></th>
<th>Human 1</th>
<th>Pig 1</th>
<th>Mouse 1</th>
<th>Human 2</th>
<th>Mouse 2</th>
<th>Marsupial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human 1</td>
<td>43.7 ± 7.6</td>
<td>51.3 ± 8.6</td>
<td>38.0 ± 6.9</td>
<td>96.0 ± 17.0</td>
<td>85.6 ± 14.8</td>
<td></td>
</tr>
<tr>
<td>Pig 1, . . .</td>
<td>0.9 ± 0.4</td>
<td>63.1 ± 10.1</td>
<td>54.6 ± 9.6</td>
<td>100.2 ± 17.1</td>
<td>88.6 ± 15.4</td>
<td></td>
</tr>
<tr>
<td>Mouse 1</td>
<td>0.9 ± 0.4</td>
<td>1.4 ± 0.5</td>
<td>79.2 ± 13.1</td>
<td>138.6 ± 30.2</td>
<td>107.7 ± 18.4</td>
<td></td>
</tr>
<tr>
<td>Human 2</td>
<td>7.5 ± 1.2</td>
<td>8.0 ± 1.2</td>
<td>7.2 ± 1.1</td>
<td>76.1 ± 12.4</td>
<td>65.9 ± 10.8</td>
<td></td>
</tr>
<tr>
<td>Mouse 2, . . .</td>
<td>14.0 ± 1.7</td>
<td>13.8 ± 1.6</td>
<td>13.9 ± 1.7</td>
<td>13.8 ± 1.6</td>
<td>135.5 ± 27.3</td>
<td></td>
</tr>
<tr>
<td>Marsupial, . .</td>
<td>3.2 ± 0.7</td>
<td>4.0 ± 0.8</td>
<td>3.5 ± 0.8</td>
<td>9.5 ± 1.3</td>
<td>46.6 ± 3.6</td>
<td></td>
</tr>
<tr>
<td><em>Ascaris</em></td>
<td>38.8 ± 3.2</td>
<td>38.6 ± 3.1</td>
<td>40.4 ± 3.3</td>
<td>43.6 ± 3.4</td>
<td>39.8 ± 3.2</td>
<td></td>
</tr>
</tbody>
</table>
of directional natural selection, or of an increased mutation rate. It does not appear to be a reflection of some factor that is specific either to mice, because the rate is increased relative to mouse Pdha-1, or to an autosomal-specific factor, because there is not a similar increase in rate in the human Pdha-2.

**Conclusions**

We have provided evidence of reticulate evolution between placental X-linked and autosomal Pgk genes, and we have shown that there is an indication of reticulate evolution between X-linked and autosomal Pdha genes. This confirms that genetic exchange can occur between paralogous genes that are not located in close proximity to each other. This possibility should be taken into account in comparative analyses of these kinds of genes. We can only speculate as to the mechanism by which the exchange occurs. One possibility is that it is mediated by the reverse transcription of mRNA demonstrated as a mechanism of gene conversion in yeast (Derr and Strathern 1993). The autosomal Pgk and Pdha genes are both intronless and are thought to have arisen by mRNA-mediated retrotransposition. It is possible that, once established, these genes recombined with DNA sequences reverse-transcribed from X-linked gene transcripts at a later date.

The relative rate comparisons show a significant increase in substitution rate in the mouse autosomal Pdha gene, although the reason for this increase cannot be discerned from the present analysis. The absence of any evidence of substitution rate difference in the other comparisons must be tempered by the realization that relatively few sites are being compared and that, in many cases, the reference gene is very divergent. There may be real differences that are undetected.

**Acknowledgments**

J.F. and I.B.J. were recipients of Melbourne University and Australian National University Postgraduate Scholarships, respectively, and H.-H.M.D. is a National Health and Medical Research Council (NH&MRC) Senior Research Fellow. This research was partly supported by a Block Grant from the NH&MRC.

**LITERATURE CITED**


———. 1990. Gene maps of marsupials (mammalian infraclass Metatheria) and monotremes (mammalian subclass...
Evolution of *Pgk* and *Pdha* Genes


Craig Moritz, reviewing editor

Accepted May 24, 1996