Modern Evolution of a Single-Copy Gene: The Immunoglobulin Cκ Locus in Wild Mice

Evelyne Jouvin-Marche,* Agnes Cuddihy,*† Sandra Butler,† J. Norman Hansen,† Walter M. Fitch,‡ and Stuart Rudikoff *

*Laboratory of Genetics, National Cancer Institute, National Institutes of Health; †Department of Chemistry and Biochemistry, University of Maryland at College Park; and ‡Department of Biological Sciences, University of Southern California, Los Angeles

The immunoglobulin kappa light-chain constant region gene (Cκ) has been cloned and sequenced from five wild mouse species. Analysis of these data has permitted an assessment of single-copy gene evolution during a limited time period as defined by the genus Mus. Sequence conservation was found to be as high (or higher) in the 5′ and enhancer regions as in the coding region. The pattern of substitutions throughout these genes suggests that parallel evolution has occurred frequently and that substitutions at replacement sites have not decreased significantly, owing to saturation during this period of ~10 Myr. Phylogenetic relationships have been determined among these wild species as well as among members of the genus Rattus.

Introduction

Many attempts have been made to study gene evolution in higher vertebrates by analysis of the same gene(s), e.g., globin and preproinsulin, among different species (Efstratiadis et al. 1980; Perler et al. 1980). However, in most of these instances, the species employed are widely separated in evolution so that estimates of divergence times based, for example, on fossil records may reflect considerable error, which is virtually impossible to correct for. A second major complication relates to the complexity of the families in which a specific gene is found. Many gene families contain such high copy numbers that it is extremely difficult to follow the evolution of a single member. Furthermore, the larger the gene family the greater the potential for complex genetic events, such as recombination or conversion, that may obscure mutational patterns and frequencies found in single-copy genes.

To avoid some of these difficulties in addressing questions of vertebrate gene evolution, we have studied the single-copy murine Cκ gene, which encodes the constant region of the immunoglobulin kappa light chain. This system offers the advantage that evolution of the genus Mus spans only ~10 Myr, and many of the earliest arising species are still in existence today. This genus has clearly been one of the most successful in modern times, and colonies are in existence that contain representatives of nearly all major taxonomic groups (Potter 1986), providing access to essentially the entire murine evolutionary spectrum.

1. Key words: single-copy gene, evolution, wild mice, phylogeny.

Address for correspondence and reprints: Dr. Stuart Rudikoff, Laboratory of Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892.

© 1988 by The University of Chicago. All rights reserved.
0737-4038/88/0505-0003$02.00

500
Table 1
Origins of Wild Mice

<table>
<thead>
<tr>
<th>Species</th>
<th>Subgenus</th>
<th>Geographic Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mus pahari</td>
<td>Coelomys</td>
<td>Tak Province, Thailand</td>
</tr>
<tr>
<td>M. saxicola</td>
<td>Pyromys</td>
<td>Mysore</td>
</tr>
<tr>
<td>M. minutoides</td>
<td>Nannomys</td>
<td>Nairobi</td>
</tr>
<tr>
<td>M. cookii</td>
<td>Mus</td>
<td>Tak Province, Thailand</td>
</tr>
<tr>
<td>M. spretus</td>
<td>Mus</td>
<td>Azrou, Morocco</td>
</tr>
<tr>
<td>M. musculus domesticus</td>
<td>Mus</td>
<td>United States</td>
</tr>
</tbody>
</table>

Material and Methods: Molecular Cloning and Sequence Determination

Recombinant-DNA libraries were constructed from five wild mouse species in the bacteriophage lambda J1 as described elsewhere (Jouvin-Marche and Rudikoff 1986). Libraries were screened with a cDNA probe encoding the \( V_\kappa \) and \( C_\kappa \) regions from MOPC 167 (Joho et al. 1980). Positive clones were rescreened at a wash stringency of 65 \( ^\circ \)C, 0.2 × standard saline citrate, 0.1% sodium dodecyl sulfate, using a 2.8-kb HindIII-BamHI fragment containing the germ-line \( C_\kappa \) sequence (Coleclough et al. 1981). Phage clones were restriction mapped, and appropriate fragments were obtained for subcloning into M13 vectors (Messing and Vieira 1982). Sequences were determined, by the chain termination method of Sanger (1977), in both directions. Pairwise sequence comparisons were made using the NUCALN program of Wilbur and Lipman (1984).

Results and Discussion

The present study was undertaken to examine substitution patterns in a single-copy gene during a short evolutionary period and to define phylogenetic relationships among a number of wild mouse species. A consensus taxonomy for the genus Mus does not presently exist, but, for the purpose of this discussion, we will employ the classification of Marshall (1981), which divides the genus into four subgenera—although the question has been raised as to whether the different “subgenera” in fact deserve ranking as separate genera (Bonhomme 1986). The subgenera Coelomys, Pyromys, and Nannomys contain the oldest wild species representing early divergence within the genus, and the subgenus Mus contains the more recently evolved commensal species, including the modern house mouse as well as most laboratory species. The various subgenera are both geographically and genetically isolated, as chromosome numbers differ and crossbreeding does not occur. It is thus possible in this system to observe evolution of the same gene in isolated populations within the genus.

Sequence Analysis

The \( C_\kappa \) locus was analyzed in five species of wild mice, including representatives of each of the subgenera (table 1). The \( C_\kappa \) gene was cloned from recombinant bacteriophage libraries (Jouvin-Marche and Rudikoff 1986) and subjected to nucleic acid sequence determination. Sequences were determined from \( \sim 1,000 \) bp upstream of the coding region to 214 bp into the 3' untranslated region. The 5' segment is part of the intron between the \( J_\kappa 5 \) and \( C_\kappa \) coding regions and contains the kappa chain enhancer, which is highly conserved in mouse, human, and rabbit (Emorine et al. 1983). All of the \( C_\kappa \) coding regions contained normal termination codons and splice acceptor sites, indicating that they are functional genes.
The nucleotide sequence from each mouse species was aligned to that of *M. saxicola* (fig. 1). Except for the *M. saxicola* gene, there is an overall conservation in length. In the intervening sequence between the enhancer and the coding region, divergence is due to single substitutions and single base insertions or deletions, whereas 5' to the enhancer, small segmental deletions or insertions are present as well as a large 122-bp insertion in the *M. saxicola* sequence. Length variation in the 3' untranslated region is due to a different number of repeats of the sequence (CCT) which, for example, is found eight times in *M. pahari* and five times in *M. spretus*.

The striking feature of the aligned sequences is the great similarity along the entire length. The most highly conserved region is the enhancer, as seen in every pairwise comparison between species (table 2). The coding region is also well conserved,
with identities ranging from 90.5% between *M. pahari* and *M. minutoides* to 99.1% between *M. musculus domesticus* and *M. spreptus*. As expected, similarity decreases as more distant species are compared. Apart from the presence of insertions and deletions, there is little difference in the similarity of the noncoding segments, whether they are
<table>
<thead>
<tr>
<th></th>
<th><em>Mus musculus domesticus</em></th>
<th><em>M. spretus</em></th>
<th><em>M. cookii</em></th>
<th><em>M. saxicola</em></th>
<th><em>M. minutoides</em></th>
<th><em>M. pahari</em></th>
<th><em>Rattus norvegicus</em></th>
<th><em>R. leucopus cocktournensis</em></th>
<th><em>R. colletti</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. m. domesticus</em></td>
<td></td>
<td>99.1</td>
<td>96.3</td>
<td>95.3</td>
<td>94.1</td>
<td>92.5</td>
<td>88.4</td>
<td>88.4</td>
<td>89.4</td>
</tr>
<tr>
<td></td>
<td>(99.5)</td>
<td>(99.0)</td>
<td>(96.6)</td>
<td>(98.5)</td>
<td>(96.1)</td>
<td>(95.6)</td>
<td>(95.6)</td>
<td>(95.6)</td>
<td>(95.6)</td>
</tr>
<tr>
<td><em>M. spretus</em></td>
<td>98.5</td>
<td>95.9</td>
<td>95.0</td>
<td>94.1</td>
<td>92.8</td>
<td>88.6</td>
<td>88.6</td>
<td>88.6</td>
<td>89.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(99.0)</td>
<td>(96.6)</td>
<td>(98.1)</td>
<td>(95.2)</td>
<td>(95.2)</td>
<td>(95.2)</td>
<td>(95.2)</td>
<td>(95.2)</td>
</tr>
<tr>
<td><em>M. cookii</em></td>
<td>97.7</td>
<td>97.4</td>
<td>94.4</td>
<td>92.8</td>
<td>92.8</td>
<td>88.4</td>
<td>88.4</td>
<td>88.4</td>
<td>88.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(95.7)</td>
<td>(97.6)</td>
<td>(94.7)</td>
<td>(94.7)</td>
<td>(94.7)</td>
<td>(94.7)</td>
<td>(94.7)</td>
<td>(94.7)</td>
</tr>
<tr>
<td><em>M. saxicola</em></td>
<td>93.8</td>
<td>93.8</td>
<td>94.7</td>
<td>94.4</td>
<td>94.1</td>
<td>87.2</td>
<td>87.5</td>
<td>86.2</td>
<td>86.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(94.7)</td>
<td>(97.1)</td>
<td>(94.7)</td>
<td>(94.7)</td>
<td>(94.7)</td>
<td>(94.7)</td>
<td>(94.7)</td>
<td>(94.7)</td>
</tr>
<tr>
<td><em>M. minutoides</em></td>
<td>93.2</td>
<td>93.3</td>
<td>94.0</td>
<td>94.9</td>
<td>90.9</td>
<td>86.9</td>
<td>86.9</td>
<td>86.9</td>
<td>86.6</td>
</tr>
<tr>
<td><em>M. pahari</em></td>
<td>93.1</td>
<td>93.0</td>
<td>93.8</td>
<td>94.0</td>
<td>94.5</td>
<td>86.9</td>
<td>86.9</td>
<td>86.9</td>
<td>86.6</td>
</tr>
<tr>
<td><em>R. norvegicus</em></td>
<td>88.9</td>
<td>88.9</td>
<td>89.6</td>
<td>89.8</td>
<td>89.6</td>
<td>97.2</td>
<td>97.2</td>
<td>94.4</td>
<td>96.3</td>
</tr>
<tr>
<td><em>R. l. cocktournensis</em></td>
<td>88.8</td>
<td>88.6</td>
<td>89.0</td>
<td>90.0</td>
<td>90.0</td>
<td>98.0</td>
<td>98.0</td>
<td>98.0</td>
<td>98.0</td>
</tr>
<tr>
<td><em>R. colletii</em></td>
<td>89.0</td>
<td>88.9</td>
<td>89.1</td>
<td>89.8</td>
<td>89.5</td>
<td>98.4</td>
<td>98.4</td>
<td>98.5</td>
<td>98.4</td>
</tr>
</tbody>
</table>

**Note.**—Percent identity was calculated as (no. of positions where differences occur × 100)/total no. of nucleotide positions compared, and an insertion or deletion was counted as one nucleotide difference. Identities for the coding region and enhancer (in parentheses) are given in the upper right triangle, and those for noncoding regions are given in the lower left triangle. Number of bases compared: coding, 321; enhancer, 206; noncoding, 1,083. ( . . . ) Indicates that enhancer region sequences from these species were not available for comparison.
5' of the enhancer, between the enhancer and the coding region, or in the 3' untranslated region. The enhancer has been shown to be essential for expression of the kappa light chain in B cells (Queen and Stafford 1984) and therefore would be under strong selection, so this sequence was not included as part of the noncoding region. In many comparisons, the sequence conservation of the noncoding regions is equal to or greater than that found in the coding region. This same conservation of noncoding sequences has been reported among rat species and among different isotypes and allotypes in rabbit (Akimenko et al. 1984; Frank et al. 1984).

When the nucleotide sequences of all mouse species are compared with those of three rat species, the coding regions are conserved in length but the 5' and 3' noncoding regions have insertions and deletions—although even in these regions nucleotide sequence is preserved. None of the insertions and deletions are shared between rat and mice, indicating that the murine subgenera evolved from a common ancestor subsequent to the separation of rat and mouse. Nineteen of the total 33 insertion/deletions are identical in sequence to adjacent sequences, as if unequal crossing-over were a principle source of length change (Jones and Kafatos 1982). One of the insertion/deletions is almost certainly the result of unequal crossing-over occurring twice, if one base difference is allowed. At position -87 of R. norvegicus in the sequence GTGTCTGT GTGTCTGT CTATAACATGT CTATAACATGC, the central two strings represent the insertion/deletion that matches the two adjacent strings.

Parallel Evolution

Throughout the coding region, instances are found in which nucleotide substitutions are either specific to a particular species or are shared among several mouse species and even with the three rat species. Shared nucleotides can occur either because species evolved from a common ancestor or because of parallel evolution; that is, there is selective bias that a given base substitution will occur at a particular site. An example of common ancestral origin is seen at nucleotide position 151, where most rat and mouse species express G but where A and T are found in M. minutoides and M. saxicola, respectively. On the other hand, parallel evolution would explain position 101, where G is the common nucleotide but R. colletii, M. m. domesticus, and M. saxicola share an A. No phylogenetic data so far uncovered suggest that these three are monophyletic—i.e., have a common ancestor separate from the other species. Among 47 codons that have undergone at least one substitution, 21 nucleotide changes can be explained by parallel evolution, including two—nucleotides 40-42 and 82-84—that exhibit parallelism only at the amino acid level. Thirteen of these 21 occurrences result in amino acid changes. In the noncoding region, there are 16 instances of potential parallelism. Nonrandom replacements such as those described may result from (1) transitions being more likely to occur than transversions (Brown et al. 1982), (2) replacements that result in conservative amino acid changes being more likely to be tolerated than replacements that produce radical changes, and (3) nonrandom codon usage (Grosjean and Fiers 1982) or (G+C) and (A+T) pressure (Jukes and Bhushan 1986). Fifteen of the changes in the coding region and all of the changes in the noncoding region consistent with parallel evolution can occur by transitions, which account for 64% of all nucleotide substitutions. Of the four transversions occurring in the coding region (nucleotide positions 165, 171, 227, and 280), all result in amino acid replacements. There is no evidence of significant bias in codon usage, and the coding region is not unusually GC or AT rich.

An alternative explanation for these shared nucleotides would suggest that a pool
I. 506 Jouvin-Marche et al.

of polymorphic ancestral C<sub>ε</sub> genes may have existed which was disseminated among
the various subgenera during the relatively short time of their divergence. Such as-
sortment could explain substitutions shared by mice from two subgenera, i.e., nu-
cleotide 239, but the overall pattern of shared replacements would require subsequently
highly complex recombination events to produce the genotypes actually represented
in the obtained sequences, so parallel evolution may be the more likely explanation.

Phylogeny

To elucidate taxonomic relationships among the putative subgenera and the genus
Rattus, the aligned sequence data were used to construct a most parsimonious tree
(fig. 2). Several alternative trees that are not significantly different are possible. The
greatest ambiguity is concentrated around the branch points of the subgenera, so that,
for example, only one extra substitution is added if M. saxicola is joined to the pahari-
minutoides pair instead of branching off from all five other mice together. Two more
substitutions would be added if saxicola were joined to the M. cookii-spretus-domesticus
triad, which is representative of the subgenus Mus. To join minutoides or pahari to
this triad would cost three and five additional substitutions, respectively. A second
ambiguous branch point is in the rat lineage, where only one extra substitution would
join R. leucopus cooktownensis to R. norvegicus instead of to R. colletti and where
five extra substitutions would join colletti to norvegicus instead of to R. l. cooktownensis.
All other observed changes in the tree structure require 15 or more additional nucleotide
substitutions. Thus, it is apparent that (1) the Mus subgenera are more closely related
to each other than to the genus Rattus and (2) all four subgenera have diverged within
a relatively short evolutionary time span and the exact order of divergence of these
subgenera remains uncertain. When the genealogic tree is redrawn so that branch
lengths are proportional to nucleotide substitutions, it is obvious that given species
are accumulating change at different rates. For example, in the pahari branch there
are 61.0 substitutions, compared with 48 in minutoides. This result indicates a varying
substitution rate in related species.

The estimation of divergence times has been approached by a second method
which uses the frequency of silent substitutions (k<sub>s</sub>) to calculate the evolutionary
distance between a given pair of genes. All genes undergo a relatively high level of
substitution at silent sites, and it has been suggested that these changes are under little
selection and tend to accumulate at a constant rate (Miyata et al. 1980). The latter
assumption—a constant rate of substitution during the evolution of Mus and Rattus—
may generally be true, but there are clearly exceptions, as noted above, which must
be kept in mind in the following discussion. The calculated values for k<sub>s</sub> are given in
the upper right triangle of table 3. The mean k<sub>s</sub> for rat/mouse C<sub>ε</sub> divergence is
22.0, in agreement with the k<sub>s</sub> values calculated for several other genes, e.g., myelin
basic protein (Takahashi et al. 1985), 20.4; interleukin 3 (Cohen et al. 1986), 23.2;
MHC E<sub>β</sub> (Roberts and McMastcr 1985), 20.6; and actin β (Alonso et al. 1986),
23.0. When the calculated k<sub>s</sub> values were used, a rate (k) in the range of 8.60–15.0
× 10<sup>−9</sup> substitutions/site/year was obtained. This rate is one of the highest noted (Li
et al. 1985) and may reflect accelerated rodent evolution (Fitch and Atchley 1985;
Britten 1986; Catzeflis et al. 1987) as well as a high rate of substitution in immuno-
globulin genes (Li et al. 1985). The resulting calculated divergence times for the Mus
species, based on the assumption that the separation of Mus and Rattus occurred 8–
14 Myr ago (Mya) (Jacobs and Pilbeam 1980), are as follows: M. pahari (5–8.4 Mya);
M. minutoides (3.4–6.0 Mya); M. saxicola (3–5 Mya); M. cookii (1.7–3 Mya); and
Evolution of a Single-Copy Gene

AVERAGE NUMBER OF CORRECTED NUCLEOTIDE SUBSTITUTIONS

<table>
<thead>
<tr>
<th>Species</th>
<th>SUBSTITUTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. saxicola</td>
<td>63.0</td>
</tr>
<tr>
<td>M. pahari</td>
<td>61.0</td>
</tr>
<tr>
<td>M. minutoides</td>
<td>48.0</td>
</tr>
<tr>
<td>M. cookii</td>
<td>28.8</td>
</tr>
<tr>
<td>M. spretus</td>
<td>7.6</td>
</tr>
<tr>
<td>M.m. domesticus</td>
<td>8.4</td>
</tr>
<tr>
<td>R. norvegicus</td>
<td>22.1</td>
</tr>
<tr>
<td>R.I. cooktownensis</td>
<td>13.6</td>
</tr>
<tr>
<td>R. colletti</td>
<td>20.4</td>
</tr>
</tbody>
</table>

FIG. 2.—Phylogeny of murids as determined from kappa light-chain immunoglobulin sequences. All possible unrooted topologies were examined to determine, for each, the minimum number of nucleotide substitutions required to account for their descent from a common ancestor (Fitch 1971) to the sequences whose alignment is shown in fig. 1. The numbers on the branches are the minimum number of substitutions required by the tree topology shown, which was the most parsimonious of all 135,135 trees and whose total length was 301 substitutions. Nonintegral values arise from averaging over multiple equally parsimonious substitutional paths. Branch lengths were corrected for unequal numbers of branches by the method of Fitch and Bruschi (1987). Branches are placed at a level representing the weighted average number of corrected substitutions along all descendant branches (lower scale). The broken circles enclose successive nodes within which alternative branching orders do not significantly increase the number of substitutions required, and hence the branching order in these regions should be regarded as tentative. The tree was rooted between mice and rats. The scales to the right of the tree provide two time axes, depending on whether the divergence of mice and rats is taken as having occurred 8 or 14 Mya. This assumes that the average rate of substitution has been reasonably uniform over all lineages.

M. spretus (1–2 Mya). These divergence times are, with the exception of those for M. saxicola, in good agreement with those obtained from the genealogic tree. The above data are somewhat in conflict with electrophoretic-mobility studies (Bonhomme 1986), which have suggested that the various subgenera are apparently no more related to each other than to Rattus. While it is possible that the Cx locus may be a special case, it would seem more likely that the electrophoretic techniques are not sufficiently sensitive to permit discrimination in this instance.
Table 3
Frequencies of Synonymous ($k_s$) and nonsynonymous ($k_a$) Substitutions in the $C_x$ Coding Region

<table>
<thead>
<tr>
<th></th>
<th>$M$. musculus</th>
<th>$M$. spretus</th>
<th>$M$. cookii</th>
<th>$M$. saxicola</th>
<th>$M$. minutoides</th>
<th>$M$. pahari</th>
<th>Rattus norvegicus</th>
<th>R. leucopus cooktownensis</th>
<th>R. colletii</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$. m. domesticus</td>
<td>0.032</td>
<td>0.049</td>
<td>0.083</td>
<td>0.102</td>
<td>0.145</td>
<td>0.218</td>
<td>0.229</td>
<td>0.269</td>
<td></td>
</tr>
<tr>
<td>$M$. spretus</td>
<td>0.004</td>
<td>0.081</td>
<td>0.083</td>
<td>0.102</td>
<td>0.144</td>
<td>0.218</td>
<td>0.228</td>
<td>0.270</td>
<td></td>
</tr>
<tr>
<td>$M$. cookii</td>
<td>0.036</td>
<td>0.032</td>
<td>0.101</td>
<td>0.118</td>
<td>0.151</td>
<td>0.217</td>
<td>0.226</td>
<td>0.279</td>
<td></td>
</tr>
<tr>
<td>$M$. saxicola</td>
<td>0.040</td>
<td>0.044</td>
<td>0.049</td>
<td>0.082</td>
<td>0.088</td>
<td>0.207</td>
<td>0.217</td>
<td>0.256</td>
<td></td>
</tr>
<tr>
<td>$M$. minutoides</td>
<td>0.057</td>
<td>0.053</td>
<td>0.062</td>
<td>0.053</td>
<td>0.141</td>
<td>0.293</td>
<td>0.281</td>
<td>0.320</td>
<td></td>
</tr>
<tr>
<td>$M$. pahari</td>
<td>0.065</td>
<td>0.061</td>
<td>0.059</td>
<td>0.087</td>
<td></td>
<td>0.251</td>
<td>0.262</td>
<td>0.332</td>
<td></td>
</tr>
<tr>
<td>R. norvegicus</td>
<td>0.104</td>
<td>0.099</td>
<td>0.109</td>
<td>0.104</td>
<td>0.109</td>
<td>0.120</td>
<td>0.046</td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>R. l. cooktownensis</td>
<td>0.103</td>
<td>0.098</td>
<td>0.103</td>
<td>0.112</td>
<td>0.118</td>
<td>0.024</td>
<td></td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>R. colletii</td>
<td>0.080</td>
<td>0.084</td>
<td>0.095</td>
<td>0.093</td>
<td>0.116</td>
<td>0.108</td>
<td>0.058</td>
<td></td>
<td>0.045</td>
</tr>
</tbody>
</table>

**NOTE.**—The frequency of substitution at synonymous and nonsynonymous sites was calculated using the program of Li et al. (1985). Values obtained were essentially the same as those obtained with the method of Perler et al. (1980), except for comparisons between $M$. spretus and $M$. cookii, where biased values were obtained owing to substitutions in the second position of isoleucine when using the method of Perler et al. Synonymous-site frequencies are presented in the upper right triangle, and nonsynonymous-site frequencies are in the lower left triangle. The average number of synonymous sites was 65, and that of nonsynonymous sites was 253.
Selective Pressure

The frequency of substitution at both silent ($k_s$) and replacement ($k_a$) positions can further be employed to assess the importance of selection on sequence conservation. Proteins that are highly conserved—and, by implication, under strong selection—have few substitutions at replacement sites and $k_s/k_a$ is high, whereas proteins that evolve rapidly have almost as many substitutions at replacement sites as they have at silent sites and the $k_s/k_a$ approaches 1.0. From the data in table 3, $k_s/k_a$ ratios for the murine species were found to range from 1.34 to 2.59. The mean $k_s/k_a$ ratio of $C_\kappa$ from other mammalian species including human, rabbit, and rodent—species that are thought to have radiated 80 Mya—is 3.21 (Li et al. 1985). Thus, the $k_s/k_a$ has not increased dramatically during mammalian evolution, suggesting that the relative rates of substitution at synonymous and nonsynonymous sites remains similar and that the $k_s$ has not decreased owing to saturation. Calculation (Kimura 1980) of the mean substitution frequency for all pairwise comparisons in the enhancer region revealed a lower value (0.030) than that for replacement substitutions in the coding region (0.051), indicating a stronger selection on the enhancer region.

One of the more interesting comparisons possible from the $C_\kappa$ data involves intronic sequences, as this is one of the few systems in which such sequences are sufficiently similar to permit meaningful analysis. If there were no selection on non-coding regions, the frequency of substitutions per site would be expected to be at least as great as the frequency of synonymous substitutions in the coding region. This is not the case, as the mean for the intron (0.052) is considerably lower than the mean for $k_s$ (0.100) and more closely approximates that of $k_a$ (0.051). It should be cautioned that the calculation of $k_s$ weights twofold and fourfold positions differently and that no such discrimination is possible in the noncoding intron. Notwithstanding this reservation, there appears to be a conservation of the intron on the order of that of the coding region. Possible mechanisms for conservation may be (1) selective pressure due to an as yet unknown function associated with this region, (2) more efficient repair mechanisms, or (3) a decrease in the intrinsic mutation frequency due to protection of the DNA in its particular chromatin conformation. Any proposed mechanism would have to account for the fact that, among mice and rats, sequence similarity is conserved but several deletions and insertions are present.

It appears that studies such as these will prove valuable in resolving phylogenetic relationships and also provide insights into the questions of mutational frequencies and rates. Furthermore, the wild-mouse system should be an appropriate medium for testing mathematical models describing the above-mentioned parameters of molecular evolution and also should as serve as a model for short-term mammalian evolution.

Acknowledgment

We thank Dr. Michael Potter for his continued interest and support of this work.

LITERATURE CITED


MAX, E. E., J. R. MAIZEL, and P. LEDER. 1981. The nucleotide sequence of a 5.5-kilobase


WESLEY M. BROWN, reviewing editor

Received March 10, 1988; revision received May 16, 1988