Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 5321–5325, June 1991
Evolution

Involucrin gene of tarsioids and other primates: Alternatives in evolution of the segment of repeats

PHILIPPE DJIAN AND HOWARD GREEN

Department of Cellular and Molecular Physiology, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115

Contributed by Howard Green, March 18, 1991

ABSTRACT **The involucrin genes of the prosimian primates and of the anthropoid primates possess nonhomologous segments of repeats located at two different sites, P and M, within the coding region. The involucrin gene of the tarsioids alone contains repeats at both sites, for it derived repeats at site P from a common ancestor of tarsioids and prosimians and a repeat at site M from a later common ancestor of tarsioids and anthropoids. After their divergence from the tarsioids, the anthropoids added many more repeats to site M and excised the older segment of repeats from site P; in contrast, the tarsioids stopped adding repeats at site M, retained the earlier segment of repeats at site P, and enlarged it. In the revision of their involucrin genes, the two lineages followed alternative routes. The mechanisms by which the revisions took place have been subject to abrupt onset or termination.**

The genes for involucrin (an epidermal protein and substrate of transglutaminase) are quite different in the anthropoid and prosimian primates (1). In both taxa, a large part of the coding region consists of a segment of tandem repeats, but there are two kinds of segments of repeats. Prosimians (1, 2), like nonprimate mammals (3), contain a premodern segment of repeats, located not far from the 5′ end of the coding region at a site designated as P. Anthropoid primates contain a modern segment of repeats located closer to the 3′ end of the coding region (4–10) at a site designated as M; these repeats were generated by successive duplications of a 10-codon sequence and the earlier segment of repeats at site P was deleted (1).

The position of tarsiers in primate phylogeny has been controversial (refs. 11–14 and pp. 17 and 25 of ref. 15). They have been grouped with the lemurs and lorises (14), with the anthropoids (12, 16–21), or in a taxon separate from both (22). Evidence derived from the sequences of protein and DNA (23–27) has supported the classification of tarsioids with anthropoids in a haplorhine suborder.

We now report the nucleotide sequence of the involucrin gene of the tarsioids.* Its repeat structure indicates that although the tarsioids and the prosimians had a common primate ancestor, the tarsioids and the anthropoids had a more proximate common ancestor. After divergence of the tarsioids from the anthropoids, the segment of repeats at site M and the segment of repeats at site P evolved differently in the two lineages. This was the result of differently utilized mechanisms of gene revision.

MATERIALS AND METHODS

Restriction maps were prepared from the DNA of livers of three animals—two *Tarsius bancanus* and one *Tarsius syrichta*. As shown in Fig. 1, the restriction maps are quite similar, though not identical. The coding regions of clones derived from both specimens of *T. bancanus* were sequenced, either completely (clone 1) or partially (clone 2). A clone of *T. syrichta* was partially sequenced to confirm the essentially identical nature of its segment of repeats.

Liver of *T. bancanus* no. 1 was obtained from the National Zoological Park (Washington, DC), through the courtesy of Richard Montali. Livers from *T. bancanus* no. 2 and *T. syrichta* were obtained from the Duke University Primate Center (Durham, NC), through the courtesy of Frances White and Ruby Ange. Genomic DNA was prepared according to ref. 28. A restriction map was generated by using, as a probe, a mixture of the lemur and gibbon involucrin genes (1, 9). Hybridization and washing conditions were those previously described (1), except that the last washes were in 1× standard saline citrate (SSC) at 60°C. A 7-kbp *Eco*RI fragment containing the whole *T. bancanus* involucrin gene was cloned in λZAP II (Stratagene). The recombinant plasmid was rescued from the λ phage according to the supplier's recommendations. For sequencing, the involucrin gene was progressively digested with the nuclease BAL-31 (29). The overlapping DNA fragments obtained were cloned in M13 (30) and sequenced by chain termination (31), using T7 DNA polymerase, kindly provided by Stanley Tabor (32).

RESULTS AND DISCUSSION

General Features of the Coding Region. The nucleotide sequence of the involucrin coding region of *T. bancanus* is shown in Fig. 2, where it is aligned with that of the human (4) and that of the lemur (*Lemur catta*), a prosimian (1). The lemur (like three other nonanthropoids) possesses a segment of repeats at site P, while the human (like eight other anthropoids) does not, but possesses instead a segment of repeats at site M. Of all the involucrin genes known, the gene of the tarsier is the only one that contains repeats at both site P and site M.

Segment of Repeats at Site P. The segment of repeats at site P in the tarsioid gene is shown in Fig. 3, where its consensus sequence is compared with consensus sequences of the corresponding repeats of other nonanthropoid mammals. The repeats in the five species are homologous. The consensus sequence of the tarsioids diverges from that of the galago and that of the lemur at only three and two nucleotide positions, respectively, whereas it diverges from the pig and dog consensus sequences at nine and seven positions, respectively. Other points of closer similarity between the segments of repeats of the tarsioids and the prosimians are the consensus repeat length and the high frequency of site-specific deletions.

On the other hand, the segment of repeats at site P of the tarsioids clearly differs from the corresponding segment of the prosimians in the following ways:

(*i*) The site-specific deletions are uniform at two CAG codons, instead of frequent at three.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*The sequence reported in this paper has been deposited in the GenBank data base (accession no. M65124).

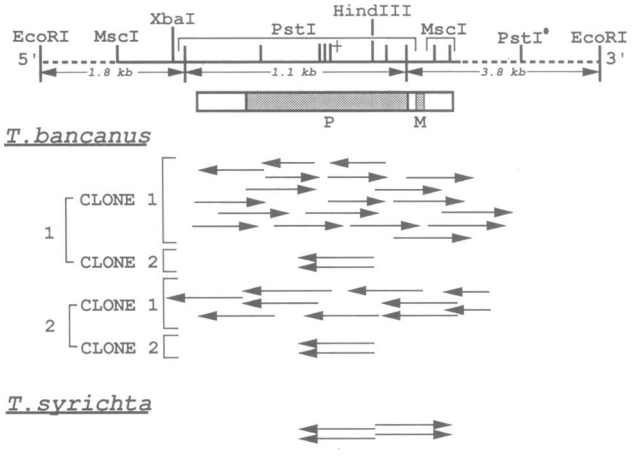

FIG. 1. Restriction map and sequencing of the tarsier involucrin gene. The coding region is shown as a box with its segments of repeats at sites P and M stippled. These two sites are unusually close together in the tarsier because of the unique deletion between them (see text). A cluster of three *Pst* I sites in site P of *T. bancanus* no. 1 is marked by a +. One of these three sites is absent from *T. bancanus* no. 2 because the repeat that contains it (corresponding to repeat 7 or 8 of the *T. bancanus* no. 1 and *T. syrichta*) has been deleted. The *Pst* I• site 3′ of the coding region is found in *T. syrichta* alone. The two broken lines indicate that the corresponding restriction fragments are not drawn to scale. Arrows represent sequenced part of involucrin subclones. kb, Kilobases.

(*ii*) Two of the repeats (7 and 8) are peculiar in that they are incomplete duplicates of an earlier repeat (number 6), and contain only 11 codons instead of the usual 14. As such repeats have not been found in any of the four nonanthropoid genes sequenced, they must have been generated in the tarsioid lineage.

Segment of Repeats at Site M. In the anthropoids, the 10-codon sequence at site M is repeated 24–63 times, depending on the species. There are two main types of repeats, A and B; these differ in their first three codons at four nucleotide positions, while the remaining seven codons are identical. Occasionally, the first three codons are deleted; such a repeat cannot be identified as A or B and has been designated as X (Fig. 4).

All four prosimians and nonprimates examined earlier possess only a single copy of the nucleotide sequence at site M: in none of the four species is there a repeat. In the tarsioids, there are two copies, one an A repeat and one an X repeat. The tarsioid A repeat possesses the four nucleotides typical of the A consensus sequence of the anthropoids and diverges from it at only 1 nucleotide out of 30. The corresponding unrepeated sequence in the prosimians and nonprimate mammals has four mismatches with the closest repeat type of anthropoids. The X repeat of the tarsioids, which is identical to the X repeat of the anthropoids, must be a duplicate of the preceding A repeat, since there are no mismatches between the X repeat and the last seven codons of the A repeat, and since no sequence of seven codons resembling an X repeat appears anywhere else in the entire coding region of the tarsioid involucrin gene (Fig. 2). There have been additions to the sequence in or near site M in other lineages, but they do not resemble the repeats of the anthropoid lineage. For example, the pig gene contains a 21-nucleotide insertion in site M, but this insertion does not appear to be a duplication of preexisting sequence (3), and the galago gene has a single 6-codon duplication 9 codons upstream of site M (2). It is therefore clear that in respect to both the nucleotide sequence at the M site and the presence of an (incomplete) repeat, the tarsioid gene shares derived features with the anthropoid gene (Figs. 4 and 5).

Coding Region Surrounding Sites P and M. The most distinctive feature of the parts of the coding region outside of sites P and M in the tarsioid gene is the deletion, shown in Fig. 2, of most of the codons lying between the two sites (the region corresponding to codons 339–394 of the lemur). Apart from the excision of the premodern segment of repeats that must have occurred in the anthropoid lineage, this is the only large deletion found so far in an involucrin gene.

The part of the tarsioid coding region located 5′ of site P can be aligned with that of anthropoids and prosimians except for a seven-codon deletion corresponding to codons 55–61 of the lemur. Similarly, the part of the coding region 3′ of site M is homologous to that of other primates. The tarsioid involucrin gene terminates with that of anthropoids, whereas the in-

```
Hum ATG TCC CAG CAA CAC ACA CTG CCA GTG ACC CTC TCC CCT GCC  (14)
Tar ATG TCC CAG CAG CAA ACA CTG CCA GTG ACC CTC CCC CCT GCC  (14)
Lem ATG TCC CAG CAA CAC ACA CTG CCA GTG ACC CTG CCC CCC ACC  (14)

Hum CTC AGT CAG GAG CTC CTC AAG ACT GTT CCT CCT CCA GTC AAT ACC CAT CAG GAG CAA ATG AAA CAG CCA ACT CCA CTG CCT CCC CCA TGC  (44)
Tar CTC AGT CAG GAA CTC CTC AAG ACG GTT CCT CCT CCG GCC AAT ACC CAG CAG GAT CAA ATG AAG CAG CCG ACT CCA TCG CCT GCC CCA TGC  (44)
Lem CTC AGT CAG GAG CTC CTC AAG AAT GTT TCT CCT CCA GCT GAC ATC CAG CAG GAG CAA AGG AAG CAG CCA ACT CCA CTG CCT GCC CCG TGC  (44)

Hum CAG AAG GTG CCT GTC GAG CTC CCA GTG GAG GTC CCA TCA AAG CAA GAG GAA AAG CAC ATG ACT GCT GTA AAG GGA CTG CCT GAG CAA GAA  (74)
Tar CAG AAG GGG CCC TCC GAA CTC CCA GTG GAG --- --- --- --- --- --- --- AAG CAC CCA GCT CCT GTA AAG CAG GTG CCC GAG CAA GAA  (67)
Lem CAG AAG GTG CTC TCT GAG CTC CCT GTA GCG GTC CCC TCA AAG CAT GAG GAG AAA CAC GCA ACT CCT GTA AAA GGG CTG CTT GAG CAA GAA  (74)

Hum TGT GAG CAA CAG CAG AAG --- GAG CCA CAG GAG CAG GAG CTG CAC --- --- --- --- --- --- --- --- --- --- CAA CAG CAC TGG GAA  (93)
Tar TGT GAA CCA CAG CAG CAG --- GAC CAC CAG GAG CAG GAA GTG CAC CTG GGA AAG CAG --- --- CAG CAG GAG CCT --- --- --- --- --- (317)
Lem TGT GGG CAG CTG CAG CAG CAG GAG CCA CAG GAG CAG GAA CTG CAC CTG GGA AAG (CAG CAG CAG) CAG CAG GAG CCA CAA CAG CAC CAG GAA (343)
                                                                P
Hum CAG CAT GAG GAA TAT CAG AAA GCA GAA AAC CCA GAG CAG CAG CTT AAG CAG GAG AAA ACA CAA AGG GAT CAG CAG CTA AAC AAA CAG CTG (123)
Tar --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Lem CAG CAT GAG GAA CGT CAG AAA GCA GAA AGC TTG GAG CAG CAA CTT GAG CAG GAG AAA GCA CAA AGG GAA GAG CAA CTG AAG GAG CAG CTG (373)

Hum GAA GAA GAG AAG AAG CTC TTA GAC CAG CAA CTG GAT CAA GAG CTA GTC AAG AGA GAT GAG CAA CTG GGA ATG AAG AAA GAG CAA CTG TTG (153)
Tar --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- CTG GGA ATG AAG GAA GAG CAG CTG TTG (326)
Lem GAA GAG AAG AAG AGG ATC TTG GAC CAG CAA CTG GAT CAA GAG GTG GCA AAG AGA TAT GAA CAA CTG GGA GTG AAG AAG GAG CAG CTG TTG (403)
  M
Hum AAG CAC CTG GAG CAG CAG GAG GGG CAG CTG GAG CAG CCT GTG TTT GCC CCA GCT CCA GGC CAG GTC CAA GAC ATT CAA CCA GCC CTG CCC (560)
Tar AAG CAC GTG GAG CAG CAG GAG GGG CAG CTG AAG CAG CCT GTA TGT ATC CCA ACA CCT GGC CAG GTC CAA GAC ATC CAG CCA GCC CAG CCC (363)
Lem CAG CCC CTG GGG CAG CAG GAG GGA CAG CTG GAG AAG CCC GTT TTT GTC CCA GCT CCT GGC CAG GTC CAA GAC ATC CAG CCA CCC CAG CCT (433)

Hum ACA AAG GGA GAA GTA TTG CTT CCT GTA GAG CAC CAG CAG CAG AAG CAG GAG GTG CAG TGG CCA CCC AAA CAT AAA TAA (585)
Tar CCA AAG GGA GAA GTC TTG CTC CCC ACA GAG AAG CAG CAG --- AAG CAG GAG GTA CAA TGG CCA CTC AAA CAA GAA TAA (387)
Lem CCA AAG GGA GAA GTC CTG CTC CCT GCA GAG CAG CAG CAA --- GAG CCA GAG GTG TAG ggg ctg ctc gaa ctt aag tac (450)
```

FIG. 2. Coding region of the involucrin gene of three primates. Sites P and M of the two segments of repeats are framed. The gene of *Lemur catta* (Lem) contains a segment of repeats only at site P, and the gene of the human (Hum) contains a segment of repeats only at site M. The gene of *T. bancanus* no. 1 (Tar) contains a segment of repeats at site P but also an incipient segment of repeats at site M. *T. bancanus* no. 2 differs at codon 50 (AAA instead of GAA). At site M in the human, the consensus sequence of the A repeats is given; the consensus of the B repeats differs at the nucleotides marked with a dot. Underlined nucleotides in the human and in the tarsier are substitutions in their respective lineages after their divergence; the lemur sequence was used as an outgroup. Numbers in parentheses are codon numbers.

Tarsius bancanus

```
 1 CAG GAG CCA GAA CTG CAG CTG GGA AGG AAG --- --- CAG CAG GAG CCA
 2 CAG GAG CAA GAA GTA CAC CCA GGA AAG CAG --- --- CAG CAG AAA CCA
 3 CAG GAA CAA GAA GCG CAT CTG GGA AAG --- --- --- AAA CAG GAG CCA
 4 CAG GGA CAG GAA GTG CAC CTG GGA AAG CAG --- --- CAA CAA AAA ACA
 5 CAG GAA CAG GAA GTG CAT CTA GGA AAG CAG --- --- CAG CAG GAG TTG
 6 CAG GAG CAG GAA GTG CAC CTG GAA AAG CAA --- --- CTG CAG GAG CCG
 7 --- --- CAG GAA GTG CAC CTG GAA AAG CAA --- --- CTG CAG GAG ---
 8 --- --- CAG GAA GTG CAC CTG GAA AAG CAA --- --- CTG CAG GAG ---
 9 CCA GAG CCG GAA TTG AAC TTG GGA AAG CAG --- --- CAG CAG GAA CCT
10 CAG GAG CAG GAA GCG TAC CTG GGA AAG CAG --- --- CAG CAG GAG CTG CCA GAA CCT
11 CAG GAC CCA GAG TTG CAC CTG GGA AAA CAG --- --- CAG CAA GAG CCT
12 CAG GAG CAG GAA GTG CAA CTG GAA AAG --- --- --- CAA CAA GAG GCT
13 CAG GAG CAG GAG TTG CAC CTG GGA AAG CAA --- --- CAG CAG GAG TCT
14 CAG GAG CAG GAA CTG CAC CTG AGA AAG CTT --- --- CAG CAG GTG CCT CAG GAG CCT
15 CAG GAC CAG GAA TTG CAC CTG GGA AAG CAA --- --- CAG CAG GAG CTG
16 CAG GAG CAG GAA GTA CAC CTG GGA AAG CAA --- --- TTG CAG GAG CCT
17 CAG GAG CAG GAA CTG CAC CTG GGA AGG CAG --- --- CAG CAG GAG CTG
18 CAG GAG GAG GAA GTG CAC
```

	Consensus	No. of repeats
Dog	CAG GAG CAG A/G AA CTG CAC▼CTG GAA CAG CAG CAG GAG CAA CAA GAG TCA	6
Pig	CAG GAG CAG GAA CTG CAT GTG GAT CAG CAG CAG CAG CAG CAA GAG TCA	13
Gal.	CAG GAG CAG G A/G A CTG CAC CTG GG A/G AAA(3Δ)CAG CAG GAG TCT	13
Lem.	CAG GAG CAG GAA CTG CAC CTG GGA AAG(3Δ)CAG CA A/G G GAG CCA	19
Tar.	CAG GAG CAG GAA GTG CAC CTG GGA AAG CAG(2Δ)CAG CAG GAG CCT	17-18
	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	

FIG. 3. Segments of repeats at site P. The nucleotide sequence of the segment of repeats of the *T. bancanus* no. 1 is given above, and the consensus sequences of the repeats of five species are given below. Overlined nucleotides in the consensus sequences of dog and pig (3), galago (*Galago crassicaudatus*) (2), and lemur are those divergent from the corresponding nucleotides of the tarsier. In the dog, the repeat length is 20 codons, the arrowhead indicating the site of an additional 4 codons. In the pig, the repeat length is 16 codons. In the galago and the lemur, 3 codons are usually deleted, making the consensus repeat length 13 codons, whereas in the tarsier, 2 codons are uniformly deleted, making the consensus repeat length 14 codons. The deletions in the tarsier have been arbitrarily placed in positions 11 and 12, but they could have been placed at positions 10 and 11, since glutamine codons are present at all three positions. An 11-codon stretch in repeat 6 (framed with broken lines) has been duplicated twice, thus generating repeats 7 and 8 (framed with solid lines). These repeats are present in both *T. bancanus* no. 1 and in *T. syrichta*. *T. bancanus* no. 2 possesses only one of the two duplicates (a total of 17 repeats instead of 18), suggesting that one of the two duplicates was deleted in its lineage. *T. bancanus* no. 2 also differs from *T. bancanus* no. 1 by a nucleotide substitution in codon 7 of repeat 5 (GTA instead of CTA), a deletion of the last codon of repeat 6, and an addition of a CCA codon in position 16 of repeat 7. Two of the tarsier repeats (10 and 14) possess 3 extra codons at their 3′ end. These codons resemble the last 3 codons of the repeats at codon positions 14–16. They could have been generated either by duplication of the codons occupying positions 14–16 of an adjacent repeat or by duplication of an entire repeat followed by deletion of the 5′ part, leaving only the 3 codons at its 3′ end.

volucrin gene of prosimians terminates seven codons earlier; early termination is likely to be a shared derived feature of the prosimian branch, since the involucrin gene of a nonprimate mammal, the dog, terminates with that of the anthropoids (3).

Evolutionary Stages in the Divergence of the Tarsioid Involucrin Gene from the Genes of Prosimians and Anthropoids. The distinctive property of the involucrin gene that gives it analytic value is the segment of repeats. That is because the segment of repeats has undergone more radical evolutionary

	No. of copies	Repeat type	Repeat sequence	No. of mismatches with closest consensus sequence of anthropoids
Anthropoids	25-64	B	consensus: GAG CTC CCA GAG CAG CAG GAG GGG CAG CTG	-
		X	consensus: --- --- --- GAG CAG CAG GAG GGG CAG CTG	-
		A	consensus: AAG CAC CTG GAG CAG CAG GAG GGG CAG CTG	-
Tarsioids	2	A	AAG CAC [G/T]TG GAG CAG CAG GAG GGG CAG CTG	1
		X	--- --- --- GAG CAG CAG GAG GGG CAG CTG	0
Galago	1	B	GAG [C]C[C] CCA [G]G[G] CAG CAG [A]AG GG[A] CAG CTG	4
Lemur	1	A	[C]AG [C]C[C] CTG [G]G[G] CAG CAG GAG GG[A] CAG CTG	4
Pig	1	(A)	[G]AG CA[G] [C]A[G] GAG▲cag cag gag ggg cag [t]tg	4
Dog	1	(A)	[G]AG CA[G] [C]A[G] G[G]G CAG CAG GAG GGG CAG CTG	4

FIG. 4. Segment of repeats at site M. Boxes enclose nucleotides divergent from those of the most similar repeat type of anthropoids. In the second codon, the second nucleotide, A, and the third nucleotide, C, have been found in nonprimates or prosimians, but the combination AC has been found only in the A repeat of tarsier and anthropoids. Both *T. bancanus* no. 2 and *T. syrichta* differ from *T. bancanus* no. 1 in the first nucleotide of the third codon of the A repeat. Arrowhead indicates an insertion of 21 nucleotides in the pig sequence.

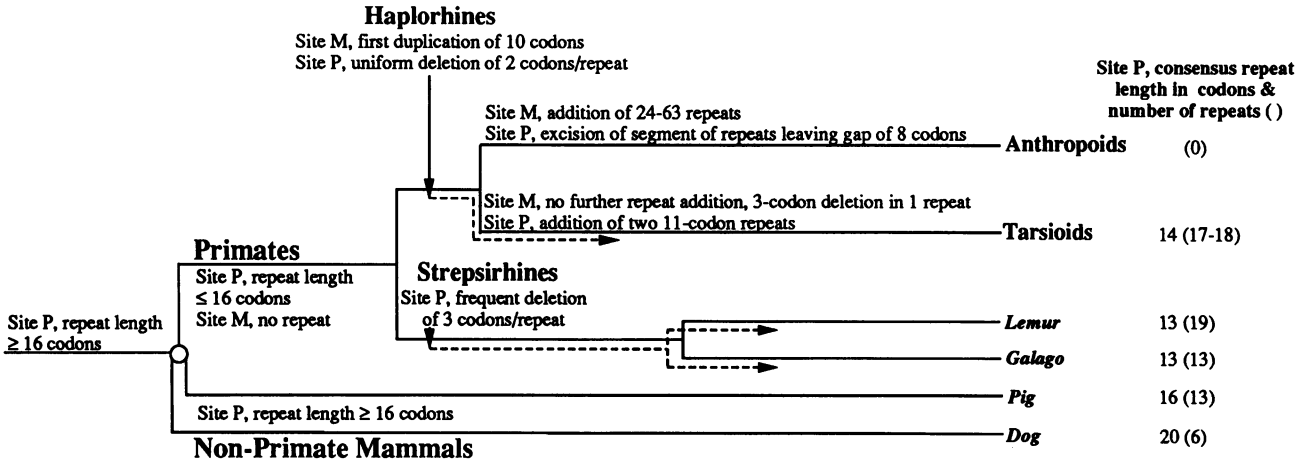

FIG. 5. Evolution of the segments of repeats in the involucrin gene of the tarsioids. The evolutionary tree is based on the repeats at sites P and M. The tarsioids are the only taxon possessing repeats at both sites. The segment of repeats at site P of the tarsioids is, in part, a retention of early features of the primates, but it has been modified. Broken lines indicate that the site-specific deletions likely occurred before and after a lineage divergence. The repeat at site M in the tarsioids reveals their common ancestry with the anthropoids. Divergence from the anthropoids was followed by arrest of further repeat addition at site M.

change than the rest of the coding region. The segment of repeats of the anthropoids is qualitatively different from that of the prosimians, and this makes it possible to distinguish the repeats acquired from a common ancestor of the prosimians and the tarsioids from the repeats acquired from a common ancestor of the anthropoids and the tarsioids. On the basis of these repeats, the following evolution can be postulated (Fig. 5):

(*i*) After its divergence from the nonprimate mammals, the common primate lineage retained a segment of repeats at site P. This segment of repeats then diverged from the homologous segment of the nonprimates.

(*ii*) Site-specific deletions at site P may have begun in the common primate lineage. But a mechanism for site-specific deletion must have been transmitted to both strepsirhine and haplorhine lineages, because there is an evident, if small, difference between the deletions in the two lineages, indicating that the control of the site specificity of the deletions was slightly different. The resulting effects on the repeat length are still visible in the prosimians and the tarsioids, but not in the anthropoids, because they no longer possess a segment of repeats at site P.

(*iii*) In the common haplorhine lineage, a number of nucleotide substitutions occurred in the 10-codon sequence at site M, making it the prototype sequence of the A repeat; this sequence was duplicated once.

(*iv*) The anthropoid lineage diverged from the tarsioid lineage and continued to add repeats at site M in number up to 64. Early in this process, the alternative repeat type B was generated by nucleotide substitutions in an A repeat.

(*v*) In the tarsioid lineage, the first 3 codons were deleted from one of the two repeats at site M. Such a deletion has also occurred occasionally in the anthropoid lineage, as both cercopithecoids and hominoids have one to three X repeats (ref. 9 and unpublished data). Since there is no X repeat in the early region of the anthropoid gene, we think it is more likely that the tarsier X repeat was generated by a subsequent 3-codon deletion in a full 10-codon duplication than by simple duplication of the last seven codons.

(*vi*) The tarsioids added no further repeats at site M and instead preserved their segment of repeats at site P. The preferred site for cross-linking of human involucrin *in vitro* by transglutaminase is known to be a glutamine of repeat 5, a B repeat in the early region (33). It is possible that the failure of the tarsioid lineage to generate a B repeat or an adequate number of repeats at site M made necessary preservation of the segment of repeats at site P. The tarsioid lineage added more repeats at site P, to a total of 18. Two of these repeats are peculiar because, in contrast to the others, they have only 11 codons instead of the usual 14. These are the most recent repeats generated in the segment, since there is no nucleotide divergence between them and repeat 6, from which they were duplicated. No other species possesses such repeats, which might be described as the expression in the tarsioid lineage of the evolutionary trend toward a shorter repeat length, in the face of inability to add more short repeats at site M.

(*vii*) After their divergence from the tarsioids, the anthropoids deleted all repeats at site P. Since the last repeat at site P was incomplete because it lacked the last 8 codons, deletion of an integral number of repeats should leave an 8-codon gap *vis a vis* the tarsier sequence. Such a gap is actually present in all anthropoids, beginning at the position corresponding to codon 88 of the human (Fig. 2). At some point, the tarsioids deleted a sequence of 55 codons immediately downstream of site P. This deletion may be analogous to the deletion in anthropoids of the entire segment of repeats at site P.

Discontinuous Evolution of the Involucrin Gene. Since the discovery of amino acid substitutions in evolution (34, 35), much attention has been given to the question of whether molecular evolution is continuous and clocklike (36–40). Although the rate of nucleotide substitution is known to have undergone changes in many lineages, including the primates (25, 41), the extreme differences in rate are less than 10-fold (39, 42), and it is clear that the process of nucleotide substitution, if not perfectly clocklike, is at least continuous.

The evolution of new morphological features does not seem to be part of a continuous process (43). For example, in contrast to the anthropoids, the extant tarsiers have morphological features that resemble those of fossils of the early Eocene (50 million years ago); since then, the evolution of those features seems to have been arrested and the tarsioids have been thought to be in a period of morphological stasis (44). The tarsioid involucrin gene shows interrupted evolution at site M, where addition of repeats was arrested at an early stage; since that time, the divergent anthropoids added many repeats.

Tarsioids are known to have retained some primitive morphological features resembling those of the prosimians (11, 12, 44). Retention of a primitive trait can be seen in the DNA at site P of the tarsioid involucrin gene, whose segment of repeats still resembles that of the prosimians. However, this segment of repeats has been differently altered in the

prosimian and tarsioid lineages by site-specific deletion and gene conversion (2). These mechanisms were terminated in the anthropoid lineage when it completely eliminated the homologous segment of repeats at site P.

In contrast to the addition and deletion of DNA, the process of nucleotide substitution in the involucrin gene continued in anthropoid and tarsioid lineages since their divergence, as was shown earlier for the globin gene (26) and for single-copy DNA (27). The involucrin coding region surrounding the segment of repeats has, in the tarsioid lineage, accumulated nucleotide substitutions (36/384 or 9.4%) at a rate not lower than in the anthropoid lineage (26/384 or 6.8% in the human) (see legend to Fig. 2). The continuous process of nucleotide substitution that has taken place in the involucrin genes of both lineages does not resemble the processes that have contributed most to the different evolution of the segments of repeats. This evolution required mechanisms of gene revision that were specifically directed to the segment of repeats and whose activity began or terminated abruptly.

In the tarsioid lineage, the arrest of repeat addition at site M and the failure to eliminate the repeats at site P are probably related, since the anthropoid lineage did the opposite in both respects. All involucrin genes studied to date possess a segment of repeats, suggesting that it is important for the function of involucrin, even if its properties can vary considerably. The trend from the nonprimates to the ancestral anthropoids was one of progressive shortening of the repeat length (3). The prosimian and tarsioid lineages did this by site-specific deletion. The haplorhine lineage began the process of short repeat addition at site M and this was continued by the anthropoids. Although the divergent tarsioids did not continue the process, they did, to a degree, follow an analogous trend by adding two short repeats at site P. This seems to be a third form of gene revision leading to reduction of repeat length.

This investigation was aided by a grant from the National Cancer Institute.

1. Tseng, H. & Green, H. (1988) *Cell* **54**, 491–496.
2. Phillips, M., Djian, P. & Green, H. (1990) *J. Biol. Chem.* **265**, 7804–7807.
3. Tseng, H. & Green, H. (1990) *Mol. Biol. Evol.* **7**, 293–302.
4. Eckert, R. L. & Green, H. (1986) *Cell* **46**, 583–589.
5. Teumer, J. & Green, H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 1283–1286.
6. Tseng, H. & Green, H. (1989) *Mol. Biol. Evol.* **6**, 460–468.
7. Djian, P. & Green, H. (1989) *Mol. Biol. Evol.* **6**, 469–477.
8. Djian, P. & Green, H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8447–8451.
9. Djian, P. & Green, H. (1990) *Mol. Biol. Evol.* **7**, 220–227.
10. Phillips, M., Rice, R. H., Djian, P. & Green, H. (1991) *Mol. Biol. Evol.*, in press.
11. Simons, E. L. (1972) *Primate Evolution: An Introduction to Man's Place in Nature* (Macmillan, New York).
12. Martin, R. D. (1990) *Primate Origins and Evolution: A Phylogenetic Reconstruction* (Princeton Univ. Press, Princeton, NJ).
13. Fleagle, J. G. (1988) *Primate Adaptation and Evolution* (Academic, San Diego, CA).
14. Simpson, G. G. (1945) *Bull. Am. Mus. Nat. Hist.* **85**, 1–350.
15. Napier, J. R. & Napier, P. H. (1985) *The Natural History of the Primates* (MIT Press, Cambridge, MA).
16. Hubrecht, A. A. W. (1908) *Q. J. Microsc. Sci.* **53**, 1–181.
17. Pocock, R. I. (1918) *Proc. Zool. Soc. London*, 19–53.
18. Hill, W. C. O. (1953) *Proc. Zool. Soc. London* **123**, 655–692.
19. Cartmill, M. & Kay, R. F. (1978) in *Recent Advances in Primatology: Evolution*, eds. Chivers, D. J. & Joysey, J. A. (Academic, London), Vol. 3, pp. 205–214.
20. Rosenberger, A. L. & Szalay, F. S. (1980) in *Evolutionary Biology of the New World Monkeys and Continental Drift*, eds. Ciochon, R. L. & Chiarelli, A. B. (Plenum, New York), pp. 139–157.
21. Aiello, L. C. (1986) in *Major Topics in Primate and Human Evolution*, eds. Wood, B., Martin, L. & Andrews, P. (Cambridge Univ. Press, New York), pp. 47–65.
22. Gingerich, P. D. (1981) *J. Hum. Evol.* **10**, 345–374.
23. de Jong, W. W. & Goodman, M. (1988) *J. Hum. Evol.* **17**, 575–582.
24. Koop, B. F., Siemieniak, D., Slightom, J. L., Goodman, M., Dunbar, J., Wright, P. C. & Simons, E. L. (1989) *J. Biol. Chem.* **264**, 68–79.
25. Koop, B. F., Tagle, D. A., Goodman, M. & Slightom, J. L. (1989) *Mol. Biol. Evol.* **6**, 580–612.
26. Beard, J. M. & Goodman, M. (1976) in *Molecular Anthropology, Genes and Proteins in the Evolutionary Ascent of the Primates*, eds. Goodman, M., Tashian, R. E. & Tashian, J. H. (Plenum, New York), pp. 239–255.
27. Bonner, T. I., Heinemann, R. & Todaro, G. J. (1980) *Nature (London)* **286**, 420–423.
28. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), p. 281.
29. Poncz, M., Soloweijczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4298–4302.
30. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
31. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
32. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
33. Simon, M. & Green, H. (1988) *J. Biol. Chem.* **263**, 18093–18098.
34. Zuckerkandl, E. & Pauling, L. (1962) in *Horizons in Biochemistry*, eds. Kasha, M. & Pullman, B. (Academic, New York), pp. 189–225.
35. Margoliash, E. (1963) *Proc. Natl. Acad. Sci. USA* **50**, 672–679.
36. Sarich, V. M. & Wilson, A. C. (1967) *Proc. Natl. Acad. Sci. USA* **58**, 142–148.
37. King, J. L. & Jukes, T. H. (1969) *Science* **164**, 788–798.
38. Kimura, M. & Ohta, T. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2848–2852.
39. Fitch, W. M. & Langley, C. H. (1976) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **35**, 2092–2097.
40. Wilson, A. C., Carlson, S. S. & White, T. J. (1977) *Annu. Rev. Biochem.* **46**, 573–639.
41. Bailey, W. J., Fitch, D. H. A., Tagle, D. A., Czelusniak, J., Slightom, J. L. & Goodman, M. (1990) *Mol. Biol. Evol.* **8**, 155–184.
42. Lake, J. A. (1991) *Trends Biochem. Sci.* **16**, 46–50.
43. Gould, S. J. & Eldredge, N. (1977) *Paleobiology* **3**, 115–151.
44. Le Gros Clark, W. E. (1960) *The Antecedents of Man: An Introduction to the Evolution of the Primates* (Quadrangle, Chicago), pp. 38–74.