Origin of Rat β-Globin Haplotypes Containing Three and Five Genes[1]

Tatjana Paunesku, Milena Stevanović, Danica Radosavljević, Radoje Drmanac, and Radomir Crkvenjakov

Institute for Molecular Genetics and Genetic Engineering, Belgrade

We have reported in rat three adult β-gene haplotypes containing either five or three genes. Detailed sequence analysis reveals that the leftmost gene is the major gene and that at the opposite end downstream lies the minor gene. All of the genes lying between them are minor-major hybrids indicating their origin by unequal crossing-over. In two haplotypes β-globin genes were found with an L1[1] element inserted directly into IVS2. The described results allow the formulation of a pathway of mutational events leading from the ancient two-β-gene rodent ancestor through a three-gene haplotype to five-gene haplotypes, one of which is postulated to have arisen in common laboratory strains since their capture in the wild.

Introduction

The superfamily of globin genes was formed by a series of gene duplication events and by the fixation of sequence changes (Czelusniak et al. 1982). The numerous haplotypes have been described consisting of restriction-fragment-length polymorphisms (Antonarakis et al. 1985). The globin haplotypes that differ in the number of genes have been reported (Rando et al. 1986; Cheng and Hardison 1988; Garner and Lingrel 1988). The structural and functional constraints during evolution presumably limit the plasticity of family organization. Thus the presence of one or at most two active adult β genes might be considered as one such constraint, since it appears to hold in the vertebrate species studied so far (Martin et al. 1983; Wood and Weatherall 1983; Harris et al. 1984; Townes et al. 1984). However, we have described in rat adult β-gene haplotypes containing either five or three genes each (Stevanović et al. 1989). The sequence data indicate that rats have one major, one minor, and different numbers (one to three) of hybrid genes located between the first two. Rat genes are also unusual in that two genes in haplotype b and one in haplotype c have a 6.5-kb long IVS2. The present paper examines different structural changes in three β-globin gene haplotypes in rat, as well as the pathway which led to their formation.

Material and Methods

Restriction Mapping

DNA was isolated from kidneys and spleens, digested, electrophoresed, blotted, and hybridized according to methods described by Herrman et al. (1986). Probes (Stevanović et al. 1989) were labeled according to the procedure described by Feinberg

1. Key words: globin genes, rat haplotypes, numerous β-globin genes, minor-major hybrid genes. Abbreviations: IVS1 = intervening sequence one; IVS2 = intervening sequence two; L1 = long interspersed sequence one; L1Rn = long interspersed sequence one in *Rattus norvegicus*.

Address for correspondence and reprints: Dr. Radomir Crkvenjakov, Institute for Molecular Genetics and Genetic Engineering, P.O. Box 794, 283 Vojvode Stepe, 11000 Belgrade, Yugoslavia.

Mol. Biol. Evol. 7(5):407–422. 1990.

0737-4038/90/0705-0002$02.00

and Vogelstein (1983). The maps were obtained in part by genomic Southern analysis and in part from genomic clones isolated from a λ-vector library (Stevanović et al. 1989). Haplotype a was found in inbred strain DA, and haplotype b was found in strains AO and Y5 and in an outbred Wistar strain. Haplotype c was found coexisting within the same Wistar strain with haplotype b (Stevanović et al. 1989).

Sequencing of Rat β-Globin Genes

Propagation of recombinant bacteriophage from the rat genomic DNA library and CsCl isolation of phage DNA were performed according to methods described by Davis et al. (1980, pp. 74–81). Fragments for subcloning were obtained by digestion of the recombinant phage DNA by the appropriate restriction enzymes and subsequent electroelution. These fragments were ligated to M13 or pUC18 vectors. Sequencing was done according to a modification of the dideoxy chain termination method of Hattori and Sakaki (1986). The partial sequence of the 3′ region of the *x* gene was obtained by sequencing the recombinant phage DNA, according to the method of Mantioletti and Schneider (1988). The necessary primers were obtained from Genetic Designs, Inc.; all chemicals and enzymes used for sequencing were from United States Biochemical, except for α-^{35}S dATP, which was from Amersham.

Sequence Comparisons

The published nucleotide sequences were compared with those obtained in the present work. The sequences were aligned by using the ALN3 (BIONET version 1.0) triple-alignment program, based on the algorithm described by Gotoh (1986). The triple alignments were then combined and improved by eye.

In the estimation of similarity, each nucleotide difference was counted as one, and indels (insertions or deletions) of 1 and 2 bp were counted as one and two points, respectively. Larger indels were assigned three points. The largest indel was 121 bp long. Similarity scores were defined as 100 minus the number of points per hundred nucleotides compared.

Results

Adult β-Globin Haplotypes and Sequence Comparisons

As a part of the study of the β-like globin family of rat, we have investigated the structures of the three adult β-globin haplotypes (Stevanović et al. 1989). Their restriction maps are shown in figure 1. The number of the adult β genes far exceeds that found in other mammals. Of the 13 genes found in the haplotypes (table 1), we have cloned six and a half from haplotypes b and c. To prove the identity of genes and to study their evolutionary origin we have sequenced either entire cloned genes or portions of them (Radosavljević and Crkvenjakov 1989; Stevanović and Crkvenjakov 1989; present paper). The sequences are shown in figure 2. All genes contain the sequences common to all functional globin genes (Myers et al. 1986; for review, see Efstratiadis et al. 1980). The sequences which are believed to be important for initiation, termination, and processing of transcript are found intact in all examined rat β-globin genes. These sequences include the CACCC box, CCAAT box, ATA box, the presumed ribosome binding site (CTTCTG), and the polyadenylation signal (AATAAA). The genes *y* (Stevanović and Crkvenjakov 1989) and *u* have a modified polyadenylation signal (AATTAA).

The sequences of all genes are very similar. To quantify the comparison, we have calculated similarity scores for each pair of genes and have represented, separately,

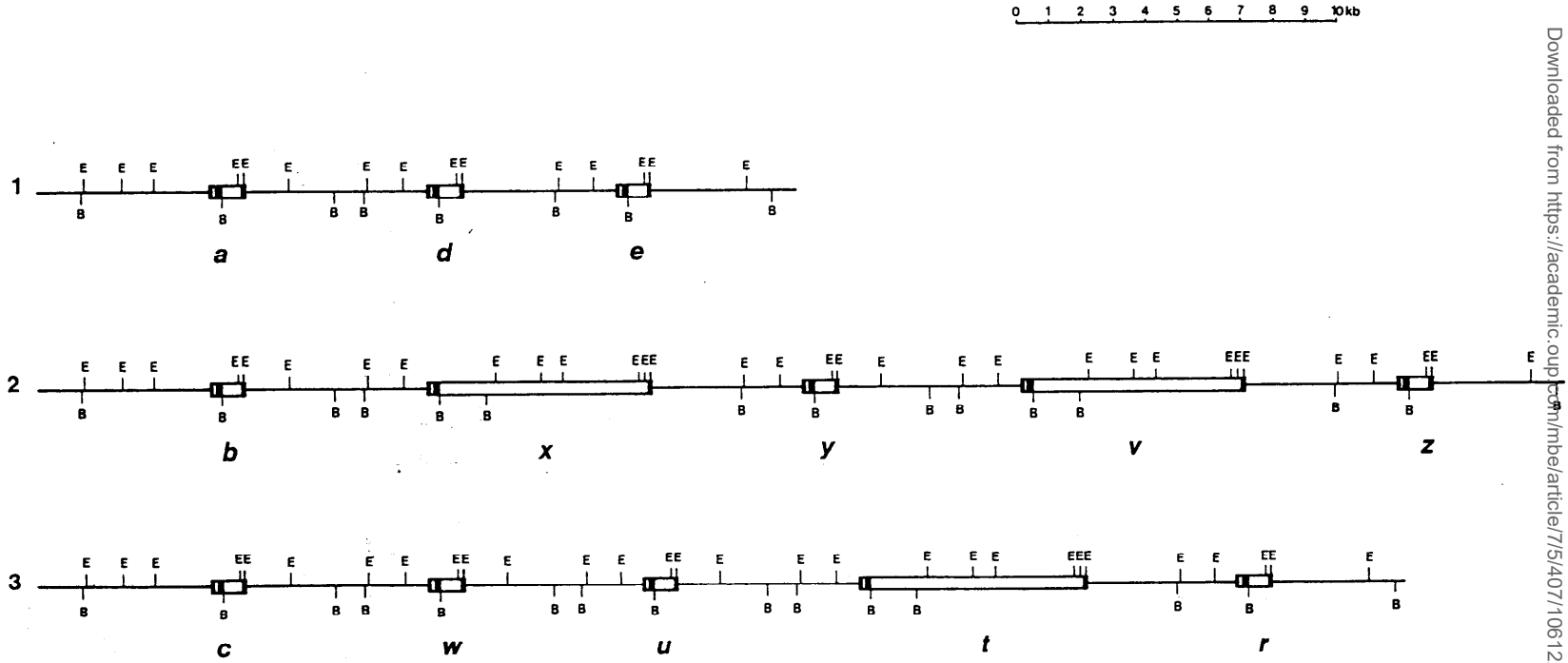

FIG. 1.—Restriction maps of three haplotypes of the β-globin gene family present in different rat strains: 1, Haplotype a. 2, Haplotype b. 3, Haplotype c. The positions of the genes are shown by the boxes, with the filled parts denoting exons. Restriction sites for *Eco*RI (E) and for *Bam*HI (B) endonucleases are indicated. The single-letter representations of genes are as given in table 1.

Table 1
Rat β-Globin Genes Studied

Strain, Haplotype	Genes[a]				
DA, a	$\beta^{maj}(a)$		$\beta^{minI}(d)$		$\beta^{minII}(e)$
Wistar (AO and Y5)[b], b	$\beta^{maj}(b, 100)$	$\beta^{minX'}(x, 54)$	$\beta^{minY}(y, 100)$	$\beta^{minX''}(v)$	$\beta^{minZ}(z, 84)$
Wistar, c	$\beta^{maj}(c)$	$\beta^{minW'}(w, 37)$	$\beta^{minW''}(u, 54)$	$\beta^{minX'}(t, 35)$	$\beta^{minZ}(r)$

[a] Italicized letters in parentheses are single-letter representations used throughout the remainder of the present paper. As an aid in remembering, *a*, *b*, and *c* are the major genes of haplotypes a, b, and c; minor genes are represented by the appropriate genotype letter where possible. The number that follows the italicized letter is the percentage, excluding the insertion, of the gene sequenced. The percentages for the *x*, *z*, *w*, *u*, and *t* genes were estimated in regard to gene *y*. All sequenced genes are from our Wistar strain, which carries both b and c haplotypes. Genes *w* and *u* are identical for 438 bp of sequence where they can be compared. Genes *z* and *r* are identical to the extent that sites for six restriction enzymes are preserved between them (Stevanović et al. 1989). Except for 13 nucleotide substitutions and eight indels, gene *z* is identical to β^{ad} (Wong et al. 1988; called *s* in the present paper) in all positions of the 1,469 bp where they can be compared. Genes *x* and *t* are identical for 546 bp of sequence where they can be compared. Gene *v* is identical to both *x* and *t* to the extent that all sites for six restriction enzymes are preserved between them (Stevanović et al. 1989).

[b] AO and Y5 inbred strains are shown in parentheses because their genes are identical to the ones from Wistar to the extent that all sites for six restriction enzymes are preserved between them (Stevanović et al. 1989).

the results for coding and noncoding portions of genes (table 2). This comparison led us to the following conclusions: genes *x* and *t* from haplotypes b and c are identical, and so are the genes *w* and *u* from haplotype c. The *y* gene is almost identical to the *w* and *u* genes, the score being 100 and 99.7 for coding and noncoding parts, respectively.

Hybrid Adult β-Globin Genes

In haplotype b the leftmost gene (*b*) is of major type, and at the opposite end downstream is the minor-type gene (*z*), according to similarity scores relative to the appropriate mouse genes. For instance, this score for mouse β-major and *b* genes is >91 in coding and IVS1 regions and >80 in the remaining noncoding parts (alignment not shown).

On the basis of the sequence-score comparisons we conclude that rats have three types of genes: major, minor, and the hybrid. For all internal genes (*x*, *y*, *w*, *u*, and *t*) the 5′ part is minor, while the 3′ untranslated and flanking regions are of the major type, with a breakpoint occurring somewhere before the 3′ untranslated regions. This is made obvious by the similarity-score comparisons of the defined smaller intragenic regions of hybrid genes *x* and *y* (*w* and *u* being nearly identical to *y*) with the major and minor genes (table 3). The shift of maximal score from minor to major sequence approximately localizes the recombination breakpoint. A more precise localization was not possible because of the scarcity of nucleotide differences between rat major and minor genes in this region for the *y* gene and because of lack of sequence in IVS2 for the *x* gene. This comparison indicates that gene *x* was formed as a result of the recombination between the major and minor gene, with the breakpoint in the IVS2, while the *y* gene has the recombination breakpoint in the third exon.

L1 Elements

The rat appears to be exceptional among mammals in possessing genes with an abnormally large IVS2 of 6.5 kb. The giant IVS2 is present in the two genes, genes *x* and *v*, in haplotype b and in the single one, gene *t*, in the c haplotype. The detailed

restriction map of a giant IVS2 has been obtained for cloned *x* gene. The map is shown in figure 3 along with the map of rat L1 element, L1Rn3 cloned by D'Ambrozio et al. (1986). There is good correspondence between the two maps, indicating the insertion of the 5.8-kb-long L1Rn in IVS2 of the *x* gene. To confirm this, the 1.3-kb *Eco*RI fragment from IVS2 of *x* was used as a probe in a Southern blot of total genomic DNA. It gave a strong repetitive signal with 1.3-kb *Eco*RI and 5-kb *Bam*HI fragments from both rat and mouse genomic DNA. Also, the 2.4-kb *Eco*RI fragment from gene *x* used as a probe in a Southern blot of the rat total genomic DNA digested with *Eco*RI and *Bam*HI gave strong signals with the repetitive 2.4-kb *Eco*RI and 5-kb *Bam*HI fragments (data not shown). All those fragments have the sizes characteristic of L1Rn elements (D'Ambrozio et al. 1986). Also, partial sequences of 329 bp from the 0.45-kb *Hin*dIII-*Bam*HI fragment, as well as the *Bam*HI 1.35-kb fragment from the *x* gene, show >95% identity with the published (D'Ambrozio et al. 1986) sequence of the 3′ region of L1Rn3 (data not shown).

Intergenic Deletions

To examine the origin of the three rat β-globin haplotypes, we analyzed in detail the pattern of *Eco*RI and *Bam*HI restriction sites (fig. 1). Those restriction sites reveal the repetitive patterns within and among the haplotypes. The sizes of the restriction fragments possessing the gene sequences are shown in table 4. The prevalence of identical size fragments—e.g., 0.18-kb and 1.4-kb *Eco*RI or the 3.5-kb *Bam*HI fragment, containing the third exon—indicates their relatively recent origin from a common ancestor gene. All the differences in *Eco*RI and *Bam*HI restriction sites among genes have been traced to the two events. One is the insertion of the L1 repetitive element into the IVS2 in genes *x*, *v*, and *t*, and the other is the deletion of 0.9 kb in the 3′ flanking region of the genes *d*, *x*, *v*, and *t*. This deletion shortens the intergenic region. The maps of the two types of intergenic spaces are shown in figure 4. The space between the end of exon 3 of one gene and the start of exon 1 of the next is contained in the following *Eco*RI fragments in the standard case: 1.4, 2.4, and 1.1 kb. In the shorter version there are 2.9-kb and 1.1-kb fragments. On the basis of the evidence of *Eco*RI fragments, we inferred that the 2.9-kb fragment arose from the 1.4-kb and 2.4-kb fragments by the deletion of 0.9 kb containing the *Eco*RI site. However, besides the *Eco*RI site, a *Bam*HI site 1.1 kb downstream from it is lost as well. Since the distance between them is longer than the overall deletion, either (*a*) two or more small deletions took place or (*b*) one of the sites was lost because of a nucleotide substitution. The latter cause obviously led either to the presence of a *Hin*dIII site close to the end exon 3 of the *x*, *v*, and *t* genes or to its absence after the *d* gene.

Apart from these changes, one is led to the conclusion of the essential monotony of the structure of the three haplotypes. The only unique *Eco*RI and *Bam*HI fragments are 5′ from the exon 3 of the major gene (*a*, *b*, and *c*) and 3′ from the exon 3 of the minor gene (*e*, *z*, and *t*), while internal fragments are duplicated and triplicated.

Discussion

Major, Minor, and Hybrid Adult β-Globin Genes

To prove the identity of cloned genes, we have sequenced portions of a majority of genes from haplotypes b and c. Structurally, they all appear capable of activity. The modified polyadenylation site AATTAA from genes *y* and *u* must be functional, since a cDNA from the rat ACI strain (Ohshita and Hozumi 1987) carries it as well. We assume that major genes are active, since the *b* gene shows in the conceptual translation

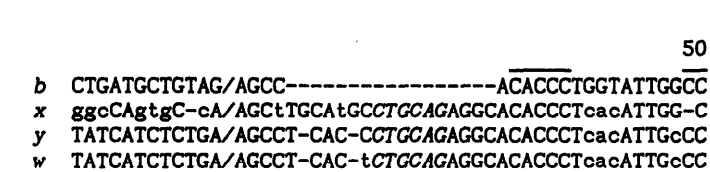

FIG. 2

```
                                                                  500
b  GCCTCTGCTATCATGGGTAACCCTAAGGTGAAGGCCCATGGCAAGAAGGT
x  ctCTCTGCTATCATGGGTAACCCccAGGTGAAGGCCCATGGCAAaAAGGT
y  GCCTCTGCTATCATGGGTAACCCccAGGTGAAGGCCCATGGCAAGAAGGT
w  GCCTCTGCTATCATGGGTAACCCccAGGTGAAGGCCCATGGCAAGAAGGT
s  GCCTCTGCTATtATGGGTAACCCccAGGTGAAGGCCCATGGCAAGAAGGT

                                                                  550
b  GATAAACGCCTTCAATGATGGCCTGAAACACTTGGACAACCTCAAGGGCA
x  GATAAACGCCTTCAATGATGGCCTGAAACACTTGGACAACCTCAAGGGCA
y  GATAAAtGCCTTCAATGATGGCCTGAAACACTTGGACAACCTCAAGGGCA
w  GATAAAtGCCTTCAATGATGGCCTGAAACACTTGGACAACCTCAAGGGCA
s  GATAAAtGCCTTCAATGATGGCCTGAAACACTTGGACAACCTCAAGGGCA

                                                                  600
b  CCTTTGCTCATCTGAGTGAACTCCACTGTGACAAGCTGCATGTGGATCGT
x  CCTTTGgcagcCTGAGTGAACTCCACTGTGACAAGCTGCATGTGGATCG
y  CCTTTGCTCATCTGAGTGAACTCCACTGTGACAAGCTGCATGTGGATCGT
w  CCTTTGCTCATCTGAGTGAACTCCACTGTGACAAGCTGCATGTGGATCG
s  CCTTTGCTCATCTGAGTGAACTCCACTGTGACAAGCTGCATGTGGATCGT

                                                                  650
b  GAGAACTTCAGGGTGAGTCTGATGGGCTCCTCACTGGGTGTCCTTCCTGT
x
y  GAGAACTTCAGGGTGAGTCTaATGGGCTCCcCACTGGGTGTCCTTCCTGT
w
s  GAGAACTTCAGGGTGAGTCTaATGGGCTCCcCACTGGGTGTCCTTCCTGT

                                                                  700
b  GGCTTTCCTGCTCAAATTCCTATCAGAAGGAAAGAGGAAGCCACTCTAGG
x
y  GGCTTTCCTGCTCAAATTCCTATCAGAAGGAAAGAGGAAGCaAtTCTAGG
w
s  GGCTTTCCTGCTCAAATTCCTATCAGAAGGAAAGAGGAAGCaAtTCTAGG

                                                                  750
b  GAGCAGTTTTGATAATGATGTGTGGATGTGCCCTGTGGAGTGTTGACAGG
x
y  GAGCAGTTTTGATgATGATGTGTGGATaTGCCCTGTGGAGTGTTGACAGG
w
s  GAGCAGTTTTGATgATGATGTGTGGATaTGCCCTGTGGAtTGTTGACAGG

                                                                  800
b  AGTCCAGTTATTTTATCCTCTATTCACAATCACTTCTCCCTCTCACTCTA
x
y  AGTCCAGTTATTTTATCCTCTATTCACAATCACTTCTCCCTCTCACTCTg
w
s  AGTCCAGTTATTTTATCCTCTATTCACAATCACTTCTCCCTCTCACTCTg

                                                                  850
b  TTATTCTATGTTGTCATTTCCTCTTTCTTTGGTAAACTTTTAATTTT-CA
x
y  TTcTTCTATGTTGTCATTTCCTCTTTCTTTGGTAAACTTTTAATTTTTCt
w
s  TTcTTCTATGTTGTCATTTtCTCTTTCTTTGGTAAACTTTTAATTTTTCt

                                                                  900
b  GATGCAGTTTTTTCTTTTTAAAAAAATTAATTAACTGACTTATTTACTTT
x
y  GtTGCAGgTTTa----------------------------------------
w
s  GtTGCAGgTTTa----------------------------------------
```

FIG. 2 (*Continued*)

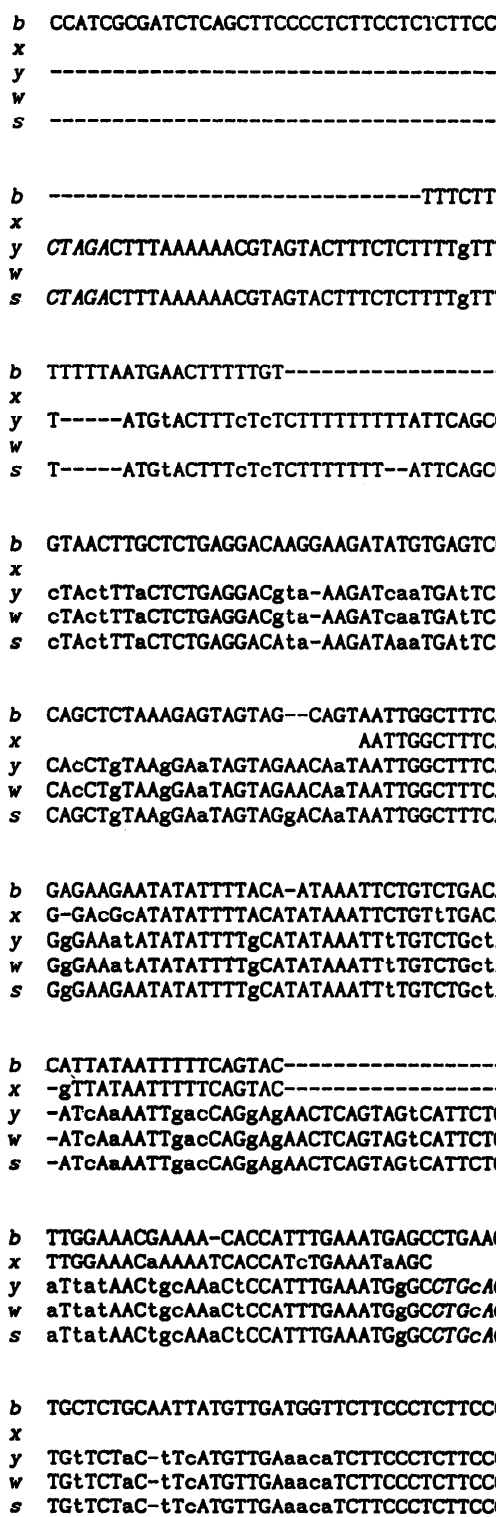

FIG. 2 (*Continued*)

```
                                                                1450
b  AACATGATTGTGATTGTGTTGGGCCACCACCTGGGCAAGGAATTCTCCCC
x
y  AAtATGATTGTGATTGTGTTGGGCCACCACCTGGGCAAGGAATTCaCCCC
w  AAtATGATTGTGATTGTGTTGGGCCACCACCTGGGCAAGGAATTCaCCCC
s  AAtATGATTGTGATTGTGTTGGGCCACCACCTGGGCAAGGAATTCaCCCC

                                                                1500
b  CTGTGCACAGGCTGCCTTCCAGAAGGTGGTGGCTGGAGTGGCCAGTGCCC
x
y  CTGTGCACAGGCTGCCTTCCAGAAGGTGGTGGCTGGAGTGGCCAGTGCCC
w  CTGTGCACAGGCTGCCTTCCAGAAGGTGGTGGCTGGAGTGGCCAGTGCCC
s  gTcTGCACAGGCTGCCTTCCAGAAGGTGGTaGCTGGAGTGGCCAGTGCCC

                                                                1550
b  TGGCTCACAAGTACCACTAAACCTCTTTTCCTGCTCTTGTCTTTGTGCAA
x
y  TGGCTCACAAGTACCACTAAACCTCTTTTCCTGCTCTTGTCTTTGTGCAA
w  TGGCTCACAAGTACCACTAAACCTCTTTTCCTGCTCTTGTCTTTGTGCAA
s  TtGCTCACAAGTACCACTAAgCCcCcTTTCCTGCT--TGTCTaTGcaCAA

                                                                1600
b  TGGTCAATTGTTCCCAAGAGAGCATCTGTCAGTTGTTGTCA-AAATGACA
x                    AAGAGAGCtTCTGTCAGTTGTTGTCA-AAATGACA
y  TGGTCAATTGTTCCCAAGAGAGCATCTGTCAGTTGTTGTCA-AAATGACA
w  TGGTCAATTGTTCCCAAGAGAGCATCTGTCAGTTGTTGTCA-AAATGACA
s  aGGTtAtgTGTcCCgtAGAGAaCAaCTGTCAacTGTgGgggGAAATGAtg

                                                                1650
b  AAGACCTTTGAAAATCTGTC---TA-CTAATAAAAGGCATTT-C--TCAC
x  AAGACCTTTGAAAATCTGTC--CTA-CTAATtAAAGGCtTTTACTTTCAC
y  AAGACCTTTGAAAATCTGTC--CTA-CTAATtAAAGGCATTTACTTTCAC
w  AAGACCTTTGAAAATCTGTC--CTA-CTAATtAAAGGCATTTACTTTCAC
s  AAGgCCTTTGggcAgCTagCTTCTATCTAATAAAtGatATTTACTTTCAt

                                                                1700
b  TGCAATGG----------------TGTGTT-AAATTAAT-GTATCTCATA
x  TGCAATGGAAATATAGTGTCTAAAaGTGTT-AAATTAATTGTgTCTCAcA
y  TGCAATGG----------------TGTGTT-AAATTAATTGTATCTCATA
w  TGCAATGG----------------TGTGTT-AAATTAATTGTATCTCATA
s  cG--ATGG----------------TGTGTTTtAATTAcTTGTgTtTCtTg

          1713
b  GAAGGGTTCATG
x  GAAGGGTTCATG
y  GAAGGGTTCATG
w  GAAGGGTTCATG
s  GAAGG-TTaATG
```

FIG. 2.—Alignment of rat adult β-globin genes. The alignment was made by applying the ALN3 algorithm (Gotoh 1986) to the groups of three sequences. After that the groups were collected together. The coding regions of the genes are underlined. Gaps are indicated by dashes. Blank spaces within these sequences represent the unsequenced regions of the genes. Sequence elements important for gene function are shown by bars above the sequences. The small letters indicate the positions where other genes differ from the rat major β-globin gene (*b*). The italicized letters indicate the positions of the restriction sites for the enzymes *Pst*I, *Bam*HI, *Kpn*I, *Xba*I, *Eco*RI, *Pst*I, and *Eco*RI. The sequence comparison of 1,713 bp begins at position −165 relative to the ATG start codon. The beginning of major-minor alignment is on our position 13 and is depicted by the slash symbol (/). The order of genes is from top to bottom: *b* = rat major gene (Radosavljević and Crkvenjakov 1989); *x*, *y* (Stevanović and Crkvenjakov 1989), and *w* (the last 359 bp of *w* are from the *u* gene) = rat minor-major hybrid genes; and *s* (Wong et al. 1988) = rat minor gene.

Table 2
Sequence-Score Comparisons

	GENE							
GENE	*b*	*x*	*t*	*y*	*w*	*u*	*z*	*s*
b ...		91.6	91.6	87.1	87.1	87.1	83.2	84.7
x ...	89.7		100	87.9	87.9	87.9	82.4	83.7
t ...	89.7	100		87.9	87.9	87.9	82.4	83.7
y ...	95.9	89.3	89.3		99.7	99.7	94.8	94.8
w ..	95.9	89.3	89.3	100		100	94.5	94.5
u ...	95.9	89.3	89.3	100	100		94.5	94.5
z ...	94.6	91.2	91.2	98.6	98.6	98.6		99.1
s ...	95.3	89.9	89.9	98.4	98.4	98.4	97.8	

NOTE.—In the lower left half of the table are the scores for coding regions; in the upper right half are the scores for noncoding regions. Scores were calculated for those positions aligned to the *b* sequence, according to data from fig. 2. For sequences *x*, *t*, *w*, *u*, and *z*, a correction was made for the absent regions. The principle involved here assumes that the ratio of rate changes in the known and unknown regions of *w* is the same as the ratio in its closest known relative, *y*. Suppose we want to know the score in the comparison of *w* and *b*, S(*w*, *b*). Let S(w^+, *b*) be the score for the region of *w* where the sequence is known. We calculate S(y^+, *b*) where the score is for that part of *y* that is homologous to the known part of *w*, w^+. Then S(*w*, *b*) = S(w^+, *b*) × S(*y*, *b*)/S(y^+, *b*), where the absence of a plus sign implies that the score is over the whole region.

the complete identity with the amino acid sequence of the most abundant adult β-globin chain, $^{I/II}\beta$ (Garrick et al. 1978). Wong et al. (1988) claim that *s* (*z*) is active, too. Whether the *x*, *v*, and *t* genes are active remains to be determined, but it is nevertheless certain that, among mammals studied so far, rats have the largest number of active adult β-globin genes.

On the basis of the sequence data, we conclude that rat has three types of β genes: major, minor, and hybrid. The similarity-score comparison (table 3) localizes the breakpoint of recombination events which created hybrid genes within sequences of the major and minor genes. We assume that unequal recombination is responsible for formation of rat haplotypes with such a large number of adult β-globin genes. Rando et al. (1986) reported a similar case for α-globin loci of sheep. These authors proposed that this situation represents an intermediate stage of coincident evolution.

Table 3
Sequence-Score Comparisons for Hybrid β-Globin Genes *x* and *y*, with Major Gene *b*, and Minor Gene *s*

Gene	5′ Flanking Region	Exon I	IVS1	Exon II	IVS2	Exon III	3′ Flanking Region
x:							
b	75.2	83.2[a]	93.3	92.5	94.5[a]	89.9[a]	92.8[a]
s	94.5	92.4[a]	93.3	94.3	83.7[a]	80.6[a]	68.3[a]
y:							
b	75.8	90.3	93.3	96.8	86.2	98.4	96.9
s	100.0	97.8	99.2	99.5	98.6	96.9	72.5

NOTE.—Sequences were aligned as in fig. 2. Scores were calculated as in table 2. Oblique lines localize the recombination region.

[a] Sequence of hybrid gene was not completed for region involved.

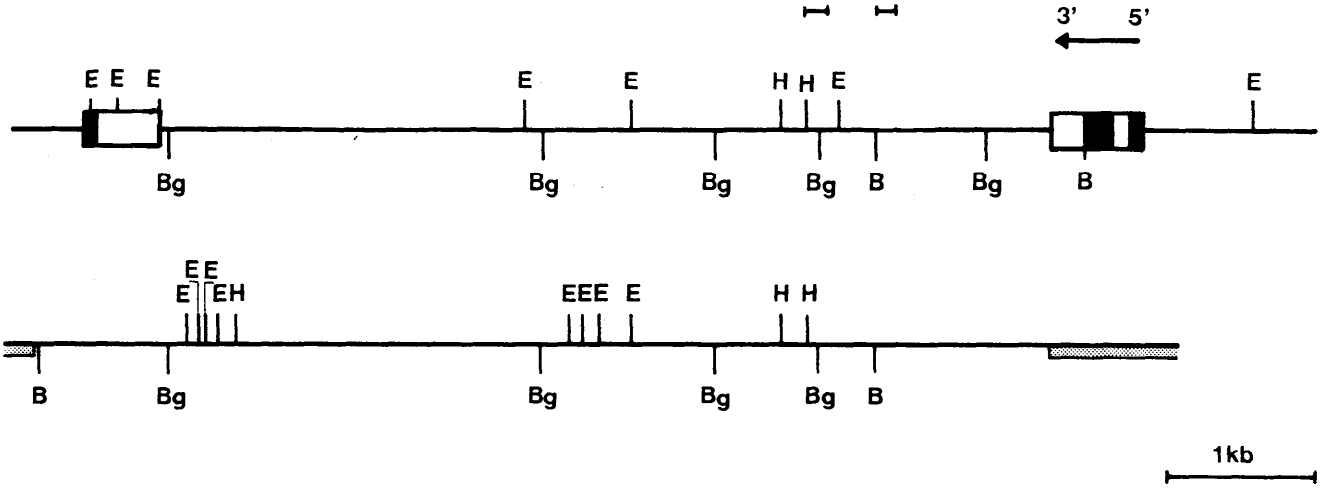

FIG. 3.—Restriction map of the *x* gene. *Upper line,* Restriction map of the *x* globin gene. The gene is presented as a box, with the filled parts denoting exons. Restriction sites *Eco*RI (E), *Bam*HI (B), *Hin*dIII (H), and *Bgl*II (Bg) are indicated. The bars above the map indicate the locations of the sequenced regions. Because of the alignment of L1 elements, the gene is presented in the reverse orientation (indicated with the arrow above the map). *Lower line,* Restriction map of the rat L1Rn 3 element, according to D'Ambrosio et al. (1986).

Table 4
Sizes (kb) of Restriction Fragments Possessing Adult Rat β-Globin Gene Sequences

Gene(s)	*Eco*RI					*Bam*HI	
a, b, c	2.7			0.18	1.4	4.4	3.5
e, z, r	1.6			0.18	3	2.3	4.5
d	1.6			0.18	2.9[a]	2.3	3.6[a]
x, v, t	2.8[b]	1.3[c]	0.7[c] 2.4[c]	0.3[b]	0.18 2.9[a]	2.3	1.35[b] 8[d]
y, w, u	1.6			0.18	1.4	2.3	3.5

NOTE.—Vertical columns contain fragments with related sequences.
[a] Changed by deletion.
[b] Changed by L1 insertion.
[c] Internal L1 fragment.
[d] Changed by both deletion and L1 insertion.

The case we found for rat β-globin genes cannot be explained in the same way. Hybrid genes—or at least one of them (*x*)—conserved their identity and regulatory elements during the long period of time, just as the major and minor genes did.

L1 Elements

L1 is a family of movable DNA sequences found in mammals (Casavant et al. 1988). The stretches of L1 elements have been found in the α-globin and β-globin gene families. L1 is a major component of the mouse β-globin complex, with 15 elements constituting ~22% of its sequence (Shehee et al. 1989). In all cases, L1 elements in α-globin or β-globin gene families are found in intergenic regions. In the present paper we show the first example of insertion of an L1 element into the globin gene itself. We localized insertion of L1Rn in IVS2 of genes *x*, *v*, and *t*. A similar situation was found in the mouse serum albumin gene, which has a 0.9-kb-long insertion of L1 element in the 12th intron (Jubier-Maurin et al. 1985).

Evolution of Rat β-Globin Genes

The described results allow us to propose a pathway of mutational events leading from the ancient two-β-gene rodent ancestor to the present rat haplotypes (fig. 5). The essential event is the unequal crossing-over between minor and major β genes that gives the three-gene haplotype, the middle gene of which is part minor and part major. Then the intergenic-distance-shortening event between the middle and last genes advanced formation of genes *d*, *x*, *v*, and *t*. At first, the deletion led to the three-gene haplotype a. It is an open question whether the original unshortened three-gene haplotype still exists. Insertion of the L1Rn is common for genes *x*, *v*, and *t*. The timing of the L1 insertion event cannot be determined, because of the high conversion and mutability rates of these sequences (Jubier-Maurin et al. 1985). In our model the insertion is arbitrarily placed in the recent past before a major-minor recombination event with two fully modified three-gene haplotypes to give the present haplotype b. Haplotype c arose from haplotype b in two successive events: a gene deletion giving a four-gene intermediate, followed by recombination which led to the duplication of the second gene.

The model is supported by several lines of evidence. First, sequencing detects three kinds of genes: the major, the minor, and the hybrid between the two.

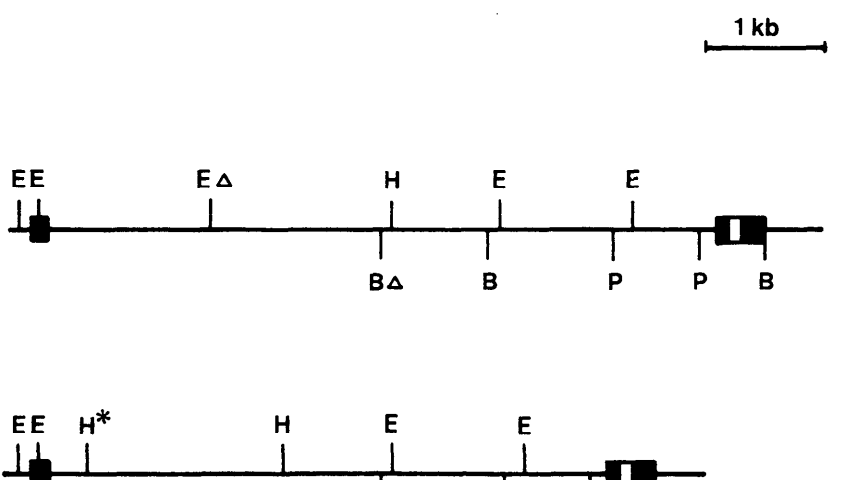

FIG. 4.—Comparison of structures of intergenic regions. *Top,* Structure of primordial intergenic region. *Bottom,* Structure of intergenic region after the 0.9-kb deletion. The black boxes denote exons. Restriction sites for endonucleases *Eco*RI (E), *Bam*HI (B), *Hin*dIII (H), and *Pst*I (P) are indicated. Δ = Restriction sites lost after the deletion; * = *Hin*dIII site present in the affected intergenic regions in haplotypes b and c. In haplotype a this restriction site does not exist.

Second, if the similarity scores are inversely proportional to the divergence time, then the oldest among the sequenced hybrid genes are the *x* and *t* genes. To examine the question of the age of the second gene of haplotype a, in the absence of sequence information, we relied on the restriction maps. Gene *d* is an old one, since it has intergenic shortening marked by a cluster of restriction sites characteristic of the *x*, *v*, and *t* genes of five-gene haplotypes. This gene is a hybrid gene, since it is marked both with the 5′ flanking *Eco*RI and *Pst*I sites of minor genes and by the absence of IVS2 *Xba*I and *Pst*I sites which are not present in major genes (Stevanović et al. 1989). The prediction of the model is that the sequence of the internal gene of haplotype a (*d*) will be very similar to the sequences of the *x* and *t* genes. The latter genes are younger, since they show the intergenic shortening and, also, the insertion of L1 element.

Third, haplotypes b and c are only partially different, since they share a 15-kb region encompassing the *v* (*t*) and *z* (*r*) genes (Stevanović et al. 1989). The hybrid genes *y*, *w*, and *u* remaining in the regions of difference of the two haplotypes are nearly identical at the sequence level. Therefore, the events creating one five-gene haplotype from another must have been the most recent ones in rat β-globin gene evolution. It seems that haplotype b is older than haplotype c. The finding of two older identical hybrid genes flanking a younger hybrid gene immediately suggests origin from a three-gene haplotype in the manner depicted in the model.

It is reasonably probable that the five-gene haplotype c was formed during captivity. The creation of a new DNA haplotype during domestication or inbreeding has not been reported before, as far as we are aware. The converse phenomenon, the occurrence of wild-type polymorphisms/haplotypes in inbred strains of mouse is common—as, for instance, for mouse β-gene haplotypes (Berry and Peters 1977). We have found an exclusively three-gene haplotype(s) in 12 animals from two pop-

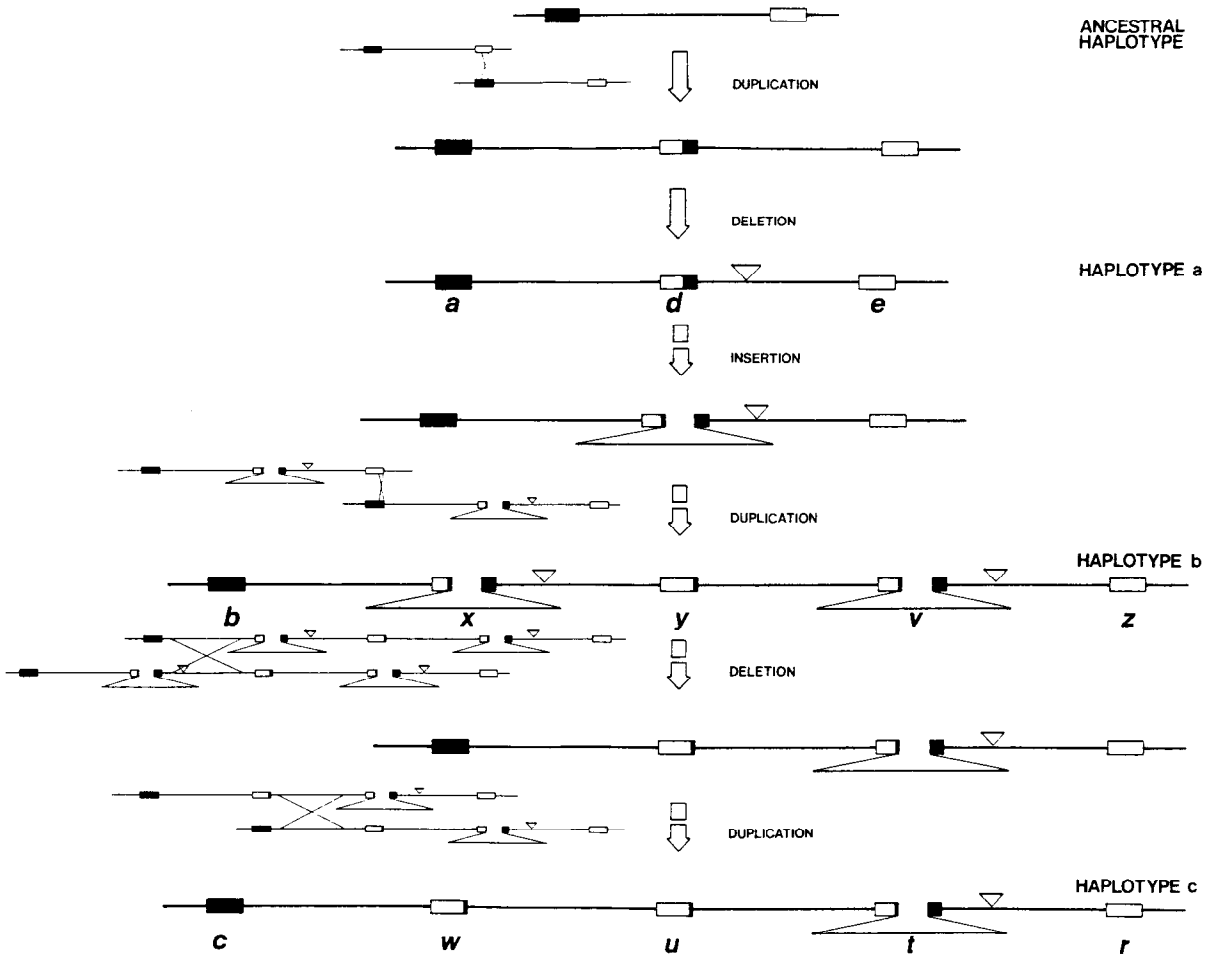

FIG. 5.—Model for the evolution of the rat β-globin gene cluster. The ancestral two-gene haplotype was formed prior to the divergence of ancestral rat and mouse. A recombination between the minor and major genes most likely formed a three-gene haplotype. After the 0.9-kb deletion it became the present haplotype a. The next step was the insertion of the L1 element in IVS2 of the second gene. The unequal recombination of two resulting chromosomes gave haplotype b in one step. The formation of haplotype c needs two more recombination events between the haplotype b chromosomes. One of them linked the major and *y* genes, and the other duplicated the *y* genes. At that moment the *y* gene was already changed in the *w* gene by a few nucleotide substitutions. The small deletion is indicated by the inverted triangle above the line; the insertion is indicated by an open triangle below the genes. The broken arrows indicate that the depicted order of events, although the most probable, is not the only one possible.

ulations from the vicinity of Belgrade (T. Paunesku, unpublished data). It is possible that the three-gene haplotype(s) originated relatively early in the evolution of genus *Rattus,* which would make this haplotype(s) common in the extant species of the genus. If transcriptional regulation evolved requiring activity of all three genes, the L1 insertion event might have reduced the activity of the target gene. Then selection against it might have favored rapid formation and fixation of a haplotype b having restored activity with three normal genes and two genes with insertions.

Acknowledgments

This research was supported in part by SIZ for Science (Serbia) and by U.S.–Yugoslav Joint Board on Scientific and Technological Cooperation grant JF 789.

LITERATURE CITED

ANTONARAKIS, S. E., H. H. KAZAZIAN, and S. H. ORKIN. 1985. DNA polymorphism and molecular pathology of the human globin gene clusters. Hum. Genet. **69**:1–14.

BERRY, R. J., and J. PETERS. 1977. Heterogeneous heterozygosities in *Mus musculus* populations. Proc. R. Soc. Lond. [B] **197**:485–503.

CASAVANT, N. C., S. C. HARDIES, F. D. FUNK, M. B. COMER, M. H. EDGELL, and C. A. HUTCHINSON III. 1988. Extensive movement of LINES ONE sequences in β-globin loci of *Mus caroli* and *Mus domesticus.* Mol. Cell. Biol. **8**:4669–4674.

CHENG, J.-F., and R. C. HARDISON. 1988. The rabbit α-like globin gene cluster is polymorphic both in the sizes of *Bam*HI fragments and in the members of duplicated sets of genes. Mol. Biol. Evol. **5**:486–498.

CZELUSNIAK, J., M. GOODMAN, D. HEWETT-EMMETT, M. L. WEISS, P. J. VENTA, and R. E. TASHIAN. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian hemoglobin genes. Nature **298**:297–300.

D'AMBROSIO, E., S. D. WAITZKIN, F. R. WITNEY, A. SALEMME, and A. V. FURANO. 1986. Structure of the highly repeated long interspersed DNA family (LINE or L1Rn) of the rat. Mol. Cell. Biol. **2**:411–424.

DAVIS, R. W., D. BOTSTEIN, and J. R. ROTH. 1980. Advanced bacterial genetics: a manual for genetic engineering. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

EFSTRATIADIS, A., J. W. POSAKONY, T. MANIATIS, R. M. LAWN, C. O'CONNELL, R. A. SPRITZ, J. K. DERIEL, B. G. FORGET, S. M. WEISSMAN, J. L. SLIGHTOM, A. E. BLECHL, O. SMITHIES, F. E. BARALLE, C. C. SHOULDERS, and N. J. PROUDFOOT. 1980. The structure and evolution of human β-globin gene family. Cell **21**:653–668.

FEINBERG, A. P., and B. VOGELSTEIN. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.

GARNER, K. J., and J. B. LINGREL. 1988. Structural organization of the β-globin locus of B-haplotype sheep. Mol. Biol. Evol. **5**:134–140.

GARRICK, L. M., R. L. SLOAN, T. W. RYAN, T. J. KLONOWSKI, and M. D. GARRICK. 1978. Primary structure of the β^{maj} chain of rat hemoglobins. Biochem. J. **173**:321–330.

GOTOH, O. 1986. Alignment of three biological sequences with an efficient traceback procedure. J. Theor. Biol. **121**:327–337.

HARRIS, S., P. A. BARRIE, M. L. WEISS, and A. J. JEFFREYS. 1984. The primate ψβ1 gene: an ancient β-globin pseudogene. J. Mol. Biol. **180**:785–801.

HATTORI, M., and Y. SAKAKI. 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152**:232–238.

HERRMAN, B., M. BUCAN, P. E. MAINS, A.-M. FRISCHAUF, L. M. SILVER, and H. LEHRACH. 1986. Genetic analysis of the proximal portion of the mouse t-complex: evidence for the second inversion within t haplotypes. Cell **44**:469–476.

JUBIER-MAURIN, V., B. J. DOD, M. BELLIS, M. PIECHACZYK, and G. ROIZES. 1985. Comparative

study of the L1 family in the genus *Mus:* possible role of retroposition and conversion events in its concerted evolution. J. Mol. Biol. **184**:547–564.

MANTIOLETTI, C., and C. SCHNEIDER. 1988. A new and fast method for preparing high quality lambda DNA suitable for sequencing. Nucleic Acids Res. 7:2873–2884.

MARTIN, S. L., K. A. VINCENT, and A. C. WILSON. 1983. Rise and fall of the delta globin gene. J. Mol. Biol. **164**:513–528.

MYERS, R. M., K. TILLY, and T. MANIATIS. 1986. Fine structure genetic analysis of a β-globin promoter. Science **232**:613–618.

OHSHITA, Y., and T. HOZUMI. 1987. Sequence of rat β-globin mRNA. Nucleic Acids Res. **15**: 1336.

RADOSAVLJEVIĆ, D., and R. CRKVENJAKOV. 1989. Genomic sequence of rat β-globin major gene. Nucleic Acids Res. **11**:4368.

RANDO, A., L. RAMUNNO, and P. MASINA. 1986. Variation in the number of α-globin loci in sheep. Mol. Biol. Evol. **3**:168–176.

SHEHEE, W. R., D. D. LOEB, N. B. ADEY, F. H. BURTON, N. C. CASAVANT, P. COLE, C. J. DAVIES, R. A. MCGRAW, S. A. SCHICHMAN, D. M. SEVERYNSE, C. F. VOLIVA, F. W. WEYTER, G. B. WISLY, M. H. EDGELL, and C. A. HUTCHINSON III. 1989. Nucleotide sequence of the BALB/c mouse β-globin complex. J. Mol. Biol. **205**:41–62.

STEVANOVIĆ, M., and R. CRKVENJAKOV. 1989. Genomic sequence of rat β-globin minor gene. Nucleic Acids Res. **12**:4878.

STEVANOVIĆ, M., T. PAUNESKU, D. RADOSAVLJEVIĆ, R. DRMANAC, and R. CRKVENJAKOV. 1989. Variant chromosomal arrangement of adult β-globin genes in rat. Gene **79**:139–150.

TOWNES, T. M., M. C. FITZGERALD, and J. B. LINGREL. 1984. Triplication of a four-gene set during evolution of the goat β globin locus produced three genes now expressed differentially during development. Proc. Natl. Acad. Sci. USA **81**:6589–6593.

WONG, W. M., V. M. S. LAM, L. Y. L. CHENG, and J. W. O. TAM. 1988. Genomic sequence of a Sprague-Dawley rat β-globin gene. Nucleic Acids Res. **16**:2342.

WOOD, W. G., and D. J. WEATHERALL. 1983. Developmental genetics of the human haemoglobins. Biochem. J. **215**:1–10.

WALTER M. FITCH, reviewing editor

Received December 19, 1989; revision received May 15, 1990

Accepted May 21, 1990