Higher Frequency of Concerted Evolutionary Events in Rodents Than in Man at the Polyubiquitin Gene VNTR Locus

Mitsuru Nenoi,* Kazuei Mita,† Sachiko Ichimura,* and Akihiro Kawano‡

**Division of Biology and Oncology, †Genome Research Group and ‡Laboratory of Animal and Plant Sciences, National Institute of Radiological Sciences, Inage-ku, Chiba 263 Japan*

Manuscript received July 28, 1997
Accepted for publication October 24, 1997

ABSTRACT

The polyubiquitin gene is an evolutionarily conserved eukaryotic gene, encoding tandemly repeated multiple ubiquitins, and is considered to be subject to concerted evolution. Here, we present the nucleotide sequences of new alleles of the polyubiquitin gene *UbC* in humans and *CHUB2* in Chinese hamster, which encode a different number of ubiquitin units from those of previously reported genes. And we analyze the concerted evolution of these genes on the basis of their orthologous relationship. That the mean of the synonymous sequence difference K_s, which is defined as the number of synonymous substitution relative to the total number of synonymous sites, within the *UbC* and *CHUB2* genes (0.192 ± 0.096) is significantly less than K_s between these genes (0.602 ± 0.057) provides direct evidence for concerted evolution. Moreover, it also appears that concerted evolutionary events have been much more frequent in *CHUB2* than in *UbC*, because K_s within *CHUB2* (0.022 ± 0.018) is much less than that within *UbC* (0.362 ± 0.192). By a numerical simulation, postulating that the major mechanism of concerted evolution in polyubiquitin genes is unequal crossing over, we estimated the frequency of concerted evolutionary events of *CHUB2* at 3.3×10^{-5} per year and that of *UbC* at no more than 5.0×10^{-7} per year.

UBIQUITIN is a highly conserved small protein of 76 amino acids functioning in the selective proteolysis of a variety of cellular proteins at the 26S proteasome (Hochstrasser 1996; Ciechanover and Schwartz 1994). In addition, ubiquitin has been shown to function in proteasome action-independent processes such as DNA repair (Bregman *et al.* 1996), endocytosis of cell surface proteins (Hicke and Riezman 1996), and NFκB signal transduction (Chen *et al.* 1996). In humans, ubiquitin is encoded by a multiple gene family composed of UbA_{52}, UbA_{80}, *UbB*, and *UbC* (Baker and Board 1991, 1987; Wiborg *et al.* 1985). The *UbB* and *UbC* genes, located on chromosome *17* (17p11.1-17p12) (Webb *et al.* 1990) and chromosome *12* (12q24.3) (Board *et al.* 1992), respectively, are termed polyubiquitin genes because they encode tandemly repeated multiple ubiquitins with no intervening spacer. The polyubiquitin genes are conserved also in rodents, and we previously isolated and sequenced the Chinese hamster polyubiquitin genes, *CHUB1* and *CHUB2*, which are the counterparts of the human *UbB* and *UbC*, respectively (Nenoi *et al.* 1992; Nenoi *et al.* 1994). Due to unequal crossing over, the number of ubiquitin units encoded by the *UbC* gene is variable among individuals, with the most frequent allele encoding nine units, and the other less frequent alleles encoding eight or seven units (Baker and Board 1989). The nucleotide sequence GNNGTGGG, which has been found to be the consensus marker sequence for a variable number of tandem repeat (VNTR) loci in humans (Nakamura *et al.* 1988), is preserved at the 3′ end of every ubiquitin-coding unit of the *UbC*. We have actually observed that the *UbC* gene alleles of HeLa cells (Nenoi *et al.* 1996) as well as the Chinese hamster polyubiquitin gene *CHUB2* of V79 fibroblasts (Nenoi *et al.* 1994) are heterogeneous in the repeat number of the ubiquitin units.

It has been suggested that the polyubiquitin genes may show strong evidence of concerted evolution (Sharp and Li 1987; Tan *et al.* 1993; Keeling and Doolittle 1995; Vrana and Wheeler 1996). This is consistent with observations of a high variability in the number of the ubiquitin-coding units in the *UbC* and *CHUB2* genes because unequal crossing over is believed to be one of the major mechanisms for concerted evolution (Dover 1982; Darnell *et al.* 1990). A greater degree of homology between repeat units within a species as compared to orthologous repeat units in different species is a diagnostic of concerted evolution. However, in the case of the polyubiquitin gene, data have not been available for all loci in each species, and it has not been clear whether the loci compared across species are orthologues or paralogues.

In this report, we will analyze the unequal crossover events that are thought to have occurred on the human

Corresponding author: Dr. Mitsuru Nenoi, Division of Biology and Oncology, National Institute of Radiological Sciences, 9-1, Anagawa-4-chome, Inage-ku, Chiba-shi 263 Japan. E-mail: m_nenoi@nirs.go.jp

UbC gene and the Chinese hamster *CHUB2* gene, and we will analyze the concerted evolution of these genes on the basis of their orthologous relationship. It will be proposed that the *UbC* gene that encodes nine ubiquitins was recently generated by an unequal crossover event between a site in the seventh ubiquitin-coding unit and the homologous site in the sixth unit of the *UbC* gene that encodes eight ubiquitins. We will also show a much higher homology between the ubiquitin-coding units within the *CHUB2* gene than within the *UbC* gene, which could be explained by a higher frequency of unequal crossover events during the evolution of the *CHUB2*.

MATERIALS AND METHODS

Isolation and sequencing of the polyubiquitin genes: HeLa cells and V79 Chinese hamster lung fibroblasts were cultured in Eagle's MEM (Nissui, Tokyo, Japan) supplemented with 10% FBS (GIBCO BRL, Rockville, MD). Fresh thymus from a Chinese hamster was generously provided by Dr. Mitsuhiro Numata (National Institute of Health, Japan). The method for isolation of the polyubiquitin genes has been described (Nenoi *et al.* 1996). Briefly, the genomic DNA was extracted from cell suspension, and the DNA fragment containing the whole ubiquitin-coding region was amplified by PCR using LA *Taq* polymerase (TAKARA, Otsu, Japan). Primers specific to the 5′ and 3′ flanking region of the *UbC* gene were designed on the basis of the reported nucleotide sequences (Wiborg *et al.* 1985; Nenoi *et al.* 1996); P.Hu1: 5′-TTGGGCAGTGCACCCGTACCTTTGG-3′, P.Hu2: 5′-GTGC AATGAAATTTGTTGAAACCTTAAAAGGGG-3′, and of the *CHUB2* gene (Nenoi *et al.* 1994); P.CH1: 5′-CCACGAATAT TTGTCATTCCTGACCTG-3′, P.CH2: 5′-GCTAAAACGAGAT CCAACACCTTTGGG-3′ (indicated in Figure 2). A series of deletion mutants was constructed by partially digesting the PCR products either with *Pvu*II (for *UbC*) or with *Bgl*II (for *CHUB2*), followed by subcloning in the pUC18 and then sequencing with an automated DNA sequencer (model 373A; Applied Biosystems, Foster City, CA).

Sequence analysis: The homology analysis between every pair of ubiquitin-coding units in the *UbC* and *CHUB2* was carried out by evaluating the synonymous sequence difference per site, K_s, which is defined as the number of synonymous substitution relative to the total number of synonymous sites (Miyata and Yasunaga 1980). K_s values were calculated as described by Miyata and Yasunaga (1980) and were corrected for multiple substitution (Kimura and Ohta 1972).

Numerical simulation for concerted evolution of the polyubiquitin gene: The synonymous sequence difference between several mammalian species excluding rodents has been estimated at 0.47 on average from 11 different genes (Kikuno *et al.* 1985). Therefore, the mean evolutionary rate of synonymous substitution can be calculated to be 2.0×10^{-9} per site per year by assuming that the mammalian divergence is 120 MYA (Easteal *et al.* 1995). The synonymous sequence difference between rodents and other mammals has been estimated at 0.66 on average from 35 different genes (Miyata *et al.* 1987). Then the mean evolutionary rate of rodent genes can be calculated to be 3.5×10^{-9} per site per year. Applying these values to polyubiquitin genes, and postulating that a major mechanism of concerted evolution is unequal crossing over (Black and Gibson 1974; Ohta 1976), we constructed a Monte Carlo simulation model code in C-language on a Sun Sparc Center2000 computer. This code simulates the evolutionary change of K_s between the ubiquitin-coding units within a given polyubiquitin gene after the divergence of human and rodents. Let $K_s(i,j;t)$ be the synonymous sequence difference between the i-th and j-th ubiquitin unit of a given polyubiquitin gene at a time t. By definition,

$$K_s(j, i; t) = K_s(i, j; t) \quad 1 \leqq i, j \leqq N$$
$$K_s(i, i; t) = 0 \qquad 1 \leqq i \leqq N$$

where N is the number of ubiquitin units encoded by the polyubiquitin gene at the time of t. Random numbers were generated to provide the dates of the unequal crossover events. The event type (unit-duplication or unit-deletion) was also determined randomly under the constraint that the unit number is kept between one and the maximum number, which was set at 20 in this study, and that the final unit number reaches that of the actual genes after the simulation. During a time interval, δt, between the adjacent events, every synonymous sequence difference increases by $v\delta t$, where v is the evolutionary rate of synonymous substitution. Then,

$$K_s(i, j; t + \delta t) = K_s(i, j; t) + v\delta t \quad 1 \leqq i, j \leqq N$$

If a unit-duplication event occurs at the n-th unit at this moment,

$$K_s(i, j; t + \delta t) = K_s(i, j; t + \delta t) \qquad 1 \leqq i, j \leqq n$$
$$K_s(i, j; t + \delta t) = K_s(i, j - 1; t + \delta t) \qquad 1 \leqq i \leqq n,\ n + 1 \leqq j \leqq N + 1$$
$$K_s(i, j; t + \delta t) = K_s(i - 1, j; t + \delta t) \qquad n + 1 \leqq i \leqq N + 1,\ 1 \leqq j \leqq n$$
$$K_s(i, j; t + \delta t) = K_s(i - 1, j - 1; t + \delta t) \quad n + 1 \leqq i, j \leqq N + 1$$

On the other hand, if a unit-deletion event occurs at the n-th unit,

$$K_s(i, j; t + \delta t) = K_s(i, j; t + \delta t) \qquad 1 \leqq i, j \leqq n - 1$$
$$K_s(i, j; t + \delta t) = K_s(i, j + 1; t + \delta t) \qquad 1 \leqq i \leqq n - 1,\ n \leqq j \leqq N - 1$$
$$K_s(i, j; t + \delta t) = K_s(i + 1, j; t + \delta t) \qquad n \leqq i \leqq N - 1,\ 1 \leqq j \leqq n - 1$$
$$K_s(i, j; t + \delta t) = K_s(i + 1, j + 1; t + \delta t) \quad n \leqq i, j \leqq N - 1$$

These calculations were repeated for the time from the man-rodent split ($t = 0$) until the present ($t = 1.2 \times 10^8$).

The sequence data presented in this article have been submitted to the DDBJ/EMBL/GenBank databases under accession numbers AB003730, AB003731, and AB003732.

RESULTS AND DISCUSSION

Polymorphism of *UbC* and *CHUB2* produced by unequal crossing over: The *UbC* gene in HeLa cells and the *CHUB2* gene in V79 cells and Chinese hamster thymus were amplified by PCR (Figure 1). The 2.5-kb fragments amplified from the HeLa DNA contain the previously reported *UbC* gene that encodes nine ubiquitins (Wiborg *et al.* 1985; Einspanier *et al.* 1987; Nenoi *et al.* 1996). The 2.1-kb fragment amplified from the V79 DNA corresponds to the *CHUB2* gene that encodes eight ubiquitins, followed by an apparently deleted and mutated ubiquitin-like polypeptide of 50-aa (Nenoi *et al.* 1994). The other three products (the 2.3-kb fragment from HeLa, the 3.3-kb fragment from V79, and the 2.8-kb fragment from the Chinese hamster thymus) were isolated and sequenced. As shown in Figure 2, every product encoded tandemly repeated ubiquitins, ex-

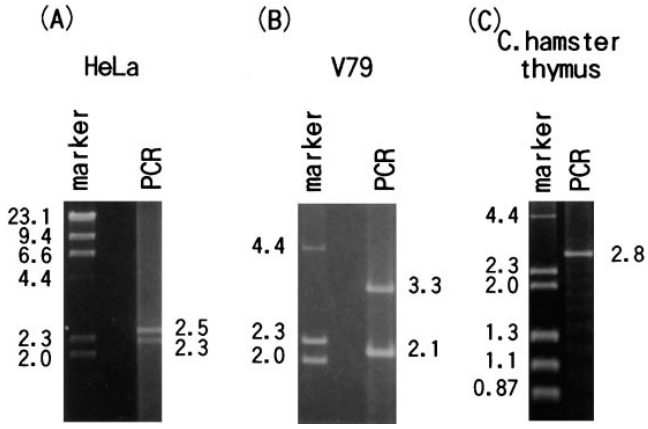

Figure 1.—Heterogeneity in the PCR products. The polyubiquitin gene *UbC* of HeLa cells (A), the *CHUB2* of V79 Chinese hamster cells (B), and the *CHUB2* gene of the cells in Chinese hamster thymus (C) were amplified by PCR with the primer sets of P.Hu1/P.Hu2 (for *UbC*) and of P.CH1/P.CH2 (for *CHUB2*) as described under materials and methods.

cept that an Ile was substituted with Val at the 61st amino acid position in the eighth unit of the 3.3-kb fragment from V79 cells (wavy underline in Figure 2B). Both of the 5′ and 3′ flanking regions precisely coincided with those of the reported *UbC* (Wiborg *et al.* 1985; Einspanier *et al.* 1987; Nenoi *et al.* 1996) and *CHUB2* gene sequence (Nenoi *et al.* 1994), indicating that these are the allele variants of the *UbC* and *CHUB2* in the repeat number of the ubiquitin unit. Hereafter, the polyubiquitin gene of HeLa cells encoding eight ubiquitins (Figure 1A) is designated as *UbC(8u)*, the gene of the V79 cells encoding 13 ubiquitins (Figure 1B) is designated as *CHUB2(13u)*, and the gene of the Chinese hamster thymus encoding 11 ubiquitins (Figure 1C) is designated as *CHUB2(11u)*. The previously reported human *UbC* gene encoding nine ubiquitins and the Chinese hamster *CHUB2* gene encoding eight ubiquitins are designated as *UbC(9u)* and *CHUB2(8u)*, respectively.

Comparing the nucleotide sequences of the *UbC* gene isolated from their independent origins, Baker and Board (1989) have identified an unequal crossover event site at 40–50 nt in the seventh and the ninth unit of the *UbC* gene encoding nine ubiquitins, that might have generated the *UbC* gene encoding seven ubiquitins. The sequence data they have used for comparison, however, seem very likely to have been deduced by a misalignment of the subcloned *Xho*I fragments (Nenoi *et al.* 1996). In fact, the region of 40–50 nt, which they marked coincides with the *Xho*I site (Baker and Board 1989), where the discontinuity of the sequence is thought to have been artificially introduced by the possible misalignment of the *Xho*I fragments. We compared the sequences of the *UbC(8u)*gene and the *UbC(9u)* gene (Figure 3), showing the position of the nucleotide discrepancy. The region from the first through the seventh unit of the *UbC(9u)* gene matches well with that of the *UbC(8u)* gene (Figure 3A). A considerable number of discrepancies can be observed from the 41 nt in the eighth unit of the *UbC(9u)* gene (9u-8-41) to the 3′ end of the last unit. However, the region from the 220 nt in the sixth unit of the *UbC(9u)* gene (9u-6-220) to the 3′ end of the last unit matches well with the region from the 220 nt in the fifth unit of the *UbC(8u)* gene (8u-5-220) to the last unit (Figure 3B). Therefore it appears that either the *UbC(8u)* gene would have been generated by an unequal crossover resulting in the deletion of one unit of the ubiquitin-coding sequence somewhere from (9u-5-220) to (9u-9-41) in the *UbC(9u)* gene (Figure 3C), or that the *UbC(9u)* gene was generated by the insertion of one unit of ubiquitin-coding sequence somewhere from (8u-6-220) to (8u-8-41) (Figure 3D). It is not clear which scenario is more likely. However, Table 1A shows the extremely low level of sequence difference between the seventh and both the sixth and eighth units of the *UbC(9u)* gene (0.059 and 0.019, respectively) compared with the differences between the other units (0.362 on average). This is what should be expected if the *UbC(9u)* gene was recently generated by an unequal crossover event between a site in the seventh ubiquitin-coding unit and the homologous site in the sixth unit of the *UbC(8u)* gene (case in Figure 3D).

Additionally, a very low sequence difference between the second and the fourth units of the *UbC(9u)* gene was observed (Table 1A). Relatively low sequence differences between the third and both the first and fifth units were also observed. This also is to be expected if two ubiquitin-coding units were inserted by an unequal crossing over between a site in the third unit and the homologous site in the first unit of the ancestral *UbC* gene. This suggests that such an event must have occurred before the *UbC(9u)* gene was ever generated.

In contrast to the human *UbC* gene, it seemed nearly impossible to estimate the site of unequal crossover events in the *CHUB2* gene only by comparing the nucleotide sequences of the *CHUB2(8u)*, *CHUB2(11u)*, and *CHUB2(13u)* because of the extremely low level of sequence differences among the ubiquitin-coding units in each of these alleles (Table 1B, data not shown). The small sequence differences within the *CHUB2* alleles seem to be caused by concerted evolution.

Evidence for the concerted evolution of polyubiquitin genes: As multiple polyubiquitin gene loci have been observed in mammals, distinguishing between orthologous and paralogous homology is of great importance when analyzing concerted evolution by comparing the degree of homology between repeats both in different species and within a locus (Sharp and Li 1987; Vrana and Wheeler 1996). However, in previous studies dealing with the concerted evolution of the polyubiquitin gene in mammals, only the presence of a higher level of homology within a species than between

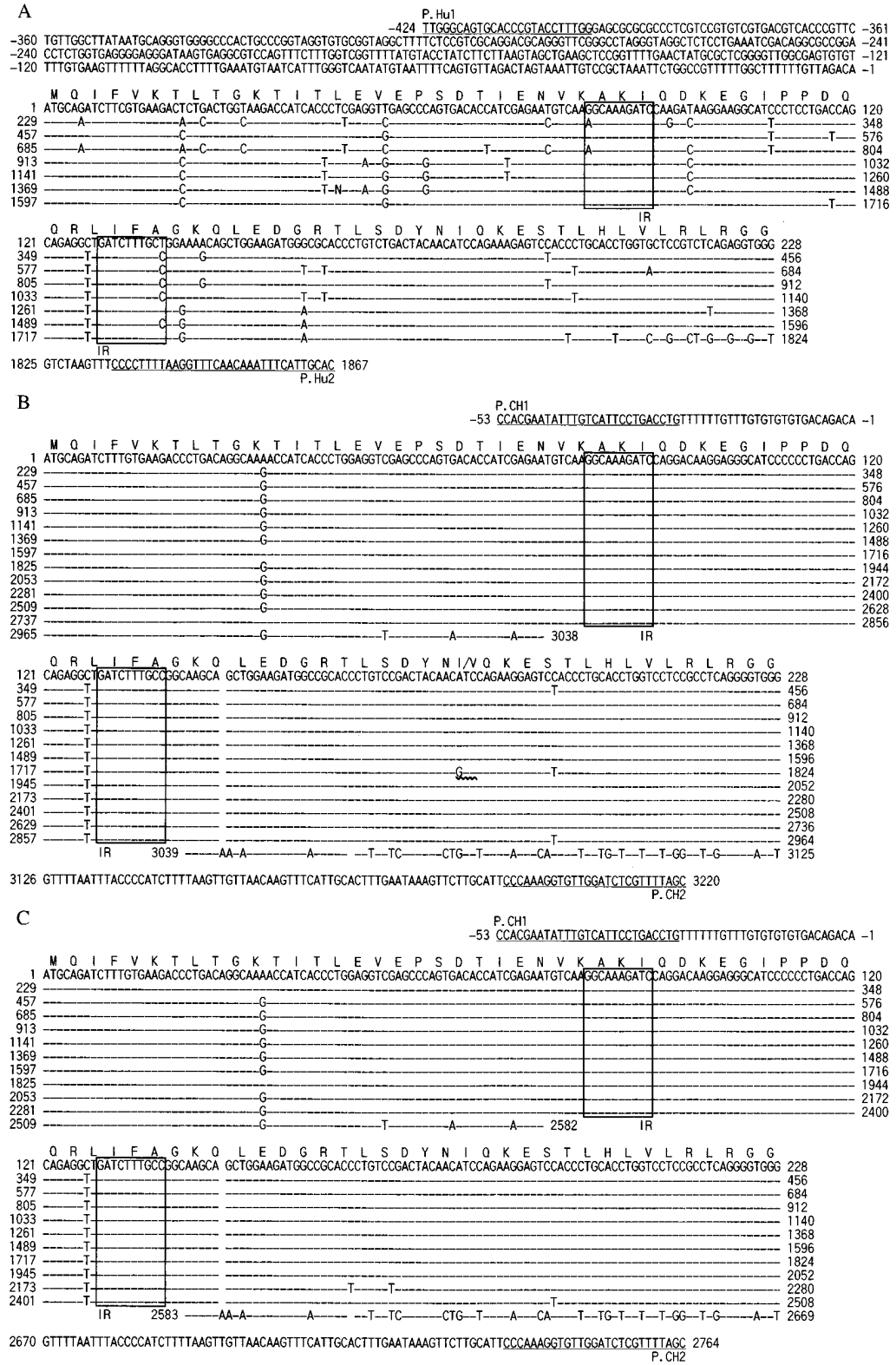

Figure 2.—Nucleotide sequence of the polyubiquitin genes. (A) the *UbC* of HeLa cells encoding eight ubiquitins (deposited in the DDBJ, EMBL, and GenBank nucleotide sequence databases under the accession number AB003730), (B) the *CHUB2* of V79 cells encoding 13 ubiquitins (accession number AB003732) and (C) the *CHUB2* of Chinese hamster thymus encoding 11 ubiquitins (accession number AB003731). Sequences are numbered from the A of the first ubiquitin initiation codon. The nucleotide sequence of the first ubiquitin-coding unit is given in full, and in the following repeats, nucleotide identity with the first unit is indicated by a dash. Primer annealing sites are underlined, and a pair of inverted repeats is boxed and designated as IR. The codon that causes amino acid substitution from Ile to Val is wavy-underlined.

species has been used as evidence of concerted evolution (Sharp and Li 1987; Tan *et al.* 1993; Keeling and Doolittle 1995).

We have previously identified a high degree of homology in the 3′ UTR between the *UbC* and the *CHUB2* (74 matches out of 88 bp when gaps were introduced), and showed also that they share a pair of inverted repeats of 10 bp in length at the same location with the same sequence (Nenoi *et al.* 1994) (boxes designated as IR in Figure 2). In addition, we isolated another Chinese hamster polyubiquitin gene, *CHUB1*, which encodes five units, and showed both that its 3′ UTR is

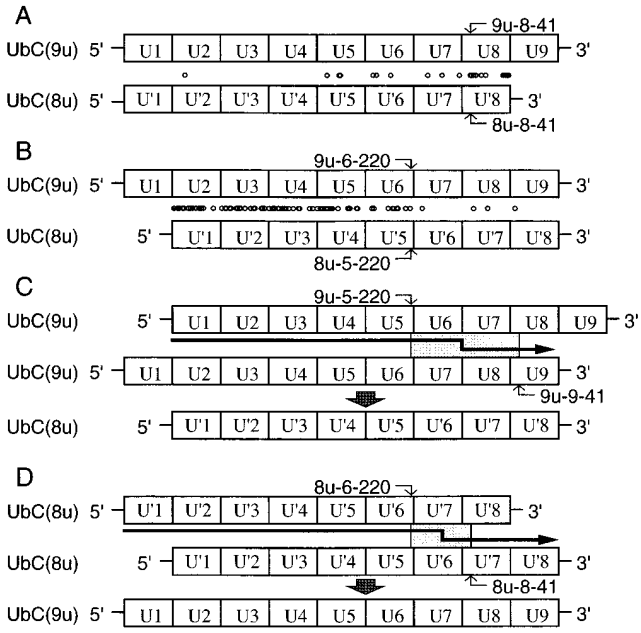

Figure 3.—Comparison of the nucleotide sequences between the *UbC(8u)* and the *UbC(9u)*, and plausible mechanisms for unequal crossing over. Ubiquitin-coding units are aligned one by one from the 5′ end (A) and from the 3′ end (B). Sites of nucleotide discrepancies are marked by circles. (C) Illustration showing a possible mechanism for the generation of the *UbC(8u)* allele by an unequal crossing over between a pair of *UbC(9u)* alleles. The possible event is considered to have occurred at a site in the hatched region. (D) Illustration showing a possible mechanism for the generation of the *UbC(9u)* allele by an unequal crossing over between a pair of *UbC(8u)* alleles.

highly homologous with that of the human polyubiquitin gene *UbB* (147 matches out of 165 bp when gaps were introduced) and that a pair of inverted repeats is conserved at a different location and with a different nucleotide sequence from those of the *UbC* and the *CHUB2* (Nenoi *et al.* 1992). Based on these facts, it is quite reasonable to assume that both the relationship between the *UbC* and the *CHUB2* and between the *UbB* and the *CHUB1* are orthologous. This assumption is further supported by the data in Table 2, which shows that the sequence differences between *UbC* and *CHUB2* and those between *UbB* and *CHUB1* are evidently smaller than those between *UbC* and *CHUB1* as well as those between *UbB* and *CHUB2*.

Consequently, the evidence for concerted evolution can be directly shown by indicating that the mean of the sequence difference within the *UbC(9u)* gene and that within the *CHUB2(11u)* gene {[(0.362 ± 0.192) + (0.022 ± 0.018)]/2 = 0.192 ± 0.096 per site, Table 1, A and B} is significantly less than the sequence difference between these genes (0.602 ± 0.057 per site, Table 2). We used the alleles of the *UbC(9u)* and *CHUB2(11u)* for the present analysis because these alleles are present in individuals, and are thought to be least affected by artificial procedures such as cell culturing. This is the first direct evidence for concerted evolution deduced from a comparison between a pair of orthologous polyubiquitin genes in mammals. In addition, it is evident that concerted evolutionary events have been much more frequent in the *CHUB2* gene than in the *UbC* gene because the sequence difference within the *CHUB2(11u)* gene (0.022 ± 0.018 per site) is much smaller than that within the *UbC(9u)* gene (0.362 ± 0.192 per site), in spite of a higher rate of synonymous substitutions in rodents than in man (Wu and Li 1985; Kikuno *et al.* 1985).

The sequence differences within the human *UbB* and the Chinese hamster *CHUB1* were also estimated (Table 3). Again, evidence of concerted evolution was apparent {[(0.187 ± 0.100) + (0.067 ± 0.035)]/2 = (0.127 ± 0.053) < (0.472 ± 0.061)}, and it appeared that there was a higher frequency of concerted evolutionary events in the *CHUB1* gene than in the *UbB* gene [(0.067 ± 0.035) < (0.187 ± 0.100)]. A similarly low level of sequence difference is observed (K_s = 0.035 ± 0.042 per site) also within a rat polyubiquitin gene (Hayashi *et al.* 1994), suggesting this is a common feature in rodents.

There are three possible reasons for such high sequence similarity within the Chinese hamster polyubiquitin genes, and possibly in other rodents as well. First, there may be a strong evolutionary constraint on the nucleotide sequence of the polyubiquitin genes in rodents, and the polyubiquitin gene alleles harboring a nucleotide substitution may have been removed from the population. However it is unlikely that such constraints are imposed only on the genes in rodents because the polyubiquitin gene is commonly conserved in all eukaryotes. In fact, the synonymous sequence differences between the *UbC* gene and the *CHUB2* gene (0.602 ± 0.057 per site) demonstrate that both genes have evolved with a comparable rate of synonymous substitution to that of other genes. Second, the concerted evolutionary events may have occurred very recently in the polyubiquitin genes in rodents. However the observed sequence similarity involving all units of the *CHUB2* has obviously required a series of events. It is very unlikely that these multiple events occurred together at once. Third, the frequency of the concerted evolutionary event in the polyubiquitin genes may have been higher especially in rodents than in other mammals. We tested this possibility for the *CHUB2* gene with the numerical simulation shown below.

Numerical simulation of the concerted evolution: Generally, concerted evolution in tandemly repeated genes is believed to occur by means of one of two mechanisms: frequent unequal crossing over or gene conversion (Darnell *et al.* 1990). Sharp and Li (1987) have suggested the possible involvement of gene conversion to explain a high sequence similarity between the second and the eighth ubiquitin-coding unit in the previously reported *UbC* gene sequence (Wiborg *et al.* 1985). However, again, their analysis was based on se-

TABLE 1
Sequence difference K_s between ubiquitin coding units within the human *UbC(9u)* gene (A) and the Chinese hamster *CHUB2(11u)* gene (B)

(A)

		UbC(9u)							
		2	3	4	5	6	7	8	9
UbC(9u)	1	0.357	0.219	0.357	0.300	0.270	0.244	0.218	0.397
	2		0.395	0.059	0.428	0.534	0.494	0.458	0.833
	3			0.428	0.149	0.303	0.332	0.303	0.403
	4				0.464	0.574	0.534	0.494	0.893
	5					0.275	0.195	0.171	0.632
	6						0.059	0.080	0.368
	7							0.019	0.400
	8								0.400

0.362 (0.192)[a]

(B)

		CHUB2(11u)									
		2	3	4	5	6	7	8	9	10	11
CHUB2(11u)	1	0.019	0.039	0.039	0.039	0.039	0.039	0.039	0.019	0.079	0.058
	2		0.019	0.019	0.019	0.019	0.019	0.019	0	0.059	0.039
	3			0	0	0	0	0	0.019	0.039	0.019
	4				0	0	0	0	0.019	0.039	0.019
	5					0	0	0	0.019	0.039	0.019
	6						0	0	0.019	0.039	0.019
	7							0	0.019	0.039	0.019
	8								0.019	0.039	0.019
	9									0.059	0.039
	10										0.059

0.022 (0.018)[a]

[a] Mean value. The value in parentheses is standard deviation.

quence data derived from a probable misalignment of the *Xho*I fragments. In the *UbC* gene sequence that we determined, these units are located closely to each other at the second and the fourth units (Nenoi *et al.* 1996). The difference between species in the repeat number of ubiquitin-coding units per locus (Nenoi *et al.* 1994; Nenoi *et al.* 1992), and evidence of within-population polymorphism in repeat number (Baker and Board 1989; Nenoi *et al.* 1996), point to a frequent occurrence of unequal crossing over within these loci.

By analogy to the analyses of concerted evolution in multigene families carried out by Black and Gibson (1974) and Ohta (1976), it can be considered that the repeated duplication and deletion of the ubiquitin-coding units by unequal crossing over may have eventually resulted in the high sequence homology observed in the present *CHUB2* gene. We have indicated this possibility by numerically simulating K_s, assuming that unequal crossover events have occurred randomly among all units and randomly in time, and that the duplication and deletion have occurred by one ubiquitin-coding unit, as supporsed by Ohta (1976). We used K_s for the present analysis because K_s is directly related to the evolutionary distance between the species (or genes, loci) being compared, irrespective of gene type and location on the chromosome (Miyata *et al.* 1980). Therefore, it is possible to apply the equal rate of synonymous substitution to that derived from a variety of rodent genes (3.5×10^{-9} per site per year) to the case of *CHUB2* gene evolution. This can be rationalized by the observation that the sequence difference between the *CHUB2* and the *UbC* [0.602 ± 0.057 per site (Table 2)] is very close to that estimated for other genes between rodents and human (0.66 ± 0.13 per site). This close correlation implies that most of the regions over the ubiquitin-coding unit of the *UbC* and the *CHUB2* genes have not been under any strong evolutionary constraint since the divergence of man and rodents, and that synonymous substitutions have been accumulated at a rate common to that of other genes (2.0×10^{-9} per site per year for genes of mammals other than rodents, 3.5×10^{-9} per site per year for rodent genes). This is further supported by the observation of Mita *et*

TABLE 2
Sequence difference K_s between ubiquitin coding units of the Chinese hamster polyubiquitin genes (*CHUB1* and *CHUB2*) and the human polyubiquitin genes (*UbB* and *UbC*)

		UbB			*UbC(9u)*								
		1	2	3	1	2	3	4	5	6	7	8	9
CHUB1	1	0.522	0.412	0.381	0.840	0.866	0.934	1.071	1.079	1.239	1.146	1.064	1.016
	2	0.559	0.516	0.481	0.785	0.928	0.871	1.154	1.003	1.146	1.064	0.990	0.946
	3	0.485	0.446	0.412	0.685	0.808	0.759	0.997	0.871	0.990	0.923	0.861	0.823
	4	0.559	0.516	0.481	0.785	0.928	0.871	1.154	1.003	1.146	1.064	0.990	0.946
	5	0.522	0.412	0.381	0.734	0.866	0.813	1.071	0.934	1.064	0.990	0.923	0.882
				0.472 (0.061)[a]									0.956 (0.128)[a]
CHUB2 (11u)	1	0.789	0.682	0.682	0.518	0.607	0.653	0.651	0.653	0.742	0.692	0.646	0.758
	2	0.808	0.698	0.698	0.563	0.569	0.616	0.611	0.616	0.700	0.653	0.609	0.715
	3	0.753	0.652	0.652	0.524	0.530	0.574	0.569	0.574	0.653	0.609	0.567	0.667
	4	0.754	0.652	0.652	0.524	0.530	0.574	0.569	0.574	0.653	0.609	0.567	0.667
	5	0.754	0.652	0.652	0.524	0.530	0.574	0.569	0.574	0.653	0.609	0.567	0.667
	6	0.754	0.652	0.652	0.524	0.530	0.574	0.569	0.574	0.653	0.609	0.567	0.667
	7	0.754	0.652	0.652	0.524	0.530	0.574	0.569	0.574	0.653	0.609	0.567	0.667
	8	0.754	0.652	0.652	0.524	0.530	0.574	0.569	0.574	0.653	0.609	0.567	0.667
	9	0.808	0.698	0.698	0.563	0.569	0.616	0.611	0.616	0.700	0.653	0.609	0.715
	10	0.658	0.567	0.567	0.524	0.530	0.574	0.569	0.574	0.653	0.609	0.567	0.667
	11	0.754	0.652	0.652	0.563	0.492	0.616	0.530	0.616	0.700	0.653	0.609	0.715
				0.689 (0.061)[a]									0.602 (0.057)[a]

[a] Mean value for all unit combinations between each gene. The value in parentheses is the standard deviation.

al. (1991) that the codon choices among synonymous codons in polyubiquitin genes generally follows the G+C content of the overall coding regions in corresponding organisms.

TABLE 3
Sequence difference K_s between ubiquitin coding units within the human *UbB* gene (A) and the Chinese hamster *CHUB1* gene (B)

(A)

		UbB		
		2	3	
UbB	1	0.167	0.295	
	2		0.099	
				0.187 (0.100)[a]

(B)

		CHUB1			
		2	3	4	5
CHUB1	1	0.099	0.058	0.103	0.079
	2		0.038	0	0.099
	3			0.038	0.058
	4				0.099
					0.067 (0.035)[a]

[a] Mean value. The value in parentheses is standard deviation.

Our simulation code can provide an estimate of the sequence difference within a polyubiquitin gene after an arbitrary number of unequal crossover events since the divergence of man and rodents [$K_s(i,j;t = 1.2 \times 10^8)$], provided that the repeat number and the sequence differences within the ancestral polyubiquitin gene are given as the initial condition. We, however, have no available data for such an initial condition. The simulation was then carried out with the initial unit number, $N_0 = 1$, 10, and 20, and the initial sequence differences, $K_s(i,j;0) = 0$ for all i and j ($1 \leq i, j \leq N_0$). Adoption of a zero initial sequence difference implies that this simulation should give the minimum estimation for $K_s(i,j;t = 1.2 \times 10^8)$, and therefore the minimum estimation for the frequency of unequal crossover events. Independent simulations were carried out 100 times for all of the event numbers and initial conditions. Figure 4A shows the sequence differences at $t = 1.2 \times 10^8$, averaged for all unit combinations (designated as $K_s(t = 1.2 \times 10^8)$ in the figure). It is evident that $K_s(t = 1.2 \times 10^8)$ decreases with an increase in the event number N_{event}, and that $K_s(t = 1.2 \times 10^8)$ is not affected by the initial unit number N_0 when $N_{event} \geq 200$. These results suggest that, to explain the extremely low sequence differences within the present *CHUB2(11u)* gene (0.022 ± 0.018, the horizontal line in Figure 4A), unequal crossover events must have occurred at least 4000 times in the *CHUB2(11u)* gene since the divergence of human and rodents 120 MYA (3.3×10^{-5} per year). Figure 4B shows the dependency of $K_s(t = 1.2 \times$

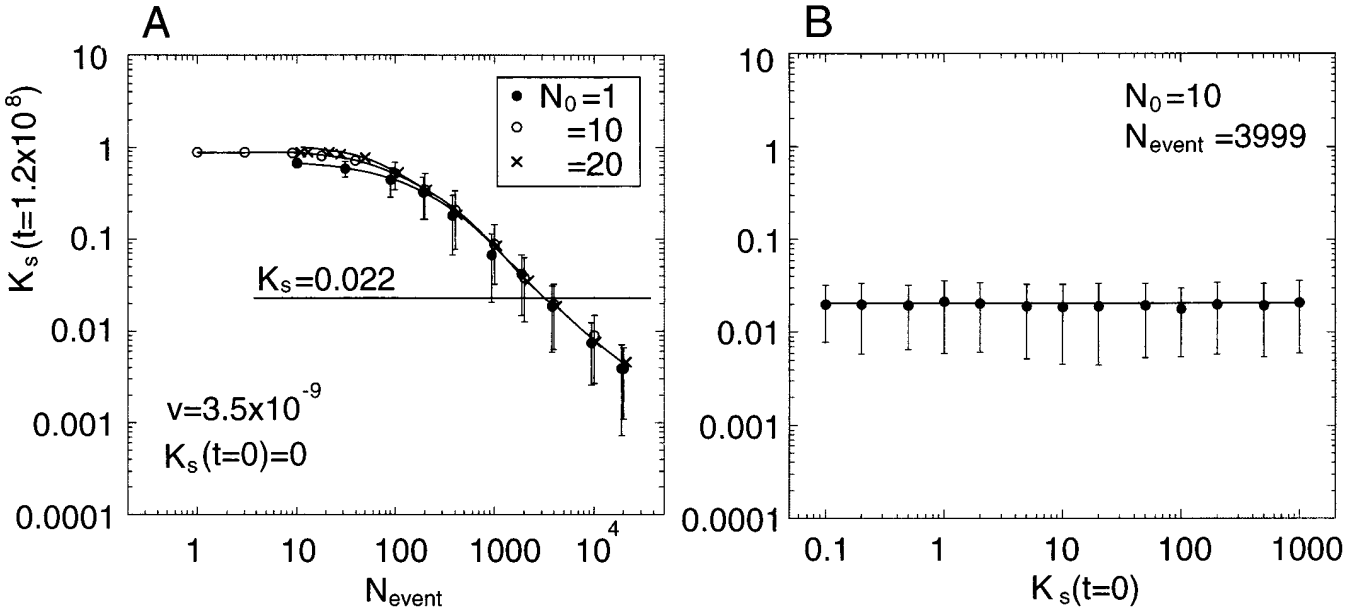

Figure 4.—Simulation for the sequence difference within the *CHUB2(11u)* gene. (A) The sequence difference $K_s(t = 1.2 \times 10^8)$ after various number of unequal crossover events N_{event} is plotted, when the rate of synonymous substitution v is assumed to be 3.5×10^{-9} per site per year, and the sequence difference within the ancestral polyubiquitin gene $K_s(t = 0)$ is postulated to be 0. The unit number of the ancestral polyubiquitin gene N_0 is postulated as in the box. The level of the sequence difference observed for the present *CHUB2(11u)* gene is indicated by a horizontal line. (B) The sequence difference after 3999 unequal crossover events is plotted, when various sequence differences are postulated for the ancestral polyubiquitin gene. Error bars represent the standard deviation.

10^8) on the initial sequence difference $K_s(i,j;t = 0)$ (a common number was postulated for all i and j ($1 \leq i, j \leq N_0$), designated as $K_s(t = 0)$ in the figure), when N_0 and N_{event} were fixed to 10 and 3999, respectively. It can be seen that $K_s(t = 1.2 \times 10^8)$ is no longer affected by $K_s(t = 0)$ after such frequent unequal crossing over as $N_{event} = 3999$. Therefore it is reasonable to conclude that the deduced value of 3.3×10^{-5} per year is not the minimum estimate, but is now the expected value for the frequency of unequal crossover event that occurred on the *CHUB2(11u)* gene.

The simulation was also carried out for the *UbC(9u)* gene, postulating $v = 2.0 \times 10^{-9}$ with the same initial conditions as those used for the *CHUB2* gene [$N_0 = 1$, 10, 20, $K_s(t = 0) = 0$]. Figure 5A shows the averaged sequence difference deduced from 100 independent simulations. In this case, $K_s(t = 1.2 \times 10^8)$ comes close to the observed sequence difference for the present *UbC(9u)* gene (0.362 ± 0.192 per site) with an event number N_{event} of no more than 60 (5.0×10^{-7} per year) even when the initial unit number N_0 was assumed to be 20. The simulation of $K_s(t = 1.2 \times 10^8)$ does not significantly depend on $K_s(t = 0)$ if $K_s(t = 0)$ is less than one (Figure 5B), which is a reasonable assumption considering that polyubiquitin genes have been subject to concerted evolution even before mammalian divergence.

The frequency of unequal crossover events was also numerically estimated for the human *UbB* gene and the Chinese hamster *CHUB1* gene (Figure 6). In this case, the simulation was carried out under the constraint that the unit number is kept between one and eight. Again, it is evident that the frequency of unequal crossing over was higher in the Chinese hamster *CHUB1* (1.7×10^{-6} per year) than in the human *UbB* (3.3×10^{-7} per year). Lower frequencies of unequal crossing over in *CHUB1* and *UbB* than those in *CHUB2* and *UbC* are considered to be due to a smaller number of ubiquitin-coding units in these genes.

Concerted evolution has been thoroughly investigated for the family of rRNA genes. In humans, the 45S pre-rRNA gene and 5S rRNA gene are present in ~250 and ~2000 copies, respectively, in tandem arrays (Darnell *et al.* 1990). Strachan *et al.* (1985) have estimated the frequency of unequal crossover events in the rRNA genes of Drosophila at 10^{-2}–10^{-4} per generation. Compared with this value, together with the consideration that our present simulation may have overestimated the frequency of unequal crossing over in the *CHUB2(11u)* gene because of neglecting the involvement of gene conversion, it can be noted that the unequal crossover events have occurred much less frequently in the *CHUB2(11u)* gene than in the rRNA genes. This agrees with the observation of Sharp and Li (1987) that the rate of concerted evolution seems to be higher in the rRNA gene family than in the ubiquitin genes of various eukaryotes although the organization of the rRNA gene array is considered to be less

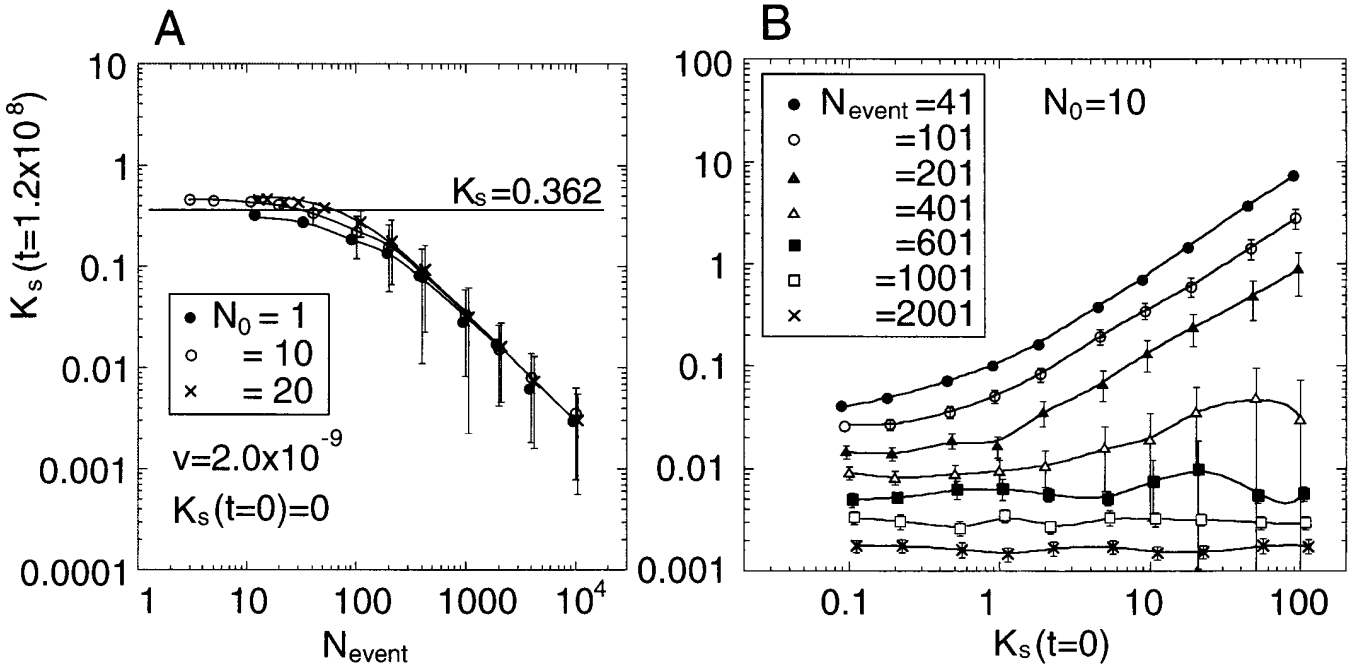

Figure 5.—Simulation for sequence differences within the *UbC(9u)* gene. (A) The sequence difference $K_s(t = 1.2 \times 10^8)$ after various numbers of unequal crossover events N_{event} is plotted, when the rate of synonymous substitution v is assumed to be 2.0×10^{-9} per site per year. The level of sequence difference observed for the present *UbC(9u)* gene is indicated by a horizontal line. (B) The dependence of $K_s(t = 1.2 \times 10^8)$ on $K_s(t = 0)$ for various numbers of unequal crossover events. Error bars represent the 0.2× standard deviation.

conducive to concerted evolution than that of a polyubiquitin locus.

A simple explanation for the different event-frequency between man and rodents is that there would be a generation time effect on the frequency of unequal crossover events. From the study of the human minisatellite loci, it has been pointed out by Jeffrey *et al.* (1988) that the large mutation events, involving the

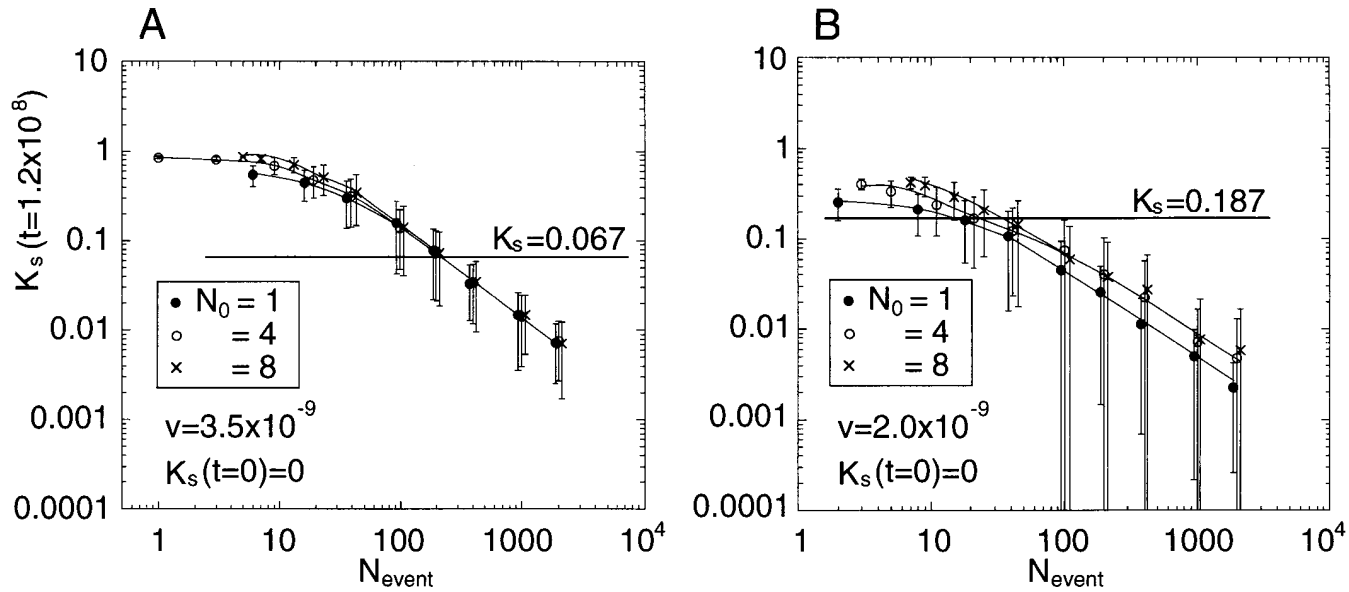

Figure 6.—Simulation for sequence differences within the Chinese hamster *CHUB1* gene (A) and the human *UbB* gene (B). The sequence difference $K_s(t = 1.2 \times 10^8)$ after various numbers of unequal crossover events N_{event} is plotted, when the rate of synonymous substitution v is assumed to be 3.5×10^{-9} per site per year for *CHUB1* and 2.0×10^{-9} per site per year for *UbB*. The level of sequence difference observed for the present polyubiquitin gene is indicated by a horizontal line. Error bars represent the standard deviation.

gain or loss of up to about 2 kb, appear to arise by means of recombinational processes at meiosis. As the generation time of rodents is considered to be 100 times shorter than that of human, it is reasonable to speculate that rodents have been more susceptible to unequal crossing over during meiosis since the divergence of these species.

The authors thank Dr. Naruya Saitou of National Institute of Genetics of Japan, for his helpful suggestions and comments on this study.

LITERATURE CITED

Baker, R. T., and P. G. Board, 1987 The human ubiquitin gene family: structure of a gene and pseudogenes from the Ub B subfamily. Nucleic Acids Res. **15:** 443–463.

Baker, R. T., and P. G. Board, 1989 Unequal crossover generates variation in ubiquitin coding unit number at the human *UbC* polyubiquitin locus. Am. J. Hum. Genet. **44:** 534–542.

Baker, R. T., and P. G. Board, 1991 The human ubiquitin-52 amino acid fusion protein gene shares several structural features with mammalian ribosomal protein genes. Nucleic Acids Res. **19:** 1035–1040.

Black, J. A., and D. Gibson, 1974 Neutral evolution and immunoglobulin diversity. Nature **250:** 327–328.

Board, P. G., M. Coggan, R. T. Baker, J. Vuust and G. C. Webb, 1992 Localization of the human *UBC* polyubiquitin gene to chromosome band 12q24.3. Genomics **12:** 639–642.

Bregman, D. B., R. Halaban, A. J. van Gool, K. A. Henning, E. C. Friedberg *et al.*, 1996 UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. Proc. Natl. Acad. Sci. USA **93:** 11586–11590.

Chen, Z. J., L. Parent and T. Maniatis, 1996 Site-specific phosphorylation of IκBα by a novel ubiquitination-dependent protein kinase activity. Cell **84:** 853–862.

Ciechanover, A., and A. L. Schwartz, 1994 The ubiquitin-mediated proteolytic pathway: mechanisms of recognition of the proteolytic substrate and involvement in the degradation of native cellular proteins. FASEB J. **8:** 182–191.

Darnell, J., H. Lodish and D. Baltimore, 1990 Eukaryotic chromosomes and genes: molecular anatomy, pp. 347–388 in *Molecular Cell Biology.* W. H. Freeman and Company, New York.

Dover, G., 1982 Molecular drive: a cohesive mode of species evolution. Nature **299:** 111–117.

Easteal, S., C. Collet and D. Betty, 1995 *The Mammalian Molecular Clock.* Springer-Verlag, Heidelberg.

Einspanier, R., H. S. Sharma and K. H. Scheit, 1987 Cloning and sequence analysis of a cDNA encoding poly-ubiquitin in human ovarian granulosa cells. Biochem. Biophys. Res. Commun. **147:** 581–587.

Hayashi, T., M. Noga and M. Matsuda, 1994 Nucleotide sequence and expression of the rat polyubiquitin mRNA. Biochim. Biophys. Acta **1218:** 232–234.

Hicke, L., and H. Riezman, 1996 Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. Cell **84:** 277–287.

Hochstrasser, M., 1996 Ubiquitin-dependent protein degradation. Annu. Rev. Genet. **30:** 405–439.

Jeffreys, A. J., N. J. Royle, V. Wilson and Z. Wong, 1988 Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature **332:** 278–281.

Keeling, P. J., and W. F. Doolittle, 1995 Concerted evolution in protists: recent homogenization of a polyubiquitin gene in *Trichomonas vaginalis.* J. Mol. Evol. **41:** 556–562.

Kikuno, R., H. Hayashida and T. Miyata, 1985 Rapid rate of rodent evolution. Proc. Jpn. Acad. **61(B):** 153–156.

Kimura, M., and T. Ohta, 1972 On the stochastic model for estimation of mutational distance between homologous proteins. J. Mol. Evol. **2:** 87–90.

Mita, K., S. Ichimura and M. Nenoi, 1991 Essential factors determining codon usage in ubiquitin genes. J. Mol. Evol. **33:** 216–225.

Miyata, T., H. Hayashida, K. Kuma and T. Yasunaga, 1987 Male-driven molecular evolution demonstrated by different rates of silent substitutions between autosome- and sex chromosome-linked genes. Proc. Jpn. Acad. **63(B):** 327–331.

Miyata, T., and T. Yasunaga, 1980 Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J. Mol. Evol. **16:** 23–36.

Miyata, T., T. Yasunaga and T. Nishida, 1980 Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA **77:** 7328–7332.

Nakamura, Y., M. Carlson, K. Krapcho, M. Kanamori and R. White, 1988 New approach for isolation of VNTR markers. Am. J. Hum. Genet. **43:** 854–859.

Nenoi, M., K. Mita and S. Ichimura, 1992 Evolutionarily conserved structure of the 3′ non-translated region of a Chinese hamster polyubiquitin gene. Biochim. Biophys. Acta **1130:** 247–252.

Nenoi, M., K. Mita, S. Ichimura and I. L. Cartwright, 1994 Novel structure of a Chinese hamster polyubiquitin gene. Biochim. Biophys. Acta **1204:** 271–278.

Nenoi, M., K. Mita, S. Ichimura, I. L. Cartwright, E. Takahashi *et al.* 1996 Heterogeneous structure of the polyubiquitin gene *UbC* of HeLa S3 cells. Gene **175:** 179–185.

Ohta, T., 1976 Simple model for treating evolution of multigene families. Nature **263:** 74–76.

Sharp, P. M., and W. H. Li, 1987 Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. J. Mol. Evol. **25:** 58–64.

Strachan, T., D. Webb and G. A. Dover, 1985 Transition stages of molecular drive in multiple-copy DNA families in *Drosophila.* EMBO J. **4:** 1701–1708.

Tan, Y., S. T. Bishoff and M. A. Riley, 1993 Ubiquitins revisited: further examples of within- and between-locus concerted evolution. Mol. Phylogenet. Evol. **2:** 351–360.

Vrana, P. B., and W. C. Wheeler, 1996 Molecular evolution and phylogenetic utility of the polyubiquitin locus in mammals and higher vertebrates. Mol. Phylogenet. Evol. **6:** 259–269.

Webb, G. C., R. T. Baker, K. Fagan and P. G. Board, 1990 Localization of the human *UbB* polyubiquitin gene to chromosome band 17p11.1-17p12. Am. J. Hum. Genet. **46:** 308–315.

Wiborg, O., M. S. Pedersen, A. Wind, L. E. Berglund, K. A. Marcker *et al.*, 1985 The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. EMBO J. **4:** 755–759.

Wu, C. I., and W. H. Li, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82:** 1741–1745.

Communicating editor: W.-H. Li