MOLECULAR PHYLOGENETICS AND EVOLUTION
Vol. 2, No. 4, December, pp. 351–360, 1993

Ubiquitins Revisited: Further Examples of Within- and Between-Locus Concerted Evolution

YING TAN,* STEPHEN T. BISHOFF,† AND MARGARET A. RILEY*

*Department of Biology, Yale University, New Haven, Connecticut 06511 and †Department of Biology, University of Massachusetts, Amherst, Massachusetts 01003

Received July 1, 1993; accepted October 7, 1993

Ubiquitin genes provide a model for studying the effects of concerted evolution on the evolution of a family of short repeated sequences. Previous work has demonstrated the occurrence of within-locus concerted evolution and raised the question of the effectiveness of between-locus concerted evolution for ubiquitin repeats. In this study comparative analysis of additional nucleotide sequences of ubiquitin tandem repeats provides further details of within-locus concerted evolution. Moreover, the availability of multiple polyubiquitin loci and ubiquitin fusion loci within a species makes possible the detection of between-locus concerted evolution. These data indicate that concerted evolution is an effective force for homogenizing repeats between, as well as within, loci.

INTRODUCTION

Concerted evolution has been detected in many repeat sequence families (Ohta, 1980; Dover, 1982; Zimmer *et al.*, 1980). The term concerted evolution refers to the observation, defined and illustrated in Fig. 1, that sequence similarity between member genes of a repeat sequence family from within a species is often higher than between members of the same family from two different species. Gene conversion and unequal crossing over have been proposed as likely mechanisms responsible for the higher level of within- versus between-species repeat similarity (Ohta, 1980; Dover, 1982). The action of concerted evolution implies that individual members of a gene family do not evolve independently, but rather evolve in concert (Li *et al.*, 1991a).

Ubiquitin genes provide a model for studying the effect of concerted evolution of short repeat sequence families. Ubiquitin is a 76-amino acid protein found in all eukaryotes (Callis *et al.*, 1990). In the cytoplasm, ubiquitin plays a major role in the nonlysosomal proteolysis of both abnormal proteins and normal proteins with rapid turnover (Hershko and Ciechanover, 1986). It also functions as a heat-shock protein during the stress response (Bond and Schlesinger, 1985). In the nucleus, ubiquitin is bound to histone 2A, implicating its role in the regulation of gene expression (Levinger and Varshavsky, 1982). Furthermore, ubiquitin is part of the receptor for lymphocyte homing (Siegelman *et al.*, 1986). Reflecting these multiple important functions, ubiquitin exhibits remarkable evolutionary conservation (Sharp and Li, 1987b).

The genomic organization of ubiquitin genes is unique among eukaryotic genes. Two distinct classes of genes encode ubiquitin (Fig. 2). The first class consists of polyubiquitin genes (poly-u), which contain tandem head-to-tail repeats of the 76-aa coding unit in a polyprotein format. The second class comprises ubiquitin fusion genes (uf52 or uf80), which encodes a single ubiquitin fused in-frame with a ribosomal protein of 52 or 76–80 amino acids (Swindle *et al.*, 1988). The exception is *Trypanosoma cruzi*, in which a 52-aa fusion protein coding sequence is attached to the 3′ end of the poly-u repeat (Swindle *et al.*, 1988; Kirchhoff *et al.*, 1988). The number of loci and the number of ubiquitin-coding repeats per locus vary considerably among the eukaryotes (Table 1). Within a species, repeats vary in nucleotide sequences, and all differences are synonymous, except in one species.

DNA sequence data from ubiquitin genes have provided an opportunity to examine the tempo and mode of evolution of a small repeated sequence family. Sharp and Li (1987a,b) examined 36 poly-u repeats obtained from eight species. Most species showed a lower level of divergence between repeats within a poly-u locus than between repeats compared across species. They argued that concerted evolution likely explains this observation. At that time, multiple poly-u loci had been sequenced in only one species (*Homo sapiens*). Sharp and Li (1987b) examined this set of poly-u repeats to assess the efficiency of within versus between poly-u locus concerted evolution. Their data suggested that concerted evolution is much less effective at homogenizing repeats between different loci than was observed for within-locus comparisons.

The availability of additional poly-u sequences

1055-7903/93 $5.00

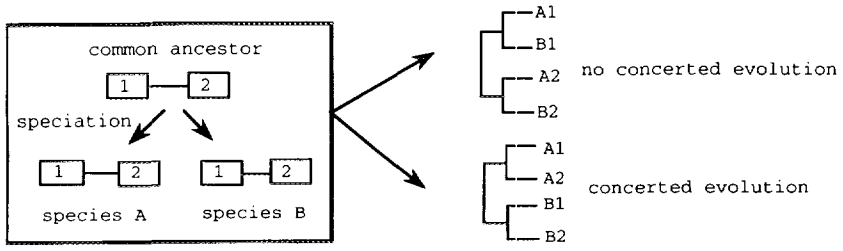

FIG. 1. Illustration of the effect of concerted evolution on phylogenetic trees. The numbered rectangles denote two members of a gene family (gene 1 and gene 2). A1 and A2 denote gene 1 and 2 in species A; B1 and B2 denote gene 1 and 2 in species B.

allows us to further explore the extent and pattern of within poly-u locus concerted evolution. Moreover, sequences of multiple poly-u loci, as well as sequences for alternative forms of ubiquitin loci, provide an opportunity to assess the importance of between-locus concerted evolution.

MATERIALS AND METHODS

Table 1 lists the ubiquitin sequences included in this study. DNA sequences were obtained from GenBank and aligned with the multiple alignment algorithm available in the DNAstar sequence analysis package (DNASTAR Inc., 1992). Pairwise distance matrices and parsimony trees were obtained using PAUP (Phylogenetic Analysis Using Parsimony, Version 3.0; Swofford, 1989).

RESULT AND DISCUSSION

Pairwise comparisons of ubiquitin DNA sequences can be made between ubiquitin repeats within a single poly-u locus, between poly-u loci, and between u-fusion loci within a species. Further, for each of these levels of comparisons, between species comparisons can be made. Table 1 provides a list of the species included in this study and the numbers and kinds of ubiquitin sequences available from the literature. Table 2 presents the average number of nucleotide differences for comparisons of ubiquitin repeats within and between poly-u loci. Table 3 gives the average number of differences for comparisons of uf52 and uf80 loci.

The most obvious feature of these data is that the average number of differences between repeats within a species is generally lower than that observed when repeats are compared across species. For example, in Table 2 the average number of differences for repeats compared within a species ranges from 1 to 57, with a mean of 23.5, while the average number of differences for repeats compared between species ranges from 15 to 76, with a mean of 48.2.

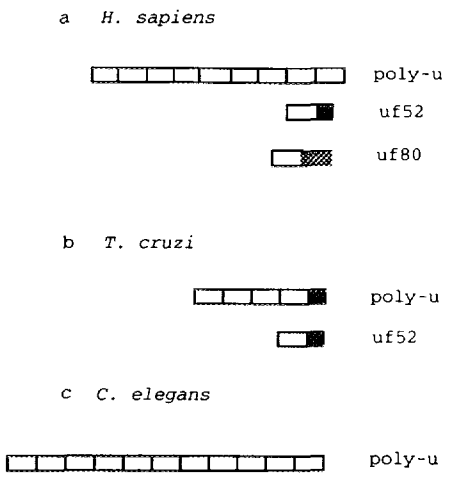

FIG. 2. Genetic organization of ubiquitin coding loci in *H. sapiens* (a), *T. cruzi* (b), and *C. elegans* (c). The open rectangles denote ubiquitin gene repeats; the hatched rectangles denote genes encoding 52aa or 80aa fusion proteins.

Concerted Evolution within Poly-u Loci

Sequences of more than one repeat for poly-u loci are available from 14 species, comprising 80 poly-u repeats in total. The results of pairwise comparisons among these sequences are presented in Table 2. For each species, the average number of nucleotide differences between repeats within a species is lower than that observed between species (Table 2). Phylogenetic trees based upon parsimony analysis were constructed separately for poly-u repeat sequences obtained from vertebrates, invertebrates, plants, and fungi (Figs. 3a, 3b, 3c, and 3d, respectively).

The higher level of similarity observed for poly-u repeat sequences from within a species versus comparisons from between species has been used as an indication of the action of concerted evolution (Sharp and Li, 1987a). Numerous examples of within species clustering are observed in the poly-u trees provided in Fig. 3. Confidence in the observed clustering has been assessed by bootstrapping (Felsenstein, 1988). This observation suggests that, for these species, ubiquitin repeats within a poly-u locus share a more recent common ancestor than do any two repeats compared between different species. This can result from (i) re-

TABLE 1

Organization of Ubiquitin Loci

Species[a]	Loci[b]	Repeats[c]	GenBank Accession No.
Homo sapiens (Hs)	3 poly-u	9 (9), 3 (3), 3 (4)	M26880, X04803, M17597
	2 uf52	1 (1), 1 (1)	X56998, X56999
	2 uf80	1 (1), 1 (1)	X63237
Mus musculus (Mm)	1 poly-u	4 (4)	X51703
Gallus gallus (Gg)	2 poly-u	4 (4), 2 (3)	M14693, M11100
	1 uf52	0 (1)	X13493
Xenopus laevis (Xl)	>2 poly-u	2 (3)	M11512
Drosophila melanogaster (Dm)	2 poly-u	3 (3)	M22428
	1 uf52	1 (1)	X53059
	1 uf80	1 (1)	M22536
Manduca sexta (Ms)	1 poly-u	2 (>2)	N/A[d]
	1 uf80	1 (1)	X53524
Caenorhabditis elegans (Ce)	1 poly-u	11 (11)	M23433
Trypanosoma brucei (Tb)	1 poly-u	1 (>1)	X14554
	1 uf52	1 (1)	X54641
Trypanosoma cruzi (Tc)	>2 poly-u	4 (4), 1 (>1)	J03945, X07451
	>2 uf52	1 (1), 1 (1)	X07451, X07452
Hordeum vulgare (Hv)	>1 poly-u	2 (3)	X04133
	2 uf80	1 (1), 1 (1)	M60176, M34707
Helianthus annuus (Ha)	3 poly-u	6 (6), 4 (4), 1 (>1)	X57005, X57003, X14333
Arabidopsis thaliana (At)	1 poly-u	5 (5)	X12853
	2 uf52	1 (1), 1 (1)	J05507, J05508
	2 uf80	1 (1), 1 (1)	J05539, J05540
Dictyostelium discoideum (Dd)	2 poly-u	5 (5), 0 (3)	M19666
	1 uf52	1 (1)	X07210
	1 uf80	1 (1)	M23750
Saccharomyces cerevisiae (Sc)	1 poly-u	5 (5)	X05731
	2 uf52	1 (1), 1 (1)	X05728, X05729
	1 uf80	1 (1)	X05730
Neurospora crassa (Nc)	1 poly-u	4 (4)	X13140
	1 uf 80	1 (1)	X15338
Chlamydomonas reinhardtii (Cr)	2 uf 52	1 (1), 1 (1)	X60826, X15427

[a] Abbreviation for each species is given in parentheses.
[b] The number of ubiquitin loci for each species is given. Poly-u denotes poly-ubiquitin locus; uf52 and uf80 denote ubiquitin fusion loci with 52 aa and 80 aa fusion proteins, respectively.
[c] The number of fully sequenced repeats per locus is given. The value in parentheses indicates the number of repeats in the entire locus.
[d] From personal communications with S. Bishoff.

cent duplication of a repeat, perhaps through a mechanism such as unequal crossing over, or (ii) gene conversion homogenizing repeats predating the divergence of the species in question.

In *H. sapiens*, one poly-u locus (*Hs*pC) has been added to the data base since the review by Sharp and Li (1987b). As shown in Fig. 3a, the repeats of the *Hs*pC locus cluster with those of the *Hs*pA locus, indicating the action of concerted evolution. Sequence comparisons between repeats of *Hs*pC reveal several features of concerted evolution within this locus. As shown in Fig. 4, *Hs*pC repeats 1, 2, and 3 bear a much higher sequence similarity to each other than to repeat 4. A single synonymous substitution distinguishes repeats 2 and 3 of *Hs*pC while 15 synonymous substitutions separate repeats 3 and 4. Given that the average rate of synonymous substitution in mammals, calibrated from 36 mammalian genes, is 4.61×10^{-9} substitutions per site per year (Li *et al.*, 1991b) and the total number of synonymous sites in a poly-u repeat is 82, we can obtain a rough estimate that the time (T) since the *Hs*pC 2 and 3 repeats most recently shared a common ancestor is 2.6 Myr. Similarly, we can estimate the time since repeat 4 shared a common ancestor with repeat 3 is 42 Myr. Closer examination of the distribution of synonymous differences between repeat 4 and other repeats reveals that most of the differences are clustered at the 3′ end of repeat 4 (Fig. 4), indicating that no exchange has occurred between the 3′ region of repeat 4 and other repeats over a long evolutionary period. A similar pattern was also observed in the human major locus (*Hs*pA), in which the 3′ region of repeat 9 has accumulated differences that distinguish it from all other repeats (Sharp and Li, 1987b). In both cases, the last repeats within *H. sapiens* poly-u loci seem isolated from exchange with other repeats.

TABLE 2

Nucleotide Differences among Poly-u Repeats

Species[a]	Poly-u[b]	*W/n* mean[c]	*W/n* range[d]	*B/n* mean[e]	*B/n* range[f]
Hs	9,3,3	18.4 ± 9.5	1 to 27	44.0 ± 11.0	15–67
Mm	4	6.7 ± 3.1	3 to 11	42.0 ± 13.0	15–71
Gg	4,2	11.3 ± 5.0	5 to 17	42.0 ± 12.0	22–69
Xl	2	11.0 ± 0.0	11	45.0 ± 11.0	22–64
Dm	3	12.7 ± 6.5	6 to 19	45.0 ± 7.1	33–64
Ms	2	40.0 ± 0.0	40	48.0 ± 6.1	33–63
Ce	11	32.2 ± 7.7	10 to 44	49.0 ± 4.8	42–64
Tc	4,1	2.0 ± 0.9	1 to 3	49.0 ± 9.9	38–76
Hv	2	17.0 ± 0.0	17	45.0 ± 9.0	31–74
Ha	6,4,1	40.0 ± 9.4	28 to 52	50.0 ± 6.3	33–72
At	5	42.3 ± 5.0	36 to 49	49.0 ± 5.6	33–64
Dd	5	20.0 ± 5.1	20 to 35	60.0 ± 6.4	44–76
Sc	5	34.0 ± 6.2	22 to 45	56.0 ± 4.6	44–68
Nc	4	41.5 ± 8.0	34 to 57	51.0 ± 7.0	34–72

[a] Abbreviations of species names indicated in Table 1.

[b] Number of fully sequenced repeats per poly-u locus for each species. For the species in which sequences of multiple poly-u loci are available, the numbers of repeats for each locus are denoted, separated by commas.

[c] The average number and standard deviation of sequence differences between the repeats within each species.

[d] The minimal and maximal numbers of sequence differences among the repeats within a species.

[e] The average number of sequence differences between the repeats of a species and those of all the other species.

[f] The minimal and maximal numbers of differences between the repeats of a species and those of all the other species.

TABLE 3

Nucleotide Differences among uf52 and uf80 Loci

Species[a]	Uf52[b]	*W/n*[c]	*B/n* mean[d]	*B/n* range[e]
Hs	2	1	45.0 ± 9.4	34–59
Tc	2	3	48.0 ± 14.0	37–75
At	2	22	52.0 ± 4.3	47–62
Sc	2	15	58.0 ± 5.6	41–65
Cr	2	1	42.0 ± 11.0	27–60
Species[a]	Uf80[b]	*W/n*[c]	*B/n* mean[d]	*B/n* range[e]
Hs	2	24	50.0 ± 5.3	41–59
Hv	2	9	45.0 ± 13.0	25–67
At	2	22	47.0 ± 8.0	36–59

[a] Abbreviations of species names indicated in Table 1.

[b] Number of uf52 or uf80 loci for each species.

[c] The number of sequences differences between ubiquitin portions of different uf52 or uf80 loci within one species.

[d] The mean and standard deviation of sequences differences between uf52 or uf80 loci of a species and those of all the other species.

[e] The minimal and maximal numbers of differences between uf52 or uf80 loci of a species and those of all the other species.

In another mammal, *Mus musculus,* all repeats within a poly-u locus cluster in the phylogenetic tree given in Fig. 3a. However, in contrast to the case seen in *H. sapiens,* the four repeats have apparently exchanged information at about the same rate. The average number of synonymous substitutions between the four repeats is 6.7 and the time since any two randomly chosen repeats shared a most recent common ancestor is estimated to be on the order of 17.6 Myr.

A more complicated pattern of within poly-u concerted evolution is observed in *C. elegans,* which has the largest poly-u locus sequenced so far. As shown in Fig. 3b, the 11 repeats of *C. elegans* cluster on the phylogenetic tree. Though the average number of differences among these 11 repeats is relatively large (Table 2), we find several examples that likely represent within-locus concerted evolution. Pairwise comparisons of the *C. elegans* repeats reveal highly variable levels of sequence similarity along the length of the locus. For example, repeats 5 and 6 and repeats 8 and 9 show lower levels of sequence divergence than more separated repeat comparisons involving any of these four repeats. This suggests that within-locus concerted evolution may occur more frequently for neighboring repeats. However, comparisons within the *Gallus gallus* poly-u locus display a higher degree of divergence between certain neighbor repeats than between repeats separated by one or more repeats.

Closer examination of the poly-u locus in *C. elegans* also reveals that the 3′ repeat is more isolated than other repeats, differing from others on average by 40.5 nucleotides. This observation suggests a pattern consistent with that observed for *H. sapiens.* It appears that the position of a repeat within a locus can affect the frequency with which a repeat is involved in information exchange with other repeats in the same locus.

Concerted Evolution between Poly-u Loci

The availability of several examples of multiple poly-u loci within species allows us to explore the extent of between-locus relative to within-locus concerted evolution. Based on comparisons of repeats from the *Hsp*A and *Hsp*B loci of *H. sapiens,* Sharp and Li (1987b) suggest that concerted evolution is much less effective in homogenizing repeats between poly-u loci. Pairwise comparisons were made for all the *Hsp*A and *Hsp*C repeats. The average number of differences is 13.1, which is about the same as that within each locus (14.0 and 11.0). Consequently, repeats from *Hsp*A and *Hsp*C loci cluster in Fig. 3a. This suggests that concerted evolution occurs between poly-u loci at least to the same degree as it occurs within the poly-u locus in *H. sapiens.*

Figure 5 illustrates the pattern of concerted evolution between the two *H. sapiens* poly-u loci. A striking observation is the near identity between the first three repeats of the *Hsp*C locus and the repeats in the central region of the *Hsp*A locus. The known sequence of *Hsp*C repeat 1 (*Hsp*C1) is identical to that of *Hsp*A repeat 4 (*Hsp*A4) while repeats *Hsp*C3 and *Hsp*A6 are identical. Only one difference exists between repeats *Hsp*C2 and *Hsp*A5. This observation argues that these repeats undergo concerted evolution between-locus as frequently as within-locus.

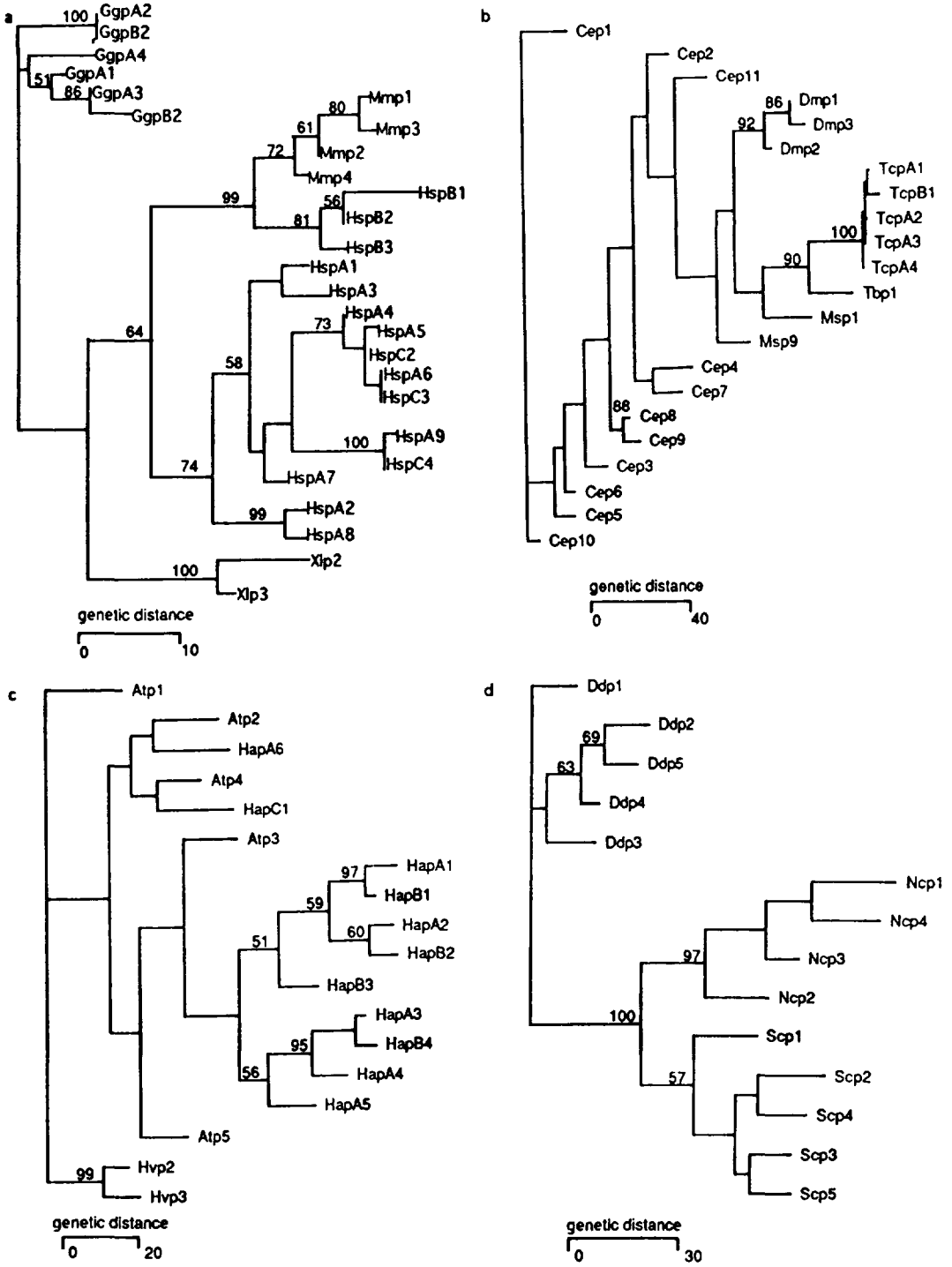

FIG. 3. Maximum parsimony trees of ubiquitin repeats from vertebrates (a), invertebrates (b), plants (c), and fungi (d). Bootstrap values for 100 replicates are shown above internal branches.

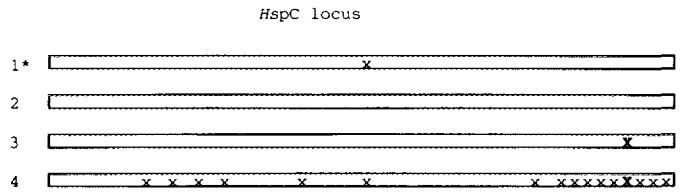

FIG. 4. Sequence comparisons among repeats from the *Hsp*C locus of *H. sapiens*. The rectangles denote ubiquitin repeats. The repeats are identical with the exception of unique (x) and nonunique (X) substitutions indicated within each rectangle. *Only 118 nucleotides of the 3′ end of this repeat have been sequenced.

The 3′ repeats of *Hsp*A and *Hsp*C loci in *H. sapiens* are nearly identical. A single substitution distinguishes between them. As discussed above, these two repeats are involved in within-locus concerted evolution at a much lower rate than other repeats. Thus, the data suggest that poly-u repeats can undergo concerted evolution at a higher frequency between loci than within a locus, a result unanticipated by previous work (Sharp and Li, 1987a,b).

Multiple poly-u loci are also available from *G. gallus*. The average number of differences between the repeats from *Ggp*A and *Ggp*B loci is 11.3 while the average numbers within *Ggp*A and *Ggp*B loci are 11 and 17, respectively. Consequently, repeats from these two loci are clustered in Fig. 3a. The pattern presented in Fig. 6 indicates that a greater similarity exists between repeats from different loci than between repeats within a locus. In particular, while repeats 2 and 3 of the *Ggp*B locus differ by 17 nucleotides, they are nearly identical or identical to repeats 3 and 2 of the *Ggp*A locus, respectively. These data provide a second example that concerted evolution occurs at a similar or even higher frequency between loci than within a locus.

Helianthus annuus is our final example of a species where multiple poly-u loci sequences are available. However, unlike *H. sapiens* and *G. gallus*, *H. annuus* does not display concerted evolution within any of its poly-u loci (Fig. 3c). Considering the relatively large number of differences between repeats within *Hap*A and *Hap*B loci and the average number of differences

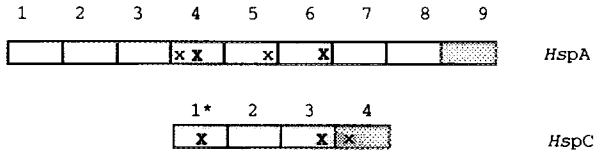

FIG. 5. The pattern of concerted evolution observed between *Hsp*A and *Hsp*C loci in *H. sapiens*. Similar shading is used to denote sequences that are identical, with the exception of unique substitutions (x) and nonunique substitutions (X) indicated within each rectangle. Repeats apparently not involved in between-locus concerted evolution are unshaded. *Only 118 nucelotides of the 3′ end of this repeat have been sequenced.

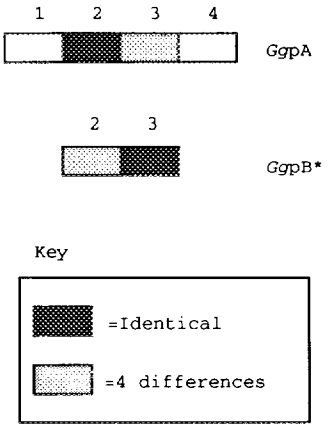

FIG. 6. The pattern of concerted evolution observed between *Ggp*A and *Ggp*B loci in *G. gallus*. Similar shading is used to denote repeats that share high DNA sequence similarity. Repeats apparently not involved in between-locus concerted evolution are unshaded. *Repeat 1 of *Ggp*B locus has not been fully sequenced.

between all the polyubiquitin repeats, *H. annuus* was initially classified as a species in which concerted evolution either does not occur or occurs at a much lower frequency than is indicated in other species. However, a more detailed examination of the *Hap*A and *Hap*B loci reveals several clear examples of between-locus concerted evolution. As shown in Fig. 7, a higher sequence similarity exists between several pairs of repeats from different loci than is observed in within-locus comparisons. For instance, repeats *Hap*A3 and *Hap*B4 differ by 9 substitutions while the smallest number of differences within these loci is 28. Consequently, each pair of these repeats from different loci cluster in the phylogenetic trees (Fig. 3c). Thus *H. annuus* provides the first example in which concerted evolution is observed between but not within a locus.

Concerted Evolution between Ubiquitin Fusion Loci

The most notable increase in ubiquitin sequence data compared to those available from the study of Sharp and Li (1987a,b) has been in the number of ubiquitin fusion loci sequences available. Fourteen uf52 and eleven uf80 sequences are currently available from the literature. These sequences allow an examination of the similarity not only between ubiquitin fusion loci but between ubiquitin fusion loci and poly-u loci as well.

Without exception, when the sequences from more than one uf52 locus or uf80 locus are available for a species, the number of differences between different uf52 or uf80 loci within a species is much lower than that observed between all possible uf52 or uf80 sequence comparisons (Table 3). Consequently, uf52 or uf80 loci within a species are always clustered in phylogenetic trees (Fig. 8).

In both *H. sapiens* and *C. reinhardii*, the two uf52 loci are nearly identical, suggesting that either there

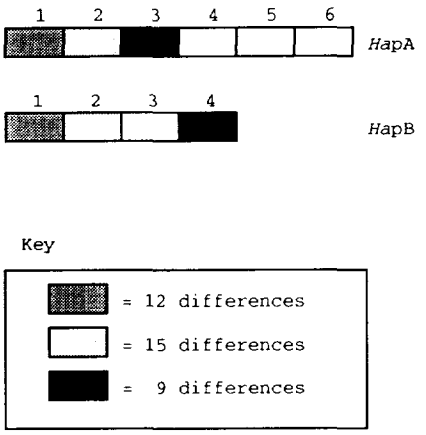

FIG. 7. The pattern of concerted evolution observed between *Hap*A and *Hap*B loci in *H. annuus*. Similar shading is used to denote repeats that share high DNA sequence similarity. Repeats apparently not involved in between-locus concerted evolution are unshaded.

has been a recent duplication producing the second copy in each species or that concerted evolution occurs frequently between these two loci. However, 24 differences exist between the two uf80 loci of *H. sapiens*, suggesting a much lower rate of concerted evolution, or a more ancient duplication. Information on chromosomal distribution of these loci may help explain the dramatic differences in the frequency of concerted evolution for these two loci.

The remaining species in which multiple ubiquitin fusion loci are available are *Hordeum vulgare, Saccharomyces cerevisiae,* and *Arabidopsis thaliana.* A common observation is that the number of differences between u-fusion loci is much lower than that observed within poly-u loci. In particular, for *H. vulgare* the number of differences between uf80 loci is 9, whereas the number within a poly-u locus is 17; for *S. cerevisiae* and *A. thaliana* the numbers are 15 versus 34.5 and 22 versus 42.9, respectively. This may suggest a much higher rate of concerted evolution between ubiquitin fusion loci than within poly-u loci. Alternatively, at least in *S. cerevisiae,* this observation may reflect a much stronger codon usage bias in u-fusion loci than in poly-u loci.

Concerted Evolution between Poly-u and u-Fusion Loci

Unlike the cases of between poly-u loci and between u-fusion loci in which a high frequency of concerted evolution is indicated in all the species where multiple loci are available, only 3 out of 9 species (excluding *T. cruzi*) display concerted evolution between poly-u and u-fusion loci. As shown in Fig. 9, for *T. brucei, S. cerevisiae,* and *H. vulgare* repeats from uf52, uf80, and poly-u loci are clustered, relative to sequences from an outgroup. For example, in *H. vulgare,* the average

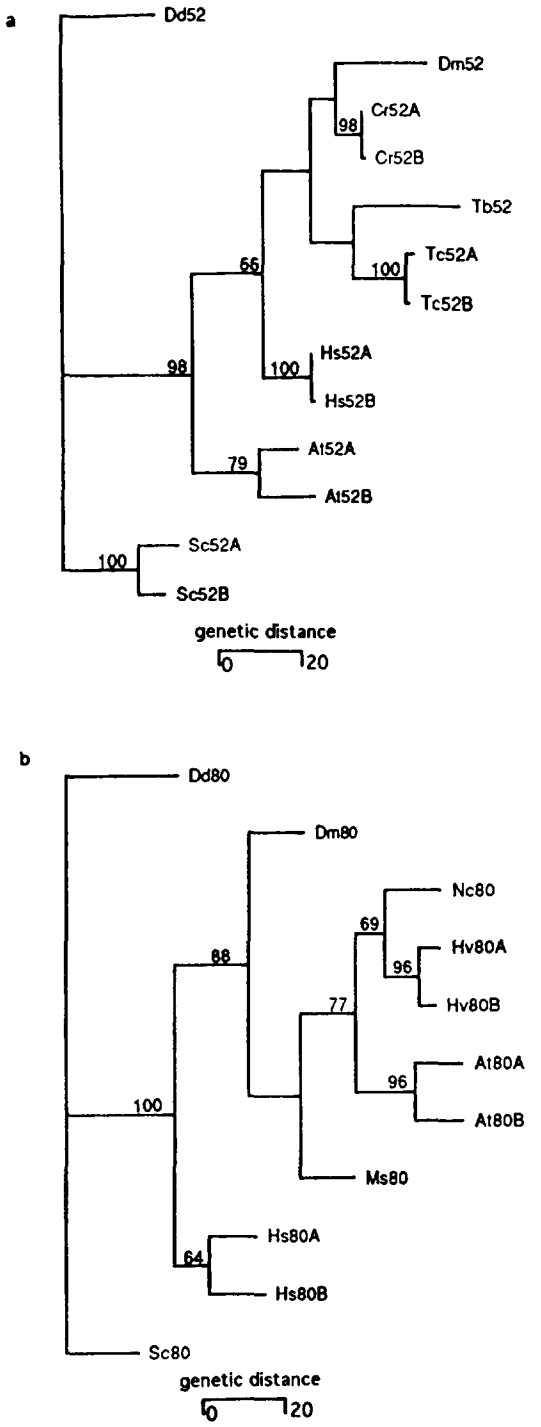

FIG. 8. Maximum parsimony trees of ubiquitin repeats of uf52 loci (a) and uf80 loci (b). Bootstrap values for 100 replicates are shown above internal branches.

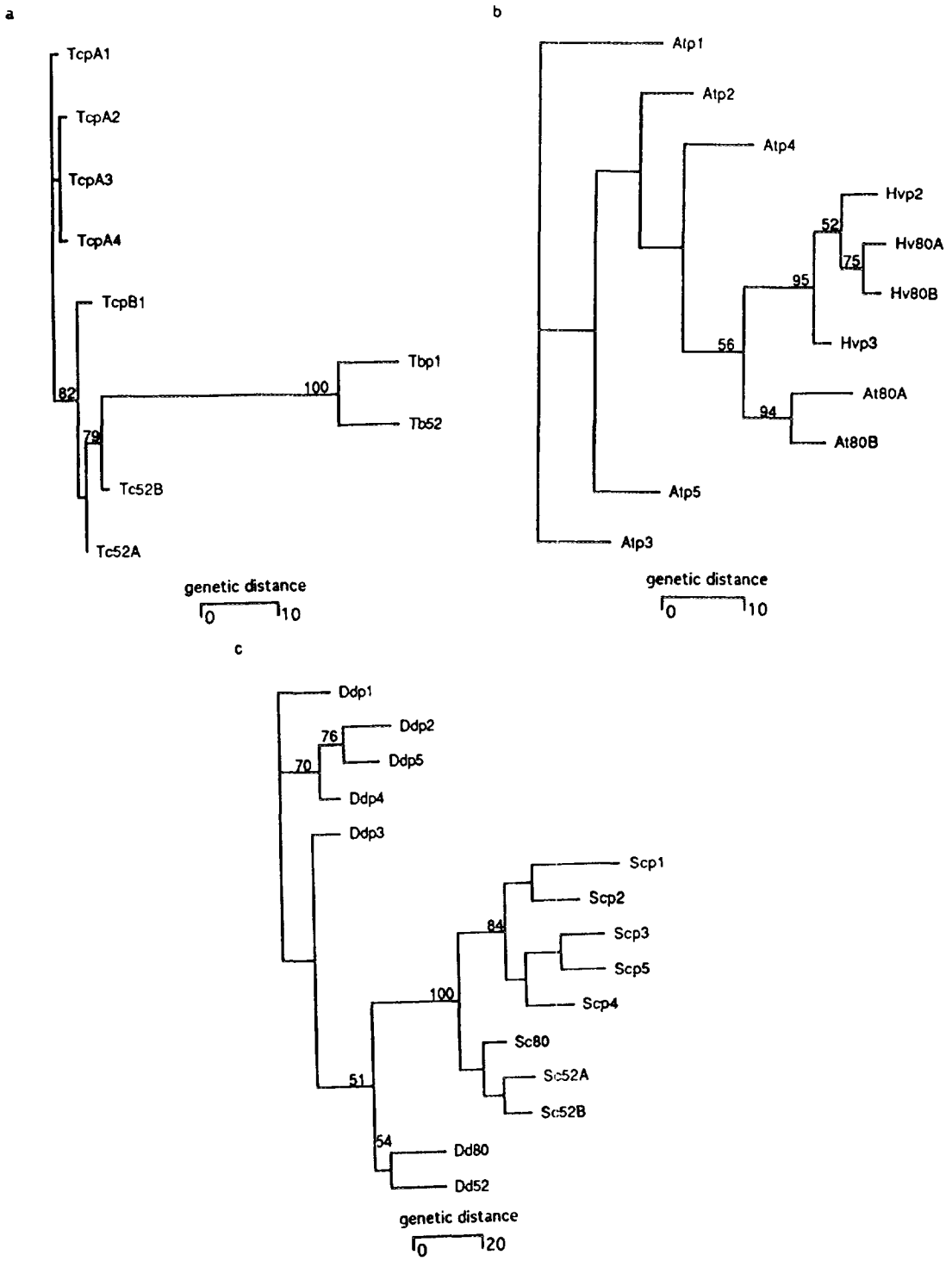

FIG. 9. Maximum parsimony trees of ubiquitin repeats of poly-u, uf52, and uf80 loci from closely related species. Bootstrap values for 100 replicates are shown above internal branches.

number of differences between uf80 and the poly-u locus is 16 while the average within poly-u locus divergence is 17. This indicates that information-exchange occurs at nearly the same rate between uf80 and the poly-u loci as occurs within the poly-u locus.

Trypanosoma cruzi, a Special Example

Because of the particular genomic organization of its ubiquitin genes, *T. cruzi* is considered a special case. in *T. cruzi* all the known ubiquitin genes are followed by a 52-aa fusion protein coding sequence (Swindle *et al.*, 1988; Kirchhoff *et al.*, 1988). Sequence comparisons illustrate remarkable similarity within a poly-u locus, between uf52 loci and between uf52 and poly-u loci (Fig. 10). The four contiguous repeats of the *Tc*pA locus are nearly identical, which suggests that these repeats exchange information frequently either by unequal crossing over and/or by gene conversion. Comparisons across uf52 loci and poly-u loci reveal an average of 4.8 substitutions while the average between uf52 loci is 4.0. These data indicate that communication between different loci occurs as frequently as within a locus. Thus all the ubiquitin repeats of *T. cruzi* seem to evolve in concert. Consequently, they all cluster in the phylogenetic tree (Fig. 9a).

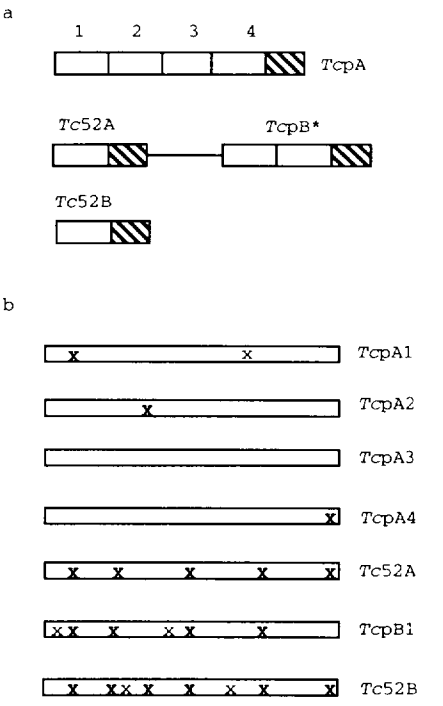

FIG. 10. Genetic organization (a) and pattern of concerted evolution observed among ubiquitin genes in *T. cruzi* (b). The open rectangles denote ubiquitin gene repeats; the hatched rectangles denote genes encoding 52aa fusion proteins; the line between *Tc*52A and *Tc*pB loci denotes the distance of about 300 bp separating these two loci (a). The rectangles denote the ubiquitin repeats, which are identical with the exception of unique (x) and nonunique substitutions (X) indicated within each rectangle (b). *Only repeat 1 of this locus has been fully sequenced.

CONCLUSION

Previous studies on the evolution of ubiquitin genes suggest that concerted evolutionary events occur rather infrequently, even within poly-u loci (Sharp and Li, 1987a,b). The present study suggests that this observation does not hold. Among the 16 species investigated in this study, representing animals, plants, and fungi, only *Manduca sexta* does not show concerted evolution of ubiquitin genes. All other species examined here provide examples of within- and/or between-locus concerted evolution. Further, we cannot conclude that concerted evolution is absent in *M. sexta*, as only two of nine repeats in this locus have been examined.

Our study suggests that not all ubiquitin repeats within a species experience the same levels of concerted evolution. For *H. sapiens*, *C. elegans*, and *G. gallus*, within poly-u locus concerted evolution seems to occur at a higher rate for some repeats than others. Some repeats in *H. sapiens*, *G. gallus*, and *H. annuus* experience a higher rate of concerted evolution between poly-u loci than they do within poly-u loci. Finally, for *H. vulgare*, *S. cerevisiae*, and *A. thaliana*, concerted evolution appears to occur at a higher rate between u-fusion loci than within poly-u loci. All these examples demonstrate that in many species the degree of concerted evolution experienced by the ubiquitin repeats is quite different.

A previous study, based on the comparisons of ubiquitin repeats from two poly-u loci of *H. sapiens*, suggested that concerted evolution is much less effective between loci. With more data for multiple poly-u and u-fusion loci, our study indicates that concerted evolution is an effective force for homogenizing repeats both between and within locus. For all species where multiple poly-u or u-fusion loci are available, concerted evolution occurs between poly-u loci or between u-fusion loci at least as often as it occurs within poly-u loci. Moreoever, in several species, concerted evolution is found to occur between poly-u and u-fusion loci at about the same rate as it occurs within poly-u loci.

To explain the different levels of concerted evolution experienced by repeats within a species, Sharp and Li (1987b) suggested that quite different chromosomal locations may contribute to the appreciable divergence between locus *Hsp*A and *Hsp*B. If chromosomal position affects the frequency of concerted evolution, then the particularly high similarity between *Hsp*A and *Hsp*C may reflect a closer chromosomal location. However, the chromosomal locations are not known for any of the ubiquitin loci. A previous investigation has shown that in *T. cruzi* more than 100 ubiquitin genes are clustered into a single 27-kb segment of the genome (Swindle *et al.*, 1988). This fact may help in understanding the above observation that exchange

among all the repeats *T. cruzi* occur at a very high rate.

While concerted evolution within and between poly-u loci may be due to unequal crossing over and/or gene conversion we argue that the mechanism responsible for between u-fusion and poly-u loci and between u-fusion loci is more likely gene conversion. Unequal crossing over events involving two u-fusion loci would produce intermediates of tandem repeats followed by a fusion protein sequence and a single fusion protein sequence, which may be selectively disadvantageous. This hypothesis is supported by the observation that genes in a poly-u-fusion format have never been found in any of the species investigated, with the exception of *T. cruzi*. Additional information about the chromosomal distribution of ubiquitin loci is needed to further explore the molecular processes creating the patterns of concerted evolution observed.

ACKNOWLEDGMENTS

We thank J. Wernegreen, M. Feldgarden, R. Wagner, and P. Sharp for comments on the manuscript. We also thank Frederick M. Cohan. This work was supported by NIH Grant GM 47471-01 to M.A.R. and GM 40458 to Larry Schwartz.

REFERENCES

Bond, U., and Schlesinger, M. (1985). Ubiquitin is a heat shock protein on chicken embryo fibroblasts. *Mol. Cell. Biol.* **5:** 949–956.

Burke, T. J., Callis, J., and Vierstra, R. D. (1988). Characterization of a polyubiquitin gene from *Arabidopsis thaliana*. *Mol. Gen. Genet.* **213:** 435–443.

Callis, J., Raasch, J. A., and Vierstra, R. D. (1990). Ubiquitin extension proteins of *Arabidopsis thaliana*. *J. Biol. Chem.* **265:** 12486–12493.

DNAstar (1992). Distributed by DNASTAR, Inc., Madison, WI.

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

Felsenstein, J. (1988). Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.* **22:** 521–565.

Hershko, A., and Ciechanover, A. (1986). The ubiquitin pathway for degradation of intracellular proteins. *Prog. Nucleic Acid Res. Mol. Biol.* **33:** 19–56.

Kirchhoff, L. V., Kim, K. S., Engman, D. M., and Donelson, J. E. (1988). Ubiquitin genes in Trypanosomatidae. *J. Biol. Chem.* **263:** 12698–12704.

Levinger, L., and Varshavsky, A. (1982). Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the *Drosophila* genome. *Cell* **28:** 375–385.

Li, W-H., and Graur, D. (1991a). Concerted evolution of multigene families. *In* "Fundamentals of Molecular Evolution," pp. 162–169, Sinaure Associates, Inc., Sunderland, Massachusetts.

Li, W-H., and Graur, D. (1991b). Rates and patterns of nucleotide substitution. *In* "Fundamentals of Molecular Evolution," pp. 69–70, Sinaure Associates, Inc., Sunderland, Massachusetts.

Ohta, T. (1980). Evolution and variation of multigene families (Lecture notes in Biomathematics 37). Springer-Verlag, Berlin.

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

Sharp, P. M. and Li, W-H. (1987b). Molecular evolution of ubiquitin genes. *Trends Ecol. Evol.* **2:** 328–332.

Siegelman, M., Bond, M. W., Gallatin, W. M., St. John, T., Smith, H. T., Fried, V. A., and Weissman, I. L. (1986). Cell surface molecule associated with lymphocyte homing is a ubiquitinated branched-chain glycoprotein. *Science* **231:** 823–829.

Swindle J., Ajioka, J., Eisen, H., Sanwal, B., Jacquemot, C., Browder, Z., and Buck, G. (1988). The genomic organization and transcription of the ubiquitin genes of *Trypanosoma cruzi*. *EMBO J.* **7:** 1121–1127.

Swofford, D. L. (1989). PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0, Illinois Natural History Survey. Champaign, Illinois.

Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W., and Wilson, A. C. (1980). Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. *Proc. Natl. Acad. Sci. USA* **77:** 2158–2162.