J Mol Evol (2000) 51:161–165
DOI: 10.1007/s002390010076

JOURNAL OF MOLECULAR EVOLUTION

Interspecific Comparison in the Frequency of Concerted Evolution at the Polyubiquitin Gene Locus

Mitsuru Nenoi,[1] Sachiko Ichimura,[1] Kazuei Mita[2]

[1] Division of Biology and Oncology, National Institute of Radiological Sciences, 9-1, Anagawa-4-chome, Inage-ku, Chiba-shi 263-8555 Japan
[2] Genome Research Group, National Institute of Radiological Sciences, 9-1, Anagawa-4-chome, Inage-ku, Chiba-shi 263-8555 Japan

Received: 18 January 2000 / Accepted: 26 April 2000

Abstract. The polyubiquitin gene, encoding tandemly repeated multiple ubiquitins, constitutes a uniquitin gene subfamily. It has been demonstrated that polyubiquitin genes are subject to concerted evolution; namely, the individual ubiquitin coding units contained within a polyubiquitin gene are more similar to one another than they are to the ubiquitin coding units in the orthologous gene from other species. However there has been no comprehensive study on the concerted evolution of polyubiquitin genes in a wide range of species, because the relationships (orthologous or paralogous) among multiple polyubiquitin genes from different species have not been extensively analyzed yet. In this report, we present the results of analyzing the nucleotide sequence of polyubiquitin genes of mammals, available in the DDBJ/EMBL/GenBank nucleotide sequence databases, in which we found that there are two groups of polyubiquitin genes in an orthologous relationship. Based on this result, we analyzed the concerted evolution of the polyubiquitin gene in various species and compared the frequency of concerted evolutionary events interspecifically by taking into consideration that the rate of synonymous substitution at the polyubiquitin gene locus may vary depending on species. We found that the concerted evolutionary events in polyubiquitin genes have been more frequent in rats and Chinese hamsters than those in humans, cows, and sheep. The guinea pig polyubiquitin gene was an intermediate example. The frequency of concerted evolution in the mouse gene was unexpectedly low compared to that of other rodent genes.

Key words: Polyubiquitin gene — Orthologue — Synonymous substitution — Synonymous sequence difference — Concerted evolution — Phylogenetic tree

Introduction

There are abundant repetitive DNA sequences in a eukaryotic genome, either dispersed throughout the genome or clustered in a chromosomal region. When units of a repetitive family are compared, greater sequence similarity is often observed within a species than between species, and the molecular process leading to this occurrence of higher similarity within a species in repetitive sequences is called concerted evolution. It may be that concerted evolution acts to slow genetic and functional diversity of duplicated genes, especially it can be viewed as a form of quality control in the production of components for macromolecular machines, such as rRNAs (Liao 1999).

Ubiquitin is a highly conserved small protein of 76 amino acids functioning in a wide range of cellular processes by degrading physiologically important regulatory proteins (Laney and Hochstrasser 1999). In all eukaryotes examined so far, ubiquitin is encoded by a multiple gene family. The polyubiquitin gene, encoding tandemly repeated multiple ubiquitins, constitutes a ubiquitin gene subfamily. It has been demonstrated that polyubiquitin

Correspondence to: M. Nenoi

genes are subject to concerted evolution (Sharp and Li 1987; Tan et al. 1993; Keeling and Doolittle 1995; Vrana and Wheeler 1996; Nenoi et al. 1998); namely, the individual ubiquitin coding units contained within a polyubiquitin gene are more similar to one another than they are to the ubiquitin coding units in the orthologous gene from other species. An unequal crossing over is thought to be one of major mechanisms for the concerted evolution of polyubiquitin genes. In fact, we observed a variability in the number of ubiquitin coding units (Nenoi et al. 1996, 1998).

By analyzing two orthologous polyubiquitin genes, *UbC* from humans and *CHUB2* from Chinese hamsters, it was demonstrated that the concerted evolutionary events have been much more frequent in Chinese hamsters than in humans, based on the observation that the sequence homology within the *CHUB2* gene is much higher than that within the *UbC* gene (Nenoi et al. 1998). However, there has been no comprehensive study on the frequency of the concerted evolutionary events at the polyubiquitin gene locus in a wide range of species, because the relationships (orthologous or paralogous) among multiple polyubiquitin genes from different species have not been extensively analyzed yet.

In this report, we present the results of analyzing the nucleotide sequence of polyubiquitin genes of mammals, available in the DDBJ/EMBL/GenBank nucelotide sequence databases, in which we found that there are two groups of polyubiquitin genes in an orthologous relationship. Based on this finding, we analyzed the concerted evolution of polyubiquitin genes from various species and compared the relative frequency of concerted evolutionary events among a variety of species by taking into consideration that the rate of synonymous substitution at the polyubiquitin gene locus may vary depending on species.

Materials and Methods

The nucleotide sequences of polyubiquitin genes used in this study are from the following species: *Homo sapiens* [human; U49869 (Baker and Board 1987), D63791 (Nenoi et al. 1996), M17597 (Einspanier et al. 1987b)], *Bos taurus* [cow; Z18245 (Wempe and Scheit 1993), M62428 (Meyers et al. 1991)], *Ovis aries* [sheep; AF038129 (Hein et al. unpublished)], *Cavia porcellus* [guinea pig; D83208 (Tukagoshi unpublished)], *Cricetulus griseus* [Chinese hamster; X60390 (Nenoi et al. 1992), AB003731 (Nenoi et al. 1998)], *Mus musculus* [mouse; X51703 (Finch et al. 1990), S40697 (Finch et al. 1992)], *Rattus norvegicus* [rat; D16554, D17296 (Hayashi et al. 1994)], and *Sus scrofa* [Pig; M18159 (Einspanier et al. 1987a)].

The homology analysis between every pair of ubiquitin-coding units either within a polyubiquitin gene or in two separate genes was carried out by evaluating the synonymous sequence difference, which is defined as the number of synonymous substitutions relative to the total number of synonymous sites (Miyata and Yasunaga 1980). The synonymous sequence difference was calculated as described by Miyata and Yasunaga (1980) and was corrected for multiple substitution (Kimura and Ohta 1972).

Phylogenetic trees were constructed for polyubiquitin genes in orthologous relationships using neighbor-joining (Saitou and Nei 1987). For the evolutionary distance between a pair of two polyubiquitin genes, we used the mean value of the synonymous sequence difference calculated for every pair of ubiquitin coding units of the respective genes. We followed the calculation scheme of Nei (1987).

Results and Discussion

Classification of Polyubiquitin Genes

In the case of analyzing the concerted evolution of a gene family, it is critical to distinguish the orthologous genes from the paralogous genes. In humans, there are two polyubiquitin genes, *UbB* on chromosome 17p11.1–17P12 (Baker and Board 1987; Webb et al. 1990) and *UbC* on chromosome 12q24.3 (Wiborg et al. 1985; Board et al. 1992), encoding three and nine ubiquitins, respectively. We have demonstrated that these two polyubiquitin genes are independently conserved in Chinese hamsters (Nenoi et al. 1992, 1994). As shown in Fig. 1, we found that the polyubiquitin genes retrieved from the database could be classified into two groups, A (*UbB*-type) and B (*UbC*-type), based on the sequence homology in the 3′ UTR.

It can be seen that all polyubiquitin genes of mammals investigated in this study show an obvious homology to either *UbB* or *UbC* in the region around the poly A signal (except in the sheep polyubiquitin gene, the 3′ UTR of which has not been sequenced yet), although the degree of homology differs between groups. It is also noticeable that the distance from the stop codon to the poly A signal is closely conserved within each of the groups. Furthermore, as shown in Table 1, the sequence homology in the coding region is higher within the group than that between groups. These facts strongly suggest that the polyubiquitin genes in each of the groups are in an orthologous relationship. Therefore we analyzed the concerted evolution of polyubiquitin genes in group A and group B separately.

Evidence for Concerted Evolution

We calculated the synonymous sequence difference for every pair of the ubiquitin coding units. The mean values of the synonymous sequence difference within a polyubiquitin gene are shown on the diagonal of Table 1, and those in a pair of separate genes are shown above the diagonal. Here the polyubiquitin genes whose coding regions are not fully sequenced were excluded from the table (cow [≧ 3], pig [≧ 3], mouse [≧ 1] in Fig. 1). When the synonymous sequence differences within each group of the gene type are compared, it can be seen that it varies between 0.105 and 0.472 for the *UbB*-type genes and 0.216 and 0.777 for the *UbC*-type genes, showing a higher homology within the *UbB*-type genes consistent with the observation of the homology in the 3′ UTR (Fig.

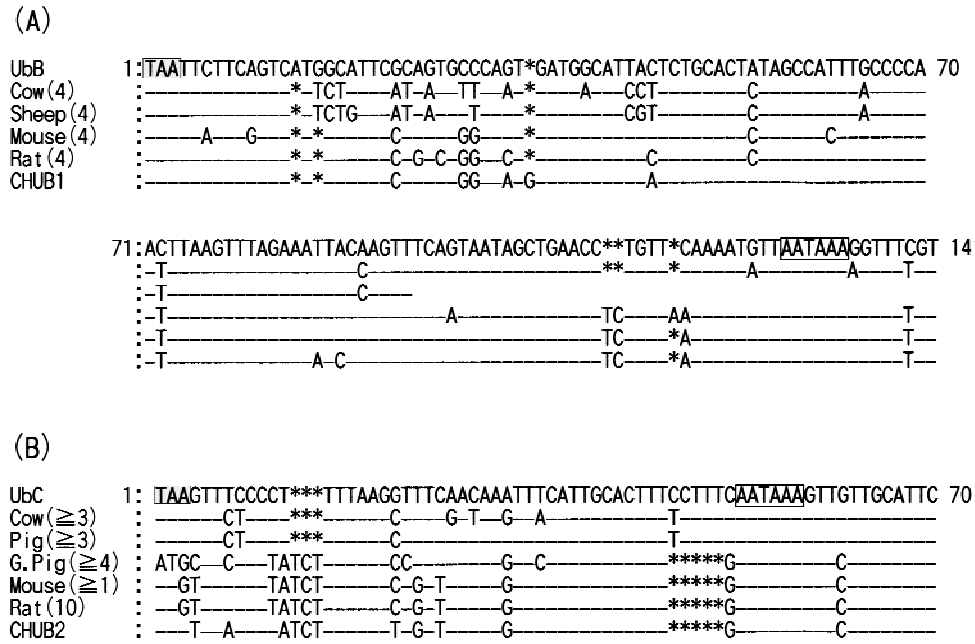

Fig. 1. Classification of polyubiquitin genes of mammals. The nucleotide sequence in the 3′ UTR is compared between various polyubiquitin genes of mammals retrieved from DDBJ/EMBL/GenBank nucleotide sequence databases. The genes homologous to the human *UbB* and *UbC* are grouped into A and B, respectively. Each gene name is described by the species name followed by the repeat number of ubiquitin coding units in brackets. *CHUB1* and *CHUB2* are the Chinese hamster genes encoding 5 and 11 ubiquitins, respectively. The genes cow (≧ 3), pig (≧ 3), and mouse (≧1) are only partially sequenced in the 3′ region. Nucleotide sequence of the human polyubiquitin gene is given in full at the top, and in the sequence of other species, the identical base is indicated by a dash. A gap is indicated by an asterisk. The termination codon is hatched, and the poly A signal is boxed.

Table 1. Synonymous sequence difference

	UbB-type						*UbC*-type			
	UbB	Cow (4)	Sheep (4)	*CHUB1*	Mouse (4)	Rat (4)	*UbC*	G. pig (4)	*CHUB2*	Rat (10)
UbB	0.187	0.304	0.325	0.472	0.434	0.456	0.969	0.689	0.689	0.636
Cow (4)	2.338	0.073	0.105	0.321	0.264	0.250	0.859	0.777	0.545	0.783
Sheep (4)	2.241	1.193	0.103	0.373	0.295	0.292	0.878	0.772	0.512	0.751
CHUB1	3.717	4.586	4.388	0.067	0.352	0.294	0.956	0.692	0.456	0.634
Mouse (4)	2.654	2.479	2.428	3.401	0.140	0.141	0.855	0.506	0.455	0.551
Rat (4)	3.663	3.704	3.539	4.558	1.396	0.062	0.930	0.563	0.442	0.600
UbC							0.362	0.677	0.602	0.777
G. pig (4)							2.989	0.091	0.489	0.439
CHUB2							3.127	8.579	0.023	0.216
Rat (10)							3.914	6.968	7.448	0.035

The mean values of the synonymous sequence difference within a polyubiquitin gene are hatched on the diagonal, and those in a pair of separate genes are above the diagonal. The ratio of $(K_{s,XX} + K_{s,YY})/2$ to $K_{s,XY}$ is boxed below the diagonal, where $K_{s,XX}$ and $K_{s,YY}$ are the synonymous sequence differences within the gene X and Y, respectively, and $K_{s,XY}$ is the synonymous sequence difference between these genes

1). However, the synonymous sequence differences between groups are significantly higher, ranging between 0.442 and 0.969, which supports the idea that the polyubiquitin genes in each of the groups are in an orthologous relationship.

The evidence for concerted evolution can be obtained from Table 1. When two polyubiquitin genes X and Y from the same group of orthologous genes are examined, one can consider that there must have been concerted evolution at least in either of the two genes X or Y, on the case that the following relation is held:

$$(K_{s,XX} + K_{s,YY})/2 < K_{s,XY}$$

where $K_{s,XX}$ and $K_{s,YY}$ are the synonymous sequence differences within the genes X and Y, respectively, and $K_{s,XY}$ is the synonymous sequence difference between these genes. The ratio of $(K_{s,XX} + K_{s,YY})/2$ to $K_{s,XY}$ is shown below the diagonal of Table 1. It can be seen that the ratio exceeds one for every pair of polyubiquitin genes in both of the orthologous gene groups, indicating that concerted evolution has occurred in all polyubiquitin genes investigated.

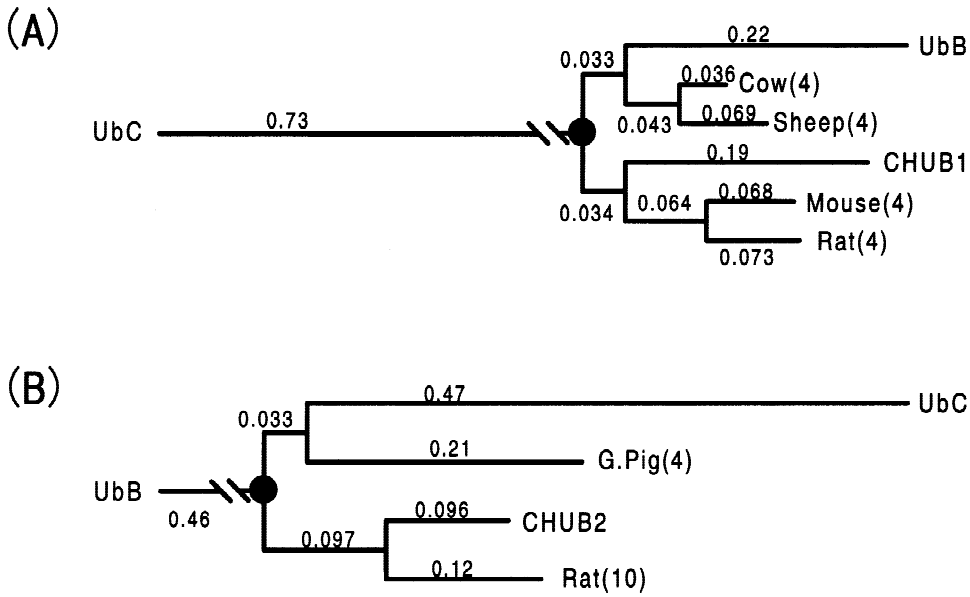

Fig. 2. Phylogenetic tree constructed for each group of polyubiquitin genes in an orthologous relationship. The human *UbC* and *UbB* genes were used as the outgroup for the group A genes (*UbB*-type) and the group B genes (*UbC*-type), respectively. The branch length is indicated for each branch in the unit of the synonymous sequence difference.

Phylogenetic Tree

Using the synonymous sequence difference, $K_{S,XY}$, as the evolutionary distance between a pair of polyubiquitin genes X and Y, we constructed a phylogenetic tree for each group of orthologous genes separately (Fig. 2). The human *UbC* and *UbB* genes were used as the outgroup for the group A genes (*UbB*-type) and the group B genes (*UbC*-type), respectively. The branch length is indicated for each branch in Fig. 2 in the unit of the synonymous sequence difference. In spite of the limited data on which the phylogenetic tree was based, the deduced trees was topologically in accordance with the detailed analysis by Kumar and Hedges (1998). It is interesting to note that the branch length was longest for human genes in both groups, implicating a relatively frequent synonymous substitution in human polyubiquitin genes. It can be seen that all species investigated in this study diverged after the time point of the man/rodent split, indicated by closed circles. We compared the concerted evolution between species in the period after the man/rodent split.

Interspecific Comparison of the Concerted Evolution

The synonymous sequence difference within a polyubiquitin gene X, $K_{s,XX}$, negatively depends on the frequency of the concerted evolutionary event at the gene X locus on one hand, and it positively depends on the synonymous substitution rate on the other. However, the contribution of the synonymous substitution rate to the synonymous sequence difference is considered to be a simple linear relation, postulating that the rate of nucleotide substitution has been constant. Therefore the net frequency of concerted evolution can be relatively measured by the ratio of the synonymous substitution rate to the synonymous sequence difference, although this ratio is not in proportion to the actual frequency.

We first estimated the branch length for every polyubiquitin gene from the time of the man/rodent split until the present from Fig. 2 (branch length in Table 2), and we then used the branch length as the parameter for the synonymous substitution rate. We calculated the ratio of the branch length to the synonymous sequence difference to measure the relative frequency of concerted evolution (Table 2).

It can be seen, in Table 2, that the value of $L_x/K_{S,XX}$ is obviously larger for rats and Chinese hamsters than for others, demonstrating a higher frequency of concerted evolution in these species. This may not be attributable to the very short generation time of rodents, because the $L_x/K_{S,XX}$ value for mice is comparable to that of other species, and that for guinea pigs is only moderately higher than that of other species. We consider that the guinea pig polyubiquitin gene may have been less conductive to concerted evolution compared to other *UbC*-type genes of rodents because of its small number of ubiquitin units encoded. The $L_x/K_{S,XX}$ value for the mouse gene was unexpectedly small compared to that of other rodents. This result suggests a peculiar stability in organization of the polyubiquitin gene in the mouse genome. In this regard, it is interesting to note an observation that sequences highly unstable in the human are often stable in the mouse (Djian 1998). For example, it has been reported that, in transgenic mice carrying ex-

Table 2. Evaluation of the parameter representing the relative frequency of concerted evolution for polyubiquitin genes of UbB-type (A) and UbC-type (B)

A

	UbB	Cow (4)	Sheep (4)	*CHUB1*	Mouse (4)	Rat (4)
Branch length (L_x)	0.25	0.11	0.15	0.22	0.17	0.17
$K_{s,\,xx}$	0.187	0.073	0.103	0.067	0.14	0.062
$L_x/K_{s,\,xx}$	1.3	1.5	1.5	3.3	1.2	2.7

B

	UbC	G. pig (4)	*CHUB2*	Rat (10)
Branch length (L_x)	0.5	0.24	0.19	0.22
$K_{s,\,xx}$	0.362	0.091	0.023	0.035
$L_x/K_{s,\,xx}$	1.4	2.6	8.3	6.3

panded repeat human androgen receptor (AR) gene, AR cDNA showed no change in repeat length with transmission, unlike that in humans (Bingham et al. 1995).

In conclusion, the concerted evolutionary events in polyubiquitin genes have been more frequent in rats and Chinese hamsters than those in humans, cows, and sheep. The guinea pig polyubiquitin gene was an intermediate example. The frequency of concerted evolution in the mouse gene was unexpectedly low compared to that of other rodent genes.

References

Baker RT, Board PG (1987) The human ubiquitin gene family: structure of a gene and pseudogenes from the Ub B subfamily. Nucleic Acids Res 15:443–463

Bingham PM, Scott MO, Wang S, McPhaul MJ, Wilson EM, Garbern JY, Merry DE, Fischbeck KH (1995) Stability of an expanded trinucleotide repeat in the androgen receptor gene in transgenic mice. Nat Genet 9:191–196

Board PG, Coggan M, Baker RT, Vuust J, Webb GC (1992) Localization of the human *UBC* polyubiquitin gene to chromosome band 12q24.3. Genomics 12:639–642

Djian P (1998) Evolution of simple repeats in DNA and their relation to human disease. Cell 94:155–160

Einspanier R, Sharma HS, Scheit KH (1987a) An mRNA encoding polyubiquitin in porcine corpus luteum: identification by cDNA cloning and sequencing. DNA 6:395–400

Einspainer R, Sharma HS, Scheit KH (1987b) Cloning and sequence analysis of a cDNA encoding poly-ubiquitin in human ovarian granulosa cells. Biochem Biophys Res Commun 147:581–587

Finch JS, Bonham K, Krieg P, Bowden GT (1990) Murine polyubiquitin mRNA sequence. Nucleic Acids Res 18:1907

Finch JS, St John T, Krieg P, Bonham K, Smith HT, Fried VA, Bowden GT (1992) Overexpression of three ubiquitin genes in mouse epidermal tumors is associated with enhanced cellular proliferation and stress. Cell Growth Diff 3:269–278

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

Keeling PJ, Doolittle WF (1995) Concerted evolution in protists: recent homogenization of a polyubiquitin gene in *Trichomonas vaginalis*. J Mol Evol 41:556–562

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

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Laney JD, Hochstrasser M (1999) Substrate targeting in the ubiquitin system. Cell 97:427–430

Liao D (1999) Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet 64:24–30

Meyers G, Tautz N, Dubovi EJ, Thiel H-J (1991) Viral cytopathogenicity correlated with integration of ubiquitin-coding sequences. Virology 180:602–616

Miyata T, Yasunaga T (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

Nei M (1987) Phylogenetic tree. In: Nei M (ed) Molecular evolutionary genetics. Columbia University Press, New York, NY, pp. 287–326

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

Nenoi M, Mita K, Ichimura S, Cartwright IL (1994) Novel structure of a Chinese hamster polyubiquitin gene. Biochim Biophys Acta 1204:271–278

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

Nenoi M, Mita K, Ichimura S, Kawano A (1998) Higher frequency of concerted evolutionary events in rodents than in man at the polyubiquitin gene VNTR locus. Genetics 148:867–876

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sharp PM, Li WH (1987) Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. J Mol Evol 25:58–64

Tan Y, Bishoff ST, Riley MA (1993) Ubiquitins revisited: further examples of within- and between-locus concerted evolution. Mol Phylogenet Evol 2:351–360

Vrana PB, Wheeler WC (1996) Molecular evolution and phylogenetic utility of the polyubiquitin locus in mammals and higher vertebrates. Mol Phylogenet Evol 6:259–269

Webb GC, Baker RT, Fagan K, Board PG (1990) Localization of the human *UbB* polyubiquitin gene to chromosome band 17p11.1–17p12. Am J Hum Genet 46:308–315

Wempe F, Scheit KH (1993) Characterization of a full-length cDNA encoding a bovine four tandem-repeat ubiquitin. Biochim Biophys Acta 1172:209–211

Wiborg O, Pedersen MS, Wind A, Berglund LE, Marcker KA, et al. (1985) The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. EMBO J 4:755–759