

# Lineage-Specific Homogenization of the Polyubiquitin Gene Among Human and Great Apes

Hiroshi Tachikui,<sup>1</sup> Naruya Saitou,<sup>2</sup> Toshiaki Nakajima,<sup>1</sup> Ikuo Hayasaka,<sup>3</sup> Takafumi Ishida,<sup>4</sup> Ituro Inoue<sup>1</sup>

<sup>1</sup> Division of Genetic Diagnosis, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan

<sup>2</sup> Division of Population Genetics, National Institute of Genetics, Mishima, Japan

<sup>3</sup> SanwaKagaku Kenkyusho Kumamoto Primate Park, Kumamoto, Japan

<sup>4</sup> Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan

Received: 28 March 2003 / Accepted: 16 July 2003

**Abstract.** Ubiquitin is a highly conserved protein, and is encoded by a multigene family among eukaryote species. The polyubiquitin genes, *UbB* and *UbC*, comprise tandem multiple ubiquitin coding units without a spacer region or intron. We determined nucleotide sequences for the *UbB* and *UbC* of human, chimpanzee, gorilla, and orangutan. The ubiquitin repeat number of *UbB* was constant (3) in human and great apes, while that of *UbC* varied: 6 to 11 for human, 10 to 12 for chimpanzee, 8 for gorilla, and 10 for orangutan. The heterogeneity of the repeat number within closely related hominoid species suggests that a lineage-specific unequal crossover and/or gene duplication occurred. A marked homogenization of *UbC* occurred in gorilla with a low level of synonymous difference ( $p_s$ ). The homogenization of *UbC* also occurred in chimpanzee and less strikingly in human. The first and last ubiquitin coding units of *UbC* were clustered independently between species, and less affected by homogenization during the hominoid evolution. Therefore, the homogenization of ubiquitin coding units is likely due to an unequal crossing-over inside the ubiquitin units. The lineage-specific homogenization of *UbC* among closely related species suggests that concerted evolution has a key role in the short-term evolution of *UbC*.

**Key words:** Ubiquitin — Polyubiquitin gene — Concerted evolution — Birth-and-death evolution

## Introduction

Ubiquitin, a 76-amino acid protein, has a very slow evolutionary rate during the eukaryote evolution (Sharp and Li 1987; Baker and Board 1991). Besides its evolutionary conservation, ubiquitin is unique in that it can covalently attach to proteins that are destined for the protein degradation pathway. Proteins marked by ubiquitin are captured by proteasomes and disassembled into amino acids. The human genome contains four ubiquitin genes, *UbA<sub>52</sub>* (Baker and Board 1991), *UbA<sub>80</sub>* (Lund et al. 1985), *UbB* (Baker and Board 1987), and *UbC* (Wiborg et al. 1985; Neno et al. 1996). *UbA<sub>52</sub>* and *UbA<sub>80</sub>* encode monomeric ubiquitin genes fused with ribosomal proteins, whereas *UbB* and *UbC* encode polyubiquitin genes. The polyubiquitin genes comprise multiple tandem ubiquitin molecules without a spacer sequence or intron, and the translated products are cleaved into monomeric ubiquitin molecules by carboxyl-terminal-specific hydroxylase (Redman and Rechsteiner 1989). Human *UbB* and *UbC* are located on 17p11-12 (Webb et al. 1990) and 12q24.3 (Board et al. 1992), respectively.

Tandem duplicated genes have a tendency to be homogenized through multiple events of interlocus

recombination and gene conversion (e.g., Kawamura et al. 1992; Kitano and Saitou 1999; Liao 1999; Gonzalez and Sylvester 2001). Coding units within polyubiquitin genes in a number of lineages are more similar to each other than to their orthologues in other species, suggesting that polyubiquitin genes are subjected to concerted evolution (Sharp and Li 1987; Tan et al. 1993; Keeling and Doolittle 1995; Vrana and Wheeler 1996; Nenoï et al. 2000). Concerted evolution is a phenomenon for multigene families when the members of a repeated sequence family are highly similar, whereas members from closely related species may differ greatly (Li 1997). Unequal cross-over is considered to be the main evolutionary mechanism for creating concerted evolution, but gene conversion may also play a role (Nei 1987).

Recently, Nei et al. (2000) observed that the coding units within polyubiquitin genes are diverged in silent substitution sites, and they proposed that the majority of polyubiquitin genes do not evolve in concert but, rather, through a birth-and-death process, in which the monomeric coding units are assumed to evolve almost independently, and protein homogeneity is attained by strong purifying selection. Comparison of ubiquitin genes for closely related species can provide a great insight to the evolution of this gene. We thus determined *UbB* and *UbC* nucleotide sequences of human, chimpanzee, gorilla, and orangutan. Nucleotide sequence similarities of ubiquitin coding units were compared among closely related species to determine whether the observed homogenization occurred by concerted evolution or birth-and-death evolution.

## Materials and Methods

### Source of Genomic DNA

Genomic DNAs of *Homo sapiens* (95 Japanese, 80 European-Americans, and 95 African-Americans), *Pan troglodytes* (12 individuals), *Gorilla gorilla* (1 individual), and *Pongo pygmaeus* (1 individual) were obtained from whole blood or cell lines. DNAs of European- and African-Americans were supplied by the Coriell Institute for Medical Research (Camden, NJ), while those of Japanese were obtained through informed consent.

### Amplification of Ubiquitin Units and Detection of Ubiquitin Repeat Number Polymorphisms

DNA fragments containing entire polyubiquitin genes were amplified by polymerase chain reaction (PCR) using LA Taq polymerase (Takara). The primer set, oligo 3 and oligo 6, designed for human *UbC* amplification, as described in Nenoï et al. (1996), was used for *UbC* amplification in great apes. For the amplification of *UbB*, a primer set was designed based on human genomic contig clone NT\_02718 (forward, 5'-AGGAAGGTTTCTTCAACTCA-AATTC-3'; reverse, 5'-TAATGACTTAAACAGCAAAGAAG-GC-3' and 5'-TGACTTAACTTGTAGCACATTACCA-3'). The product sizes of amplified *UbB* and *UbC* were measured with ag-

rose gel electrophoresis, and the numbers of ubiquitin repeats were estimated from the product size.

### Sequencing of Polyubiquitin Genes

The *UbB* sequence was determined directly using BigDye Terminator cycle sequencing on an ABI PRISM 377 DNA analyzer (Applied Biosystems). Parts of the sequences of gorilla and orangutan *UbB* could not be determined by direct sequencing, therefore TA-cloning was performed to obtain complete sequencing. The sequencing strategy for *UbC* was as follows: PCR-amplified fragments containing the entire *UbC* of various species were purified using a QIAquick PCR Purification Kit (QIAGEN) and subcloned into pGEM-T Easy TA-cloning vector (Promega). The clones containing a full-length coding sequence of *UbC* were subjected to *SaII* and *SacI* digestion and exonuclease III treatment using the Erase-a-Base System (Promega). A series of truncated clones were then constructed. The clones were sequenced from a universal primer on the vector sequence with DNA sequencer model 377 (Applied Biosystems). Sequences of the deletion clones were aligned using Sequencher (Gene Codes) and the sequence diversity was analyzed.

### Phylogenetic Tree Construction and Examination of Ubiquitin-Repeat Homogenization

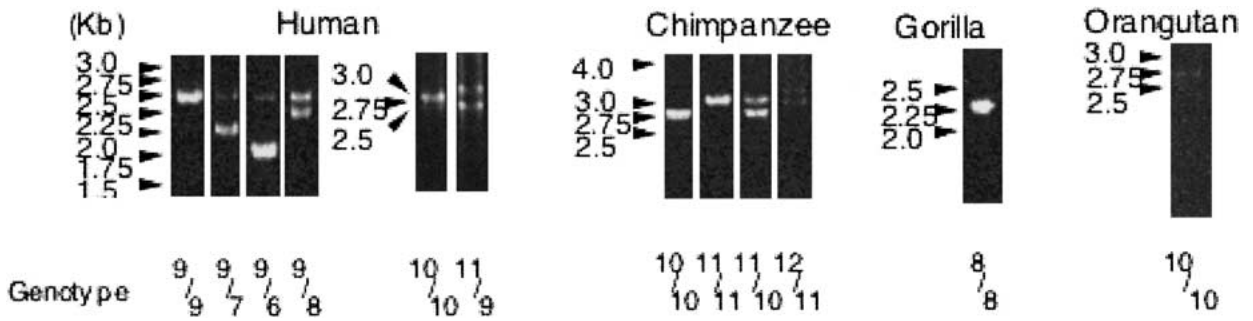
Each ubiquitin coding unit of *UbB* and *UbC* from human and great apes was separated and nucleotide diversities were calculated for both within and between species. To examine the extent of sequence divergence, the number of synonymous differences per synonymous sites ( $p_s$ ) was calculated using a modified Nei-Gojobori method (Kumar et al. 2001; Nei et al. 2000) with MEGA2 software (<http://www.megasoftware.net/>). The  $p_s$  values for all pairs of ubiquitin units were compared within and between polyubiquitin genes in all species. We obtained neighbor-joining trees (Saitou and Nei 1987) with Kimura's (1980) two-parameter model using Clustal W (Thompson et al. 1994) and visualized it using Dendromaker (<http://www.cib.nig.ac.jp/dda/timanish/dendromaker/home.html>).

## Results

### Repeat Number Polymorphisms of *UbB* and *UbC*

Two polyubiquitin genes, *UbB* and *UbC*, of chimpanzee, gorilla, and orangutan were amplified by PCR using primer sets designed for human *UbB* and *UbC*. By detecting the nucleotide similarities of the 5'-UTR regions between PCR products from various species, we confirmed that *UbB* and *UbC* were specifically amplified.

The number of ubiquitin coding units of *UbC* was estimated to be 6 to 11 for human ( $n = 270$ ), 10 to 12 for chimpanzee ( $n = 12$ ), 8 for gorilla ( $n = 1$ ), and 10 for orangutan ( $n = 1$ ) (Fig. 1 and Table 1). The repeat number 9 was predominant among three distinct human populations, African-Americans, European-Americans, and Japanese. We identified novel repeat number polymorphisms, 6 in Japanese and 10 and 11 in African-Americans. The proportion of minor alleles of *UbC* repeat number polymorphisms, i.e.,



**Fig. 1.** Analysis of ubiquitin repeat number of *UbC* among human and great apes. *UbC* of human and great apes was amplified by PCR and the PCR products were separated on 0.9% agarose gels with size standards. Six individuals for human, four for chimpanzee, one for gorilla, and one for orangutan are shown.

**Table 1.** Frequencies of repeat number polymorphisms of *UbC* among human and great apes

	<i>N</i>	Repeat allele frequency (%)						
		12	11	10	9	8	7	6
Human								
Japanese	190				94.7	2.6	1.6	1.1
European-American	160				93.8	3.8	2.5	
African-American	190		1.1	0.5	88.3	5.9	4.3	
Chimpanzee	24	4.2	29.2	66.7				
Gorilla	2					100.0		
Orangutan	2			100.0				

alleles other than repeat 9, in human were 0.053 for Japanese, 0.062 for European-Americans, and 0.117 for African-Americans, while that for Australian Caucasians is reported to be 0.127 (Baker and Board 1989). The repeat number polymorphisms of chimpanzee *UbC* showed a higher heterozygosity than those of human *UbC* (Table 1). These variations in coding unit number most likely resulted from unequal crossover of two *UbC* alleles, presumably promoted by the repetitive nature of the polyubiquitin gene. To compare sequence variations between species, we defined the representative repeat number in each species that is the most common allele: 9 for human, 10 for chimpanzee, 8 for gorilla, and 10 for orangutan.

The number of repeat of ubiquitin coding units in *UbB* was three in human and great apes (data not shown). Because of the small number of ubiquitin repeats in *UbB*, the unequal crossover might not have occurred in the great ape lineage.

#### Nucleotide Sequences Determined

The consensus sequence was defined from multiple individuals (13 humans homozygous for 9 *UbC* repeats and 5 chimpanzees homozygous for 10 *UbC* repeats) or from multiple clones (6 clones for gorilla and 5 for orangutan) to avoid PCR-induced error. Polyubiquitin gene sequences of human and great apes were deposited in the DDBJ/EMBL/GenBank

International Nucleotide Sequence Database under accession numbers AB089613–AB089616 for *UbC* and AB089617–AB089620 for *UbB*. Multiple alignments of these sequences only for variants sites are shown in Fig. 2. Because of repeat structure, each repeat was considered as a unit of comparison.

#### Phylogenetic Trees of Monomeric Ubiquitin Units

Each monomeric ubiquitin coding unit of *UbB* and *UbC* from human, chimpanzee, gorilla, and orangutan was used as unit of comparison. Because *UbB* and *UbC* are located in different chromosomes, they probably diverged long time ago. We therefore conducted a BLAST homology search (Altschul et al. 1997) using human ubiquitin B repeat unit 1 as query and obtained seven homologous nonhominoid mammalian sequences (accession numbers are AB071067 for crab-eating macaque, AF038129 for sheep, Z18245 for bovine, BC019850 for mouse, D16554 for rat, AB003731 for Chinese hamster, and AF506969 for horse). Phylogenetic trees for those sequences as well as *UbC* repeat sequences of human, chimpanzee, gorilla, and orangutans were constructed as shown in Fig. 3. Human *UbC* repeat 1 was used as outgroup. We found clustering of different mammalian orders; rodents (mouse, rat, and Chinese hamster), ungulates (sheep and bovine), primates (hominoids and crab-eating macaque). Four horse repeat sequences are so diverged that they did not

**A** 11111112222222222222  
 12244499011568901122233333  
 84725839814929543628901234

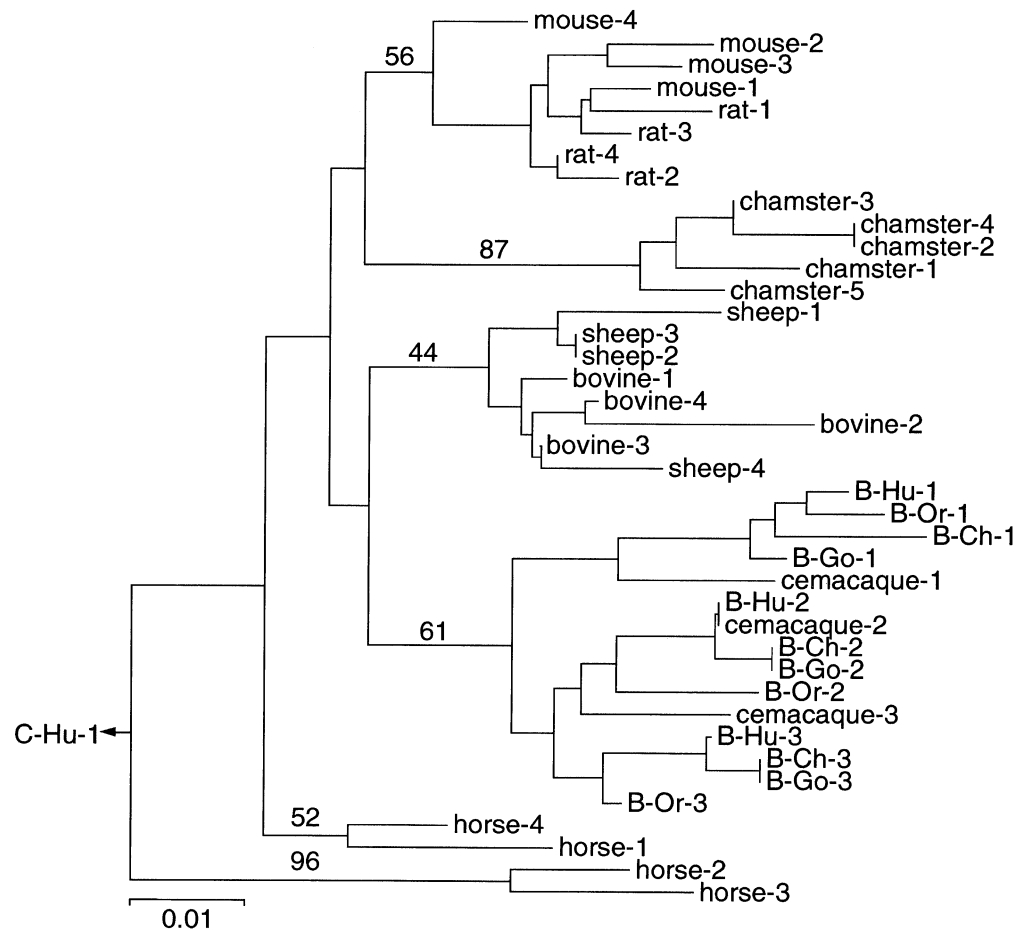
B-Hu-1 atcctgggtccctggcgat-----  
 B-Ch-1 .....ct.....t....-----  
 B-Go-1 .....c.....-----  
 B-Or-1 ..t.....a.....-----  
 B-Hu-2 gg.ga.act..c.....-----  
 B-Ch-2 ggt.ga.act..c.....-----  
 B-Go-2 ggt.ga.act..c.....-----  
 B-Or-2 gg.g.a.t.tc.a.a.....-----  
 B-Hu-3 gg.tg.aac...ca...cgctgtaa  
 B-Ch-3 gg.tg.aact..ca...cgctgtaa  
 B-Go-3 gg.tg.aact..ca...cgctgtaa  
 B-Or-3 gg.g.aac.t.ca...cgctgtaa  
 11 1 31 333

**B** 1111111111111111111122222222222222  
 122233334445566678990012223344556679990111112222333333  
 621470369258173695136287379281439284589503679258901234

C-Hu-1 gctgttgcccgtcccctgataccgcgctaagcgccccgctccatg-----  
 C-Hu-2 a.ac.c...t.c...cagc.t.t.t.c.g....t.....-----  
 C-Hu-3 .c.....g.....tt.t.c...tt...t.a.....-----  
 C-Hu-4 a.ac.c...t.c...t.ca.c.t.t.c.g....t.....-----  
 C-Hu-5 ..c.....t.agg..t..c...t.c...tt...t.....-----  
 C-Hu-6 ..c.....t.gg..t..c...t..g.a.....t...-----  
 C-Hu-7 ..c.....t.agg..t..c...t..g.a.....t...-----  
 C-Hu-8 ..c.....t.agg..t..c...t..g.a.....t...-----  
 C-Hu-9 a.c.....g.....t.t..g.a...t.tcgctgggtgtctaa  
 C-Ch-1 .....-----  
 C-Ch-2 ..ac.ca...t.c...ca.c.t.t...g....t.....-----  
 C-Ch-3 ..c.....t.....tt...t.....-----  
 C-Ch-4 a.ac.c...t.c...ca.c.t.t.c.g....t.....-----  
 C-Ch-5 ..c.....g.....tt.t.c...tt...t.....-----  
 C-Ch-6 a.ac.c...t.c...ca.c.t.t.c.g....t.....-----  
 C-Ch-7 ..c.....g.....c...t.....tt...t.....-----  
 C-Ch-8 ..c.....g.....c...t.....tt...t.....-----  
 C-Ch-9 atac.c...t.c...ca.c.t.t.c.g....t.....-----  
 C-Ch-10 ..c.....t.gg..t..c...t..g.a...t.tcgctgggtgtctaa  
 C-Go-1 .....c.....tt.....-----  
 C-Go-2 ..c.....c...ca.c...t.c.gg.....-----  
 C-Go-3 ..c.....c...ca.c...t.c.gg.....-----  
 C-Go-4 ..c.....c...ca.c...t.g.....-----  
 C-Go-5 ..c.....c...ca.c...t.g.....-----  
 C-Go-6 ..c.....t.c...c...t.c.gg.....-----  
 C-Go-7 ..c.....g...c.c...t.c.g.a...t.....-----  
 C-Go-8 ..c.c...g.....c...ta.g.a...t.tcgctgggtgtctaa  
 C-Or-1 .....g.....t.c.g.t.t.....-----  
 C-Or-2 ..c.c...g...t...c.t.t.c.g.t.tt.....-----  
 C-Or-3 ..c.....g...c.c.t.t.g.g.t.tt.....-----  
 C-Or-4 ..c.c...tt.ct.ttc.gcgt.t.c...t...tt.....-----  
 C-Or-5 ..c.c.t.t.a.t.t.agcgttat.g.g.a.c.t...t.....-----  
 C-Or-6 ..c.....t.gt.ttc.c.t.t.g.t...t.....-----  
 C-Or-7 ..ac.c.t.t.c...a.c...t.c.g.t.t.....-----  
 C-Or-8 ..c.....g...t.....t.tcg.tt.....-----  
 C-Or-9 ..ac.c...c...t.gc...t.c.g.tt.....-----  
 C-Or-10 ..c.ac...g.....c.t.t..g.a...t.tcgctagggtgtctaa

**Fig. 2.** Multiple alignments of the representative *Ubb* and *Ubc* nucleotide sequences for human and great apes. **A** *Ubb*. **B** *Ubc*. Each repeat was separated for comparison. Only variant nucleotide sites are presented, and site positions are shown above sequences.

Periods designate same nucleotide as the first sequence, while hyphen means gaps. Sites with numbers 1 and 3 shown at the bottom of alignment A are those having repeat 1- and 3-specific nucleotides, respectively.



**Fig. 3.** A phylogenetic tree for monomeric ubiquitin coding units of *UbB* among various mammalian species. Human ubiquitin C repeat 1 was used as outgroup. Bootstrap probabilities are given only for branches leading to major mammalian orders, except for

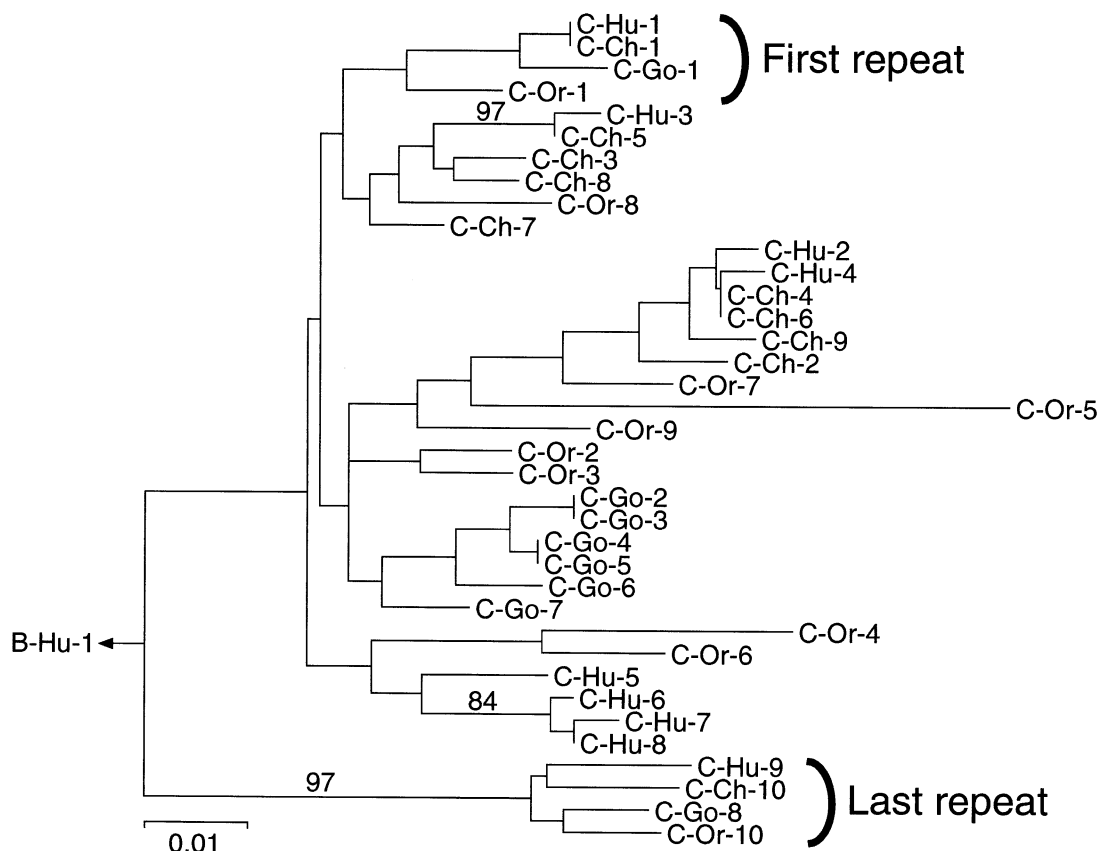
horse. B-Hu, human *UbB*; B-Ch, chimpanzee *UbB*; B-Go, gorilla *UbB*; Ub-Or, orangutan *UbB*; hamster, Chinese hamster; cemacaque, crab-eating macaque. Numbers after hyphen designate repeat number for each species.

form a monophyletic cluster. Because relatively short nucleotides (228 bp) were compared, bootstrap probabilities were not so high. In any case, it is clear that homogenization of repeats did not occur frequently, as pointed out by Nei et al. (2000).

The phylogenetic relationship of the ubiquitin B repeat 2 in primates is somewhat unusual. The human and crab-eating macaque repeat 2 are identical, while chimpanzee and gorilla repeat 2 are identical. These two pairs are clustered and orangutan repeat 2 is outside of this cluster (Fig. 3). This topology is clearly different from the species phylogeny of primates, and some kind of homogenization must have occurred during the primate evolution. It is at least very difficult to explain the sequence identity of human and macaque repeat 2 by assuming a birth-and-death process. In contrast, the phylogenetic relationship of repeats 1 and 3 is more or less concordant with the primate phylogeny, though human and orangutan were clustered in repeat 1.

We also conducted the BLAST homology search by using human ubiquitin C repeat unit 1 as query

and obtained five homologous nonhominoid mammalian sequences (accession numbers are M18159 for pig, BC025894 and BC021837 for mouse, D17296 for rat, D63782 and AB003732 for Chinese hamster, and D83208 for guinea pig). As in the case of *UbB*, all the hominoid repeat sequences became monophyletic (tree not shown). We therefore present the phylogenetic tree of hominoid ubiquitin C repeats in Fig. 4. Human *UbB* repeat 1 was used as outgroup. Ubiquitin coding units of *UbC* of the four hominoid species have complicated patterns. The last ubiquitin coding units formed a distinct cluster with bootstrap probability of 97%. The first repeat also formed a monophyletic cluster, though its bootstrap support (35%; not shown in Fig. 4) was low. In contrast, there was no monophyletic clustering for intermediate repeats. In gorilla, a high degree of homogenization of ubiquitin coding units was observed, and repeats 2–7 formed a monophyletic cluster. In chimpanzee, the homogenization of ubiquitin repeats was weaker than gorilla, however, eight intermediate repeats 2–9 belong to either one of two clusters, suggesting a certain



**Fig. 4.** A phylogenetic tree for monomeric ubiquitin coding units of *UbC* among the four hominoid species. Human ubiquitin B repeat 1 was used as outgroup. Bootstrap probabilities (%) are given only when they were higher than 80%. C-Hu, human *UbC*; C-Ch, chimpanzee *UbC*; C-Go, gorilla *UbC*; C-Or, orangutan *UbC*. Numbers after hyphen designate repeat number for each species.

homogenization. In human, partial homogenization was observed; repeats 5–8 formed one cluster, while repeats 2 and 4 are very similar. On the other hand, the homogenization of ubiquitin repeats was not extensive in orangutan; only repeats 4 and 6 made a cluster (see Fig. 4).

#### *Synonymous Differences of UbC*

We evaluated the *UbC* homogenization in gorilla and others using the following statistical analysis as described by Nei et al. (2000). The mode of evolution was tested by comparison of the proportion of non-synonymous differences ( $p_N$ ) and the proportion of synonymous differences ( $p_S$ ). The  $p_N$  of ubiquitin was zero among human and great apes, therefore only the  $p_S$  was compared. Table 2 shows intraspecific comparison of *UbC* repeats. The model of concerted evolution predicts small  $p_S$  values. Because the first and last ubiquitin coding units of *UbC* are more conserved between species and less conserved within species,  $p_S$  was also estimated after removing the first and last ubiquitin repeats (data not shown). The homogenization was remarkable when ubiquitin coding units were compared between internal units (mean  $p_S = 0.05$  in gorilla).

#### **Discussion**

Two polyubiquitin genes, *UbB* and *UbC*, were examined and each ubiquitin repeat was compared among the four hominoid species (human, chimpanzee, gorilla, and orangutan). Because the chromosomal locations of *UbB* and *UbC* are assigned to the same regions as *HOXB* (17q21.3) and *HOXC* (12q13.3), respectively (Apiou et al. 1996), the ubiquitin gene family might have arisen through genome duplication events as already shown for the *HOX* gene family (Ohno 1970; Sidow 1996; Kasahara et al. 1996; Bailey et al. 1997; Skrabanek and Wolfe 1998; Pennisi 2001). A large amount of ubiquitin appears to be advantageous, especially under stressful conditions, therefore the number of coding units per allele should gradually increase with time until a maximum is reached, which might be balanced between increased selective advantage and reduced allele stability in a species-specific manner. As shown in Table 1 and Fig. 1, the number of ubiquitin repeats in *UbC* varies even among closely related species; 9 for human, 10 for chimpanzee, 8 for gorilla, and 10 for orangutan. The repeat number varies within species, e.g., 6 to 11 for human, indicating that the repeat number itself might not be under strong selection. On

**Table 2.** Synonymous differences ( $p_S$ ) per site among human and great apes (*UbC*) using the Nei–Gojobori method

Human	1	2	3	4	5	6	7	8	Chimpanzee	1	2	3	4	5	6	7	8	9
2	0.224								2	0.194								
3	0.149	<b>0.255</b>							3	0.074	<b>0.225</b>							
4	0.224	<b>0.030</b>	<b>0.255</b>						4	0.209	<b>0.045</b>	<b>0.240</b>						
5	0.179	<b>0.285</b>	<b>0.120</b>	<b>0.285</b>					5	0.134	<b>0.240</b>	<b>0.060</b>	<b>0.225</b>					
6	0.148	<b>0.284</b>	<b>0.194</b>	<b>0.284</b>	<b>0.104</b>				6	0.209	<b>0.045</b>	<b>0.240</b>	<b>0.000</b>	<b>0.225</b>				
7	0.163	<b>0.299</b>	<b>0.209</b>	<b>0.299</b>	<b>0.090</b>	<b>0.015</b>			7	0.104	<b>0.179</b>	<b>0.060</b>	<b>0.194</b>	<b>0.090</b>	<b>0.194</b>			
8	0.148	<b>0.284</b>	<b>0.194</b>	<b>0.284</b>	<b>0.075</b>	<b>0.030</b>	<b>0.015</b>		8	0.089	<b>0.224</b>	<b>0.045</b>	<b>0.239</b>	<b>0.075</b>	<b>0.239</b>	<b>0.045</b>		
9	0.236	0.372	0.238	0.372	0.312	0.222	0.237	0.237	9	0.209	<b>0.075</b>	<b>0.240</b>	<b>0.030</b>	<b>0.225</b>	<b>0.030</b>	<b>0.224</b>	<b>0.239</b>	
									10	0.266	0.372	0.267	0.387	0.297	0.387	0.237	0.252	0.387
Gorilla	1	2	3	4	5	6	7	Orangutan	1	2	3	4	5	6	7	8	9	
2	0.148							2	0.089									
3	0.148	<b>0.000</b>						3	0.089	<b>0.060</b>								
4	0.119	<b>0.030</b>	<b>0.030</b>					4	0.253	<b>0.164</b>	<b>0.193</b>							
5	0.119	<b>0.030</b>	<b>0.030</b>	<b>0.000</b>				5	0.299	<b>0.209</b>	<b>0.239</b>	<b>0.239</b>						
6	0.134	<b>0.045</b>	<b>0.045</b>	<b>0.074</b>	<b>0.074</b>			6	0.193	<b>0.164</b>	<b>0.149</b>	<b>0.119</b>	<b>0.329</b>					
7	0.134	<b>0.074</b>	<b>0.074</b>	<b>0.074</b>	<b>0.074</b>	<b>0.089</b>		7	0.150	<b>0.135</b>	<b>0.165</b>	<b>0.209</b>	<b>0.240</b>	<b>0.269</b>				
8	0.281	0.281	0.281	0.252	0.252	0.267	0.237	8	0.104	<b>0.134</b>	<b>0.164</b>	<b>0.238</b>	<b>0.314</b>	<b>0.149</b>	<b>0.209</b>			
								9	0.135	<b>0.120</b>	<b>0.179</b>	<b>0.194</b>	<b>0.256</b>	<b>0.224</b>	<b>0.105</b>	<b>0.135</b>		
								10	0.281	0.252	0.267	0.326	0.372	0.237	0.327	0.267	0.313	

the other hand, hominoid and macaques have maintained 3 ubiquitin repeats in *UbB* (see Fig. 3).

Concerted evolution is a phenomenon of intra-specific homogenization that is observed in many repeat sequence families (Liao 1999). The high degree of intrasequence homogenization of *UbC* typically observed in gorilla (Table 2) could be due to concerted evolution, as previously reported for the Chinese hamster polyubiquitin gene (Nenoi et al. 1994, 1998). A major mechanism of concerted evolution is an unequal crossing-over of two alleles, which is a reciprocal recombination process. Another possible mechanism is gene conversion, which is a nonreciprocal transfer of genetic information between similar sequences. When we compared each ubiquitin unit of *UbC* in different species, there were high similarities in the first ubiquitin unit and the last ubiquitin unit between species (Figs. 2B and 4) and the homogenization within species occurred inside the repeat. This might be regarded as a signature of an interlocus crossing-over event between alleles. A similar tendency was also observed for the *UbB* repeats. Repeat specific nucleotides were observed only for repeats 1 and 3, not for the intermediate repeat 2 (see Fig. 2A). Therefore, we believe that unequal crossing-over has a key role in the homogenization of *UbC* among human and great apes. However, gene conversion may also play some role. Homogenization seems to occur even among the *UbC* last repeats. If we examine variant sites for *UbC* shown in Fig. 2B, the last nine consecutive sites (205–228) are mostly specific to last repeats (C-Hu-9, C-Ch-10, C-Go-8, and C-Or-10). This pattern suggests that the very end of the last repeat escaped from homogenization, while the remaining part of the last repeat experienced some kind

of homogenization. If so, this homogenization may have been caused by gene conversion, since repeat was not the unit of homogenization.

Recently, Nei et al. (2000) proposed the concept of birth-and-death evolution to explain the homogenization of multigene families, where a selective pressure has a major role in homogenization. Birth-and-death evolution is a form of evolution that assumes that new genes are created by repeated gene duplication and that some duplicate genes stay in the genome for a long time, whereas others are deleted or become nonfunctional. To distinguish concerted evolution from the birth-and-death evolution under strong purifying selection, the extent of synonymous differences within and between species was examined in this study. Intraspecific synonymous differences that are nearly as high as interspecific differences supports birth-and-death evolution. As described by Nei et al. (2000), concerted evolution can be distinguished from birth-and-death evolution by examining ubiquitin gene sequences from many pairs of closely related species. In our sequence comparison of *UbC* among closely related species, marked homogenization was observed in gorilla and the mean  $p_S$  value (0.05) was considerably low, suggesting that birth-and-death evolution is less likely. When three ubiquitin repeat pairs, [2, 3], [4, 5], and [6, 7], of chimpanzee *UbC* were combined, the mean  $p_S$  became 0.05. On the other hand, relatively high mean  $p_S$  values were observed for orangutan *UbC*, suggesting that the concerted evolution is not so strong. Therefore, concerted evolution was more responsible for homogenization of *UbC* in gorilla and chimpanzee but not so in orangutan. It suggests that lineage-specific homogenization occurred in closely related

species of hominoids. An interlocus crossing-over event could conceivably accelerate further crossing-over events so that the lineage specific interlocus crossing-over occurred at a relatively rapid pace. The lineage-specific concerted evolution as shown for *Ubc* of closely related species executes a role in the evolution of polyubiquitin genes.

*Acknowledgments.* This work was supported in part by a Research for the Future Program Grant of The Japan Society for the Promotion of Science to I.I and Grants-in-Aid for Scientific Research on priority areas from the Ministry of Education, Culture, Sport, Science, and Technology of Japan to N.S. We thank Ms. Kozue Otaka for her technical assistance.

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Apiou F, Flagiello D, Cillo CB, Malfoy B, Poupon MF, Dutrillaux B (1996) Fine mapping of human HOX gene clusters. *Cytogenet Cell Genet* 73:114–115
- Bailey WJ, Kim J, Wagner GP, Ruddle FH (1997) Phylogenetic reconstruction of vertebrate Hox cluster duplications. *Mol Biol Evol* 14:843–853
- Baker RT, Board PG (1987) The human ubiquitin gene family: Structure of a gene and pseudogenes from the *Ubb* subfamily. *Nucleic Acids Res* 15:443–463
- Baker RT, Board PG (1989) Unequal crossover generates variation in ubiquitin coding unit number at the human *Ubc* polyubiquitin locus. *Am J Hum Genet* 44:534–542
- Baker RT, Board PG (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
- 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
- Gonzalez IL, Sylvester JE (2001) Human rDNA: Evolutionary patterns within the genes and tandem arrays derived from multiple chromosomes. *Genomics* 73:255–263
- Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, Ishibashi T (1996) Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proc Natl Acad Sci USA* 93:9096–9101
- Kawamura S, Saitou N, Ueda S (1992) Concerted evolution of the primate immunoglobulin alpha-gene through gene conversion. *J Biol Chem* 267:7359–7367
- Keeling PI, Doolittle WF (1995) Concerted evolution in protists: Recent homogenization of a polyubiquitin gene in *Trichomonas vaginalis*. *J Mol Evol* 41:556–562
- Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Kitano T, Saitou N (1999) Evolution of the Rh blood group genes has experienced gene conversions and positive selection. *J Mol Evol* 49:615–626
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 17:1244–1245
- Li WH (1997) *Molecular evolution*. Sinauer Associates, Sunderland, MA
- Liao D (1999) Concerted evolution: Molecular mechanism and biological implications. *Am J Hum Genet* 64:24–30
- Lund PK, Moats-Staats BM, Simmons JG, Hoyt E, D'Ercole AJ, Martin F, Van Wyk JJ (1985) Nucleotide sequence analysis of a cDNA encoding human ubiquitin reveals that ubiquitin is synthesized as a precursor. *J Biol Chem* 260:7609–7613
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Nei M, Rogozin IB, Piontkivska H (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci USA* 97:10866–10871
- 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, Yamauchi M, Tsuji H (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
- Nenoi M, Ichimura S, Mita K (2000) Interspecific comparison in the frequency of concerted evolution at the polyubiquitin gene locus. *J Mol Evol* 51:161–165
- Ohno S (1970) *Evolution by gene duplication*. Springer-Verlag, New York
- Pennisi E (2001) Genome duplications: The stuff of evolution? *Science* 294:2458–2460
- Redman KL, Rechsteiner M (1989) Identification of the long ubiquitin extension as ribosomal protein S27a. *Nature* 338:438–440
- 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
- Skrabaneck L, Wolfe KH (1998) Eukaryote genome duplication—Where's the evidence? *Curr Opin Genet Dev* 8:694–700
- Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. *Curr Opin Genet Dev* 6:715–722
- Tan Y, Bishoff ST, Riley MA (1993) Ubiquitins revisited: Further examples of within- and between-locus concerted evolution. *Mol Phylogenet Evol* 2:260–351
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Vrana PE, 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
- Wiborg O, Pedersen MS, Wind A, Berglund LE, Marcker KA, Vuust J (1985) The human ubiquitin multigene family: Some genes contain multiple directly repeated ubiquitin coding sequences. *EMBO J* 4:755–759