ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

GENOMICS

Genomics 81 (2003) 78–84

www.elsevier.com/locate/ygeno

Evolution of the cystatin B gene: implications for the origin of its variable dodecamer tandem repeat in humans☆

Motoki Osawa,[a,*] Mika Kaneko,[a] Hidekazu Horiuchi,[a] Takashi Kitano,[b] Yoshi Kawamoto,[c] Naruya Saitou,[b] and Kazuo Umetsu[a]

[a] *Department of Forensic Medicine, Yamagata University School of Medicine, Yamagata 990-9585, Japan*
[b] *Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima 411-8540, Japan*
[c] *Laboratory Primate Center, Primate Research Institute, Kyoto University, Inuyama 484-8506, Japan*

Received 28 May 2002; accepted 21 October 2002

Abstract

The human cystatin B gene contains a variable number of 12-bp tandem repeats in its promoter region, of which the common alleles contain two or three copies and unusual expansion causes progressive myoclonus epilepsy of the Unverricht–Lundborg type. We undertook a comprehensive analysis of the genomic sequence to address the evolutionary events of this variable repeat. By examination of a contiguous genome sequence spanning 5.0 kb and linkage analysis of detected polymorphic changes, we identified six major intragenic haplotypes in unrelated Japanese subjects. The number of normal repeats was closely correlated with these alleles, indicating that changes in the array should be comparatively rare events during human evolution. To examine the origin of the repeat array further, we also analyzed five primate genomes. Repetitive polymorphism was unlikely in hominoids, and the array originated with the dodecamer itself in the course of primate evolution. The variability conceivably developed after the separation to humans.

Keywords: Cystatin B; Primates; VNTR; Molecular evolution

Introduction

Cystatin B (CSTB), also called stefin B, is a 98-amino-acid single polypeptide that is a member of the cystatin superfamily, which includes fetuin and kininogen [1]. CSTB is a cytoplasmic and nuclear component with ubiquitous expression [2]. Cystatins are tightly binding reversible inhibitors of cathepsin family cysteine proteinases, in which their interacting sites are located in the amino-terminal half of the molecule [3]. Its role is thought to be that of an inhibitor acting against proteinases leaking from lysosomes.

Progressive myoclonus epilepsy (EPM1, Unverricht–Lundborg type) is an autosomal recessive disease characterized by stimulus-sensitive myoclonic seizures and slowly progressive cerebellar ataxia [4]. Linkage analysis and disequilibrium studies mapped the gene responsible for this disease to chromosome 21q22.3 [5,6], and by Northern blot analysis Pennacchio and colleagues [7] showed reduced levels of *CSTB* mRNA in affected individuals. Missense mutations initially appeared to account for the disease, but further studies showed that the most common disorders correlate with unstable insertions of approximately 600 to 900 bp in the promoter region of *CSTB.* These insertions were found to consist of large repetitive expansions of a GC-rich dodecamer unit, the motif CCCCGCCCCGCG [8]. An extensive repeat formation of more than 30 copies remarkably reduced the transcription rate of the gene [9]. Mice lacking CSTB by means of targeted gene disruption developed ataxia and myoclonic seizures similar to the symptoms observed for EPM1 patients [10]. Pathological features of these mice displayed degradation of cerebellar

☆ Sequence data from this article have been deposited with the DDBJ/EMBL/GenBank Data Libraries under Accession Nos. AB083085 to AB083089, AB083416, and AB083417.

* Corresponding author. Fax: +81-23-628-5273.
E-mail address: mosawa@med.id.yamagata-u.ac.jp (M. Osawa).

0888-7543/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0888-7543(02)00010-1

granule cells, in which characteristic signs of apoptosis, such as condensed nuclei and fragmented DNA, were evident.

In unaffected individuals, Lalioti et al. [11] identified the dodecamer tandem repeat as a polymorphism, of which normal alleles contained two or three copies. The frequency of the triple-repeat allele was 0.66 in the normal Caucasian population. Polymorphic microsatellites consisting of simple units of less than 5 bp are widely dispersed over the mammalian genome. In contrast, minisatellite arrays, which consist of 10- to 100-bp-long elementary units, are differentiated from microsatellites by several properties such as cluster at the telomeric side [12]. Although it is classified as a minisatellite, this dodecamer repeat in *CSTB* is unique in a couple of respects. The normal repeat array is very short compared with that of minisatellites, for which the common overall size ranges from 0.5 to 30 kb [13]. Second, minisatellites usually exhibit internal variation among constituent repeat units, but the repeats in the dodecamer array appear to be uniform [11]. Since this repeat array more closely resembles pathogenic microsatellites rather than typical minisatellites, we prefer the term "variable number of tandem repeats" (VNTR) for it. One approach to investigating the significance of VNTR is to assess its evolutionary origin and sequence divergence by comparing intra- and interspecific sequence differences [14]. In the present study, we performed intragenic haplotyping through an analysis of the whole genomic sequence of *CSTB* to examine the relationship of the VNTR to human alleles. After comparing the human sequence with those of five related primates, we speculated from an evolutionary perspective how VNTR arose and developed at this locus.

Results

VNTR analysis

The variable dodecamer tandem repeat is located from 175 to 210 bp upstream of the ATG translation initiation codon. Amplification of the region using the DF1 and DR1 primers yielded products of 176 and 188 bp (Fig. 1). For heterozygotes, the smaller of the two bands became predominant, but heterozygosity was clearly evident, as judged by the appearance of heteroduplex bands in the polyacrylamide gels. The study of 189 unrelated Japanese subjects revealed complete consistency, including 15 individuals homozygous for the three-copy repeat, 73 heterozygotes, and 101 homozygotes for the two-copy repeat. Otherwise, no apparent variants were observed. Frequencies of the three- and two-copy alleles, as calculated from the genotype frequencies, were 0.27 and 0.73, respectively. This population group is statistically in Hardy–Weinberg equilibrium with regard to the copy number of the *CSTB* VNTR ($\chi^2 = 0.126$, $P < 0.05$).

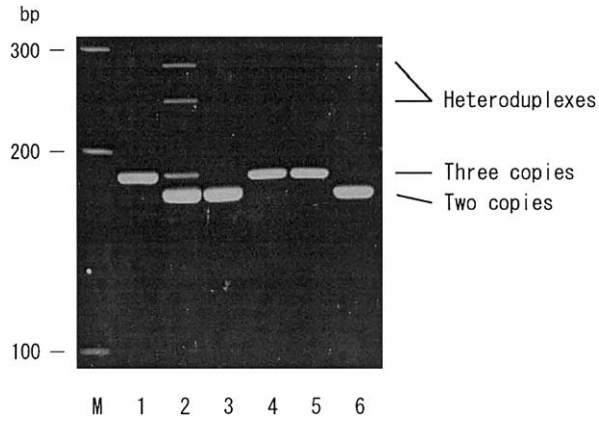

Fig. 1. Genotyping of the *CSTB* tandem dodecamer repeat by PCR using the primers DF1 and DR1. After nondenaturing electrophoresis in a 7.5% polyacrylamide gel, amplified products were stained with SYBR gold. Lanes: M, molecular markers; 1, the homozygous triple repeat in a human subject; 2, the heterozygote (human); 3, the homozygous double repeat (human); 4, the corresponding region in chimpanzee; 5, gorilla; and 6, bonobo.

Haplotype analysis

To find nucleotide changes in the human *CSTB* locus, we sequenced the gene, which spans 2.7 kb in length and contains three small exons. The analyzed sequence was extended to the 5′-flanking region because only a small number of variations were observed in the initial region. A full sequence of 5.0 kb was determined for a total of 15 subjects (30 chromosomes) chosen from individuals homozygous for either of the two VNTR types. In the target region, 16 single-nucleotide variation sites were found, none of which is located inside an open reading frame or at an intron splicing junction (Fig. 2). Differences encountered more than twice were classified as nonunique polymorphic changes, which accounted for 10 sites, including 4 transition-type substitutions and 6 transversions. The other 6 changes were regarded as unique singletons that represented a rarer single change. To type 8 of the 10 polymorphic sites, we performed PCR-RFLP on the samples of 189 subjects and examined their patterns of linkage disequilibrium (Table 1). Most of the polymorphic variations were in complete linkage disequilibrium with one another. Variability of the VNTR was found to be closely linked to these variations as well.

From the results obtained, we verified that *CSTB* could be subdivided into six distinct intragenic haplotypes and designated these alleles *CSTB*1* to **6* (Table 2). Homozygotes for *CSTB*2* to **6* alleles were involved in our samples, which assisted us in the assignment of haplotypes. The genotypes of 182 subjects could be described as combinations of the six alleles. The genotypes of the remaining seven likely consist of distinct minor alleles. When the allele frequencies were estimated, we found that the *CSTB*3* allele was predominant in the test group. A test of the Hardy–Weinberg equilibrium for the six alleles from the

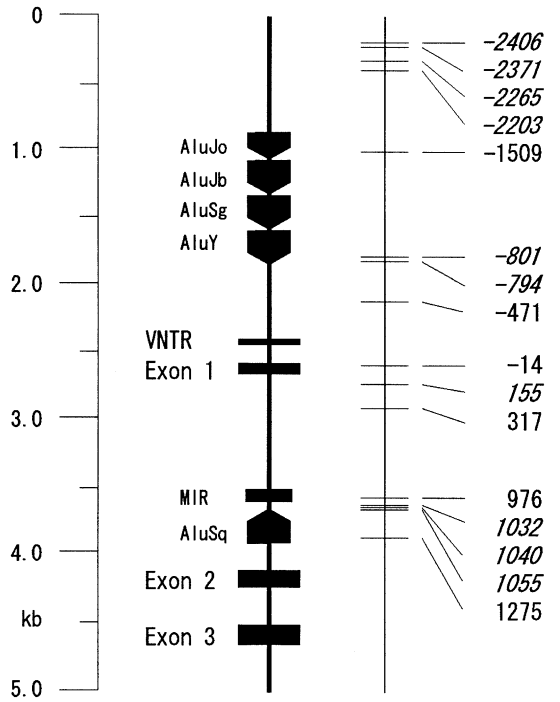

Fig. 2. Schematic diagram of the human *CSTB* and locations of identified nucleotide changes. The numbers on the right indicate nucleotide positions of variations detected in an analysis of 30 chromosomes, in which polymorphic changes are italicized. The variable positions in the 5′-flanking region are numbered by reference to the position of the ATG translation-initiating codon. The structure of human *CSTB*, which includes transposable elements, is shown in the middle.

examined samples showed that their frequencies do not significantly depart from the expected proportions ($\chi^2 = 11.98$, $P < 0.05$). The overall nucleotide diversity (π) across the entire gene was estimated to be 0.0007.

When the relationship between the *CSTB* alleles and the VNTR was examined, the number of repeat units was closely related to a given allele. *CSTB*1* to **4* alleles have only the two-copy array, and the triple repeat was observed only in *CSTB*5* and **6* alleles. In particular, more than 90% of *CSTB**6 alleles had the triple repeat. In a preliminary experiment using Caucasian samples, *CSTB*6* and its derivative alleles were predominant, which is associated with the higher frequency of the triple repeat.

The VNTR sequence in primate CSTB

To compare the human DNA sequence with those of closely related primates, we also determined the orthologous sequence for the chimpanzee, bonobo, gorilla, orangutan, and Japanese macaque genomes. As shown in Fig. 3, the VNTR sequence of the chimpanzee, gorilla, and orangutan genomes consists of an incomplete three-time repeat that includes a C to T substitution in the second repeat unit, which was identical among the hominoids. While we examined nine distinct chimpanzee subjects (18 chromosomes), the sequence was consistent, indicating that the repeat region is likely monomorphic. Also of interest is the fact that the complete two-time array was observed to be homozygous in the bonobo genome (Fig. 1, lane 6). As only one bonobo subject (two chromosomes) was available, we do not know if intraspecific variation exists. In the macaque, the corresponding region showed a more incomplete three-time array, in which the third repetitive part was of the consensus motif and a total of four nucleotide substitutions and an indel were observed in the first and middle repeats. These organizations in the primate genomes presumably indicate a stable ancestral precursor state at this locus.

Evolutionary relationship

In the phylogenetic network (Fig. 4), constructed for the human *CSTB* alleles, we observed complex branching in the evolution from the chimpanzee. It includes possible parallel substitutions or recombination events at this locus. The primate *CSTB* sequences indicate which of the nucleotides segregating in the human alleles is ancestral, as indicated in Table 2. Judging from conserved nucleotides at the corresponding positions, *CSTB*1* allele appears to be more ancestral than the others in the course of human evolution.

When the full *CSTB* sequences of the four primate genomes are compared with the human sequence, those of the chimpanzee, bonobo, gorilla, and orangutan differ by 1.3, 1.2, 2.1, and 4.0%, respectively, from the human sequence. In the open reading frame, one or two substitutions are evident among the human and four primates, one of which is nonsynonymous at a position corresponding to amino acid residue 63 (Asp to Glu) only in the chimpanzee. The

Table 1
Pairwise linkage-disequilibrium between *CSTB* polymorphic sites

	T−2406C	G−2371C	C−2265G	A−2203C	T−801C	G−794A	VNTR	T155C
G−2371C	1.00							
C−2265G	1.00	1.00						
A−2203C	−0.77	−1.00	1.00					
T−801C	0.69	0.95	−0.47	−0.71				
G−794A	0.69	0.95	−0.47	−0.71	1.00			
VNTR	0.80	0.90	−0.28	0.67	0.98	0.98		
T155C	−1.00	−1.00	−1.00	−1.00	1.00	1.00	−1.00	
G1055T	0.97	1.00	1.00	−0.96	0.95	0.95	0.96	−1.00

Table 2
Human *CSTB* alleles and their frequency based on 10 SNPs, accompanied by the corresponding nucleotides in the chimpanzee gene

Position	*1	*2	*3	*4	*5	*6	Chimp
−2406	T	T	T	C	T	C	T
−2371	G	G	G	G	G	C	G
−2265	C	C	G	G	G	G	C
−2203	A	A	C	C	C	A	A
−801	T	C	T	T	C	C	T
−794	G	A	G	G	A	A	G
155	T	C	T	T	T	T	T
1032[a]	G	G	G	G	G	T	ND[b]
1040[a]	G	G	G	G	G	T	ND[b]
1055	G	G	G	G	G	T	G
No. (%)	21 (5.8)	53 (14.6)	159 (43.7)	12 (3.3)	45 (12.4)	74 (20.3)	
VNTR 3/2[c]	0/17	0/40	0/110	0/8	9/15	21/2	

[a] The nucleotides at positions 1032 and 1040 were estimated by direct sequencing of 30 DNA specimens in which all were completely linked to the nucleotide at position 1055
[b] ND, not determined. Since the vicinities of these sites included indels, it was difficult to identify the corresponding nucleotide in the chimpanzee sequence.
[c] The number in each allele was judged from 111 subjects homozygous for either repeat of the VNTR.

phylogenetic tree is constructed from the primate sequences, as shown in Fig. 3. The open reading frame of the macaque gene revealed that 14 nucleotide substitutions, of which 5 are nonsynonymous changes at amino acid positions 9, 18, 71, 84, and 96, had occurred.

Discussion

Trinucleotide repeat expansion has been identified as causing a number of neurodegenerative disorders, including Huntington disease and fragile X syndrome. EPM1 represents the first disorder attributed to the instability of a simple repeat unit—a dodecamer—other than a triplet, which is often CAG or CGG [15]. Riggins et al. [16] reported that several genes expressed in the brain contain GC-rich repeats upstream of the transcription start site. The expansion of such repeats usually results in a reduction of transcription. Normal and pathogenic ranges of repeat size, and their allelic distributions in population groups, differ among loci. The normal repeat number of 2 or 3 copies observed in the *CSTB* locus is smaller than that for pathogenic triplets, which can range widely: for example, the number of repeats at the fragile X syndrome locus is 6 to ~40 copies [17]. The gene frequency of the triple repeat accounted for 0.27 in a normal Asian group, as found in this series of experiments. Lalioti et al. [11] reported that this frequency in Africans and Caucasians is 0.53 and 0.66, respectively. Being commonly variable, the number of repeats in the normal array appears to vary among ethnic groups.

Analysis of a contiguous sequence of *CSTB* and of linkage of the observed variations comprehensively demonstrated the extent of allelic divergence and the relationship of allelic variations with the copy number of the repeat. Six distinct alleles, *CSTB*1* to **6*, were defined by intragenic haplotyping of 10 closely linked polymorphic sites observed over 5.0 kb, and we found a close relationship between these alleles and the VNTR. The persistence of this association might indicate that changes in the repeat number are comparatively rare events in human evolution. Moreover, when the human alleles are compared with that of the

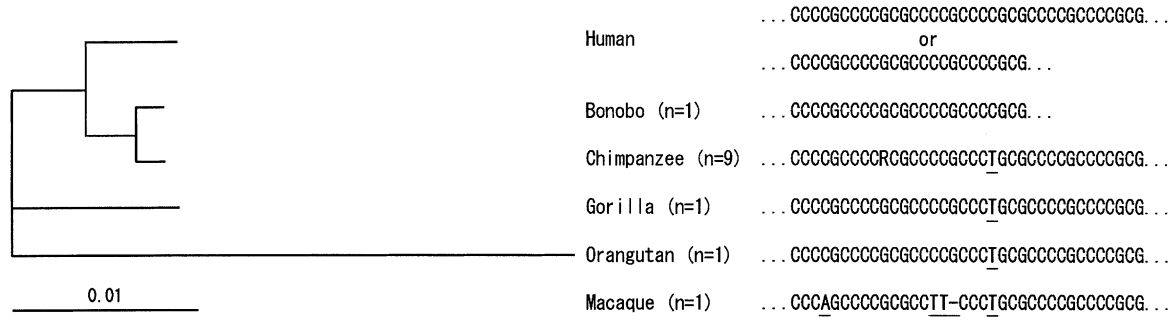

Fig. 3. Multiple sequence alignment of the VNTR sequence and dendrogram of *CSTB* in humans and five related primates. Nucleotide substitutions are underlined, and a gap relative to the human sequence is indicated by a dash. In the chimpanzee, another nucleotide change from G to A, indicated with R, was observed in only one chromosome. The phylogenetic tree constructed by the neighbor-joining method is shown at left. The scale bar corresponds to 0.01 substitutions per site.

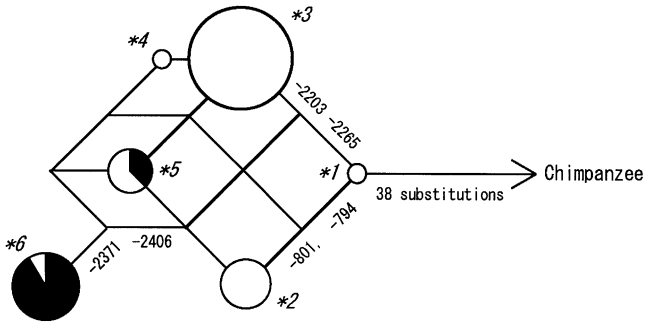

Fig. 4. Estimated phylogenetic network for human *CSTB*. To simplify the network, we constructed it from the six nonunique variable sites in the 5′ flanking region of 2.5 kb. The edge length corresponds to a nucleotide difference that is denoted by the position number beside it. Grids indicate possible parallel substitutions or recombination events. Circles indicate the observed alleles whose designations are shown in Table 2. The filled-in black and empty parts in the circles represent the proportions of triple and double repeats, respectively, of the alleles.

chimpanzee gene, *CSTB*1* appears to be the most ancestral allele. The finding that this allele has only two copies in the repeat array is possibly interpreted to mean that the double repeat is older in the human lineage. The increase in repeat number to 3 may have separately occurred in *CSTB*5* and *6 alleles with dispersion into populations.

The repeat segment of the chimpanzee gene was monomorphic in the search of nine subjects and that of the macaque sequence contained five changes that were likely static as well. Although the smaller population size of hominoids may affect the result [18], these findings conclusively suggest that the VNTR arose after the human–chimpanzee split. Concerning the evolutionary origin of repetitive DNA, a stepwise mutation model has been proposed for microsatellites [19]. It is, however, unclear how the VNTR originated, or if variability was intrinsic to the 12-bp unit from the beginning. Since the basic sequence motif resides in the last part of the segment in the macaque sequence, its simple repeat structure should begin with the dodecamer itself in the course of primate evolution. Deletion of the middle part from the incomplete triple repeat observed in the chimpanzee, gorilla, and orangutan genes, possibly by unequal crossing over in meiosis, may have resulted in a set of two motifs, which is strongly supported by the appearance of the perfect two-copy repeat in the bonobo sequence. While microsatellite variability already existed in the early hominoid ancestor [20], the incidence of dodecamer repetitiveness is evolutionarily a recent event. Gray and Jeffreys [21] reported a similar human-specific polymorphism for the MS32 minisatellite. At the MS32 locus, an enlarged hypervariable array is observed in humans in contrast to the several stable copies observed in hominoids.

Lalioti et al. [11] identified alleles with 12 and 13 copies of the dodecamer that were transmitted unstably to offspring, a situation called the "premutational state." These intermediate alleles are not associated with the clinical symptoms of EPM1. Expansion takes place over generations to produce approximately 30 to 75 copies, with clinical consequences. The mechanism of array enlargement has not been elucidated, but it is likely due to strand slippage or unequal crossing over [22]. Recent investigations of the secondary substructure of the dodecamer DNA showed that it can engage in tetraplex formation, with the production of folded hairpins, which causes hypervariability in triplet-repeat disease genes [23,24]. Errors should occur by unfolding of the hairpin structure during discontinuous DNA replication at the 3′ end. As shown by extensive studies, the normal repeat number is seemingly restricted as a minimum to 2 or 3 copies. We speculate that once the VNTR gained additional repeat units, the allele became unstable beyond a critical threshold due to the tendency toward tetraplex formation and was finally lost from the human DNA pool.

In conclusion, based on a comparative sequence analysis of the CSTB locus, we proposed a model to explain how VNTR has evolved. Such analyses of the genome will provide further insight into the evolutionary origins of humans and various diseases.

Materials and methods

Materials

Peripheral blood was obtained from 189 unrelated Japanese individuals. All volunteers gave informed and written consent to this DNA analysis. DNA of one chimpanzee (*Pan troglodytes*), one bonobo (also called the pygmy chimpanzee, *Pan paniscus*), one orangutan (*Pongo pygmaeus*), and one Japanese macaque (*Macaca fuscata*) was obtained from the Primate Research Institute, Kyoto University. The DNA of one gorilla (*Gorilla gorilla*) was obtained from the Ueno Zoo, Tokyo. The other eight DNA samples for chimpanzee were obtained from the Kumamoto Laboratory, Sanwa Kagaku Kenkyusho, Kumamoto. Restriction enzymes were purchased from MBI Fermentas and Toyobo.

PCR amplification and tandem repeat genotyping

To obtain the region containing VNTR, we performed PCR using the following oligonucleotides: DF1, CCCGGAAAGACGATACCAG (5′ to 3′ for all primers shown), and DR1, GGCACTTTGGCTTCGGAG. Amplification from 20 ng DNA was performed in a volume of 20 μl with HotStar *Taq* polymerase (Qiagen) in a supplemented Q solution mixture. The reaction proceeded in a GeneAmp 9700 cycler (PE Biosystems) at 95°C for 15 min for denaturation, followed by 35 rounds of temperature cycling, at 95°C for 20 s, 55°C for 20 s, and 72°C for 40 s. The products were electrophoresed in a 7.5% nondenaturing polyacrylamide gel, followed by staining with SYBR gold (Molecular Probes).

Contiguous sequencing

Overlapping segments of *CSTB* were amplified using six pairs of oligonucleotide primers designed according to the sequence (AF208234) published by Pennacchio et al. [7]: CGGTTCAAACTGTCTGCTGC and CACACAGTCCCTTTGGTTTAGG, GCAGGTGGTCCAGTCTAACAGAA and GAAGGCAGGGTTCTCTGGTG, GGCTCACACCTGTAATCCCAG and GGCTGGTATCGTCTTTCCG, DF1 and CCTTATCACTGGGGACCTCTTG, CCGACTACAGTTTGGCCAAG and TCAAGCTGGGACCTCACCTAG, and GGGAATCCAAGAAGCCACTG and CTGGAGGGGAAGGACCTG. This contiguous region spans a total of 5012 bp, including exons, introns, and 5′ flanking region. After purification by means of a spin column (Qiagen), the PCR products were labeled with a BigDye terminator cycle sequencing kit and then sequenced using an ABI Prism 310 autoanalyzer (PE Biosystems). A total of 15 DNA specimens were fully sequenced by this approach. The genomic sequence was also determined for the chimpanzee, bonobo, gorilla, orangutan, and Japanese macaque by this system. Several additional primers were necessary for the analysis of the orangutan and macaque genes.

Detection of polymorphic substitutions

To detect polymorphic changes in the sequenced region, we employed the restriction fragment length polymorphism (RFLP) method. Nucleotide changes at nucleotide positions −2406, −2265, −801, and 1055 were detected by digestion of the PCR products with *Fok*I, *Hin*fI, *Alu*I, and *Hin*fI, respectively, followed by polyacrylamide gel electrophoresis. To examine four other specific changes at nucleotide positions −2371, −2203, −794, and 155, we subsequently performed nested PCR using the restriction site-incorporated oligonucleotide primers GTAAAGCAAGTCCTGAATAGA, TGTACGGGTCCTGCCGCT, GAGGCGGAGCTTGCGG, and CAGGGGTGCGCAGCGTGG, respectively, in combination with the appropriate opposite-strand primer, as described above. The products were then digested by *Mbo*I, *Bse*LI, *Hph*I, and *Bal*I, respectively. Since changes at nucleotide positions 1032 and 1040 were not detectable by the RFLP method, these were determined by direct DNA sequencing of up to 30 DNA specimens. Haplotypes were constructed from nucleotide changes identified among the samples as previously described [25].

Statistical and phylogenetic analysis

Hardy–Weinberg tests were performed on the observed VNTR and allele frequencies with significance based on a standard observed–expected χ^2 with $df = 1$ and 25, respectively. Pairwise linkage disequilibrium (D') was estimated according to the maximum-likelihood procedure of Thompson et al. [26], in which D' was calculated by $D/D_{\max}$ or $D/D_{\min}$. Nucleotide diversities (π) were estimated by the method of Nei [27], available at http://www.bio.ub.es/~julio/DnaSP.html. A phylogenetic tree for primate *CSTB* was constructed by the neighbor-joining method using the orangutan sequence as an outgroup [28]. A phylogenetic network of human *CSTB* alleles was constructed by the procedure of Saitou and Yamamoto [29]. Because of the complexity of the network constructed from all 10 substitution events, the analysis was simplified by using only changes in the 5′ flanking region of 2505 bp in this study.

References

[1] V. Turk, W. Bode, The cystatins: protein inhibitors of cysteine proteinases, FEBS Lett. 285 (1991) 213–219.

[2] M. Riccio, Nuclear localization of cystatin B, the cathepsin inhibitor implicated in myoclonus epilepsy (EPM1), Exp. Cell Res. 262 (2001) 84–94.

[3] M.T. Stubbs, The refined 2.4 A X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction, EMBO J. 9 (1990) 1939–1947.

[4] R. Norio, M. Koskiniemi, Progressive myoclonus epilepsy: genetic and nosological aspects with special reference to 107 Finnish patients, Clin. Genet. 15 (1979) 382–398.

[5] A.E. Lehesjoki, Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping, Hum. Mol. Genet. 2 (1993) 1229–1234.

[6] R.G. Lafrenière, P.J. de Jong, G.A. Rouleau, A 405-kb cosmid contig and *Hin*dIII restriction map of the progressive myoclonus epilepsy type 1 (EPM1) candidate region in 21q22.3, Genomics 29 (1995) 288–290.

[7] L.A. Pennacchio, Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1), Science 271 (1996) 1731–1734.

[8] M.D. Lalioti, Identification of mutations in cystatin B, the gene responsible for the Unverricht–Lundborg type of progressive myoclonus epilepsy (EPM1), Am. J. Hum. Genet. 60 (1997) 342–351.

[9] M.D. Lalioti, H.S. Scott, S.E. Antonarakis, Altered spacing of promoter elements due to the dodecamer repeat expansion contributes to reduced expression of the cystatin B gene in EPM1, Hum. Mol. Genet. 8 (1999) 1791–1798.

[10] L.A. Pennacchio, Progressive ataxia, myoclonic epilepsy and cerebellar apoptosis in cystatin B-deficient mice, Nat. Genet. 20 (1998) 251–258.

[11] M.D. Lalioti, Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy, Nature 386 (1997) 847–851.

[12] G. Vergnaud, The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence, Genomics 11 (1991) 135–144.

[13] A.J. Jeffreys, R. Neumann, V. Wilson, Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis, Cell 60 (1990) 473–485.

[14] A. Fujiyama, Construction and analysis of a human–chimpanzee comparative clone map, Science 295 (2002) 131–134.

[15] C.T. Ashley, S.T. Warren, Trinucleotide repeat expansion and human disease, Annu. Rev. Genet. 29 (1995) 703–728.

[16] G.J. Riggins, Human genes containing polymorphic trinucleotide repeats, Nat. Genet. 2 (1992) 186–191.

[17] K. Snow, Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population, Am. J. Hum. Genet. 53 (1993) 1217–1228.
[18] D.C. Rubinsztein, Microsatellite evolution—evidence for directionality and variation in rate between species, Nat. Genet. 10 (1995) 337–343.
[19] A.M. Valdes, M. Slatkin, N.B. Freimer, Allele frequencies at microsatellite loci: the stepwise mutation model revisited, Genetics 133 (1993) 737–749.
[20] E.E. Eichler, Evolution of the cryptic FMR1 CGG repeat, Nat. Genet. 11 (1995) 301–308.
[21] I.C. Gray, A.J. Jeffreys, Evolutionary transience of hypervariable minisatellites in man and the primates, Proc. R. Soc. London B 243 (1991) 241–253.
[22] B. Charlesworth, P. Sniegowski, W. Stephan, The evolutionary dynamics of repetitive DNA in eukaryotes, Nature 371 (1994) 215–220.
[23] T. Saha, K. Usdin, Tetraplex formation by the progressive myoclonus epilepsy type-1 repeat: implications for instability in the repeat expansion diseases, FEBS Lett. 491 (2001) 184–187.
[24] P.V. Jithesh, P. Singh, R. Joshi, Molecular dynamics studies of trinucleotide repeat DNA involved in neurodegenerative disorders, J. Biomol. Struct. Dyn. 19 (2001) 479–495.
[25] M. Osawa, Haplotype analysis of the human α2-HS glycoprotein (fetuin) gene, Ann. Hum. Genet. 65 (2001) 27–34.
[26] E.A. Thompson, S. Deeb, D. Walker, A.G. Motulsky, The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes, Am. J. Hum. Genet. 42 (1988) 113–124.
[27] M. Nei, Molecular Evolutionary Genetics (1987, Columbia Univ. Press, New York.
[28] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[29] N. Saitou, F. Yamamoto, Evolution of primate ABO blood group genes and their homologous genes, Mol. Biol. Evol. 14 (1997) 399–411.