Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 338 (2004) 47–54
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

Evolution of the tandem repeats in thymidylate synthase enhancer region (*TSER*) in primates

Yun-wu Zhang[a,b,1,2], Huai-rong Luo[a,c,1], Oliver A. Ryder[b], Ya-ping Zhang[a,c,*]

[a] *Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, PR China*
[b] *Center for Reproduction of Endangered Species, San Diego Zoo, P.O. Box 120551, San Diego, CA 92112, USA*
[c] *Laboratory for Conservation and Utilization of Bioresources, Yunnan University, Kunming 650091, PR China*

Received 2 February 2004; received in revised form 21 April 2004; accepted 17 May 2004
Available online 17 July 2004
Received by Takashi Gojobori

Abstract

The upstream regulatory region of the human thymidylate synthase gene (thymidylate synthase enhancer region, *TSER*) is length polymorphic, attributable to variable numbers of tandemly repeated copies of a 28-bp fragment. It has been found that *TSER* length polymorphism is correlated to malignancy risk. To further our understanding of the origin and evolution of *TSER*, this region was investigated among different primates, including hominoids, two subfamilies of the Old World monkeys (OWMs): colobines and cercopithecines, and two species of the New World monkeys (NWMs). In addition to humans, our results show that length polymorphism in *TSER* is also present in some primate populations, although it appears that this region is length monomorphic in many other primates. We identified three unique repeat motifs in *TSER* and defined them as R1, R2, and R3, respectively, starting from the 3′ end. The same repeat motifs from different species are more similar to each other than different repeat motifs within same species are. Such a paraphyletic pattern suggests that divergence of the three repeat motifs predated divergence of the OWMs/hominoids and the NWMs. The most recent common ancestor (MRCA) of hominoids and the OWMs probably possessed triple repeats but now double and triple repeats are two dominant types in hominoids and the OWMs. In addition, our results show that each of the three repeat motifs may be lost independently. We have also found clues that recombination was involved in formation of tandem repeat polymorphism in *TSER*.

Keywords: Repeat motif; Length polymorphism; Population; Recombination

1. Introduction

Thymidylate synthase (TS) (EC2.1.1.45) catalyzes the reductive methylation of dUMP during the synthesis of dTMP and is essential in regulating a balanced supply of the four DNA precursors for DNA replication. Because TS serves as a key enzyme in folate metabolism, it is the primary target of cancer chemotherapy drugs such as 5-fluorouracil (5-FU, Rustum et al., 1997; Ulrich et al., 2002). It has been reported that the upstream regulatory region of the human thymidylate synthase (TS enhancer region, *TSER*) is length polymorphic, containing two, three, four, five, or nine tandemly repeated copies of a 28-bp sequence in different populations (Horie et al., 1995; Luo et al., 2002; Marsh et al., 1999, 2000). An inverted repeat sequence is also present upstream the tandemly repeated region and can pair with each of the tandem repeat motifs to form stable stem-loop structures that may be interconvertible (Takeishi et al., 1989). *TSER* length polymorphism has been found to contribute to the *TS* gene expression level (Horie et al., 1995; Kawakami et al., 1999), and several studies have shown evidence that *TSER* length polymorphism may

Abbreviations: *TSER*; thymidylate synthase enhancer region; TS; thymidylate synthase; PCR; polymerase chain reaction; OWMs; Old World monkeys; NWMs; New World monkeys; MRCA; most recent common ancestor; SSM; slipped-strand mispairing.

* Corresponding author. Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, PR China. Tel.: +86-871-5190761; fax: +86-871-5195430.
E-mail address: zhangyp1@263.net.cn (Y. Zhang).
[1] These authors contributed equally to the work.
[2] Present address: Center for Neuroscience and Aging, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA.

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.05.018

modify the risk of cancers including colorectal adenomas and leukemias (Adleff et al., 2004; Skibola et al., 2002; Ulrich et al., 2002). This unusual repeat sequence structure in *TSER* has been found in a rhesus monkey, but not in rodents, suggesting it originated as a result of duplication or of duplication followed by deletion of the specific sequence in the 5′-terminal region of the human *TS* gene during the evolutionary divergence of rodents and primates (Horie et al., 1995, 1998). To further our understanding of the evolution of this region, we investigated *TSER* in different primate species by PCR and sequencing, and compared the findings with those obtained from humans.

2. Materials and methods

We amplified and sequenced *TSER* from 31 different nonhuman primates, including hominoid species, species of the two subfamilies of the Old World monkeys (OWMs)—colobines and cercopithecines—and two species of platyrrhines (or the New World monkeys, NWMs). Whenever possible, a population of 10 or more individuals from one species was utilized for PCR and the products were analyzed on 4% agarose gels to identify length polymorphism. For other species, due to limited samples, only one or several individuals were studied (Table 1). For hominoids and the Old World monkeys, PCR primers and conditions were the same as those in Luo et al. (2002). For the New World monkeys, we used another PCR primer pair (*TSER*f2: 5′-ATCCCGCTGAGCAGGAAGAG-3′ and *TSER*r2: 5′-TGCCGAATACCGACAGGGTG-3) instead. Sequencing was performed on an ABI 377 sequencer or 3100 genetic analyzer (Applied Biosystems). For each length difference variant allele identified by electrophoresis within any primate population that had a sample size of 10 or more individuals, we sequenced at least two independent PCR products from different animals. For all other primate species, we sequenced all of them (GenBank accession nos. AY532281–AY532316). Sequences were aligned manually. The 28-bp repeat motif was defined according to Takeishi et al. (1989). Although definition of the repeat motif can be modified by shifting the fragment to the 3′-end by 1 or several nucleotides, it does not appreciably change our results. The UPGMA method embedded in MEGA version 2.1 (Kumar et al., 2001) was used for gene tree construction.

3. Results

3.1. Tandem repeats exist in the TSER in different primates

After screening *TSER* from primate species previously unstudied, alleles with one, two, three, or four copies of the repeat motif were identified in different primate taxa (Table 1). These tandem repeat motifs are 28-bp long except the

Table 1
TSER length difference variation in primates

Species	Chromosomal numbers studied	Allele type	Repeat motif	Frequency (%)
Hominoids				
Homo sapiens[a]	–	2R	(Rh)1R1h	19–40
		3R	(Rh)2R1h	60–81
		4R	(Rh)3R1h	<7
		5R	(Rh)4R1h	<4
		9R	(Rh)8R1h	<2
Pan troglodytes	28	3R	R3R2R1	100
Pan paniscus	44	3R	R3R2R1	100
Gorilla gorilla	112	1R	R2	5.4
		2R	R2R1	84.8
		3R	R3R2R1	9.8
Pongo pygmaeus pygmaeus	40	3R	R3R2R1	100
Pongo pygmaeus abelii	40	3R	R3R2R1	100
Hylobates leucogenys	2	3R	R3R2R1	
Hylobates hoolock	2	3R	R3R2R1	
Hylobates moloch	2	3R	R3R2R1	
Hylobates lar	2	3R	R3R2R1	
Symphalangus syndactylus	4	2R	R2R1	
Colobines				
Colobus polykomos	2	3R	R3R2R1	
Nasalis larvatus	2	2R	R3R2	
Pygathrix nemaeus	20	2R	R3R2	100
Rhinopithecus roxellenae	60	2R	R3R2	100
Rhinopithecus bieti	2	2R	R3R2	
Rhinopithecus avunculus	2	2R	R3R2	
Presbytis johnii	2	3R	R3R2R1	
Presbytis francoisi	6	2R	R3R1	16.7
		3R	R3R2R1	83.3
Presbytis phayrei	4	2R	R3R1	25
		3R	R3R2R1	75
Cercopithecines				
Macaca mulatta	64	3R	R3R2R1	100
Macaca fascicularis	2	2R	R3R1	
Macaca silenus	20	3R	R3R2R1	100
Macaca arctoides	2	3R	R3R2R1	
Maraca nemestina	2	3R	R3R2R1	
Macaca assamensis	14	3R	R3R2R1	50
		4R	R3R1R2R1	50
Macaca thibetana	24	4R	R3R1R2R1	100
Papio cynocephalus	2	3R	R3R2R1	
Mandrillus leucophaeus	22	2R	R3R1	72.7
		3R	R3R2R1	27.3
Allenopithecus nigroviridis	2	2R	R3R1	
Platyrrines				
Ateles paniscus	2	2R	R3R1	
Pithecia pithecia	2	1R	R3	

[a] Human allele frequency data were from other reports (Luo et al., 2002; Marsh et al., 1999, 2000).

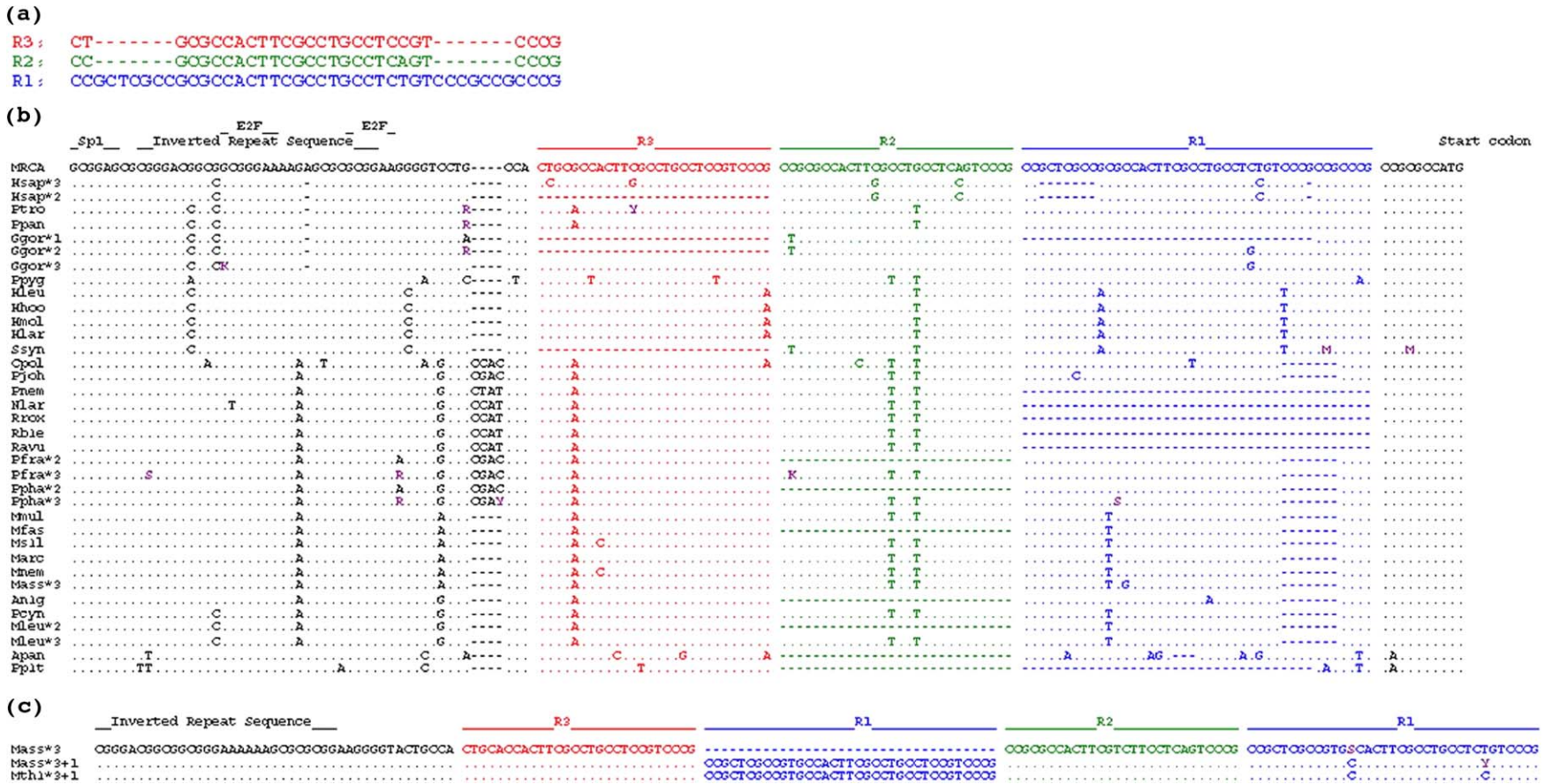

Fig. 1. *TSER* sequence alignment. (a) Sequence definitions and comparisons of the three repeat motifs from the most recent common ancestor of hominoids and the OWMs (MRCA): R1 (in blue), R2 (in green), and R3 (in red). Dashes represent nucleotide deletions. (b) Primate *TSER* sequence alignments. Here we use the first letter of the genus name in capitals and the first three letters of the species name in small capitals to represent each species. MRCA denotes the most recent common ancestor of hominoids and OWMs. The locations of the inverted repeat sequence and each repeat motif are according to Takeishi et al. (1989). The location of the Sp1-binding site is according to Horie and Takeishi (1997). The locations of the two potential E2F binding sites are according to Dong et al. (2000). Dots represent identical nucleotides. SNPs found within species are represented by standard codes (in purple). (c) The unique pattern of the allele type R3R1R2R1 found in *M. assmensis* and *M. thibetana*, comparing to the allele type R3R2R1 found in *M. assamensis*. Regions behind repeat motif R1 were not shown.

ones at the 3′-end (defined as R1, see below), which have different deletions compared to repeat motif R1 of the most recent common ancestor (MRCA) of the OWMs/hominoids that is 42-bp long (Fig. 1).

3.2. The patterns of TSER tandem repeats in the Old World monkeys/hominoids and their most recent common ancestor (MRCA)

Double (2R) and triple (3R) repeats are two dominant allele types in hominoids and the OWMs (Table 1). By comparing the three repeat motifs of triple repeat alleles in these primates, we found that although highly conservative, each of the three repeat motifs from same individual/species can be distinguished by unique nucleotide substitutions and/or indels; and these unique substitutions and/or indels were shared by corresponding repeat motifs of different primates (Fig. 1). Because it is unlikely that triple repeats were derived from double repeats by gaining one copy of the repeat motif and then followed by convergent evolution among different species, the most recent common ancestor (MRCA) of hominoids and the OWMs probably possessed triple repeat alleles. Using the NWMs' sequences as reference, we inferred a possible sequence of the MRCA of hominoids and the OWMs and then defined each repeat motif of the MRCA as R1, R2, and R3, respectively, starting from the 3′ end (Fig. 1). We also aligned every motif from all the primates studied here and constructed a gene tree using the UPGMA method. As shown in Fig. 2, in all the studied nonhuman primates, repeat motifs were derived from R1, R2, or R3 and clustered correspondingly. The only exception is human's repeat motifs Hsap2 and Hsap3 (both are a copy of Rh, see below), which were clustered within the group of repeat motif R1. These repeat motifs are unidirectionally tandem with a pattern of 5′-R3R2R1-3′, except that in two macaque species, *Macaca assamensis* and *Macaca thibetana*, there was one additional repeat motif R1 inserted between repeat motifs R3 and R2. In addition, we found that in some primates, one or two of the three repeat motifs were lost and each of the repeat motifs can be lost without affecting the other two repeat motifs (Table 1; Fig. 1).

3.3. The pattern of TSER tandem repeats in humans

In humans, the last repeat motif was derived from repeat motif R1 of the MRCA by one nucleotide substitution (T to C at site 29 of R1) plus two deletion events (a 7-bp fragment "GCTCGCC" at sites 3–9 and a single nucleotide deletion of "G" at site 35 of R1), and thus classified as R1 h. For the other repeat motifs in humans, they are identical and have two nucleotide differences from either the MRCA repeat motif R2 or repeat motif R3, so we could not decide their

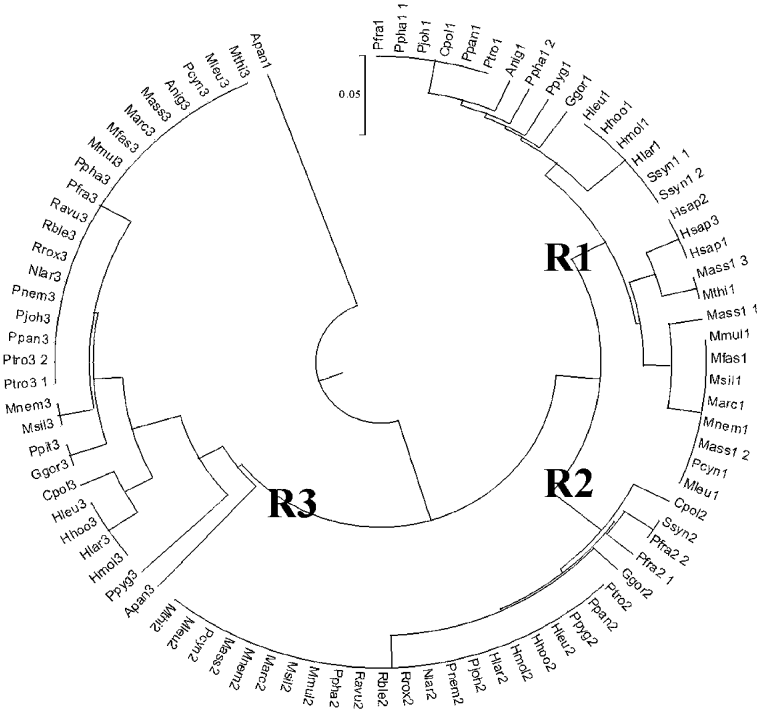

Fig. 2. Gene tree construction of *TSER* repeat motifs. Every repeat motif from primates was used for gene tree construction using UPGMA embedded in MEGA software. Denotations have the same meanings as in Fig. 1. The numbers following the denotations represent each motif's corresponding position to that of the MRCA of the OWMs/hominoids. Occasionally, for one specific repeat motif in one primate species, there were two or more SNP alleles, and we used additional numbers following them to differentiate each SNP allele. Note that repeat motifs were clustered corresponding to their relative position, except that Hsap2 and Hsap3 were clustered within the group of repeat motif R1.

exact origin and defined them as Rh. Thus, the tandem pattern of repeat motifs in humans is 5′-(Rh)*n*R1h-3′, where *n* equals 1, 2, 3, 4, or 8, as found in different human populations so far (Horie et al., 1995; Luo et al., 2002; Marsh et al., 1999, 2000).

3.4. TSER length polymorphism also exists in some nonhuman primates

In addition to humans, length polymorphism was also identified in five primate species: gorillas (*Gorilla gorilla*), Phayre's langur (*Presbytis phayrei*), Francois langur (*Presbytis francoisi*), drills (*Mandrillus leucophaeus*), and Assam macaque (*M. assamensis*). However, for chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), orangutans (*Pongo pygmaeus pygmaeus* and *Pongo pygmaeus abelli*), douc langurs (*Pygathrix nemaeus*), golden monkeys (*Rhinopithecus roxellenae*), rhesus monkeys (*Macaca mulatta*), lion-tail monkeys (*Macaca silenus*), and Tibetan monkeys (*M. thibetana*), we have studied at least 10 individuals from each species and did not find length polymorphism within these species. Thus, it seems that this region has been fixed as either allele-type 3R or 2R in these species (Table 1; Fig. 1).

For each length difference variant allele identified by electrophoresis, at least two samples from the same species were sequenced whenever available. We found single nucleotide polymorphisms (SNPs) in some primate species (Fig. 1).

4. Discussion

4.1. TSER polymorphism in primates

Length polymorphism of the tandemly repeated sequences near the initiation start site of thymidylate synthase has been identified in humans (Horie et al., 1995; Luo et al., 2002; Marsh et al., 1999, 2000). In an early study, Kaneda et al. (1987) found that only one repeat motif in humans is sufficient for insuring the translation of *TS* mRNA and suggested that the translation efficiency of the *TS* mRNA was unlikely to be affected by a change in length of the repeated region. However, later experiments showed that expression activity of the *TS* gene with a double repeat (2R) was lower than that with either a triple repeat (3R) or a quintuple repeat (5R) in humans. Thus, *TSER* length polymorphism may contribute to the efficiency of expression of the gene, without notable phenotypic effects in humans (Horie et al., 1995; Kawakami et al., 1999; Luo et al., 2002). On the other hand, transient expression experiments showed that the enhancing activity of quadruple or quintuple repeats was not much different from that of triple repeat (Kaneda et al., 1987; Luo et al., 2002). Recently, several studies have shown correlations between *TSER* genotype and malignancy risk: One study demonstrated that triple repeat homozygotes (3R/3R) conferred much higher protection on leukemia than double repeat homozygotes (2R/2R) (Skibola et al., 2002). In another study, it was found that 3R/3R homozygotes had lower risk to colorectal cancer when associated with high folate intake, while 2R/2R homozygotes were protected at low folate intake (Ulrich et al., 2002). In the present study, we found *TSER* repeat number variations within and between different nonhuman primates. It would be interesting to examine whether the *TS* gene expression activity is also affected by these polymorphisms. The sequence alignment (Fig. 1) showed that loss of any of the repeat motifs, R1, R2, or R3, is not lethal; and the existence of only one repeat motif in some primates provides evidence that one repeat motif might be sufficient for insuring the translation of *TS* mRNA.

TSER is polymorphic in some primates while monomorphic in others. Within polymorphic populations, the frequencies of different allele types are different (Table 1). Marsh et al. (1999) have suggested migration and dietary variations as possible explanations for different *TSER* polymorphism frequencies in human populations. It is also possible that such factors could have played an important role in *TSER* polymorphism frequency distribution in other primate populations and/or fixation of certain allele type in some nonhuman primate species. On the other hand, historical evolutionary events may also be involved in the formation of such observed allele-type patterns in these primates. For example, it is suggested that modern humans originated from an African stock after a bottleneck of moderate size followed by a range expansion out of Africa (Excoffier, 2002); this bottleneck may have helped to shape the unique tandem pattern 5′-(Rh)nR1h-3′ in human *TSER*.

4.2. Origin/evolution of TSER allele types and mechanisms responsible for TSER length polymorphism

Each of the three unique repeat motifs are more similar among different species than they are to different repeat motifs within same individual/species (Fig. 2), suggesting that divergence of the three repeat motifs predated divergence of the OWMs/hominoids and the NWMs. Consistent with this, although we only acquired two sequences from the NWMs, they also possess this unique repeat motif structure and have different repeat numbers. Because this repeat motif structure was not found in rodents (Deng et al., 1986), these results suggest that this unique repeat motif structure at *TSER* originated after primates divergence from rodents but earlier than divergence between the NWMs and the OWMs/hominoids. There is also a possibility that *TSER* length polymorphism already existed in the MRCA of the NWMs and the OWMs/hominoids.

Based on the allele types of *TSER* observed in these primates, a network was drawn to hypothesize the origin of different *TSER* allele types observed in the OWMs/hominoids (Fig. 3): all double repeats (including R3R2, R3R1, and R2R1) might have originated from the ancestral type

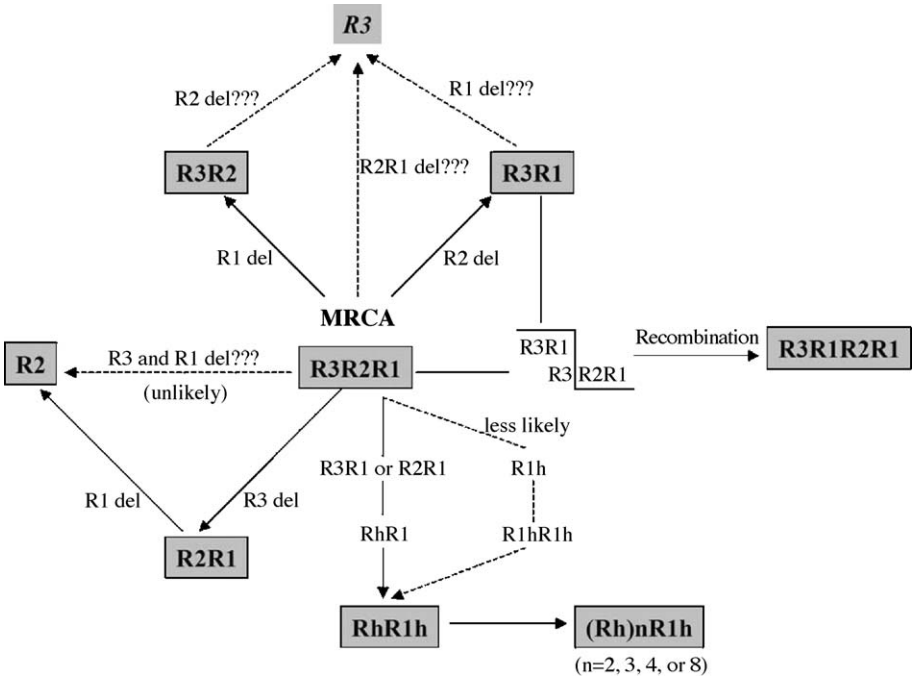

Fig. 3. Evolutional network to hypothesize origins of *TSER* allele types found in primates. Allele types found in primates were grayed in the background. The mono repeat allele R3 found in the NWMs was italicized. "Del" denotes deletion. Note that the allele-type R3R1R2R1 was originated from recombination between the allele types R3R2R1 and R3R1. The human-allele-type RhR1h was originated first by deletion of one repeat motif (either R2 or R3), then followed by nucleotide substitutions and deletions. Less likely, unlikely, or questionable origination pathways were indicated with dash lines.

R3R2R1 by deletion of any repeat motif. A mono repeat (R2) observed in gorillas probably originated from R2R1 (also observed in gorillas) by one deletion event, rather than originated from the ancestral-type R3R2R1 by two deletion events. Consistent with such a parsimony principle, at the second nucleotide of repeat motif R2, both the gorilla R2 and R2R1 allele types have a thymine, while the gorilla R3R2R1 allele type has a cytosine; and at the fourth nucleotide upstream repeat motif R3 (not including a 4bp deletion), the gorilla R2 allele type is monomorphic with an adenine, the gorilla R2R1 allele type is polymorphic with either an adenine or a guanine, while the gorilla R3R2R1 allele type is monomorphic with a guanine (Fig. 1). All of these suggest that R2 was derived from R2R1 by a single deletion rather than from R3R2R1, which requires not only two deletion events but also two nucleotide substitutions. Because we do not know the ancestral type of the NWMs, the origin of another mono repeat (R3) observed in *Pithecia pithecia* was presumed from different possibilities: R3 might have originated from R3R2R1, R3R2, or R3R1 by the deletion of R2R1, R2, or R1, respectively.

Two types of mechanisms are responsible for length difference variations at molecular level: recombination and slipped-strand mispairing (SSM) during DNA replication. Here our results suggested a role for recombination in generating *TSER* length difference variation: In two macaque species, *M. assamensis* and *M. thibetana*, there was one additional repeat motif R1 inserted between repeat motifs R3 and R2, forming the 4R allele R3R1R2R1. The origin of this unique allele type could not be explained by sole slipped-strand mispairing from the allele-type R3R2R1, as found in *M. assamensis*. Based on a reasonable assumption that the allele-type R3R2R1 in *M. assamensis* may generate the allele-type R3R1 by a deletion of R2, as found in some other macaque species, we proposed a possible explanation; that is, the allele-types R3R2R1 and R3R1 once coexisted in the ancestral species of *M. assamensis* and *M. thibetana*; and the 4R allele R3R1R2R1 was generated by interchromosomal recombination between the two allele types (Ji et al., 2000), as shown in Fig. 3. On the other hand, we cannot exclude the possibility that SSM also played a role in generating other *TSER* allele types found in other species by sole deletion of repeat motifs.

Comparison of the allele-type (Rh)2R1h in humans to the allele-type R3R2R1 in the MCRA and other hominoids revealed that at least two nucleotide substitutions were needed to reach repeat motif Rh from repeat motif R3 and another two nucleotide substitutions at different sites were also needed to reach Rh from R2 (Fig. 1). It is unlikely that changes occurring at different nucleotide sites in different repeat motifs coincidentally generated the same result (Rh). One plausible explanation is that one repeat motif (either R2 or R3) was lost during the origin of modern humans, while the other one evolved into Rh. Then, the changes leading the MRCA repeat motif R1 to human R1h, especially the deletion of a fragment "GCTCGCC" (nucleotide sites 3–9 in R1), which made human R1h have nucleotides identical to

those of Rh at the beginning region, triggered recombination or SSM, resulting in the observed length polymorphism in humans. Alternatively, based on the fact that human repeat motif Rh (Hsap2 and Hsap3 in Fig. 2) was clustered within the group of repeat motif R1, there is another possibility that all human repeat motifs might have originated from repeat motif R1 of the MRCA. But this is less likely because in this case, ancestral human repeat motifs R2 and R3 needed to be lost first, then human repeat motif R1h duplicated, and R1h at the 5′-end evolved into Rh by one nucleotide substitution (C to G at site 19 of the MRCA R1) and one convergent fragment deletion (sites 32 to 38 of the MRCA R1), which was also found in the OWMs (Fig. 1).

4.3. Regulation of the TS gene expression at posttranscriptional level

There is evidence that the *TS* gene expression is primarily controlled at the posttranscriptional level in both humans and mice (Ayusawa et al., 1986; Jenh et al., 1985; Johnson, 1994). Takeishi et al. (1989) found an inverted repeat sequence 14 nucleotides upstream from the first tandem repeat motif in humans. The complementarity between each of the tandem repeat motifs and the inverted repeat is 75%. Conformation stability analysis suggested that this inverted repeat is capable of forming a stable, stem-looped structure with any of the tandem repeat motifs (Takeishi et al., 1989). Thus, one possible role this unique tandem repeat in the downstream from the cap site plays at a posttranscriptional step might be the formation and interconversion of different stem-loop structures in the 5′-terminal region of the human *TS* mRNA (Horie et al., 1995; Takeishi et al., 1989). Our results showed that this inverted sequence was also present in and relatively conserved among primates, in addition to the existence of tandem repeat motifs in primates. These results imply that regulation of the *TS* gene expression at posttranscriptional level in other primates might be similar to that in humans.

4.4. Regulation of the TS gene expression at transcriptional level

Although expression of the *TS* gene is mainly regulated at the posttranscriptional level (Ayusawa et al., 1986; Jenh et al., 1985; Johnson, 1994), regions essential for the regulation that are dependent on the stage of the cell cycle have been identified both in the first intron and in the 5′-flanking region of the human *TS* gene, suggesting the involvement of the promoter region in the cell-cycle-dependent regulation of the gene (Kaneda et al., 1992; Takayanagi et al., 1992). Neither the human *TS* gene nor the mouse *TS* gene has a TATA box or a CAAT box at the 5′-flanking region (Jolliff et al., 1991; Takeishi et al., 1989). Horie and Takeishi (1997) identified an Sp1-binding motif in the human *TS* promoter region as "GAGGCGGAG". This motif is highly homologous to the Sp1-binding site of the mouse *TS* gene and of the rat *TS* gene (Jolliff et al., 1991; Lee and Johnson, 2000). Although our sequences do not cover the entire Sp1-binding motif, the overlapping region (the first six nucleotides in our sequence alignment) showed complete sequence conservation between human and other primates (Fig. 1). This result implies the existence of the Sp1-binding motif in other primates. The fact that the Sp1-binding motif is conserved in humans (Horie and Takeishi, 1997), mice (Jolliff et al., 1991), rats (Lee and Johnson, 2000), rhesus monkeys (Horie et al., 1998), and all other primates studied in our present work suggests that the Sp1-binding motif might be the core sequence of the promoter that maintains the basic transcription activity of *TS* genes.

Horie et al. (1995) discussed one of the possible mechanisms for the enhancing effect of this unique tandem repeat structure as that an unidentified nuclear factor binds to this region and controls the expression of the human *TS* gene. However, they failed to detect any DNA-binding factor that bound to the repeated sequence (Horie et al., 1992). Potential E2F-binding sites were found in the inverted repeat region and it seemed like these sites acted as negative promoter elements (Dong et al., 2000; Horie et al., 1995). Our comparison showed that these sites were also highly conserved throughout primates (Fig. 1).

In the present study, we demonstrated that besides humans, *TSER* was also length polymorphic in many nonhuman primates but fixed in others. The frequencies of length difference variant allele types are different among primates. Recombination has been found involved in the generation of one allele type, although we cannot exclude the possibility that SSM is also involved in forming other allele types. Further functional study will be important in showing whether these allele types may affect expression levels of the *TS* gene in different nonhuman primates, as that found in humans.

Acknowledgements

The authors would like to thank Leona Chemnick for help on DNA samples, Jennie Lau for proofreading the manuscript, and anonymous reviewers for helpful comments. This work was supported by the Chinese Academy of Sciences, the National Nature Science Foundation of China, Alice B. Tyler Perpetual Trust, and the Zoological Society of San Diego.

References

Adleff, V., Hitre, E., Koves, I., Orosz, Z., Hajnal, A., Kralovanszky, J., 2004. Heterozygote deficiency in thymidylate synthase enhancer region polymorphism genotype distribution in Hungarian colorectal cancer patients. Int. J. Cancer 108, 852–856.

Ayusawa, D., Shimizu, K., Koyama, H., Kaneda, S., Takeishi, K., Seno, T., 1986. Cell-cycle-directed regulation of thymidylate synthase messenger

RNA in human diploid fibroblasts stimulated to proliferate. J. Mol. Biol. 190, 559–567.

Deng, T.L., Li, D.W., Jenh, C.H., Johnson, L.F., 1986. Structure of the gene for mouse thymidylate synthase. Locations of introns and multiple transcriptional start sites. J. Biol. Chem. 261, 16000–16005.

Dong, S., Lester, L., Johnson, L.F., 2000. Transcriptional control elements and complex initiation pattern of the TATA-less bidirectional human thymidylate synthase promoter. J. Cell. Biochem. 77, 50–64.

Excoffier, L., 2002. Human demographic history: refining the recent African origin model. Curr. Opin. Genet. Dev. 12, 675–682.

Horie, N., Takeishi, K., 1997. Identification of functional elements in the promoter region of the human gene for thymidylate synthase and nuclear factors that regulate the expression of the gene. J. Biol. Chem. 272, 18375–18381.

Horie, N., Nozawa, R., Takeishi, K., 1992. Identification of cellular differentiation-dependent nuclear factors that bind to a human gene for thymidylate synthase. Biochem. Biophys. Res. Commun. 185, 127–133.

Horie, N., Aiba, H., Oguro, K., Hojo, H., Takeishi, K., 1995. Functional analysis and DNA polymorphism of the tandemly repeated sequences in the 5′-terminal regulatory region of the human gene for thymidylate synthase. Cell Struct. Funct. 20, 191–197.

Horie, N., Wakabayashi, M., Katai, H., Takeishi, K., 1998. Cloning and characterization of the promoter region of monkey thymidylate synthase gene. Biol. Pharm. Bull. 21, 1020–1023.

Jenh, C.H., Geyer, P.K., Johnson, L.F., 1985. Control of thymidylate synthase mRNA content and gene transcription in an overproducing mouse cell line. Mol. Cell. Biol. 5, 2527–2532.

Ji, Y., Eichler, E.E., Schwartz, S., Nicholls, R.D., 2000. Structure of chromosomal duplicons and their role in mediating human genomic disorders. Genome Res. 10, 597–610.

Jolliff, K., Li, Y., Johnson, L.F., 1991. Multiple protein–DNA interactions in the TATAA-less mouse thymidylate synthase promoter. Nucleic Acids Res. 19, 2267–2274.

Johnson, L.F., 1994. Posttranscriptional regulation of thymidylate synthase gene expression. J. Cell. Biochem. 54, 387–392.

Kaneda, S., Takeishi, K., Ayusawa, D., Shimizu, K., Seno, T., Altman, S., 1987. Role in translation of a triple tandemly repeated sequence in the 5′-untranslated region of human thymidylate synthase mRNA. Nucleic Acids Res. 15, 1259–1270.

Kaneda, S., Horie, N., Takeishi, K., Takayanagi, A., Seno, T., Ayusawa, D., 1992. Regulatory sequences clustered at the 5′ end of the first intron of the human thymidylate synthase gene function in cooperation with the promoter region. Somatic Cell Mol. Genet. 18, 409–415.

Kawakami, K., Omura, K., Kanehira, E., Watanabe, Y., 1999. Polymorphic tandem repeats in the thymidylate synthase gene is associated with its protein expression in human gastrointestinal cancers. Anticancer Res. 19, 3249–3252.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA 2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

Lee, Y., Johnson, L.F., 2000. Transcriptional control elements of the rat thymidylate synthase promoter: evolutionary conservation of regulatory features. Exp. Cell Res. 258, 53–64.

Luo, H.R., Lu, X.M., Yao, Y.G., Horie, N., Takeishi, K., Jorde, L.B., Zhang, Y.P., 2002. Length polymorphism of thymidylate synthase regulatory region in Chinese populations and evolution of the novel alleles. Biochem. Genet. 40, 41–51.

Marsh, S., Collie-Duguid, E.S.R., Li, T., Liu, X., McLeod, H.L., 1999. Ethnic variation in the thymidylate synthase enhancer region polymorphism among Caucasian and Asian populations. Genomics 58, 310–312.

Marsh, S., Ameyaw, M.M., Githang'a, J., Indalo, A., Ofori-Adjei, D., McLeod, H.L., 2000. Novel thymidylate synthase enhancer region alleles in African populations. Human Mutat. 16, 528.

Rustum, Y.M., Harstrick, A., Cao, S., Vonhoefer, U., Yin, M.B., Wilke, H., Seeber, S., 1997. Thymidylate synthase inhibitors in cancer therapy: direct and indirect inhibitors. J. Clin. Oncol. 15, 389–400.

Skibola, C.F., Smith, M.T., Hubbard, A., Shane, B., Roberts, A.C., Law, G.R., Rollinson, S., Roman, E., Cartwright, R.A., Morgan, G.J., 2002. Polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and risk of adult acute lymphocytic leukemia. Blood 99, 3786–3791.

Takayanagi, A., Kaneda, S., Ayusawa, D., Seno, T., 1992. Intron 1 and the 5′-flanking region of the human thymidylate synthase gene as a regulatory determinant of growth-dependent expression. Nucleic Acids Res. 20, 4021–4025.

Takeishi, K., Kaneda, S., Ayusawa, D., Shimizu, K., Gotoh, O., Seno, T., 1989. Human thymidylate synthase gene: isolation of phage clones which cover a functionally active gene and structural analysis of the region upstream from the translation initiation codon. J. Biochem. 106, 575–583.

Ulrich, C.M., Bigler, J., Bostick, R., Fosdick, L., Potter, J.D., 2002. Thymidylate synthase promoter polymorphism, interaction with folate intake, and risk of colorectal adenomas. Cancer Res. 62, 3361–3364.