Variations in short tandem repeats deduced on the basis of the number of repeats and the relationship of these variations with longevity

Liu Hui · Yu Weijian · Deng Xuelian · Liu Qigui

Received: 28 March 2010 / Accepted: 14 June 2010 / Published online: 29 June 2010
© American Aging Association 2010

Abstract The objective of this study was to investigate the quantitative characteristics of short tandem repeat (STR) variations deduced on the basis of the number of STRs that are beneficial for human survival. The longevity group included 60 nonagenarian subjects, and the control group included 250 reference adults (age, 20–50 years). Alleles of 15 Combined DNA Index System STR loci were determined using a commercial polymerase chain reaction kit. An STR with the highest frequency distribution in a population (control group) was considered as a conservative STR, and the number of core unit repeats of this STR allele was considered as the median repeat number in the STR locus (STR_m). The absolute difference between the STR_m and the number of core unit repeats of other STR alleles can be considered as the quantitative marker of variation for that particular STR allele (*M* value). The mean *M* values of CSF1TPO in the longevity group were significantly higher than those in the control

group ($P < 0.05$). These findings appear to suggest that at least one of the STR loci may be associated with longevity. The *M* value of STR may be a new and high-efficacy genetic marker.

Keywords Longevity · STR · Genetic marker · Variation

Introduction

Microsatellites or short tandem repeats (STRs) consist of tandemly repeated DNA units with lengths ranging from two to six nucleotides (Edwards et al. 1991, 1992; Fan and Chu 2007). STRs are named according to the number of repeats; different STR alleles can occur in a single STR locus, thereby resulting in STR polymorphism. Generally, an STR polymorphism is considered the result of gene mutation (Butler 2006; Eckert and Hile 2009; Laghi et al. 2008). Accordingly, an STR with the highest frequency distribution in a certain population is considered a conservative STR, and an STR with relatively low or very low frequency distribution is considered a result of gene mutation and natural selection. Furthermore, the number of repeats can be considered as the quantitative marker of genetic variation.

Human longevity is determined by both genetic and environmental factors; several susceptibility

L. Hui (✉) · Y. Weijian · D. Xuelian
College of Medical Laboratory, Dalian Medical University,
Dalian 116044, People's Republic of China
e-mail: liuhui60@mail.dlptt.ln.cn

L. Qigui
Department of Hygienic Statistics, Dalian Medical
University,
Dalian, China

genes act together with environmental factors to produce the longevity phenotype (Browner et al. 2004; Vijg and Suh 2005). Because of the variable number of highly polymorphic tandem repeats in humans, STR assessment is a powerful tool in human genetics. STR polymorphisms are also the popular allelic markers used in investigations of longevity and multifactorial diseases of the elderly (Tan et al. 2002; Yamaya et al. 2003). The STR assessment method for the identification of longevity-associated genes is solely based on the differences in the frequency of each STR frequency between the longevity group and the control group and does not consider the number of repeats of the STRs as a quantitative marker of genetic variation. One of the rapidly developing fields in genetic research is the identification of the genes involved in longevity and multifactorial diseases of the elderly. However, the progress in the identification of the longevity-associated genes that provide protection from common diseases or slow the aging process has been impeded by inconsistent findings (Kim et al. 2007; Liu et al. 2007). This lack of progress can be partly attributed to the unavailability of high-efficacy genetic analysis. The use of quantitative measures, which are statistically more powerful than qualitative measures, will improve the efficacy of genetic analysis.

Combined DNA Index System (CODIS) STR loci, including D8S1179, D3S1358, D2S1338, D19S433, vWA, D18S51, FGA, D21S11, D7S820, CSF1TPO, TH01, D13S317, D16S539, TPOX, and D5S818, are being currently used by forensic laboratories and the US national CODIS. As a reliable assay performed using commercial multiplex kits, the procedure was carried out at half-atomized condition in accordance to the international standards, and the results were assessed by using standard software; therefore, error due to artificial performance and judgment was remarkably reduced (Budowle and Sprecher 2001; Buse et al. 2003; Moretti et al. 2001). CODIS STR loci were chosen in this study because of their reliability for genotypic analysis. We are interested in understanding whether the quantitative characteristics of the deduced STR variations, which are deduced on the basis of the number of STRs, are associated with human survival. There have been no previous reports indicating that the tandem repeat number of STRs provides initial suggestive genetic evidence of longevity.

Materials and methods

Subjects

In this study, we included 310 northern Han Chinese individuals (620 chromosomes) residing in Dalian, China. Sixty nonagenarian subjects (30 men and 30 women) who expressed interest in participating in the study were recruited in the longevity group. A total of 250 reference adults (125 men and 125 women; age, 20–50 years) were selected from a random population of the same city and recruited in the control group in our study. The exclusion criteria were exacerbation of various chronic diseases 4 weeks before the study and long-term use of medication.

Sample preparation and genotyping

DNA was extracted from 3 ml of peripheral blood by the Chelex 100 procedure (Moretti et al. 2001). Polymerase chain reaction (PCR) was performed using the AmpFLSTR Profiler plus PCR Amplification Kit (Perkin Elmer, Foster City, CA, USA) as per the manufacturer's recommendations; the PCR reaction volume was 50 μ l, and a 9600 Perkin Elmer thermal cycler was used. Amplification products (1.5 μ l) were added to 10 μ l formamide and 1 μ l of an internal size standard (Genescan-500 ROX; Applied Biosystems). The samples were heat-denatured at 95°C for 5 min and chilled for 5 min in an ice-water bath before performing capillary electrophoresis using an ABI 310 automated sequencer (Applied Biosystems). We used the GeneScan Analysis 2.1 software (Applied Biosystems) to determine fragment sizes. Allele identification was achieved by comparing the size of the amplified fragments with the allelic ladders included in the reagent set, and the alleles were labeled according to the international nomenclature using the Genotyper Software package (Perkin Elmer).

Statistical analysis

The number of core nucleotide unit repeats in the STR loci is considered to change continuously and gradually. Therefore, an STR with the highest frequency distribution in a certain population (control group) is considered as a conservative STR, and the number of core unit repeats of this STR allele is

Fig. 1 Representative results of STR alleles were assessed by standard software (*the number in the pane represents the number of core unit repeats along with the name of the STR allele, such as D8S1179-12*)

considered as the original repeat number in the STR locus. Thus, the absolute difference between the original repeat number and the number of core unit repeats of other STR alleles at that STR locus can be considered as the quantitative marker of variation for the STR allele. We used the median repeat number instead of the original repeat number because the original number was purely hypothetical. Thus, the degree of variation can be calculated by the following equation:

$$M = |\text{STRn} - \text{STRm}|.$$

In the above equation, M represents the degree of variation, STRn the number of core unit repeats of the STR allele, and STRm the median repeat number for a certain STR locus. The larger the M value, the larger is the degree of variation of the STR allele.

In each individual, each STR locus has two STR alleles; the sum of the M values of these two STR alleles indicates the variation at the STR locus for an individual. The degree of correlativity between variation and longevity may also change continuously and gradually; thus, the degree of correlation with longevity may increase when the M value increases for a certain STR that is partially associated with longevity.

Two types of STRs are normally identifiable from a single STR locus in an individual. The two alleles were considered identical (homozygous) if only one type of STR was found in a locus. Allele frequencies in the control group were calculated to estimate the median repeat number (STRm) for each STR locus. The distributions of the genotypes of these polymorphisms were assessed using exact tests to determine whether they followed the Hardy–Weinberg equilibrium (Lim et al. 2009). M values were calculated for each STR locus in an individual. A nonparametric test (Mann–Whitney U test) was used to assess the difference between the mean M values for the longevity and control groups. The difference was considered statistically significant when the P value was <0.05 (two-tailed test). The calculations were performed using the Statistical Package for Social Sciences 13.0 software for Windows.

Results

The raw data for alleles in the four STR loci, which were examples for 15 STR loci, were obtained using the GeneScan Analysis software. As shown in Fig. 1,

Table 1 Results of the Hardy–Weinberg equilibrium test for the 15 STR loci in the two groups

Locus	P value		Locus	P value	
	Longevity	Control		Longevity	Control
D8S1179	0.588	0.742	D2S1338	0.922	0.998
D21S11	0.948	0.790	D19S433	0.958	0.829
D7S820	0.983	0.580	vWA	0.862	0.060
CSF1TPO	0.980	0.976	TPOX	0.943	0.696
D3S1358	0.997	0.691	D18S51	0.976	0.998
TH01	0.773	0.577	D5S818	0.425	0.812
D13S317	0.966	0.980	FGA	0.888	0.795
D16S539	0.978	0.625			

Table 2 Allele frequencies for STR loci in the control population

STR	Repeat number	Freq. (%)	STR	Repeat number	Freq. (%)	STR	Repeat number	Freq. (%)	STR	Repeat number	Freq. (%)
D8S1179	9	0.2	D3S1358	12	0.2	D19S433	11	0.8	D18S51	10	0.2
	10	8.8		14	4.0		12	6.2		11	0.2
	11	9.2		15	38.4		12.2	0.2		12	2.2
	12	14.4		16	29.8		13	27.8		13	22.6
	13	25.6		17	20.4		13.2	4.2		14	19.2
	14	17.8		18	6.8		14	21.8		14.2	0.2
	15	14.6		19	0.4		14.2	10.8		15	21.4
	16	8.0		6	14.4		15	9.2		16	13.6
D21S11	17	1.4	TH01	7	23.0	vWA	15.2	13.0	FGA	17	7.4
	27	0.2		8	5.2		16	0.8		18	3.0
	28	4.0		9	49.4		16.2	4.2		19	3.0
	28.2	1.2		9.3	5.0		17.2	0.6		20	1.8
	29	25.4		10	3.0		18.2	0.4		21	2.4
	29.2	0.8	D13S317	8	26.6		13	0.2		22	1.6
	30	32.2		9	13.8		14	24.6		23	0.2
	30.2	1.4		10	15.4		15	3.4		24	0.4
	31	9.2		11	20.4		16	16.2		25	0.4
	31.2	6.4		12	19.8		17	27.8		26	0.2
	32	3.4		13	3.0		18	19.0		16	0.8
	32.2	10.6		14	0.6		19	8.0		18	2.0
	33	0.6		15	0.4		20	0.8		19	4.8
	33.2	4.2	D16S539	8	0.4	TPOX	8	45.2		20	4.4
	34.2	0.4		9	26.6		9	13.8		21	12.0
D7S820	8	15.4		10	14.0		10	3.2		22	15.6
	9	6.4		11	26.2		11	34.8		22.2	0.6
	10	16.8		12	22.2		12	2.4		23	23.8
	11	34.8		13	8.8		13	0.2		23.2	1.2
	12	21.4		14	1.6		14	0.4		24	16.4
	13	4.6		15	0.2	D5S818	7	0.6		24.2	1.4
	14	0.6	D2S1338	16	1.0		9	8.2		25	11.2
CSF1TPO	8	0.6		17	7.0		10	22.4		25.2	0.2
	9	6.4		18	9.8		11	34.0		26	4.4
	10	19.2		19	17.0		12	21.0		27	1.0
	11	23.6		20	9.4		13	13.4		28	0.2
	12	39.8		21	2.0		14	0.2			
	13	8.2		22	4.2		16	0.2			
	14	1.8		23	24.2						
	15	0.4		24	16.2						
				25	7.4						
				26	1.8						

Boldface type represents the median repeat number defined in an STR locus; the number after the decimal point in non-integer repeat numbers represents the number of non-integer core unit repeats

Table 3 Intergroup variation in the M values at different STR loci (M value) [mean (SD)]

Locus	Longevity	Control	P
D8S1179	3.20±1.55	2.73±1.59	0.039
D21S11	2.28±1.20	2.02±1.40	0.131
D7S820	2.38±1.35	2.16±1.54	0.234
CSF1TPO	2.47±1.31	1.93±1.36	0.003
D3S1358	1.77±1.32	1.94±1.33	0.301
TH01	1.83±1.54	1.98±1.58	0.364
D13S317	4.32±2.43	4.13±2.33	0.554
D16S539	3.43±1.90	3.54±2.02	0.736
D2S1338	5.03±3.20	4.78±2.87	0.774
D19S433	2.01±1.48	2.22±1.41	0.371
vWA	3.13±1.61	2.70±1.77	0.070
TPOX	2.37±2.14	2.75±2.00	0.139
D18S51	4.65±2.86	4.58±3.33	0.644
D5S818	1.95±1.05	1.81±1.14	0.415
FGA	3.12±1.80	2.95±2.00	0.304

Bonferroni correction, $\alpha'=0.05/15=0.0033$

the alleles were labeled using the Genotyper Software package.

For each locus, the Hardy–Weinberg equilibrium was tested by comparing the observed genotype numbers with those expected under the hypothesis of panmixia (Hardy–Weinberg equation); no deviations from the Hardy–Weinberg equilibrium were observed in the longevity and control groups, as shown in Table 1.

The allele frequencies and the median repeat numbers of 15 STR loci are shown in Table 2. The comparison of variations in the STR locus (M value) in the two groups is shown in Table 3. The mean M value of CSF1TPO in the longevity group was significantly higher than those in the control group ($P<0.05$). There were no significant intergroup differences in the M values for other STR loci.

Discussion

We believe that the process of gene mutation and natural selection involves continuous and gradual changes. A quantitative marker is necessary to describe the process and the degree of genetic variation in order to identify genes, especially minor genes, associated with complex diseases or longevity.

We have developed a quantitative marker of genetic variation on the basis of the properties of STRs. The primary aim of this study was to investigate whether our new method or quantitative marker is feasible for longevity analysis.

The main characteristics of longevity-associated genes are reflected in nonagenarian individuals (Anselmi et al. 2009; Kim et al. 2007; Liu et al. 2007). Our results indicated that the mean M values of the STR locus CSF1TPO in the subjects of the longevity group were significantly higher than those in the control group ($P<0.05$). The statistical tests did not reveal any deviations from the Hardy–Weinberg equilibrium. These results imply that (1) the selected STR loci are suitable for the genetic analysis of the population in our study, (2) the STR loci CSF1TPO might be associated with longevity, and (3) elderly individuals, especially nonagenarians, are prone to carrying STRs with large variations. The M value of an STR may be a new, quantitative, and high-efficacy genetic marker.

Although STRs themselves do not express any functional proteins, variations in the STRs could reflect the influence of relevant factors on the genome. Investigations of the possible contribution and role of the M value of an STR in longevity could demonstrate the application of STR analysis for the quantification of the influence of external factors on the genome. Further study is needed to illustrate the detailed mechanisms and improve the analytical efficacy of this approach.

References

- Anselmi CV, Malovini A, Roncarati R, Novelli V, Villa F, Condorelli G, Bellazzi R, Puca AA (2009) Association of the FOXO3A locus with extreme longevity in a southern Italian centenarian study. *Rejuvenation Res* 12(2):95–104
- Browner WS, Kahn AJ, Ziv E, Reiner AP, Oshima J, Cawthon RM, Hsueh WC, Cummings SR (2004) The genetics of human longevity. *Am J Med* 117(11):882–883
- Budowle B, Sprecher CJ (2001) Concordance study on population database samples using the PowerPlex 16 kit and AmpFISTR Profiler Plus kit and AmpFISTR COfiler kit. *J Forensic Sci* 46(3):637–641
- Buse EL, Putinier JC, Hong MM, Yap AE, Hartmann JM (2003) Performance evaluation of two multiplexes used in fluorescent short tandem repeat DNA analysis. *J Forensic Sci* 48(2):348–357

- Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 51(2):253–265
- Eckert KA, Hile SE (2009) Every microsatellite is different: intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol Carcinog* 48(4):379–388
- Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 49:746–756
- Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric repeat loci in four human population groups. *Genomics* 12:241–253
- Fan H, Chu JY (2007) A brief review of short tandem repeat mutation. *Genomics Proteomics Bioinformatics* 5(1):7–14
- Kim KI, Na JE, Kang SY, Cho YS, Choi DJ, Kim CH, Kim HS, Oh BH, Choi YH, Kwon IS, Park SC (2007) Impact of NAD(P)H oxidase p22 phox gene polymorphism on vascular aging in Korean centenarian and nonagenarian. *Int J Cardiol* 123(1):18–22
- Laghi L, Bianchi P, Malesci A (2008) Differences and evolution of the methods for the assessment of microsatellite instability. *Oncogene* 27(49):6313–6321
- Lim EJ, Lee HY, Sim JE, Yang WI, Shin KJ (2009) Genetic polymorphism and haplotype analysis of 4 tightly linked X-STR duos in Koreans. *Croat Med J* 50(3):305–312
- Liu H, Wang B, Liu D, Cheng Y, Yu W, Liang X, An W (2007) Molecular genetic studies on relationships among longevity, diseases, and HLA-DRB1/DQB1 allelic polymorphism. *Exp Aging Research* 33:123–125
- Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B (2001) Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *J Forensic Sci* 46(3):647–660
- Tan Q, Bellizzi D, Rose G, Garasto S, Franceschi C, Kruse T, Vaupel JW, De Benedictis G, Yashin AI (2002) The influences on human longevity by HUMTHO1.STR polymorphism (Tyrosine Hydroxylase gene). A relative risk approach. *Mech Ageing Dev* 123(10):1403–1410
- Vijg J, Suh Y (2005) Genetics of longevity and aging. *Annu Rev Med* 56:193–212
- Yamaya M, Nakayama K, Ebihara S, Hirai H, Higuchi S, Sasaki H (2003) Relationship between microsatellite polymorphism in the haem oxygenase-1 gene promoter and longevity of the normal Japanese population. *J Med Genet* 40(2):146–148