Biochem Genet
DOI 10.1007/s10528-007-9105-y

Number of STR Repeats as a Potential New Quantitative Genetic Marker for Complex Diseases, Illustrated by Schizophrenia

Hui Liu · Weijian Yu · Xuebin Wang · Fang Fang · Guang Yang · Jie Zhou · Xiaohua Liang · Wanxin An

Received: 18 July 2006 / Accepted: 24 February 2007

Abstract It has proved difficult to find strong and replicable genetic linkages for complex diseases, since each susceptibility gene makes only a modest contribution to onset. This is partly because high-efficacy genetic markers are not usually available. The aim of this article is to explore the possibility that the total number of tandem repeats in one STR locus, rather than the frequencies of different alleles, is a higher efficacy quantitative genetic marker. DNA samples were collected from schizophrenic patients and from a control population. Alleles of the short tandem repeats (STR) loci D3S1358, vWA, and FGA were determined using the STR Profiler Plus PCR amplification kit. The two groups did not differ statistically in the frequencies of alleles at the D3S1358, vWA, or FGA loci. However, a significant difference was obtained in the vWA locus when the total number of core unit repeats was compared between the schizophrenia and control groups (33.28 ± 2.61 vs. 32.35 ± 2.58, $P < 0.05$). It seems that the number of STR repeats may be a new, quantitative, and higher efficacy genetic marker for directly indicating genetic predisposition to complex hereditary diseases such as schizophrenia.

Keywords Genetics · Genetic marker · Short tandem repeats · Quantitative STR · Schizophrenia

H. Liu (✉) · W. Yu · G. Yang · J. Zhou · X. Liang · W. An
College of Medical Laboratory, Dalian Medical University, Dalian 116027, P.R. China
e-mail: liuhui60@mail.dlptt.ln.cn

X. Wang · F. Fang
Dalian Seventh Hospital, Dalian 116023, P.R. China

Introduction

The analysis of genetic linkage has been highly successful in mapping the genes responsible for Mendelian diseases. During the past decade, attempts have been made to extend this approach to multifactorial disorders and other health-related traits. It has proved difficult, however, to find strong and replicable linkages, since a significant feature of a complex disease is the modest contribution of each susceptibility gene to its onset (Sham 2001; Bracken 2005). Our linkage studies with short tandem repeats (STR) in schizophrenia have also been discouraging because of inconsistent findings and weak signals, but this research has indicated that STR may provide a quantitative and higher efficacy genetic marker.

STR or microsatellites consist of tandemly repeated DNA units ranging from two to six nucleotides (Edwards et al. 1991, 1992). The repeat units in D3S1358, vWA, and FGA, which are the subjects of the present investigation, are AGAT, TCTA, and TCTT, respectively (Perez-Lezaun et al. 1997). They afford one of the most powerful tools in human genetics because the number of tandem repeats shows highly polymorphic variability. We are interested in understanding whether the total number of tandem repeats in one STR locus, rather than the frequencies of different alleles, is associated with schizophrenia in our population. There have been no previous reports that the tandem repeat number of STR provides initial suggestive genetic evidence of schizophrenia or any other complex disease.

Materials and Methods

Subjects

The 57 Chinese patients (114 chromosome sets), including 31 males and 26 females with the mean age of 45 ± 7.3 years and meeting the Diagnostic and Statistic Manual of Mental Disorders (DSM-IV) criteria for diagnosis of schizophrenia, were recruited from the Dalian Seventh Hospital (psychiatric hospital) in Dalian, China. Exclusion criteria: (1) Those complicated with other severe systemic diseases; (2) Those dependent on medication or alcohol. Patients were evaluated by two or more senior psychiatrists and a consensus was obtained. They were all chronic patients with at least 10 years duration of illness. The patients and their relatives signed informed consent before volunteering for this study. The control subjects were 123 healthy blood donors (246 chromosome sets), including 65 males and 58 females with the mean age of 37 ± 8.2 years, from Dalian, China.

Genotyping

DNA was extracted from 3 ml of peripheral blood by the Chelex100 procedure (Walsh et al. 1991). PCR analysis was performed using an AmpFISTR Profiler plus

PCR Amplification Kit (Perkin Elmer, Foster City, CA) under conditions recommended by the manufacturer in a reaction volume of 50 µl, using a 9600 Perkin Elmer thermal cycler. Amplification products (1.5 µl) were added to 10 µl formamide and 1 µl 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 capillary electrophoresis using an ABI 310 automated sequencer (Applied Biosystems). Genescan Analysis 2.1 software (Applied Biosystems) was used to determine fragment sizes. Alleles were identified by comparing the amplified fragments with the allelic ladders included in the reagent set, and alleles were labeled according to international nomenclature conventions using the Genotyper Software package (Perkin Elmer).

Statistical Analysis

Two types of STR are normally identifiable in one subject from one STR locus. Two alleles were considered identical (homozygous) if only one type of STR was found in a locus. Allele frequencies were calculated. The distributions of genotypes of these polymorphisms were examined using the chi-square test to determine whether they followed the Hardy–Weinberg equilibrium (Soares-Vieira et al. 2002). The allele frequency distributions of these polymorphisms were compared between patients and control subjects using the multiple chi-square test. The number of tandem repeats in two alleles at one STR locus was added as a quantitative genetic parameter. For D3S1358-14;16, for instance, in one subject the quantitative STR parameter was 30 (14 + 16). Differences between numbers of repeats were compared using the independent-samples *t*-test and were considered significant when the *P* values were less than 0.05. All analyses were performed using the SPSS statistical software package.

Results

The raw data for alleles in the three STR loci were obtained using Genescan Analysis software. As shown in Fig. 1, the alleles were labeled by the Genotyper Software package; the number of core unit repeats identified with the allele is used as the name of the STR allele. 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 control group (Table 1). There was no statistical difference between the two groups with respect to the allele frequencies of the D3S1358, vWA, or FGA loci (Table 2).

The comparison between the total numbers of core STR unit repeats in the schizophrenia and control groups is shown in Table 3. A significant difference was observed in the vWA locus.

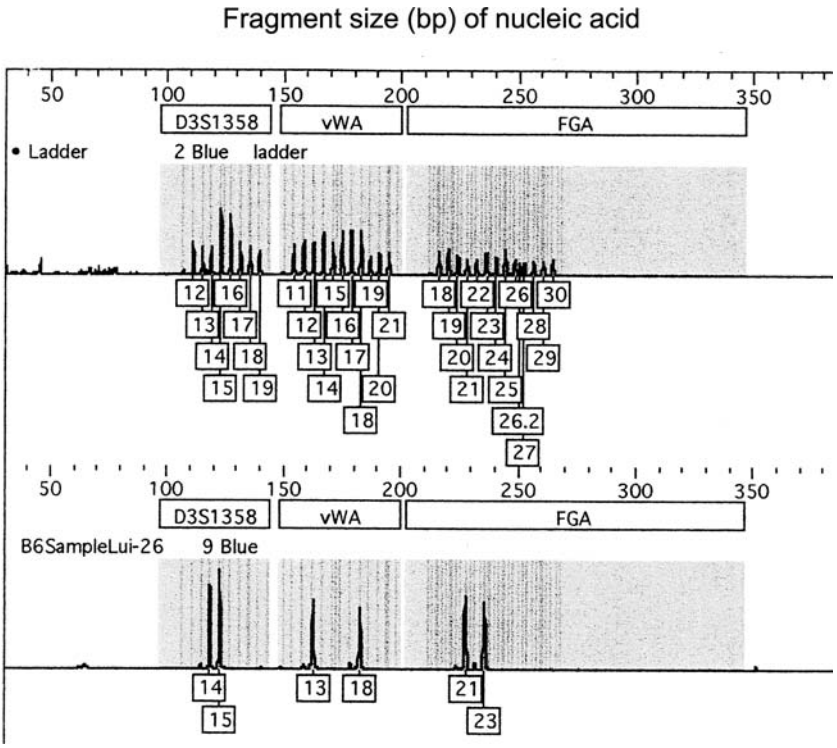

Fig. 1 Polymorphism of D3S1358, vWA, and FGA loci, analyzed by the Genotyper Software package. Each allele is identified by the amplified fragment size (abscissa) and labeled with the number of core unit repeats (the number in the square). The number of core unit repeats serves as the name of the STR allele

Table 1 Hardy–Weinberg equilibrium in control group

Locus	χ^2	df	*P*
D3S1358	23.29	26	>0.5
vWA	23.05	34	>0.75
FGA	46.71	64	>0.9

Discussion

Schizophrenia is a typical complex disorder to which an as-yet-unknown number of genes contribute, interacting with each other and the environment (Thaker and Carpenter 2001). Twin studies confirm that schizophrenia is highly heritable (McGuffin et al. 2003), but the genes have remained elusive in spite of the rapid pace of development of molecular technology and the expansion of genome sequence information. This is partly because high-efficacy genetic markers are not available. Single nucleotide polymorphisms (SNP) are currently popular allelic

Table 2 Allele distribution for three STR loci in schizophrenia and control groups

Allele	D3S1358		vWA		FGA	
	SCH	Control	SCH	Control	SCH	Control
13	0.0000	0.0041	0.0000	0.0041		
14	0.0965	0.0447	0.2105	0.3333		
15	0.3333	0.3699	0.0351	0.0488		
16	0.3772	0.3089	0.1491	0.1382		
17	0.1140	0.2073	0.2719	0.1951	0.0000	0.0041
18	0.0702	0.0528	0.2105	0.1748	0.0439	0.0122
19	0.0088	0.0122	0.0877	0.0894	0.0702	0.0447
20			0.0351	0.0163	0.0351	0.0407
21					0.1053	0.1098
22					0.1140	0.1992
23					0.3158	0.2358
24					0.1754	0.1992
25					0.1053	0.1138
26					0.0351	0.0285
27					0.0000	0.0122
P	0.128		0.245		0.223	

Table 3 Comparison of core STR unit repeats in schizophrenia and control groups

Locus	Group	*N*	Number of repeats		*P*
			Mean	SD	
D3S1358	SCH	57	31.30	2.01	0.077
	Control	123	31.77	1.48	
vWA	SCH	57	33.28	2.61	0.028
	Control	123	32.35	2.58	
FGA	SCH	57	45.09	2.85	0.388
	Control	123	45.46	2.57	

markers (Brohede et al. 2005; Nguyen et al. 2004), but their value is largely qualitative. They may not be satisfactory indicators of complex diseases because of their low quantitative efficacy. These considerations led us to investigate new, potentially higher efficacy genetic markers.

Short tandem repeat DNA sequences have been used to analyze hereditary linkage balance in studies of the location of genes associated with complex diseases such as schizophrenia, and some candidate genes have been screened out (Hattori et al. 2001; Kaufmann et al. 1998). If the total number of core unit repeats in STR loci provides a quantitative indicator rather than a qualitative indicator (allele

frequency in STR locus), the efficacy of analysis will be improved because a quantitative measure is statistically more powerful than a qualitative one.

We consider that changes in the number of core unit repeats in STR loci are occurring continuously and gradually; therefore, the correlation between the number of core unit repeats and virulence genes of a complex disease is also continuous and gradually changing. The morbidity risk may increase when the combination of two close numbers of core unit repeats in one STR locus occurs, if some STR is in partial linkage with some complex diseases, and consequently the random combination of two core unit repeats in one STR locus may change. This is the basic reason for using the total number of core unit repeats in STR loci as quantitative genetic markers.

A reliable assay for D3S1358, vWA, and FGA genotypes was adopted in our investigation; these loci are currently used by forensic laboratories and the U.S. National Combined DNA Indexing System (CODIS) with commercial multiplex kits, including the AmpFlSTR Profiler Plus system (Buse et al. 2003; Moretti et al. 2001). The procedure was conducted under conditions that accord with international standards, and the results were assessed using standard software. Therefore, errors due to artifacts and operator judgments were minimized (Budowle and Sprecher 2001; Budowle et al. 2001). The results for these three loci obtained using the AmpFlSTR Profiler Plus kit are shown in Fig. 1. All three loci were highly polymorphic, as shown in previous publications (Wang et al. 2003). According to the statistical tests (Table 1), no deviations from the Hardy–Weinberg equilibrium were detected in the control group. These results imply that reasonable data were obtained in this study. Quantitative comparisons of allele frequencies between the schizophrenia and control groups are shown in Table 2. The two groups did not differ statistically in the allele frequencies of the D3S1358, vWA, or FGA loci. A significant difference was obtained, however, in the vWA locus when the total numbers of core STR unit repeats were compared between the schizophrenia and control groups (Table 3). It seems that the number of STR repeats may be a new, quantitative, and higher efficacy genetic marker.

These investigations of the possible contribution and potential role of the quantitative STR genetic marker in schizophrenia may exemplify the different approaches needed for clinical genetic diagnosis. We expect increasing integration between genetics, epidemiology, and clinical trials through the sharing of valuable data among laboratories, leading to informative designs that directly analyze genetic predispositions to compound hereditary diseases such as schizophrenia, with quantitative STR as higher efficacy genetic markers. Further study is needed to develop this prospect.

Acknowledgments The study was supported by grant NSFC-39670273 from the National Natural Science Foundation of China.

References

Bracken MB (2005) Genomic epidemiology of complex disease: the need for an electronic evidence-based approach to research synthesis. Am J Epidemiol 162(4):297–301

Brohede J, Dunne R, McKay JD, Hannan GN (2005) PPC: an algorithm for accurate estimation of SNP allele frequencies in small equimolar pools of DNA using data from high density microarrays. Nucleic Acids Res 33(17):e142

Budowle B, Sprecher CJ (2001) Concordance study on population database samples using the PowerPlex 16 kit and AmpFlSTR Profiler Plus kit and AmpFlSTR COfiler kit. J Forensic Sci 46(3):637–641

Budowle B, Masibay A, Anderson SJ, Barna C, Biega L, Brenneke S, Brown BL, Cramer J, DeGroot GA, Douglas D, Duceman B, Eastman A, Giles R, Hamill J, Haase DJ, Janssen DW, Kupferschmid TD, Lawton T, Lemire C, Llewellyn B, Moretti T, Neves J, Palaski C, Schueler S, Sgueglia J, Sprecher C, Tomsey C, Yet D (2001) STR primer concordance study. Forensic Sci Int 124(1):47–54

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

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

Hattori E, Ebihara M, Yamada K, Ohba H, Shibuya H, Yoshikawa T (2001) Identification of a compound short tandem repeat stretch in the 5′-upstream region of the cholecystokinin gene, and its association with panic disorder but not with schizophrenia. Mol Psychiatry 6(4):465–470

Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD, Meyer J, Zambuto CT, Schmitt K, Matise TC, Harkavy Friedman JM, Hampe C, Lee H, Shore D, Wynne D, Faraone SV, Tsuang MT, Cloninger CR (1998) NIMH genetics initiative millennium schizophrenia consortium: linkage analysis of African-American pedigrees. Am J Med Genet 81(4):282–289

McGuffin P, Tandon K, Corsico A (2003) Linkage and association studies of schizophrenia. Curr Psychiatry Rep 5(2):121–127

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

Nguyen TH, Liu C, Gershon ES, McMahon FJ (2004) Frequency Finder: a multi-source web application for collection of public allele frequencies of SNP markers. Bioinformatics 20(3):439–443

Perez-Lezaun A, Calafell F, Mateu E, Comas D, Bosch E, Bertranpetit J (1997) Allele frequencies for 20 microsatellites in a worldwide population survey. Hum Hered 47:189–196

Sham P (2001) Shifting paradigms in gene-mapping methodology for complex traits. Pharmacogenomics 2(3):195–202

Soares-Vieira JA, Billerbeck AE, Pinto EM, Iwamura ES, Bilharinho de Mendonca B, Otto PA (2002) Population genetics of nine short tandem repeat loci: allele frequency distribution in a Brazilian population sample. Am J Forensic Med Pathol 23(2):186–190

Thaker GK, Carpenter WT Jr (2001) Advances in schizophrenia. Nat Med 7:667–671

Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506–513

Wang W, Jia H, Wang Q, Cai Z, Wei L, Wang D, Bittles AH (2003) STR polymorphisms of "forensic loci" in the northern Han Chinese population. J Hum Genet 48(7):337–341