Am. J. Hum. Genet. 70:625–634, 2002

Mutation Patterns at Dinucleotide Microsatellite Loci in Humans

Qing-Yang Huang,[1,2] Fu-Hua Xu,[1,2] Hui Shen,[1,2] Hong-Yi Deng,[1,2] Yong-Jun Liu,[1,2] Yao-Zhong Liu,[1,2] Jin-Long Li,[1,2] Robert R. Recker,[1] and Hong-Wen Deng[1,2,3]

[1]Osteoporosis Research Center and [2]Department of Biomedical Sciences, Creighton University, Omaha, and [3]Laboratory of Molecular and Statistical Genetics, College of Life Sciences, Hunan Normal University, Changsha, Hunan, Peoples' Republic of China

Microsatellites are a major type of molecular markers in genetics studies. Their mutational dynamics are not clear. We investigated the patterns and characteristics of 97 mutation events unambiguously identified, from 53 multigenerational pedigrees with 630 subjects, at 362 autosomal dinucleotide microsatellite loci. A size-dependent mutation bias (in which long alleles are biased toward contraction, whereas short alleles are biased toward expansion) is observed. There is a statistically significant negative relationship between the magnitude (repeat numbers changed during mutation) and direction (contraction or expansion) of mutations and standardized allele size. Contrasting with earlier findings in humans, most mutation events (63%) in our study are multistep events that involve changes of more than one repeat unit. There was no correlation between mutation rate and recombination rate. Our data indicate that mutational dynamics at microsatellite loci are more complicated than the generalized stepwise mutation models.

Introduction

Microsatellites, also known as "short sequence repeats" (SSRs), are short (2–6 bp), tandemly repeated DNA sequences that are ubiquitously interspersed in eukaryotic genomes (Tautz and Renz 1984; Goldstein and Schlötterer 1999). Because of their high variability, codominance, and ubiquity, microsatellites have been exploited extensively as DNA markers for genetic mapping (Dib et al. 1996; Broman et al. 1998), for linkage analyses (Deng et al. 2001), and for population and evolutionary genetics studies (Di Rienzo et al. 1994, 1998; Cooper et al. 1999; Goldstein and Schlötterer 1999).

Despite the widespread use of microsatellites, their evolutionary dynamics are still poorly understood. Microsatellite variation appears to be a complex phenomenon that is influenced by DNA slippage, mismatch-repair efficiency, selection, length constraint, and other factors. Factors influencing microsatellite mutation rates include repeat number, repeat type, flanking sequence, recombination rate, sex, and age (reviewed by Schlötterer 2000). Unraveling the mutational dynamics of microsatellites is crucial in a number of research fields. Microsatellites are increasingly used for population and evolutionary studies (Di Rienzo et al. 1994, 1998; Cooper et al. 1999; Goldstein and Schlötterer 1999; Gonser et al. 2000). Analysis of population divergence and interspecific phylogenetic reconstruction require specification of a distance measure. A number of distance measures have been proposed specifically for microsatellite data (Goldstein et al. 1995; Slatkin 1995). These distances are based on the assumption that microsatellites roughly follow the stepwise mutation model (SMM). This assumption is critical for the accuracy in the distance computation. In addition, instability of microsatellites is implicated in a number of human genetic disorders (Ashley and Warren 1995; Sutherland and Richards 1995; Rubinsztein 1999). Hence, elucidation of the mutational dynamics would help us in understanding the molecular mechanism of these genetic diseases.

The predominant mutation mechanism of microsatellites is DNA replication slippage. The gain/loss of repeat units in a microsatellite is assumed to be caused by strand displacement of the nascent DNA strand, followed by an out-of-register pairing (Levinson and Gutman 1987). The simplest model (i.e., the SMM) assumes that the size of a microsatellite changes by only one repeat unit per mutation. Moreover, the probabilities of addition and deletion are identical and constant across alleles (Ohta and Kimura 1973). Evidence based on population distribution of alleles and pedigree analyses shows that most mutations are compatible with this simplest model (Weber and Wong 1993; Kayser et al. 2000), but deviations from it are not exceptional (Valdes et al. 1993; Di Rienzo et al. 1994, 1998; Amos and Rubinstzein 1996; Primmer et al. 1996; Colson and Goldstein 1999; Nielsen and Palsbøll 1999; Palsbøll et al. 1999; Ellegren 2000*b*; Xu et al. 2000). In recent

Received October 1, 2001; accepted for publication December 3, 2001; electronically published January 15, 2002.

Address for correspondence and reprints: Dr. Hong-Wen Deng, Osteoporosis Research Center, Creighton University, 601 North 30th Street, Suite 6787, Omaha, NE 68131. E-mail: deng@creighton.edu

0002-9297/2002/7003-0008$15.00

years, several modifications of the basic SMM have been made. Such modifications included introduction of a two-phase stepwise mutation model (TPM) (Di Rienzo et al. 1994), allowance for mutation-rate variation between loci and introduction of allele-length ceiling (Feldman et al. 1997), biased mutational-process models (Kimmel et al. 1996; Zhivotovsky et al. 1997), and size limitation by deletion or mutation (Garza et al. 1995; Kruglyak et al. 1998, 2000).

Since mutations are generally rare, it is often difficult to get detailed insight into the patterns and characteristics of mutation events. The most straightforward and conclusive way to study microsatellite mutations is the direct detection of mutation events in pedigree genotyping that may yield detailed information about mutation process. When this approach is used, in humans, autosomal microsatellite mutation rates are estimated to be in the range of 10^{-2} to 10^{-4} (Weber and Wong 1993; Brinkmann et al. 1998; Henke and Henke 1999; Sajantila et al. 1999; Xu et al. 2000), and Y-chromosomal microsatellite mutation rates are estimated to be 2×10^{-3} to 4×10^{-3} (Heyer et al. 1997; Bianchi et al. 1998; Sajantila et al. 1999; Kayser et al. 2000; Holtkemper et al. 2001). Our purpose here is to investigate mutation patterns and characteristics at 362 autosomal dinucleotide microsatellite loci in humans.

Subjects and Methods

Subjects

The study was approved by the Creighton University institutional review board. Fifty-three pedigrees with 630 subjects (248 males and 382 females) from two to four generations were genotyped. Each pedigree was ascertained through a proband having bone mineral density $\leqslant -1.28$ at the hip or spine. Only healthy people were included in the study. The exclusion criteria were detailed by Deng et al. (2001). The subjects were whites of European origin. All of the study subjects signed informed-consent documents before entering the project.

Genotyping

For each subject, blood (20 cc) was drawn into lavender cap (EDTA-containing) tubes and were stored chilled (at ~4°C). DNA was extracted by means of a Puregene DNA Isolation Kit (Gentra Systems) following the procedures detailed in the kit. DNA was genotyped using fluorescently labeled markers. The 400 dinucleotide microsatellite markers we started with are commercially available through Perkin Elmer Applied Biosystems, have an average population heterozygosity of ~0.79, and are spaced, on average, ~8.6 cM between adjacent markers throughout the human genome (ABI PRISM Linkage Mapping Sets Version 2). The PCR was performed on PE 9700 thermocyclers (GeneAmp PCR System 9700, Applied Biosystems). PCR cycling conditions followed those suggested in the ABI PRISM Linkage Mapping Sets Version 2. Genotyping was performed using an Applied Biosystems automated DNA-sequencing system (Model 377; Perkin Elmer-ABI) running the Genescan and GENOTYPER software programs for allele identification and sizing. GenoDB, a genetic database management system developed by us (Li et al. 2001), was employed for allele binning (including establishing allele-binning criteria and converting allele sizes to distinct allele numbers) and data formatting for PedCheck (O'Connell and Weeks 1998). Three hundred sixty-two autosomal dinucleotide microsatellite markers were successfully genotyped. The GenoDB and the associated experimental and data-analysis procedures (including marker labeling and allele binning) are very efficient in generating high-throughput and high-quality genotype data (Li et al. 2001). The rate of missing and erroneous genotyping data, after multiple rounds of repeat work and checking of the conformation to Mendelian inheritance patterns within pedigrees, is ~0.3%. This low genotyping-error rate assumes that the genotype data used are of high accuracy and high quality.

Among the 362 autosomal microsatellites, the repeat sequences of D2S206, D7S798, D10S249, D11S987, D11S968, D13S217, D13S218, D15S165, D17S849, and D22S315 are not available in the GenBank database. Among these markers, only D7S987 and D22S315 had mutations detected in this study. All other microsatellites markers were of CA repeats. Therefore, we were unable to evaluate the impact that the sequence of the dinucleotide repeat had on the expansion and deletion rates, because of the limited data.

Identification of Mutations and Their Origin

We checked each of the 53 pedigrees for Mendelian discrepancies, using the PedCheck software package (see StatGen Web site). For each Mendelian discrepancy found, we pursued repeated genotyping at least two additional times for both parents and offspring involved in mutation for confirmation. We constructed the family haplotypes, using SimWalk2 (see StatGen Web site). The parental origin and the mutating allele were identified by use of the haplotype information. If alleles of the same length were transmitted from both parents that are equally possible as potentially mutating alleles, then the parental origin of mutation was treated as unknown. When neither of the two alleles from the predicted haplotypes of mutated offspring was consistent with the observed offspring's genotype, the mutation was regarded as uninformative. In three cases, the same mutation was observed in more than one offspring, a fact that may be due to germline mosaicism (Xu et al. 2000);

to be conservative in the counting of mutation events, only one mutation was counted. To avoid pitfalls that may arise from the existence of possible null alleles, at loci where mutation was observed in more than one offspring, and both parent and offsprings were homozygous, only mutations in parents heterozygous for two amplifying alleles were considered (Ellegren 2000*a*). There are eight cases (five loci) in which null alleles exist in this study.

Analysis of Mutation Data

To facilitate data analyses across microsatellite loci, standardized allele sizes were derived as by Xu et al. (2000) and Ellegren (2000*b*), on the basis of the distribution of allele frequencies. Standard allele size is defined as the sum of the population frequencies of all shorter alleles and one-half the frequency of the allele of interest. Standardized allele size ranges from 0 to 1, with alleles close to 0 being the shortest and those close to 1 being the longest (Ellegren 2000*b*; Xu et al. 2000). For some mutation events, it was impossible to obtain information on all factors addressed and analyzed in this study. Thus, the number of observations does not always add up to 97 in individual categories (see table 1). Recombination rates at different microsatellite markers in the human genome (table 2) were based on those of Payseur and Nachman (2000), calculated separately for the GB4 radiation hybrid map and Morton's map by use of the whole-chromosome method (see Human Recombination Rates Web site).

Results

At 362 autosomal dinucleotide microsatellite loci, we unambiguously identified 97 verified mutations (table 2). The 97 verified mutations were distributed at 68 marker loci and 19 chromosomes. The number of mutations varied considerably among chromosomes. For the markers we employed, chromosomes 3 and 7 have the largest number of mutations (11 each), followed by chromosome 2 (10). We did not observe any mutation on chromosomes 10, 11, and 21. Although we found several chromosomal regions where mutations occur more frequently than others (e.g., D2S325–D2S338, D3S1279–D3S1262, D6S262–D6S308, D7S515–D7S530, and D13S153–D13S170), *t*-tests revealed no significant difference in recombination rates for mutating loci versus all the microsatellite loci on the GB4 map or Morton's map. On the GB4 map, the mean recombination rate of the genomic regions of all the microsatellite markers was 1.46 cM/Mb (Payseur and Nachman 2000), and that of our mutating loci was 1.56 cM/Mb. The difference is not significant (*t*-test, $P = .258$). On Morton's map, the mean recombination rate of the genomic regions of all the microsatellite markers was 1.37 cM/Mb (Payseur and Nachman 2000), and that of mutating loci was also 1.37 cM/Mb. Of the 68 mutating microsatellite loci, we observed one mutation at 45 loci, two mutations at 18 loci, three mutations at 4 loci (D3S1614, D5S400, D7S636, and D16S423), and four mutations at 1 locus (D6S308). Of the 87 mutations whose parental origin was unambiguously determined, the total number of paternal mutations was 40, and that of maternal mutations was 47 (table 1). Hence, there seems to be no sex difference regarding mutation rates at dinucleotide microsatellite loci in humans ($\chi^2 = 0.41$; $P = .52$).

The distribution of the mutation steps according to various categories is shown in table 1. One striking feature of our data is the frequent multistep mutation events (63%), contrasting with the previous studies in humans, which found that <15% of the mutations were multistep (Brinkmann et al. 1998; Kayser et al. 2000; Xu et al. 2000). In addition, several mutations involving more than five steps were observed. The distribution of the mutations according to step numbers roughly follows a negative exponential distribution, so that the number of mutation events decreases with the mutation steps involved.

Distribution of the mutation counts by standardized

Table 1

Distribution of the Mutation Steps

	No. of Mutation Events Observed for Step										
Mutation Events	1	2	3	4	5	6	7	8	9	10	Total
All	36	25	13	7	4	7	2	1	0	2	97
Paternal origin	16	12	2	4	0	3	1	0	0	2	40
Maternal origin	13	12	11	3	3	4	1	0	0	0	47
Expansion	12	10	5	4	2	3	0	1	0	0	37
Contraction	24	15	8	3	2	4	2	0	0	2	60

Note.—Step *n* involves changes of *n* repeat units in mutations. Since it was impossible to obtain information on all factors analyzed in this study, as noted in the Subjects and Methods section, the number of observations does not always add up to 97 in each category. Notably, most mutation events (63%) are multistep mutations, and several mutations involving more than five steps were observed.

Table 2

97 Microsatellite Mutations, as Revealed by Haplotype Analyses

Chromosome and Marker	Parental Origin	Contraction vs. Expansion[a]	Mutation Step	Mutant Allele	Standardized Allele Size	Recombination Rate on Map	
						GB4	Morton's
1:							
D1S238	Mother	−	1	3	.17		.87
D1S214	Father	−	7	9	.897	2.53	1.29
D1S218	Father	−	1	6	.415	1.21	.86
D1S230	Mother	−	2	5	.473		.95
D1S230	Mother	−	3	5	.473		.95
2:							
D2S126	Mother	+	2	8	.415		.87
D2S142	Father	+	4	1	.107	1.19	.75
D2S325	Mother	+	6	7	.516		.82
D2S325	Father	+	1	3	.17		.82
D2S2211	Mother	−	2	6	.669	2.25	
D2S335	Father	+	4	3	.151	1.43	
D2S335	Mother	+	1	6	.508		
D2S338	Mother	+	2	6	.456	2.71	.94
D2S396	Father	+	1	4	.246	2.42	.91
D2S396		+	1	6	.571	2.42	.91
3:							
D3S1263	Mother	+	5	2	.09	2	1
D3S1614	Mother	+	2	8	.657	1.18	1.11
D3S1614	Mother	+	3	3	.01	1.18	1.11
D3S1614	Father	+	4	5	.077	1.18	1.11
D3S3681	Mother	−	6	9	.735		
D3S1262	Mother	−	4	8	.948	1.48	1.26
D3S1262		−	1	8	.948	1.48	1.26
D3S1304		−	1	7	.508	2.14	1.02
D3S1304	Mother	−	1	8	.747	2.14	1.02
D3S1267	Mother	−	7	16	.986	.83	.95
D3S1279	Mother	−	5	9	.929	.98	1.06
4:							
D4S405	Mother	−	1	9	.883	.47	1.09
D4S1535	Father	+	2	4	.584		1.21
D4S402	Father	−	10	18	.865	1.14	.93
5:							
D5S400	Mother	−	2	9	.86		.85
D5S400	Mother	+	3	8	.678		.85
D5S400	Father	−	6	6	.263		.85
D5S407	Father	−	2	10	.902		.88
D5S422	Father	−	1	10	.767	1.31	.8
D5S408	Mother	−	1	6	.519	1.5	.89
6:							
D6S262	Mother	−	2	8	.631	1.32	.83
D6S262	Father	+	1	3	.16	1.32	.83
D6S257	Father	−	3	15	.993	.8	.75
D6S308	Mother	−	3	4	.615	1.66	1
D6S308	Father	+	2	1	.109	1.66	1
D6S308	Mother	−	3	4	.615	1.66	1
D6S308	Mother	−	3	4	.615	1.66	1
7:							
D7S513	Mother	+	2	7	.404		
D7S530	Father	−	2	5	.668	1.32	1.06
D7S486	Father	−	1	6	.478	1.05	.98
D7S486	Father	−	1	8	.904	1.05	.98
D7S493	Father	+	2	10	.577	1.33	1.22
D7S515		−	1	10	.612		.82
D7S515		−	1	12	.949		.82
D7S636	Mother	−	1	15	.897		
D7S636		−	1	13	.706		

(*continued*)

Table 2 (continued)

Chromosome and Marker	Parental Origin	Contraction vs. Expansion[a]	Mutation Step	Mutant Allele	Standardized Allele Size	Recombination Rate on Map	
						GB4	Morton's
D7S636	Mother	−	1	13	.706		
D7S798	Mother	−	1	6	.344		1.52
8:							
D8S264	Mother	−	3	8	.683	3.27	2.01
9:							
D9S171	Father	−	1	10	.979	1.13	1.32
D9S171	Father	−	6	9	.891	1.13	1.32
D9S157	Mother	+	5	2	.225	1.29	1.49
12:							
D12S310	Mother	+	3	4	.082	1.67	1.25
D12S346	Mother	+	4	7	.255	1.41	1.04
D12S346	Father	−	10	17	.995	1.41	1.04
D12S351	Father	−	1	6	.905	1.11	.92
13:							
D13S170	Mother	−	2	12	.966	1.19	.97
D13S153	Father	+	1	5	.423	.77	.58
D13S156	Father	+	1	9	1	1.02	.67
D13S156	Father	−	1	6	.773	1.02	.67
D13S1265	Father	−	1	11	.49		
14:							
D14S276	Father	−	4	6	.936		4.42
D14S280		+	2	3	.312	2.27	
D14S280	Mother	−	1	6	.976	2.27	
D14S288	Mother	+	6	6	.269	1.03	3.73
15:							
D15S978		+	8	4	.427		.98
D15S1002	Mother	+	1	7	.455		
16:							
D16S423	Father	−	2	9	.864		1.37
D16S423	Mother	+	6	3	.348		1.37
D16S423	Mother	+	3	6	.67		1.37
17:							
D17S787	Father	−	1	7	.767	.6	1.21
D17S785	Father	−	2	8	.875	1.53	
D17S921	Mother	+	1	2	.208	1.25	1.64
D17S921	Mother	+	2	2	.208	1.25	1.64
18:							
D18S1102	Mother	+	1	6	.452		
D18S53	Father	−	2	6	.718		1.76
D18S63	Father	−	1	12	.956	3.12	2.89
D18S474	Father	−	3	9	.734	1.13	.66
D18S474	Father	+	1	10	.796	1.13	.66
D18S1161		−	5	6	.357	3.51	2.46
19:							
D19S418	Father	−	2	4	.433	3.37	2.72
D19S220	Mother	+	3	7	.34		1.26
D19S221	Mother	+	1	3	.022	1.41	.55
D19S221	Mother	−	2	10	.903	1.41	.55
20:							
D20S173	Father	+	2	8	.827	2.25	6.7
D20S178	Mother	−	4	5	.71		3.59
D20S186	Father	−	2	9	.668	1.24	1.12
D20S889	Father	−	2	18	.993		2.52
D20S889	Father	−	6	14	.902		2.52

(*continued*)

Table 2 (continued)

Chromosome and Marker	Parental Origin	Contraction vs. Expansion[a]	Mutation Step	Mutant Allele	Standardized Allele Size	Recombination Rate on Map	
						GB4	Morton's
22:							
D22S315		−	1	10	.527		2.95
D22S315	Mother	−	2	10	.527		2.95
D22S420	Mother	−	3	7	.919	1.35	2.75

Note.—Mutant alleles are numbered within loci sequentially, according to their relative size. Blank cells indicate uncertain or unknown information.

[a] + = expansion; − = contraction.

allele sizes is shown in figure 1. We can see that: (1) The number of mutant alleles that had lost repeat units exceeded the number with gains (60 losses vs. 37 gains); the difference is significant ($\chi^2 = 4.99$; $P = .025$). Summing the entire repeat gains and losses, there was a pure loss of 59 repeat units. (2) Mutation shows a nonrandom distribution among alleles within loci. Mutations at alleles longer than a standardized allele size of 0.5 are significantly more common than mutations at shorter alleles (61 vs. 36; $\chi^2 = 5.94$; $P = .015$). (3) In the standardized allele size range of 0.4–0.6, the rates of expansion and contraction are almost equal. Mutations in alleles shorter than 0.4 are biased toward expansion, whereas mutations in alleles longer than 0.6 are biased toward contraction. Notably, there is a significant negative relationship between the magnitude and the direction of microsatellite mutations and standardized allele sizes ($r^2 = 0.34$; $P < .001$) (fig. 2), reflecting a size-dependent mutation bias in the magnitude and direction.

There were a total of 1,380 parent-offspring relationships analyzed. The mutation rate in our study is 1.94×10^{-4} (i.e., 97/499,560). It is within the lower range of the previous estimates of the microsatellite mutation rate, which range from 10^{-2} to 10^{-4} (Weber and Wong 1993; Brinkmann et al. 1998; Henke and Henke 1999; Sajantila et al. 1999; Xu et al. 2000). Our estimate of the mutation rate was conservative and an underestimate. Uninformative mutations (defined earlier) were not counted, and some pedigrees are not complete (e.g., are without parents) in our genotyping. Hence, in this study, we choose to study only informative mutations, and we focus on mutation process, rather than on the mutation rate.

Discussion

Size-Dependent Mutational Bias and Microsatellite Constraints

Whether microsatellites evolve directionally has been the focus of much controversy (Ellegren 2000*a*). Several studies have suggested that microsatellite mutations are biased toward expansion (Amos et al. 1996; Primmer et al. 1996, 1998; Brinkmann et al. 1998; Cooper et al. 1999; Ellegren 2000*b*; Kayser et al. 2000). We did not observe an overall bias favoring expansion. Instead, there is a size-dependent mutational bias. Within a locus, longer alleles lose repeat units more often than do shorter alleles, and shorter alleles tend to gain repeat units more often than do longer alleles (fig. 1). The overall result was a net loss of 59 repeat units in our data, contrasting with earlier results (Amos et al. 1996; Primmer et al. 1996, 1998; Cooper et al. 1999; Ellegren 2000*b;* Kayser et al. 2000). Two large family studies of autosomal microsatellites also show a slightly higher (but not significant) number of losses versus gains of repeats (Brinkmann et al. 1998; Sajantila et al. 1999). The observed bias toward repeat expansion is problematic in the context of a length ceiling, since this bias would lead to infinite growth (Ellegren 2000*a*). A few studies indicated that, in other species, contraction might be more common among long alleles than short alleles (Wierdl et al. 1997; Schlötterer et al. 1998; Ellegren 2000*b;* Harr and Schlötterer 2000), consistent with our observation in humans for dinucleotide microsatellite loci. Xu et al. (2000) found that the rate of contraction mutation increases exponentially with allele size at human tetranucleotide repeat loci, whereas the rate of expansion mutation is constant across the entire allele distribution. However, our results indicated not only allele size–dependent contraction rate but also the allele size–dependent expansion rate, providing the first piece of empirical evidence for the latter in humans.

A long-standing question in the evolution of microsatellite loci is what mechanisms prevent infinite growth (Ellegren 2000*a;* Schlötterer 2000). Our data set suggests that, within loci, long alleles are more prone to decrease in size than shorter alleles. There is a strong, statistically significant negative relationship between the magnitude of microsatellite mutation (magnitude and direction) and standardized allele size (fig. 2), supporting the theory of biased mutations as a potential mechanism for the ob-

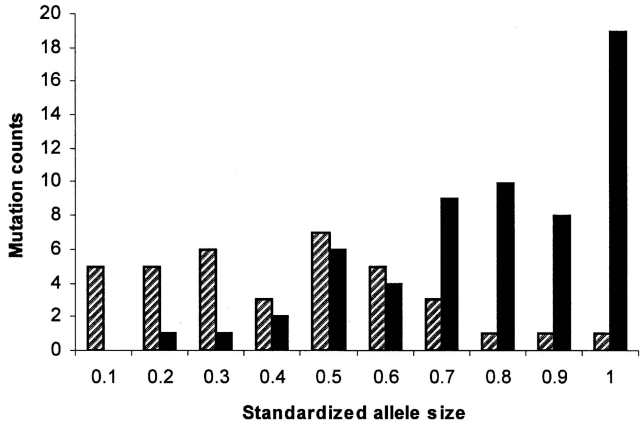

Figure 1 Distribution of microsatellite mutation events by standardized allele sizes. Hatched bars are for expansions, and filled bars are for contractions. Mutations at alleles longer than a standardized allele size of 0.5 are significantly more common than mutations at shorter alleles (61 vs. 36; $\chi^2 = 5.94$; $P = .015$). In the standardized allele-size range of 0.4–0.6, the rates of expansion and contraction are almost equal. Mutations in alleles shorter than 0.4 are biased toward expansion, whereas mutations in alleles longer than 0.6 are biased toward contraction.

served constraints on microsatellite allele sizes (Zhivotovsky et al. 1997; Falush and Iwasa 1999). A similar pattern has also been seen in the pipefish (Jones et al. 1999) and the barn swallow (Primmer et al. 1998). Selection (Garza et al. 1995) and a balance between point mutations and slippage events (Kruglyak et al. 1998, 2000) have also been used to explain why infinite growth does not occur. A future challenge is to develop new models of microsatellite evolution that may integrate the effects of mutation biases, selection, and the balance between point mutations and slippage events in the evolution of microsatellites.

Distribution of the Mutation Size at Microsatellite Loci

The SMM has often been used to model microsatellite evolution (Valdes et al. 1993; Di Rienzo et al. 1994). Two problems with this model are (1) that it does not converge to a stationary distribution and (2) that it cannot explain the absence of very long alleles (Xu et al. 2000). Although most mutations are compatible with this simplest model, deviations from it also occur. Several studies showed that the observed frequencies of multistep changes is 0%–14% for humans (Weber and Wong 1993; Amos and Rubinstzein 1996; Brinkmann et al. 1998; Ellegren 2000*b;* Kayser et al. 2000; Xu et al. 2000). In this study, we observed a much larger proportion of multistep mutations (63%). These are important observations, because they seriously undermined the assumptions of the classical SMM. Our result is based on dinucleotide repeats in humans. It has yet to be demonstrated, in different species and repeat types, how general these observations are.

Observations of the multistep mutations are not uncommon. Di Rienzo et al. (1994) observed, in humans, allele distributions at 8 of 10 dinucleotide-repeat microsatellite loci that were consistent with the occurrence of multistep mutations when compared with the null expectations under a strict SMM. Nielsen and Palsbøll (1999) estimated the frequency of multistep mutations at nine microsatellite loci in different baleen whale populations and found significant deviations from the null expectations under a strict SSM. Their results are consistent with multistep mutations at two loci. Colson and Goldstein (1999) studied the mutations at 19 microsatellite loci in *Drosophila melanogaster* and found that only seven loci had their size variation among species consistent with the occurrence of strictly stepwise mutations. The observed frequencies of multistep changes are, respectively, 18% at the *HrU9* locus for swallows (Primmer et al. 1996), 46% for Australian lizards (Gardner et al. 2000), 68% for zebrafish (Shimoda et al. 1999), and 74% for green turtles (Fitzsimmons 1998). One reason why multistep mutations are rarely recognized in humans may be that screening for microsatellite mutations is often performed with the pooling of single-locus data from a large number of families, rather than by analysis of a large number of loci in the same families, as has been done here (Kayser et al. 2000).

Computer simulation results suggest that the underlying mutational mechanism for generating new micro-

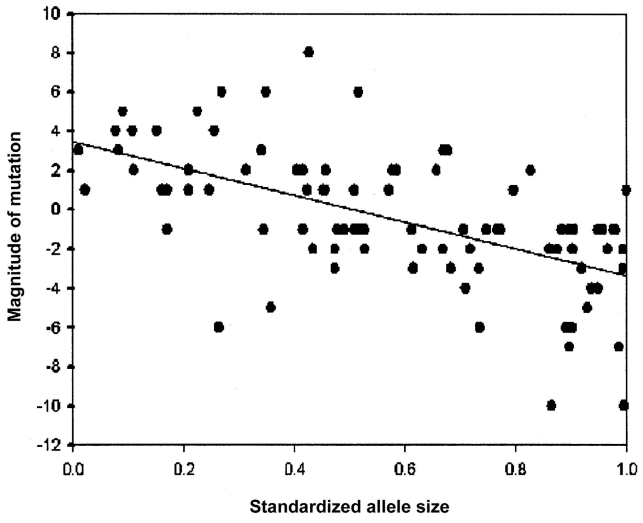

Figure 2 Magnitude (change in the number of repeat units) and direction (positive for expansion and negative for contraction) of human microsatellite mutations in relation to standardized allele sizes. There is a significant negative relationship between the magnitude and the direction of microsatellite mutations and standardized allele sizes ($r^2 = 0.34$; $P < .001$).

satellite alleles is likely to be multistep. The distribution of the size differences between adjacent ranked alleles in the Généthon sample (Dib et al. 1996) is very closely matched by the simulated data under the multiple-step mutation models (MMM), whereas the SMM underestimates the frequency of large differences in size between adjacently ranked alleles (Farrall and Weeks 1998).

The results from these studies and from the present study indicate that many microsatellite mutation events are more complex than is represented by generalized SMMs. It is not clear whether larger mutations reflect the difference of the mutational mechanism, such as unequal crossing-over (Garza et al. 1995; Richard and Pâques 2000), since large jumps in repeat length are uncharacteristic of slippage. The increase in variance of the microsatellite-specific statistics caused by multistep mutations may have a considerable impact on the accuracy of studies of natural populations (Palsbøll et al. 1999). Therefore, one should be cautious when inferring population or phylogenetic relationships from microsatellite size data alone, or microsatellites that more closely match the assumptions of SSM should be used in the distance measures that only assume SMM.

Microsatellite Mutation and Recombination Rate

A positive correlation between recombination rate and nucleotide variability is well established (Begun and Aquadro 1992). However, conflicting results concerning the relationship between microsatellite variation and recombination rate have been reported. Although Schug et al. (1998) documented a strong positive association between microsatellite variation and recombination rate in *D. melanogaster,* other studies failed to confirm this relationship (Michalakis and Veuille 1996; Harr et al. 1998; Payseur and Nachman 2000). In this study, although we found several regions where mutations occur more frequently than in others, there is no significant difference in recombination rate for mutation loci versus all loci in the GB4 map or Morton's map. This finding is consistent with the notion that mutations that reduce or eliminate most types of recombination in *E. coli* and yeast do not affect microsatellite instability (Levinson and Gutman 1987; Wierdl et al. 1997). The mutational mechanism of microsatellites is hence likely to be independent of recombination.

In summary, characterization of 97 mutation events at 68 autosomal dinucleotide microsatellite loci identified from 53 pedigrees through 362 dinucleotide microsatellite loci in humans have provided a number of important insights into the patterns and characteristics of microsatellite mutations in humans. A size-dependent mutation bias may be one of the forces that constrain the variation at microsatellite loci. There was no correlation between mutation rate and recombination rate. However, there is a statistically significant negative relationship between the magnitude and direction of mutation and standardized allele size. Moreover, most mutation events are multistep mutations. The results from the current studies, as well as those of other studies, suggest that many mutation events at microsatellite loci do not follow the SMM. A future challenge is to develop novel models of microsatellite evolution that integrate the effects of mutation biases and the balance between point mutations and slippage events.

Acknowledgments

We thank Jian Chen, Hai-Tao Zhang, Qiang-Hua Xia, Jian Li, Xuan Zhang, Xu-Tao Deng, Dong-Bon Lai, and Yan Zhou for their involvement in genotyping. DNA-extraction efforts by the technicians in the Osteoporosis Research Center of Creighton University are greatly appreciated. The investigators were partially supported by Creighton University and by grants from the Health Future Foundation, the National Institutes of Health (K01 AR02170-01, R01 AR45349-01, R01 GM60402-01A1, and P01 DC01813-07), the State of Nebraska Cancer and Smoking Related Disease Research Program (LB598), the Tobacco Settlement Fund (LB692), the United States Department of Energy (DE-FG03-00ER63000/A00), the National Science Foundation of China, Hunan Normal University, and the Ministry of Education of the People's Republic of China.

Electronic-Database Information

URLs for data in this article are as follows:

Applied Biosystems, http://www.appliedbiosystems.com/ (for microsatellite markers used in genotyping)
GenBank, http://www.ncbi.nlm.nih.gov/
Human Recombination Rates, http://eebweb.arizona.edu/nachman/publications/data/microsats.html
StatGen, http://watson.hgen.pitt.edu/register/soft_doc.html (for Pedcheck and SimWalk2 software)

References

Amos W, Rubinstzein DC (1996) Microsatellites are subject to directional evolution. Nat Genet 12:13–14

Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC (1996) Microsatellites show mutational and heterozygote instability. Nat Genet 13:390–391

Ashley CT, Warren ST (1995) Trinucleotide repeat expansion and human disease. Annu Rev Genet 29:703–728

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356:519–520

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Margnac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, López-Camelo JS (1998) Characterisation of ancestral and derived Y-chromosome haplotypes of new world native populations. Am J Hum Genet 63:1862–1871

Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62:1408–1415

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861–869

Colson I, Goldstein DB (1999) Evidence for complex mutation at microsatellite loci in Drosophila. Genetics 152:617–627

Cooper G, Burroughs NJ, Rand DA, Rubinsztein DC, Amos W (1999) Markov chain Monte Carlo analysis of human Y-chromosome microsatellite provides evidence of biased mutation. Proc Natl Acad Sci USA 96:11916–11921

Deng HW, Xu FH, Conway T, Deng XT, Davies KM, Li JL, HY Deng, Recker RR (2001) Is population bone mineral density variation linked to the marker D11S987 on chromosome 11q12-13? J Clin Endocrinol Metab 86:3735–3741

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines K, Barch DH (1998) Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics 148: 1269–1284

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutation processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Ellegren H (2000*a*) Microsatellite mutatations in the germline: implications for evolutionary inference. Trends Genet 16: 551–558

——— (2000*b*) Heterogeneous mutation processes in human microsatellite DNA sequences. Nat Genet 24:400–402

Falush D, Iwasa Y (1999) Size-dependent mutability and microsatellite constraints. Mol Biol Evol 16:960–966

Farrall M, Weeks DE (1998) Mutational mechanisms for generating microsatellite allele-frequency distributions: an analysis of 4,558 markers. Am J Hum Genet 62:1260–1262

Feldman MW, Bergman A, Pollock DD, Goldstein DB (1997) Microsatellite genetic distances with range contraints: analytic description and problems of estimation. Genetics 145: 207–216

Fitzsimmons NN (1998) Single paternity of clutches and sperm storage in the promiscuous green turtle (*Chelonai mydas*). Mol Ecol 7:575–584

Gardner MG, Bull CM, Cooper SJB, Duffield GA (2000) Microsatellite mutations in litters of the Australian lizard *Egernia stokesii*. J Evol Biol 13:551–560

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Mol Biol Evol 12:594–603

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evalution of genetic distances for use with microsatellite loci. Genetics 139:463–471

Goldstein DB, Schlötterer C (eds) (1999) Microsatellites: evolution and applications. Oxford University Press, London

Gonser R, Donnelly P, Nicholson G, DiRienzo A (2000) Microsatellite mutations and inferences about human demography. Genetics 154:1793–807

Harr B, Schlötterer C (2000) Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genomewide underrepresentation. Genetics 155:1213–1220

Harr B, Zangerl B, Brem G, Schlöterer C (1998) Conservation of locus specific microsatellite variability across species: a comparison of two *Drosophila* sibling species *D. melanogaster* and *D. simulans*. Mol Biol Evol 15:176–184

Henke L, Henke J (1999) Mutation rate in human microsatellites. Am J Hum Genet 64:1473

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799–803

Holtkemper U, Rolf B, Hohoff C, Forster P, Brinkmann B (2001) Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. Hum Mol Genet 10:629–633

Jones AG, Rosenqvist G, Berglund A, Avise JC (1999) Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. Genetics 152:1057–1063

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588

Kimmel M, Chakraborty R, Stivers R, Deka R (1996) Dynamics of repeat polymorphisms under a forward-backward model: within- and between-population variability at microsatellite loci. Genetics 143:549–555

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA 95:10774–10778

——— (2000) Distribution and abundance of microsatellite in the yeast genome can be explained by a balance between slippage events and point mutations. Mol Biol Evol 17:1210–1219

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Li J, Deng H, Lai D, Xu F, Chen J, Gao G, Recker R, Deng H (2001) Toward high-throughput genotyping: dynamic and automatic software for manipulating large-scale genotype data using fluorescently labeled dinucleotide markers. Genome Res 11:1304–1314

Michalakis Y, Veuille M (1996) Length variation of CAG/CAA trinucleotide repeats in natural population of *Drosophila melanogaster* and its relation to the recombination rate. Genetics 143:1713–1725

Nielsen R, Palsbøll PJ (1999) Single-locus tests of microsatellite evolution: Multistep mutations and constraints on allele size. Mol Phylogenet Evol 11:477–484

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266

Ohta T, Kimura M (1973) The model of mutation appropriate to calculate the number of electrophoretically detectable alleles in a genetic population. Genet Res 22:201–204

Palsbøll PJ, Bérubé M, Jørgensen H (1999) Multiple levels of single-strand slippage at cetacean tri- and tetranucleotide repeat microsatellite loci. Genetics 151:285–296

Payseur BA, Nachman MW (2000) Microsatellite variation and recombination rate in the human genome. Genetics 156: 1285–1298

Primmer CR, Ellegren H, Saino N, Møller AP (1996) Directional evolution in germline microsatellite mutations. Nat Genet 13:391–393

Primmer CR, Saino N, Møller AP, Ellegren H (1998) Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. Mol Biol Evol 15:1047–1054

Richard G, Pâques F (2000) Mini- and microsatellite expansions: the recombination connection. EMBO Rep 1:122–126

Rubinsztein DC, Amos B, Cooper G (1999) Microsatellite and trinucleotide-repeat evolution: evidence for mutational bias and different rates of evolution in different lineages. Philos Trans R Soc Lond B Biol Sci 354:1095–1099

Sajantila A, Lukka M, Syvänen A-C (1999) Experimentally observed germline mutations at human micro- and minisatellite loci. Eur J Hum Genet 7:263–266

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. Chromosoma 109:365–371

Schlötterer C, Ritter R, Harr B, Brem G (1998) High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. Mol Biol Evol 15:1269–1274

Schug MD, Hutter CM, Noor MA, Aquadro CF (1998) Mutation and evolution of microsatellites in *Drosophila melanogaster*. Genetica 102/103:359–367

Shimoda N, Knapik EW, Ziniti J, Sim C, Yamada E, Kaplan S, Jackson D, de Sauvage F, Jacob H, Fishman MC (1999) Zebrafish genetic map with 2000 microsatellite markers. Genomics 58:219–232

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462

Sutherland GR, Richards RI (1995) Simple tandem DNA repeats and human genetic disease. Proc Natl Acad Sci USA 92:3636–3641

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127–4138

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769–779

Xu X, Peng M, Fang Z, Xu X (2000) The direction of microsatellite mutations is dependent upon allele length. Nat Genet 24:396–399

Zhivotovsky LA, Feldman LA, Grishechkin SA (1997) Biased mutation and microsatellite variation. Mol Biol Evol 14:926–933