J Hum Genet (2001) 46:448–455

ORIGINAL ARTICLE

Masaki Mizutani · Toshimichi Yamamoto · Kouji Torii
Hirokazu Kawase · Takashi Yoshimoto · Rieko Uchihi
Miwa Tanaka · Keiji Tamaki · Yoshinao Katsumata

Analysis of 168 short tandem repeat loci in the Japanese population, using a screening set for human genetic mapping

Received: March 26, 2001 / Accepted: April 25, 2001

Abstract We devised a multiplex polymerase chain reaction (PCR) amplification and loading system for the convenient typing of 168 short tandem repeat (STR) polymorphic markers in a commercially available screening primer set for human linkage analysis. We genotyped all these 168 STR loci with 32 healthy unrelated Japanese, calculated allele frequencies at each STR locus, and performed three kinds of tests for Hardy-Weinberg equilibrium (HWE). Significant deviations from HWE in all three tests were observed at only three loci, and the average heterozygosity in the Japanese (0.733) was slightly lower than that in Caucasians (0.773). We also examined 32 Caucasians at some selected loci, to be compared with Japanese. Some markers showed greatly different heterozygosities or allelic distributions in Japanese and Caucasian populations. In two groups of STRs, those with and without irregular alleles (or inter-alleles), the former had a higher proportion of bimodal allelic distribution and possessed more alleles per locus than the latter. However, no significant differences in the observed and expected heterozygosities, or in the powers of discrimination, were found between the two groups. The present basic study of allele frequency databases of these STRs will contribute to further applications in forensic science and human genetics.

Key words Short tandem repeat · Polymorphism · Multiplex · Heterozygosity · Allele frequency · Japanese · Population

M. Mizutani · T. Yamamoto (✉) · K. Torii · H. Kawase · T. Yoshimoto · R. Uchihi · M. Tanaka · Y. Katsumata
Department of Legal Medicine and Bioethics, Graduate School of Medicine, Nagoya University, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550, Japan
Tel. +81-52-744-2119; Fax +81-52-744-2121
e-mail: yamachan@med.nagoya-u.ac.jp

K. Tamaki
Department of Legal Medicine, Sapporo Medical University, School of Medicine, Sapporo, Japan

Introduction

Short tandem repeat (STR) polymorphisms are widely used in various scientific fields, such as linkage analyses and diversity studies in human genetics, and personal identification and paternity tests in forensic science (Wijmenga et al. 1990; Schellenberg et al. 1992; Bowcock et al. 1994; Micka et al. 1996; Thomson et al. 1999), because they have relatively high heterozygosities and are more easily genotyped than other chromosomal markers. More than one million STRs (one/2 kb) are widely distributed in the whole human genome (International Human Genome Sequencing Consortium 2001), and most of them consist of dinucleotide repeats, in particular, CA repeats (Gyapay et al. 1994). Although these dinucleotide STRs are most commonly used for genetic linkage mapping, there are some problems in genotyping, caused by their artifactual "stutter" bands, particularly in the sizing of heterozygous samples and/or mixed samples. Therefore, tetranucleotide STRs showing smaller "stutter" bands are mainly used in forensic practice (Micka et al. 1996; Thomson et al. 1999; Yamamoto et al. 1999; Sweet and Hildebrand 1999).

Dubovsky et al. created a set of STRs for the genetic linkage mapping of disease loci by combining highly polymorphic STRs that are distributed evenly throughout the human genome (Dubovsky et al. 1995). The first version of this set included a high percentage (49%) of dinucleotide STRs. The latest version of the screening set, the Human MapPair screening set 8A, contains 168 STR markers, spaced at an average genetic distance of 25 cM throughout the entire human genome. Most of the dinucleotide STRs were replaced by tetranucleotide STRs in this set, and this set consists of a high percentage (83%) of tetranucleotide STRs, allowing greater accuracy in the determination of their repeat sizes.

We investigated allele frequency distributions and heterozygosities at all 168 STR markers in the screening set 8A in the Japanese population. The results could be useful for the genetic mapping of human diseases, for forensic

applications, and for studies of population genetics, such as the origin of the Japanese.

Materials and methods

Blood samples were obtained, with written informed consent, from 32 healthy unrelated Japanese volunteers (undergraduate students at our medical school who lived in or near Nagoya city). DNA was extracted from each sample as described previously (Tamaki et al. 1991). The 168 markers were genotyped using each primer set in the screening set 8A (Research Genetics, Huntsville, AL, USA), each one of which is labeled with any one of three different fluorescent-colored dyes (FAM, HEX, and TET). We divided these primer sets into groups that could be amplified in a single PCR tube (Table 1), and then performed multiplex PCR amplification. The PCR mixture (10 μl) contained 1 × PCR Gold Buffer, 1.5 mM $MgCl_2$, 200 μM of each dNTP, 2.4 pmol of each STR primer set, 0.25 unit AmpliTaq Gold polymerase (Perkin-Elmer, Norwalk, CT, USA), and 10 ng of template DNA. Thermal cycling protocols began with 95°C for 11 min and then 26 cycles at 94°C for 45 s, 57°C for 45 s, and 72°C for 60 s, followed by 72°C for 15 min. One μl of each of the PCR set products was mixed into 24.5 μl of deionized formamide and 0.5 μl of GeneScan-500 [TAMRA] Size Standard (PE Applied Biosystems, Foster City, CA, USA), according to the loading panel shown in Table 1. For example, for loading panel 1, 1 μl of each of the PCR set 1, 2, 3, and 4 products was mixed together into the formamide and standard mixture. Allels were separated in the POP-4 polymer (PE Applied Biosystems) with the GS STR POP 4 (1 mL) C Module, by capillary electrophoresis, using the Genetic Analyzer 310 (PE Applied Biosystems). Allele sizes for each STR were determined on the basis of sizing data analyzed by GeneScan Analysis software (PE Applied Biosystems). For some representative loci, we also examined 64 Caucasian chromosomes to calculate allele frequencies at each locus, and compared each allele frequency in Japanese and Caucasians, using the software GENEPOP 3.1b (Raymond and Rousset 1995). The Caucasian DNA samples (from the United Kingdom) were kindly provided by Dr. Yuri E. Dubrova, at the University of Leicester.

Tests for Hardy-Weinberg equilibrium (HWE) were carried out using a homozygosity test (Weir 1992), a likelihood ratio test (Chakraborty et al. 1991), and an exact test (Guo and Thompson 1992). As statistical properties, the observed heterozygosity, the unbiased estimates of expected heterozygosity (Edwards et al. 1992), and the power of discrimination (Fisher 1951) were calculated. Heterozygosities in Caucasians were obtained from the Research Genetics website (http://www.resgen.com), which does not give details about the source of these samples. The sequence data for each STR were obtained from the Genome Data base (GDB; http://gdbwww.gdb.org), the Cooperative Human Linkage Center (CHLC; http://lpg.nci.nih.gov/CHLC) and other organizations, and served for our reference to determine sizes at each STR locus.

Results and discussion

The 168 STR markers were grouped into 103 temporary multiplex PCR sets for genotyping by multiloading performed 21 times (Table 1). From the result of sizing, we temporarily named the shortest allele observed in the Japanese population at each locus as allele 1, and the larger alleles following at every repeat as alleles 2, 3, 4, 5, . . . in order, taking into consideration the sequence data from the websites noted above. The irregular alleles observed were named according to an established recommendation (Gill et al. 1997). Using this temporary nomenclature for alleles, we genotyped and calculated allele frequencies at each locus. From these frequencies, three kinds of tests (the homozygosity test, likelihood ratio test, exact test) were performed to check deviation from HWE. The allele frequencies deviated from HWE (P values fell to less than 0.05) for 8 loci in the homozygosity test, 11 loci in the likelihood ratio test, and 13 loci in the exact test. However, it was only at 3 loci (D7S3051, D19S586, and D1S1588) that significant deviations from HWE were observed with all three tests. Furthermore, of the 8 loci whose P values were less than 0.05 in the homozygosity test, 5 showed very low observed heterozygosities (<0.60), resulting in large differences between the observed and expected heterozygosities. These results may have been caused by a contingent event, sampling error, or population substructure. An investigation using more samples would, presumably, explain this result.

Table 1 summarizes the size ranges, numbers of alleles, statistical properties of the 168 markers in the Japanese population, and the observed heterozygosity in Caucasians obtained from the Research Genetics website. The observed heterozygosity at 158 autosomal STRs ranged from 0.38 to 0.94 in the Japanese population, varying more widely than among Caucasians. The mean observed heterozygosity was 0.733, slightly below that in Caucasians (0.773). In some markers, the heterozygosity in the Japanese population was very different from that in the Caucasian population. For example, heterozygosities at D1S1588 and D18S481 were 0.38 and 0.91 in Japanese, against 0.68 and 0.76, respectively, in Caucasians. Spearman's correlation coefficient (Siegel 1956) between Japanese and Caucasians was 0.372, indicating considerably different allele distributions in the two populations. These results suggest the necessity of making a database for each population. We also calculated the expected heterozygosity at each locus, and found a mean value of 0.747. The variance of the expected heterozygosities (SD, 0.086) was significantly smaller than that of the observed heterozygosities (SD, 0.111). Figure 1 summarizes the distribution of the observed and expected heterozygosities in Japanese and the observed heterozygosity in Caucasians. Comparison of the Japanese allelic distributions with those of Caucasians at some representative loci showed a significant difference ($P < 0.05$) at almost all loci, using the software GENEPOP 3.1b (examples are shown in Fig. 2). Two alleles (alleles 1 and 2) of the D7S1823 locus are much more frequent in Japanese

Table 1. Multiplex PCR sets and statistical properties at each locus in the Japanese population

Loading Panel	PCR set	Locus	Marker	Repeat unit	Irregular allele	Size range	Allele	PD	Expected H.	Observed H.	Observed H. (Caucasian)
		D3S2418	ATA22E01	Tri		80–110	7	0.81	0.66	0.59	0.71
	1	D4S2361	ATA2A03	Tri		149–164	6	0.87	0.75	0.75	0.74
		D1S2134	GATA72H07	Tetra		253–289	5	0.86	0.74	0.81	0.84
		D5S1456	GATA11A11	Tetra		188–208	6	0.88	0.77	0.78	0.78
0	2	D6S1056	GATA68H04	Tetra		232–256	7	0.92	0.85	0.94	0.85
		D20S481	GATA47F05	Tetra		235–247	4	0.81	0.66	0.66	0.83
	3	D10S1239	GATA64A09	Tetra		155–179	5	0.78	0.60	0.59	0.75
		D3S2387	GATA22G12	Tetra	+	170–214	11	0.94	0.86	0.84	0.86
	4	D18S481	AFM321xc9	Di		181–193	7	0.87	0.78	0.91	0.76
	5	D2S1400	GGAA20G10	Tetra		108–116	3	0.64	0.52	0.59	0.66
		D1S1679	GGAA5F09	Tetra		143–171	8	0.92	0.85	0.81	0.84
	6	D7S3051*	GATA137H02	Tetra		144–176	9	0.89	0.79	0.59	0.75
1		D17S1308	GTAT1A05	Tetra		294–310	5	0.75	0.60	0.56	0.66
	7	D7S2846	GATA31A10	Tetra		169–189	6	0.86	0.72	0.69	0.76
		D10S677	GGAA2F11	Tetra		197–221	7	0.93	0.84	0.78	0.81
	8	D11S4464	GATA64D03	Tetra		226–250	7	0.90	0.76	0.66	0.78
		D1S1660	GATA48B01	Tetra		227–247	6	0.88	0.81	0.88	0.78
		D11S1999	GATA23F06	Tetra		104–132	8	0.87	0.70	0.72	0.80
	9	D14S617	GGAA21G11	Tetra		136–174	7	0.89	0.76	0.72	0.78
		D3S2460	GATA68F07	Tetra		154–166	4	0.83	0.69	0.72	0.76
		D15S655	ATA28G05	Tri		133–148	4	0.73	0.59	0.66	0.72
2	10	D7S1823	GATA30D09	Tetra		205–233	8	0.92	0.82	0.91	0.85
		D5S1505	GATA62A04	Tetra		246–270	7	0.92	0.84	0.78	0.80
	11	D5S2500	GATA67D03	Tetra	+	151–175	8	0.87	0.76	0.72	0.82
		D5S820	GATA6E05	Tetra		182–206	6	0.91	0.78	0.72	0.77
	12	D5S1725	GATA89G08	Tetra		184–208	7	0.88	0.79	0.75	0.77
	13	D16S2616	ATA41E04	Tri		114–132	6	0.82	0.64	0.56	0.69
		D6S474	GATA31	Tetra		149–165	5	0.81	0.66	0.47	0.77
	14	D18S858	ATA23G05	Tri		190–205	5	0.80	0.64	0.53	0.75
3		D11S2000	GATA28D01	Tetra	+	196–236	18	0.95	0.92	0.84	0.87
	15	D18S878	GATA7E12	Tetra		158–186	7	0.91	0.82	0.91	0.76
		D5S2488	ATA20G07	Tri		211–238	7	0.86	0.70	0.66	0.74
	16	D19S591	GATA44F10	Tetra		92–108	5	0.81	0.72	0.72	0.74
	17	D17S1293	GGAA7D11	Tetra	+	264–296	9	0.93	0.85	0.84	0.83
		DXS6789	GATA31F01	Tetra		115–147	5				
	18	D7S1824	GATA32C12	Tetra		162–190	7	0.81	0.68	0.56	0.82
4			GGAAT1B07	Penta		181–196	4				
		DXS9896	GATA124E07	Tetra		186–234	8				
	19	D22S689	GATA21F03	Tetra		203–227	7	0.88	0.74	0.72	0.76
		DYS389	GATA30F10	Tetra		244–256	4				
		D22S683	GATA11B12	Tetra	+	159–203	14	0.92	0.82	0.69	0.90
	20	D19S246	Mfd232	Tetra		182–222	7	0.88	0.74	0.78	0.82
		D15S642	GATA27A03	Tetra	+	196–212	7	0.87	0.74	0.66	0.81
5	21	D19S254	Mfd238	Tetra		112–140	7	0.91	0.79	0.72	0.75
	22	D9S1118	GATA71E08	Tetra		137–173	8	0.91	0.82	0.84	0.81
	23	D22S420	AFM217xf4	Di		144–156	7	0.88	0.75	0.69	0.77
	24	D2S1384	GATA52A04	Tetra		132–156	7	0.86	0.71	0.78	0.80
		D14S606	GATA30A03	Tetra		260–276	5	0.71	0.48	0.53	0.73
	25	D5S807	GATA3A04	Tetra		164–204	6	0.86	0.73	0.63	0.76
6		D8S1128	GATA21C12	Tetra		230–258	7	0.82	0.68	0.72	0.76
	26	D15S643	GATA50G06	Tetra	+	198–224	10	0.94	0.85	0.84	0.86
		D9S934	GATA64G07	Tetra		202–230	8	0.90	0.77	0.81	0.76
	27	D2S1399	GGAA20G04	Tetra		133–177	12	0.94	0.87	0.88	0.80
	28	D2S1356	ATA4F03	Tri		222–249	8	0.91	0.80	0.78	0.76
	29	D14S306	GATA4B04	Tetra		187–211	7	0.91	0.79	0.72	0.79
		D19S586*	GATA23B01	Tetra		231–247	5	0.65	0.63	0.88	0.73
	30	D1S552	GGAT2A07	Tetra		212–256	5	0.81	0.66	0.72	0.72
		D15S822	GATA88H02	Tetra		234–302	16	0.94	0.87	0.88	0.77

Table 1. Continued

Loading Panel	PCR set	Locus	Marker	Repeat unit	Irregular allele	Size range	Allele	PD	Expected H.	Observed H.	Observed H. (Caucasian)
7	31	D21S2052	GATA129D11	Tetra		120–148	8	0.91	0.82	0.84	0.77
	32	D14S1426	GATA136B01	Tetra	+	129–157	11	0.89	0.79	0.81	0.78
	33	D1S1609	GATA50F11	Tetra		175–203	8	0.88	0.77	0.72	0.80
	34	D18S843	ACT1A01	Tri		181–193	5	0.83	0.74	0.84	0.75
	35	D20S480	GATA45B10	Tetra		278–306	8	0.92	0.85	0.91	0.76
		D4S1627	GATA7D01	Tetra		176–200	7	0.91	0.81	0.75	0.81
	36	F13A1	SE30	Tetra	+	178–188	4	0.73	0.57	0.53	0.78
		D1S1665	GATA61A06	Tetra		220–252	7	0.89	0.77	0.72	0.74
	37	D5S2845	GATA134B03	Tetra		137–161	6	0.89	0.76	0.69	0.66
8		D4S2639	GATA90B10	Tetra		158–190	9	0.92	0.82	0.81	0.85
	38	D7S1842	GGAA6D03	Tetra		123–147	7	0.91	0.79	0.72	0.83
	39	D17S1303	GATA64B04	Tetra		221–241	6	0.88	0.77	0.81	0.70
	40	D20S470	GGAA7E02	Tetra		273–313	9	0.95	0.89	0.84	0.87
	41	D12S1045	ATA29A06	Tri		75–90	5	0.86	0.79	0.94	0.80
		D3S2432	GATA27C08	Tetra		131–155	7	0.90	0.79	0.72	0.83
	42	D21S1432	GATA11C12	Tetra		129–145	5	0.83	0.71	0.75	0.63
		D1S1597	GATA27E01	Tetra		156–176	6	0.87	0.75	0.78	0.71
9	43	D16S753	GGAA3G05	Tetra		149–169	6	0.90	0.78	0.72	0.79
		D11S1984	GGAA17G05	Tetra		179–203	7	0.91	0.82	0.78	0.79
	44	D2S1391	GATA65C03	Tetra		117–129	4	0.76	0.63	0.72	0.79
	45	D11S1392	GATA6B09	Tetra		188–216	7	0.86	0.71	0.66	0.77
	46	D11S2359	ATA27C09	Tri	+	210–229	7	0.88	0.76	0.69	0.73
	47	D2S1394	GATA69E12	Tetra		159–175	5	0.85	0.71	0.72	0.70
		D5S1470	GATA7C06	Tetra		163–203	10	0.92	0.81	0.81	0.82
10	48	D11S2371	GATA90D07	Tetra		173–205	6	0.87	0.75	0.78	0.67
	49	D7S2212	GATA87D11	Tetra		187–203	5	0.77	0.58	0.63	0.73
	50	D13S285	AFM309va9	Di		80–106	12	0.94	0.85	0.81	0.81
		D20S171	AFM046xf6	Di		128–140	7	0.87	0.78	0.88	0.78
	51	ACTC	ACTC	Di		64–90	9	0.93	0.81	0.78	0.87
11		D9S158	AFM073yb11	Di		213–225	7	0.83	0.74	0.81	0.69
	52	D8S264	143xd8	Di		121–141	10	0.93	0.84	0.78	0.83
		D3S2427	GATA22F11	Tetra	+	200–250	13	0.95	0.89	0.88	0.87
	53	D1S549	GATA4H09	Tetra		167–191	7	0.89	0.77	0.81	0.77
		D16S764	GATA42E11	Tetra		98–110	4	0.76	0.58	0.56	0.70
	54	D4S2366	GATA22G05	Tetra		116–136	6	0.90	0.80	0.69	0.79
		D16S539	GATA11C06	Tetra		146–162	5	0.88	0.80	0.75	0.76
	55	D2S1328	GATA27A12	Tetra		134–158	7	0.79	0.64	0.56	0.75
12		D3S4545	GATA164B08	Tetra		200–236	6	0.86	0.72	0.66	0.82
	56	D12S372	GATA4H03	Tetra		172–188	5	0.86	0.73	0.75	0.76
		D2S2976	GATA165C07	Tetra	+	194–228	10	0.74	0.48	0.53	0.85
	57	PAH	PAH	Tetra		226–250	6	0.86	0.71	0.59	0.80
		D10S1435	GATA88F09	Tetra		251–271	6	0.86	0.74	0.75	0.75
		D8S373	UT721	Tetra		190–222	8	0.94	0.83	0.72	0.78
	58	D8S1113	GGAA8G07	Tetra		212–232	6	0.80	0.65	0.56	0.81
		DXS7132	GATA72E05	Tetra		280–296	5				
	59	D18S877	GATA64H04	Tetra		113–133	6	0.86	0.70	0.69	0.68
13	60	D1S1588*	ATA2E04	Tri		113–134	4	0.74	0.58	0.38	0.68
	61	D21S2055	GATA188F04	Tetra		115–203	12	0.95	0.85	0.88	0.88
	62	DXS6814	GGAT3F08	Tetra		160–180	4				
	63	D2S1363	GATA23D03	Tetra		166–206	8	0.84	0.70	0.81	0.79
	64	D9S922	GATA21F05	Tetra	+	256–272	6	0.87	0.75	0.75	0.78
			GATA172D05	Tetra		106–126	5				
	65	D6S1017	GGAT3H10	Tetra		154–178	4	0.77	0.69	0.75	0.68
		D10S1225	ATA24F10	Tri		179–191	5	0.85	0.74	0.75	0.76
		D12S375	GATA3F02	Tetra		163–183	6	0.86	0.77	0.88	0.74
14	66	D4S1625	GATA107	Tetra		191–203	4	0.81	0.66	0.66	0.74
		D2S2972	GATA176C01	Tetra	+	217–237	7	0.89	0.76	0.66	0.77

Table 1. Continued

Loading Panel	PCR set	Locus	Marker	Repeat unit	Irregular allele	Size range	Allele	PD	Expected H.	Observed H.	Observed H. (Caucasian)
	67	D12S395	GATA4H01	Tetra	+	222–246	9	0.87	0.79	0.78	0.76
		D6S1277	GATA81B01	Tetra		281–309	6	0.84	0.72	0.81	0.72
	68	D10S1248	GGAA23C05	Tetra		238–258	6	0.91	0.80	0.72	0.75
	69	D7S1819	GATA24F03	Tetra		163–187	7	0.86	0.76	0.63	0.73
		D13S894	GATA86H01	Tetra		186–198	4	0.78	0.59	0.63	0.64
	70	D16S3253	GATA22F09	Tetra		166–198	7	0.87	0.75	0.84	0.71
		D15S657	GATA22F01	Tetra		329–353	7	0.92	0.85	0.88	0.82
15	71	D10S2325	GAAT5F06	Penta		110–155	9	0.93	0.84	0.91	0.85
	72	D3S1744	GATA3C02	Tetra		133–165	9	0.91	0.82	0.69	0.80
	73	D14S592	ATA19H08	Tri		216–240	9	0.90	0.82	0.84	0.68
	74	D6S2439	GATA163B10	Tetra		222–250	7	0.90	0.79	0.75	0.87
	75	D14S1280	GATA31B09	Tetra		281–297	5	0.85	0.71	0.66	0.70
	76	D12S1064	GATA63D12	Tetra		170–202	9	0.88	0.78	0.75	0.82
		D1S534	GATA12A07	Tetra	+	194–218	9	0.90	0.77	0.81	0.83
16	77	DXYS154		Di		225–251	10				
		D4S2431	GGAA19H07	Tetra	+	224–264	12	0.87	0.76	0.78	0.82
	78	D17S928	AFM217yd10	Di		131–159	14	0.94	0.87	0.84	0.79
	79	D21S1446	GATA70B08	Tetra	+	206–224	5	0.81	0.64	0.59	0.69
	80	D1S1612	GGAA3A07	Tetra		102–130	8	0.89	0.80	0.81	0.83
		D4S1652	GATA5B02	Tetra		133–145	4	0.78	0.58	0.56	0.82
	81	D17S1301	GATA28D11	Tetra		139–159	6	0.83	0.69	0.69	0.65
		D19S433	GGAA2A03	Tetra	+	195–213	10	0.90	0.78	0.69	0.77
17	82	D8S1119	ATA19G07	Tri		171–191	7	0.88	0.80	0.78	0.80
		D9S925	GATA27A11	Tetra		180–196	5	0.90	0.79	0.72	0.82
	83	D10S1432	GATA87G01	Tetra	+	159–179	7	0.88	0.74	0.66	0.74
	84	D3S2390	GATA31E08	Tetra		227–251	7				
	85	D7S3046	GATA118G10	Tetra		321–349	8	0.92	0.82	0.81	0.81
		D10S1426	GATA73E11	Tetra		154–174	5	0.83	0.70	0.66	0.74
	86		GATA184A08	Tetra		168–200	9	0.92	0.86	0.91	0.78
		D1S518	GATA7C01	Tetra		188–212	7	0.91	0.78	0.66	0.84
	87	D10S1213	GGAA5D10	Tetra		85–129	10	0.88	0.72	0.81	0.80
		D6S1053	GATA64D02	Tetra		294–314	6	0.90	0.80	0.75	0.81
18	88	D12S1042	ATA27A06	Tri		114–132	7	0.93	0.83	0.75	0.81
	89	D8S1132	GATA26E03	Tetra		138–166	8	0.93	0.87	0.88	0.86
	90	D3S1766	GATA6F06	Tetra		212–232	6	0.90	0.78	0.78	0.76
	91	D4S2394	ATA26B08	Tri		233–251	5	0.76	0.63	0.75	0.79
		D8S1106	GATA23D06	Tetra		134–146	4	0.83	0.68	0.75	0.73
	92	D13S796	GATA51B02	Tetra		146–166	6	0.89	0.79	0.78	0.80
		D12S391	GATA11H08	Tetra		207–247	11	0.93	0.86	0.88	0.88
	93	D13S317	GATA7G10	Tetra		175–195	6	0.90	0.82	0.91	0.79
		D9S2169	GATA62F03	Tetra		274–290	5	0.83	0.68	0.66	0.64
19	94	D9S910	ATA18A07	Tri		101–125	6	0.82	0.68	0.59	0.66
	95	D3S4529	GATA128C02	Tetra		147–163	5	0.88	0.78	0.81	0.72
	96	D8S1477	GGAA20C10	Tetra		157–189	7	0.92	0.81	0.69	0.86
	97	D13S787	GATA23C03	Tetra		247–263	5	0.77	0.61	0.44	0.72
		D6S1027	ATA22G07	Tri		110–134	7	0.83	0.70	0.53	0.77
	98	D16S2624	GATA81D12	Tetra		127–147	6	0.81	0.65	0.59	0.70
		D20S482	GATA51D03	Tetra		140–164	7	0.87	0.73	0.59	0.68
	99	D17S1290	GATA49C09	Tetra		172–208	8	0.88	0.80	0.88	0.84
20	100	D2S2968	GATA178G09	Tetra		177–189	4	0.64	0.48	0.41	0.72
	101	D18S844	ATA1H06	Tri		177–198	6	0.87	0.76	0.81	0.76
	102	DXS6810	GATA69C12	Tetra		214–238	5				
	103	D13S793	GATA43H03	Tetra		248–272	7	0.89	0.80	0.69	0.77

*Significant deviations ($P < 0.05$) from Hardy-Weinberg equilibrium (HWE) were observed in all three tests used (see text for details)
PD, Power of discrimination; H, heterozygosity

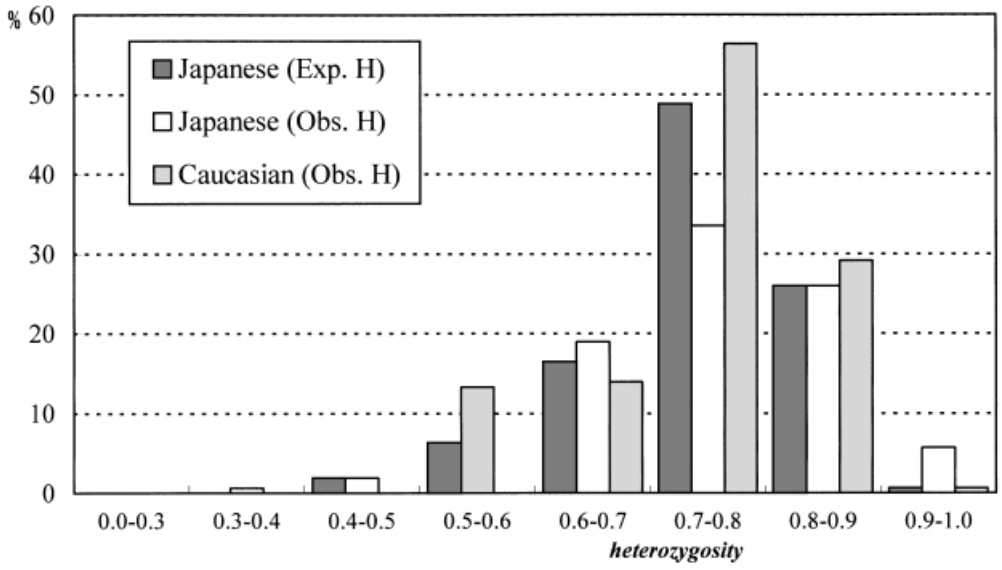

Fig. 1. Distribution of expected and observed heterozygosities (*Exp. H* [*black bars*] and *Obs. H* [*white bars*]) in the Japanese population and the observed heterozygosity (*Obs. H* [*gray bars*]) in the Caucasian population

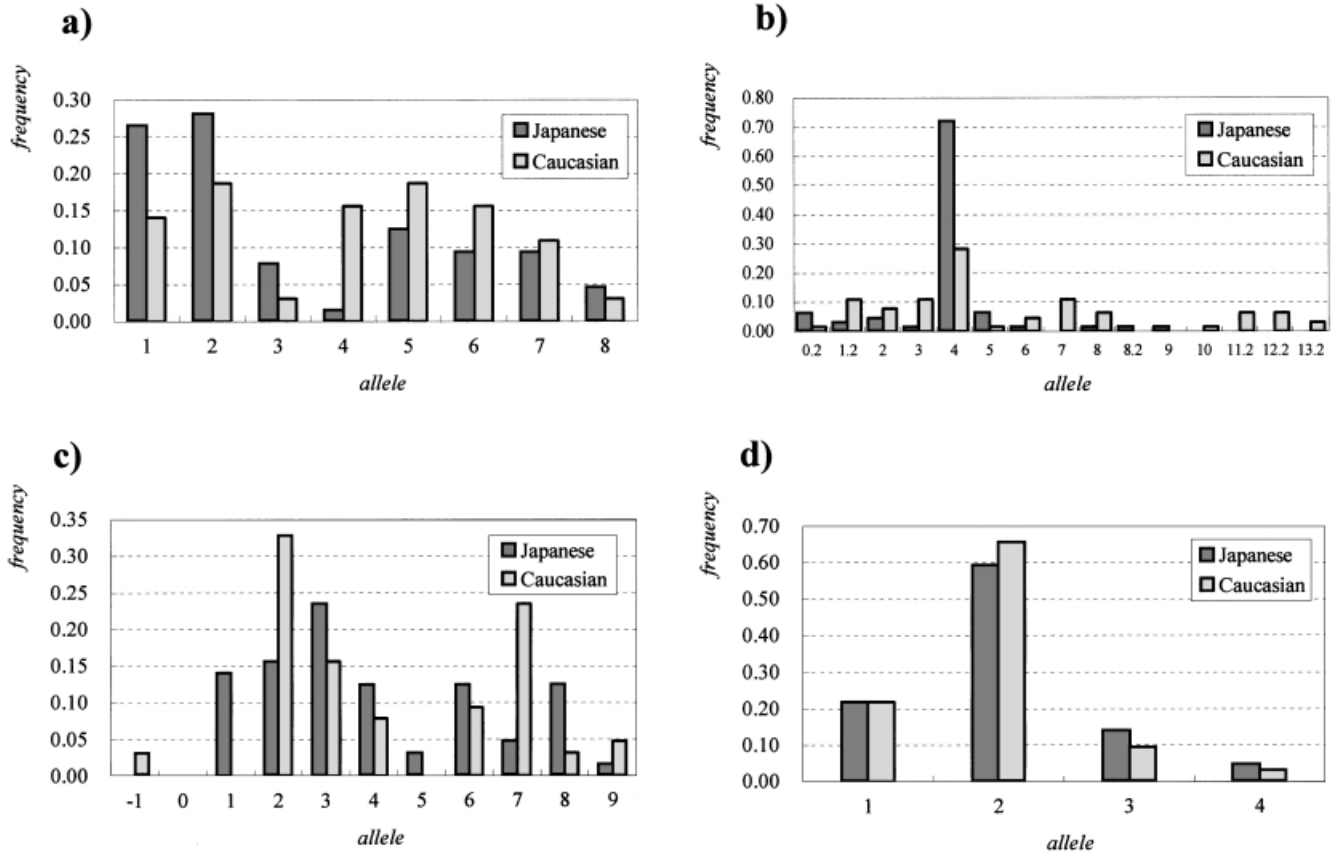

Fig. 2a–d. Examples of allele distributions in representative loci in the Japanese (*black bars*) and Caucasian (*gray bars*) populations. **a** At D7S1823; **b** at D2S2976; **c** at GATA184A08; **d** at D13S894

than in Caucasians. The most frequent allele, allele 4, at D2S2976 reached a frequency of 72% in Japanese, but the frequency was only 28% in Caucasians. The heterozygosity at this locus was only 0.53 in Japanese, against 0.84 in Caucasians. At the GATA184A08 marker, the Japanese alleles were distributed more equally than the Caucasian alleles. On the other hand, D13S894 showed very similar allelic distributions in both populations. At the three loci D7S1823, D2S2976, and GATA184A08, significant differences ($P < 0.05$) were observed between Japanese and Caucasians, while no significant differences were seen at the D13S894 locus.

We divided 130 autosomal tetranucleotide STRs in the screening set into two groups; those with and those without irregular alleles. One group consisted of 19 loci with irregular alleles found in this study, and the other group consisted of the other 111 loci, without irregular alleles. Comparison of the allele frequency distributions (Table 2) revealed that most of the loci with irregular alleles (79%) had two or more peaks of allele frequency distribution, while most of the loci without irregular alleles (76%) showed single-peak allele frequency distribution. The number of alleles per locus was also larger in loci with irregular alleles (9.5) than in loci without irregular alleles (6.6). There were, however, no significant differences among observed heterozygosities, expected heterozygosities, and powers of discrimination between the two groups. Figure 3 shows a typical example of the allele distribution locus in each group. Although the number of alleles at D3S2427 showing bimodal distribution was about twice more than that at D6S1056, showing almost equal allele frequency at each allele except allele 1, both loci had very high observed heterozygosities (0.88 at D3S2427 and 0.94 at D6S1056). Also, no significant differences concerning the proportion of loci departing from HWE were found between STR loci with and without irregular alleles. Therefore, STR loci without irregular alleles are preferable for use in practice because of their simplicity and convenience in typing.

The present study provides basic data for investigating linkage analyses in genetic diseases and for use in forensic practice (e.g., in personal identification and paternity tests). Consequently, this database would be useful for selecting STR loci suitable for forensic applications in the Japanese population, and for constructing an optimal multiplex STR typing system. Furthermore, a future comparative study of these STR loci in Japanese and other ethnic populations would give precise information about their genetic relationships.

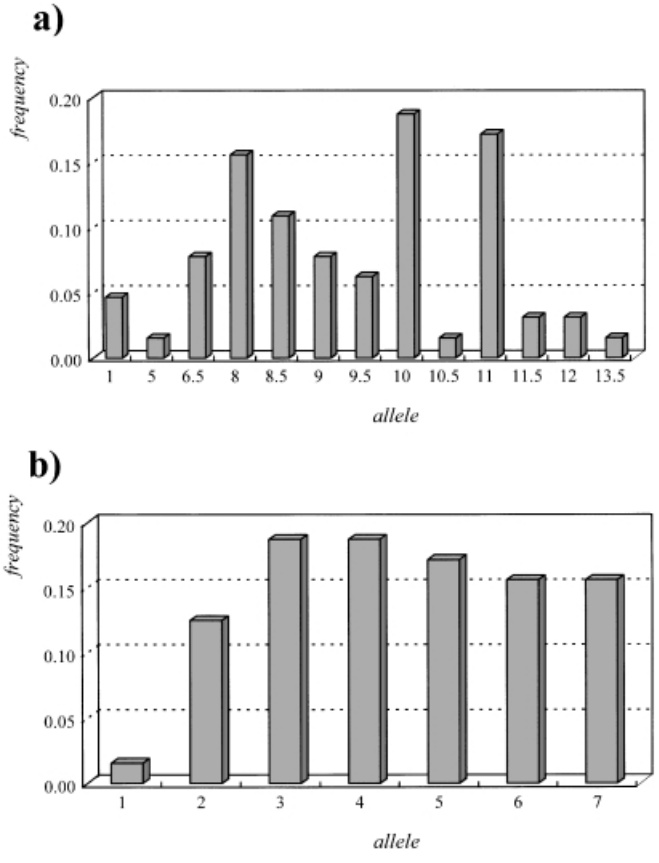

Fig. 3a,b. Typical examples of allele distribution in loci with and without irregular alleles, **a** at D3S2427 and **b** at D6S1056, in the Japanese population. The Exp. H, Obs. H, and power of discrimination values were 0.89, 0.88, and 0.95, respectively, for D3S2427; and 0.85, 0.94, and 0.92, respectively, for D6S1056

Table 2. Correlation between irregular alleles and properties of markers

Irregular allele	With	Without
Number of markers	19	111
Bimodal markers (%)[a]	79	24
Mean number of alleles per locus	9.47	6.62
Observed heterozygosities	0.732	0.732
Expected heterozygosities	0.763	0.745
Power of discrimination	0.880	0.864
Departure from HWE (%)[b]	7.0	6.3

[a] Proportion of loci with bimodal allele distribution
[b] Proportion of loci with *P* value of less than 0.05 in the three tests for HWE

Acknowledgments We are grateful to Dr. Yuri E. Dubrova for providing Caucasian DNA samples. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan.

References

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Chakraborty R, Fornage M, Gueguen R, Boerwinkle E (1991) Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) DNA fingerprinting: approaches and applications. Birkhauser Verlag, Berlin, pp 127–143

Dubovsky J, Sheffield VC, Duyk GM, Weber JL (1995) Sets of short tandem repeat polymorphisms for efficient linkage screening of the human genome. Hum Mol Genet 4:449–452

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

Fisher RA (1951) Standard calculations for evaluating a blood-group system. Heredity 5:95–102

Gill P, Brinkmann B, d'Aloja E, Anderson J, Bar W, Carracedo A, Dupuy B, Eriksen B, Jangblad M, Johnsson V, Kloosterman AD, Lincoln P, Morling N, Rand S, Sabatier M, Scheithauer R, Schneider P, Vide MC (1997) Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. Forensic Sci Int 87:185–192

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J (1994) The 1993–94 Genethon human genetic linkage map. Nat Genet 7:246–339

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Micka KA, Sprecher CJ, Lins AM, Theisen Comey C, Koons BW, Crouse C, Endean D, Pirelli K, Lee SB, Duda N, Ma M, Schumm JW (1996) Validation of multiplex polymorphic STR amplification sets developed for personal identification applications. J Forensic Sci 41:582–590

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Schellenberg GD, Bird TD, Wijsman EM, Orr HT, Anderson L, Nemwns E, White JA, Bonnycastle L, Weber JL, Alonso E, Potter H, Heston LL, Martin GM (1992) Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. Science 258:668–671

Siegel S (1956) Nonparametric statistics for the behavioral sciences. McGraw-Hill, New York

Sweet D, Hildebrand D (1999) Saliva from cheese bite yields DNA profile of burglar: a case report. Int J Legal Med 112:201–203

Tamaki K, Yamamoto T, Uchihi R, Katsumata Y, Kondo K, Mizuno S, Kimura A, Sasazuki T (1991) Frequency of HLA-DQA1 alleles in the Japanese population. Hum Hered 41:209–214

Thomson JA, Pilotti V, Stevens P, Ayres KL, Debenham PG (1999) Validation of short tandem repeat analysis for the investigation of cases of disputed paternity. Forensic Sci Int 100:1–16

Weir BS (1992) Independence of VNTR alleles defined as fixed bins. Genetics 130:873–887

Wijmenga C, Frants RR, Brouwer OF, Moerer P, Weber JL, Padberg GW (1990) Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. Lancet 336:651–653

Yamamoto T, Uchihi R, Nozawa H, Huang XL, Leong YK, Tanaka M, Mizutani M, Tamaki K, Katsumata Y (1999) Allele distribution at nine STR loci — D3S1358, vWA, FGA, TH01, TPOX, CSF1PO, D5S818, D13S317 and D7S820 — in the Japanese population by multiplex PCR and capillary electrophoresis. J Forensic Sci 44:167–170