SHORT COMMUNICATION

Ene-choo Tan · Huimin Wu · Rita Yong · Selen Tan
Joyce Chang · Linda Gan · Eric Yap

Heterozygosities and allelic frequencies of a set of microsatellite markers used for genome-wide scans in a Chinese population

Received: July 1, 2002 / Accepted: August 23, 2002

Abstract Microsatellite (short tandem repeat) markers are useful tools for genetic linkage analysis because of their high frequency of occurrence in eukaryotic genomes, ease of typing, and high polymorphism content. To establish a panel of microsatellite markers useful for genome-wide screens in the Chinese population, we determined the heterozygosities and allelic frequencies of a widely used set of 285 markers in 208 individuals of Han Chinese descent. Although the median heterozygosity level in our Chinese population was 0.72, only 63.6% of these markers have a heterozygosity of at least 0.7, compared with 90.8% in the original Caucasian sample. The significant difference in heterozygosity and allelic frequencies between populations suggests that markers should be optimally selected for each study population to maximize information content and power.

Key words Microsatellites · Genome-wide scan · Alleles · Heterozygosity · Chinese

Introduction

Positional cloning with DNA markers has been instrumental in the cloning and identification of many genes involved in diseases for which the exact mechanism of disease, the biomolecules involved, and the biochemical defect are not

known. Such linkage analysis with pedigrees was previously performed primarily with single nucleotide polymorphism (SNP) markers using the restriction fragment length polymorphism (RFLP) method. This approach has been successful in identifying genetic defects involved in several mendelian disorders caused by single genes.

In contrast to SNPs, the use of microsatellites in genetic mapping is relatively recent. Short tandem repeat (STR) markers such as microsatellites are superior to SNP markers in being more polymorphic and hence more informative because STR markers typically have more than ten alleles each, whereas SNP markers are usually biallelic. Hence, the use of STR markers offers greater statistical power for the detection of linkage, whether by parametric analysis of large families, or by nonparametric analysis of sib-pairs, for which SNPs are not suitable. Genome-wide scans using microsatellite markers have been used in the search for susceptibility loci for many mendelian diseases as well as for complex disorders such as asthma, schizophrenia, bipolar disorder, and diabetes mellitus (e.g., Collaborative Study on the Genetics of Asthma 1997; Xu et al. 2001; Faraone et al. 1998; DeLisi et al. 2002; Polymeropoulos and Schaffer 1996; Detera-Wadleigh et al. 1999; Ghosh et al. 2000; Watanabe et al. 2000; Mori et al. 2002; Vaessen et al. 2002). Because there are multiple genes of varying effects involved in complex diseases, these linkage studies require the collection, ascertainment, and genotyping of large number of samples at significant cost and effort. Therefore, it would be useful to maximize the informative nature of the markers used, hence optimizing the effective sample size for genetic linkage.

As the use of microsatellites for linkage mapping becomes more widespread, sets of primers designed for whole genome or chromosome-specific mapping have become commercially available. These primers are typically spaced at regular intervals 10–20 cM apart. They are especially suitable for high-throughput screening, having been optimized for both multiplex polymerase chain reaction (PCR) and fragment analysis. However, the markers are usually selected based on heterozygosity information and recombination frequency for Caucasians and may not be suitable for other ethnic groups. We describe here the genotype data for

E. Tan · H. Wu · R. Yong · S. Tan · E. Yap (✉)
Defence Medical Research Institute, Defence Science and
Technology Agency Clinical Research Centre, 10 Medical Drive,
Singapore 117597, Republic of Singapore
Tel. +65-67792184; Fax +65-67791677
e-mail: nmiv3@nus.edu.sg

J. Chang · L. Gan · E. Yap
Singapore Eye Research Institute, Republic of Singapore

E. Yap
Department of Community, Occupational and Family Medicine,
National University of Singapore, Republic of Singapore

The first two authors contributed equally to this work.

a population of Chinese descent for a genome-wide panel covering 285 markers from the 22 autosomes and demonstrate that they are less polymorphic and informative compared with available data for Caucasian populations. These findings would be useful for future genetic linkage as well as for linkage disequilibrium studies in specific Asian populations.

Subjects and methods

Subjects

Subjects in this study were 208 unrelated individuals from 120 nuclear families. All were second- or third-generation Chinese Singaporeans descended from the Han. The majority originated from the Southern provinces of Fujian and Guangdong of China. Written informed consent was obtained from all subjects.

STR marker panel

The marker set used was the C290 GENPAK Human Microsatellite Set purchased from GENPAK (Brighton, UK). It comprised mostly dinucleotide markers based on the original set developed by Reed et al. (1994) in which markers were selected based on heterozygosity and relative distance. The set had an average heterozygosity of 0.8 (at least 0.70 for 97% of the markers) and an intermarker distance of less than 30 cM (<4% had an intermarker distance of >20 cM).

PCR amplification

Ten milliliters of venous blood was collected in ethylenediaminetetraacetate tubes from all subjects. Genomic DNA was extracted using the QIAGEN (Hilden, Germany) blood and tissue kit according to the manufacturer's recommendations. Amplification of each DNA fragment was performed in 96-well or 384-well formats in a total of 10 µl for each pair of primers. The reaction mixture consisted of 1 X AmpliTaq Gold Buffer, 1 unit AmpliTaq Gold (Applied Biosystems, Foster City, USA), 0.2 mM deoxyribonucleoside triphosphates, 1.0–2.5 mM MgCl₂, 5 pmole of each primer, and 20 ng genomic DNA. PCR conditions were 12 min at 95°C, 30–35 cycles of denaturation at 95°C for 30 s, annealing at between 55°C and 60°C for 30 s, and extension at 72°C for 30 s in either MJ Research PTC-100/200 (MJ Research, Waltham, MA, USA) or GeneAmp 9700 (Applied Biosystems, Foster City, CA, USA) thermal cyclers.

Fragment analysis

For each pair of primers, PCR products were randomly selected from a few samples and checked by agarose gel electrophoresis before being assembled into panels for multiplex analysis according to the fluorescent dye used and the

size range. The fluorescent-labeled fragments were electrophoresed through ABI 377 and analyzed using GeneScan Analysis (version 2.1) and Genotyper DNA fragment analysis (version 2.1) software from Applied Biosystems. Some of the electrophoresis was also done on the MegaBACE 1000 (Molecular Dynamics, Sunnyvale, USA) and the results analyzed with Genetic Profiler (Version 1.1). Comparative genotyping of a common set of STRs revealed high correlation between the two systems (data not shown).

Statistical analysis

The relative positions of the markers and genetic distances used were obtained from the publicly available genetic maps at the Genethon (www.genethon.com) and the Marshfield Medical Clinic (www.marshmed.org) web sites. Allelic frequencies for each marker were computed by counting the alleles in unrelated individuals. Heterozygosity of markers and possible error in genotyping or allele calling was checked by comparing the expected against the observed number of heterozygotes using the Hardy-Weinberg equilibrium test in the LINKAGE UTILITY program (<ftp://linkage.rockefeller.edu/software/linkage>). The observed polymorphism information content of markers was calculated with SPLINK (David Clayton, Cambridge Institute, <http://www-gene.cimr.cam.ac.uk/clayton/software/>), using data from the nuclear families. Marker information for Caucasians and Japanese were obtained from the CEPH database (<http://www.cephb.fr>) and published data (Ikari et al. 2001; Mizutani et al. 2001), respectively.

Results

Allelic frequency distribution

PCR for all 285 markers was performed on 208 unrelated individuals. Those with ambiguous genotypes or unsuccessful amplification after three attempts were not included in the analysis. All loci met Hardy-Weinberg equilibrium. The number of alleles ranged from 4 (D1S484 and D11S219) to 20 (D18S21 and D21S167) and 21 (D3S1298).

Intrapopulation variation

Heterozygosity data for all loci in the study population (with available data for Caucasians and Japanese) are summarized in Fig. 1. The level of heterozygosity ranged from 0.28 (D10S212) to 0.92 (D12S1670), with the average at 0.72 (SD ± 0.1) and a median of 0.74. Seven markers (3.2%) had a heterozygosity of 0.5 or less, and 61 markers (23.8%) had a heterozygosity of at least 0.8.

Most markers (74.1%) had between 8 and 14 alleles. Five markers (1.8%) had 5 alleles or fewer, and 25 markers

Fig. 1. Distribution of short tandem repeat marker heterozygosities in Chinese (this study, 285 markers), Japanese (Ikari et al. 2001; Mizutani et al. 2001, 136 markers), and Caucasian (<http://www.cephb.fr>, 285 markers) populations

Fig. 2. Difference in heterozygosity between ethnic groups. Figures were obtained by subtracting the heterozygosity value in Chinese from the corresponding value for Caucasians or Japanese for the same marker. Positive values indicate higher heterozygosity in Caucasians or Japanese and negative values indicate higher heterozygosity in Chinese

(8.8%) had 16 or more alleles. The average number of alleles is 10.7 (SD $\pm$ 3.13), with the median at 10.

Interpopulation variation

Detailed comparisons with available data for Caucasians and Japanese in terms of heterozygosity and allele number for each of the markers are presented in Table 1. Figures for all 285 markers for Caucasians were obtained from publicly available data from the CEPH web site ([http://](http://www.cephb.fr)

www.cephb.fr). A comparison of the results with the Japanese population was made for a subset of 136 markers for which data were available (Ikari et al. 2001; Mizutani et al. 2001).

The difference in the level of heterozygosity among the three populations was also observed. The differences in heterozygosities for all 285 markers between Caucasians and Chinese and for 136 markers between Japanese and Chinese are summarized in Fig. 2. Most markers were more heterozygous in Caucasians. Heterozygosity for 19.6% of

Table 1. Observed number of alleles and heterozygosity for 285 STR markers in the Chinese population, in comparison with CEPH and the Japanese population

Chr	Marker	Position (cM)	Success (%)	NA (CEPH)	NA (Chinese)	Het (CEPH)	Het (Japanese)	Het (Chinese)	PIC (Chinese)
1	D1S228	32.4	91.4	7	8	0.76	0.59	0.75	0.70
	D1S470	57.5	92.7	6	7	0.76		0.78	0.75
	D1S255	71.1	91.8	7	7	0.74	0.25	0.45	0.42
	D1S197	85.1	92.3	8	8	0.80	0.88	0.68	0.65
	D1S209	109.0	91.8	9	11	0.79	0.66	0.74	0.70
	D1S216	124.9	88.2	16	16	0.89	0.72	0.86	0.84
	D1S207	133.6	92.3	13	12	0.83	0.81	0.78	0.76
	D1S252	155.1	89.1	8	12	0.81	0.88	0.77	0.74
	D1S484	173.9	93.4	4	4	0.63	0.72	0.70	0.65
	D1S196	186.4	88.6	5	6	0.73	0.66	0.60	0.55
	D1S238	202.7	91.6	12	14	0.85	0.75	0.82	0.79
	D1S229	241.6	90.9	8	7	0.77		0.65	0.59
	D1S103	245.1	91.1	11	12	0.80		0.87	0.85
2	D2S281	12.6	77.2	9	8	0.81	0.47	0.66	0.62
	D2S131	32.9	97.1	10	12	0.85		0.78	0.77
	D2S165	50.7	99.0	13	14	0.85	0.84	0.83	0.82
	D2S177	62.8	98.3	12	13	0.84		0.83	0.82
	D2S380	88.0	99.0	11	12	0.84		0.65	0.62
	D2S139	106.0	99.3	9	15	0.81	0.84	0.81	0.79
	D2S160	127.4	99.3	7	9	0.78	0.72	0.67	0.63
	D2S114	147.1	95.6	10	16	0.84		0.84	0.83
	D2S142	166.3	99.3	6	8	0.76	0.66	0.65	0.61
	D2S326	184.1	98.8	10	14	0.85	0.84	0.86	0.85
	D2S152	194.4	98.5	9	14	0.77		0.81	0.80
	D2S72	204.4	97.3	8	8	0.79		0.54	0.51
	D2S126	228.8	96.1	8	15	0.81	0.75	0.78	0.76
	D2S159	236.1	99.3	6	11	0.76		0.72	0.70
	D2S206	248.3	97.8	9	13	0.79	0.69	0.79	0.77
	D2S125	269.5	99.0	7	10	0.82	0.59	0.80	0.78
3	D3S1297	2.5	92.5	8	11	0.82	0.75	0.71	0.67
	D3S1560	12.9	90.5	9	12	0.81		0.83	0.81
	D3S1263	30.9	92.5	9	15	0.86	0.97	0.89	0.88
	D3S1298	56.7	92.0	12	21	0.87	0.84	0.91	0.90
	D3S1300	79.0	92.7	9	11	0.82	0.69	0.83	0.81
	D3S1261	98.8	92.7	12	13	0.84		0.60	0.58
	D3S1271	117.7	92.7	5	6	0.73	0.56	0.62	0.54
	D3S1303	140.0	92.5	10	16	0.77		0.84	0.82
	D3S1576	156.1	92.5	8	9	0.75		0.74	0.71
	D3S1279	174.3	92.5	10	13	0.84	0.75	0.74	0.72
	D3S1282	186.4	92.7	7	11	0.78		0.77	0.74
	D3S1262	207.2	92.7	8	11	0.79	0.63	0.69	0.66
	D3S1314	220.3	92.3	11	12	0.86		0.75	0.72
	D3S1265	228.0	92.5	10	12	0.84		0.75	0.72
4	D4S412	3.7	92.7	6	11	0.76	0.72	0.70	0.65
	D4S1582	22.0	92.7	6	9	0.79		0.79	0.76
	D4S403	24.9	93.2	7	9	0.76	0.50	0.54	0.50
	D4S418	43.9	93.0	9	10	0.77		0.64	0.59
	D4S405	56.7	92.7	11	7	0.85	0.75	0.76	0.72
	D4S398	71.5	93.0	11	11	0.82	0.81	0.79	0.76
	D4S231	99.5	93.0	7	9	?		0.73	0.69
	D4S406	115.8	92.7	12	9	0.87	0.78	0.68	0.64
	D4S430	125.1	92.7	9	11	0.80		0.78	0.75
	D4S1565	143.1	92.0	7	10	0.75		0.70	0.66
	D4S413	157.9	93.0	12	16	0.84	0.59	0.61	0.59
	D4S415	185.0	92.7	10	13	0.80	0.66	0.72	0.69
	D4S408	199.0	93.0	6	9	0.76		0.62	0.57
	D4S426	211.0	93.0	6	9	0.76	0.72	0.68	0.63
5	D5S432	21.4	92.7	6	9	0.66		0.81	0.79
	D5S416	27.9	92.7	6	10	0.77	0.72	0.68	0.66
	D5S268	31.5	93.2	7	9	0.64		0.80	0.77
	D5S419	39.5	93.0	11	15	0.80	0.78	0.84	0.83
	D5S426	51.6	92.7	8	12	0.79	0.81	0.75	0.72
	D5S407	65.0	93.0	11	14	0.86	0.88	0.85	0.84
	D5S428	95.4	93.0	7	9	0.76	0.78	0.72	0.68
	D5S409	109.3	92.7	9	9	0.67		0.61	0.57
	D5S421	121.4	92.7	8	8	0.82	0.41	0.67	0.64
	D5S210	147.5	92.0	11	10	0.78		0.71	0.68

Table 1. Continued

Chr	Marker	Position (cM)	Success (%)	NA (CEPH)	NA (Chinese)	Het (CEPH)	Het (Japanese)	Het (Chinese)	PIC (Chinese)
6	D5S410	156.0	93.0	9	8	0.79	0.57	0.57	0.53
	D5S422	163.9	92.7	10	14	0.84	0.66	0.81	0.79
	D5S400	174.3	93.0	10	17	0.81	0.91	0.87	0.86
	D5S408	195.8	93.0	7	14	0.73	0.69	0.77	0.74
	D6S470	17.7	92.7	7	10	0.79	0.75	0.71	0.67
	D6S260	30.7	92.5	11	18	0.84		0.89	0.88
	D6S273	44.9	92.7	6	8	0.75		0.76	0.73
	D6S276	44.9	92.5	13	11	0.83	0.94	0.75	0.72
	D6S291	49.8	93.0	11	7	0.70		0.75	0.71
	TNFA	56.8	92.7	7	13	0.27		0.76	0.73
	D6S286	90.0	92.7	13	14	0.78		0.78	0.75
	D6S268	114.8	92.7	7	9	0.74	0.68	0.78	0.74
	D6S287	122.0	92.3	12	10	0.85	0.63	0.61	0.58
	D6S262	129.8	92.0	10	10	0.82	0.78	0.62	0.60
	D6S314	144.4	93.0	8	11	0.80		0.64	0.60
	D6S308	145.5	92.7	6	8	0.75	0.66	0.68	0.64
	D6S290	155.3	93.0	6	10	0.70		0.56	0.53
	D6S305	166.6	92.5	9	16	0.83	0.78	0.80	0.78
	D6S264	179.1	93.0	6	8	0.69	0.41	0.44	0.42
	D6S281	201.1	92.5	8	14	0.68	0.78	0.80	0.78
7	D7S531	4.8	98.3	7	10	0.78	0.78	0.73	0.69
	D7S513	28.0	93.9	15	17	0.83	0.81	0.89	0.88
	D7S493	35.0	98.3	13	16	0.88	0.78	0.72	0.69
	D7S629	36.7	99.5	7	8	0.73		0.64	0.60
	D7S484	55.6	99.5	7	9	0.73	0.81	0.87	0.85
	D7S519	70.5	98.1	6	9	0.80	0.63	0.68	0.64
	D7S502	79.6	97.8	10	12	0.84	0.85	0.85	0.83
	D7S669	90.9	98.1	8	11	0.80	0.88	0.82	0.80
	D7S524	98.0	99.5	7	7	0.73		0.56	0.52
	D7S527	109.0	97.8	8	14	0.74		0.80	0.79
	CFTR	113.5	99.3	10	10	0.89		0.55	0.50
	D7S486	125.3	99.3	12	9	0.80	0.63	0.74	0.71
	D7S530	136.4	98.5	7	11	0.78	0.69	0.74	0.71
	D7S684	149.6	97.6	9	11	0.80	0.78	0.83	0.82
	D7S483	167.6	99	10	10	0.81		0.76	0.74
	D7S550	180.8	98.5	13	13	0.82	0.78	0.70	0.66
8	D8S504	0.0	92.7	5	7	0.71	0.75	0.63	0.60
	D8S503	15.5	92.5	8	9	0.73		0.50	0.48
	D8S552	25.8	92.3	8	9	0.79		0.65	0.62
	D8S261	35.8	91.8	8	9	0.78		0.53	0.50
	D8S258	40.3	92.0	5	6	0.70	0.38	0.69	0.63
	D8S283	60.0	91.6	10	11	0.78	0.69	0.62	0.59
	D8S285	70.6	93.2	8	9	0.77	0.78	0.76	0.72
	D8S260	78.8	92.5	10	10	0.81	0.84	0.82	0.79
	D8S286	93.5	91.8	7	11	0.80		0.74	0.70
	D8S273	101.0	93.6	7	7	0.79		0.74	0.70
	D8S88	110.3	94.3	13	13	0.84		0.84	0.83
	D8S257	110.3	91.1	7	9	0.70		0.70	0.66
	D8S281	122.6	93.6	6	8	0.64		0.55	0.52
	D8S198	128.2	93.9		15	0.78		0.75	0.71
	D8S284	142.7	92.0	8	18	0.82	0.84	0.83	0.81
	D8S272	152.5	88.6	10	10	0.81	0.69	0.77	0.74
9	D9S144	18.6	98.8	10	9	0.77		0.54	0.50
	D9S156	29.5	94.4	15	13	0.79		0.63	0.57
	D9S171	42.0	99.3	9	8	0.79	0.28	0.35	0.32
	D9S147E	53.4	99.3	7	8	0.78		0.68	0.61
	D9S43	54.2	99.0	12	9	0.76		0.77	0.73
	D9S15	66.5	98.8	8	7	0.77		0.64	0.57
	D9S175	68.8	99.0	15	16	0.84	0.72	0.62	0.61
	D9S176	104.9	98.3	9	10	0.81		0.78	0.75
	D9S154	125.6	74.8	11	13	0.84		0.77	0.75
	D9S159	142.7	94.4	9	12	0.79		0.80	0.78
	D9S118	152.4	98.8	13	13	0.83		0.65	0.63
	D9S158	163.0	97.1	9	12	0.71	0.81	0.60	0.57
10	D10S189	17.3	92.5	4	5	0.72	0.63	0.75	0.71
	D10S191	36.3	92.7	12	15	0.81	0.72	0.84	0.82
	D10S197	50.5	92.7	6	8	0.74	0.75	0.76	0.73
	D10S220	72.5	93.0	10	11	0.82	0.78	0.77	0.73

Table 1. Continued

Chr	Marker	Position (cM)	Success (%)	NA (CEPH)	NA (Chinese)	Het (CEPH)	Het (Japanese)	Het (Chinese)	PIC (Chinese)
11	D10S537	93.8	92.5	9	10	0.82	0.63	0.77	0.73
	D10S201	105.9	92.5	9	17	0.82	0.84	0.83	0.81
	D10S192	131.2	92.7	8	14	0.77	0.88	0.84	0.83
	D10S190	147.2	92.7	8	11	0.82	0.94	0.83	0.81
	D10S217	167.2	93.0	9	13	0.80	0.91	0.85	0.83
	D10S212	180.7	92.7	6	7	0.70	0.38	0.28	0.27
	D11S922	0.0	88.6	25	19	0.92	0.84	0.84	0.83
	TH	7.5	88.4	7	5	0.78		0.69	0.65
	D11S569	17.9	90.0	9	9	0.84		0.75	0.70
	D11S899	28.9	86.6	11	9	0.70		0.79	0.76
	D11S904	41.1	91.8	8	7	0.82	0.72	0.72	0.68
	D11S907	51.4	90.5	9	6	0.72		0.74	0.70
	D11S903	65.5	87.0	8	6	0.74		0.68	0.65
	D11S1765	71.3	88.0	10	8	0.79		0.63	0.59
	D11S1883	75.0	88.0	10	9	0.73		0.75	0.71
	D11S913	77.4	90.2	6	4	0.57		0.65	0.58
	INT2	81.7	89.5	11	9	0.82		0.76	0.72
	D11S916	87.0	90.5	16	9	0.72		0.72	0.70
	D11S901	97.7	88.4	7	8	0.82	0.75	0.68	0.62
	D11S35	108.9	88.2	8	8	0.82		0.62	0.57
	D11S927	118.4	90.7	10	9	0.85		0.84	0.82
	D11S925	135.3	90.0	13	11	0.84		0.78	0.75
	D11S934	144.2	90.0	15	11	0.84	0.44	0.66	0.63
	D11S910	160.0	90.2	10	6	0.71	0.72	0.76	0.72
	D11S968	168.5	88.9	8	9	0.81	0.38	0.65	0.60
12	D12S99	0.0	86.6	14	11	0.83	0.81	0.83	0.81
	D12S77	8.6	87.7	12	13	0.87		0.71	0.69
	D12S364	13.9	86.6	14	14	0.86	0.81	0.82	0.80
	D12S87	37.8	88.0	10	9	0.79		0.65	0.61
	D12S85	48.8	90.9	14	8	0.67	0.84	0.80	0.78
	D12S339	51.9	91.6	13	9	0.84		0.89	0.88
	D12S92	81.4	88.4	8	8	0.75		0.67	0.64
	D12S1708	90.8	92.3	9	5	0.72		0.79	0.76
	D12S1670	92.4	84.5	27	16	0.80		0.92	0.92
	D12S81	93.2	92.7	11	11	0.81		0.75	0.72
	D12S88	94.0	90.7	19	10	0.84		0.89	0.87
	D12S351	95.3	93.2	10	6	0.74	0.63	0.80	0.77
	D12S95	96.8	88.0	10	6	0.77		0.81	0.79
	D12S327	98.9	92.7	13	14	0.85		0.78	0.76
	D12S1716	102.8	92.7	9	6	0.74		0.74	0.71
	D12S1706	105.0	93.0	7	11	0.88		0.63	0.58
	D12S346	105.8	92.3	18	10	0.83	0.66	0.80	0.77
	D12S306	106.4	93.0	9	10	0.84		0.75	0.73
	D12S1727	107.8	90.7	12	6	0.70		0.83	0.81
	D12S1607	109.1	93.0	11	8	0.75		0.73	0.69
	D12S318	111.7	91.8	9	9	0.71		0.74	0.71
	D12S338	113.1	88.0	10	8	0.76		0.72	0.69
	D12S366	137.0	87.7	11	8	0.80		0.85	0.84
	D12S342	150.4	87.5	11	10	0.73		0.74	0.70
	D12S97	170.3	87.5	7	7	0.68		0.41	0.39
13	D13S192	15.0	93.4	24	15	0.90		0.89	0.88
	D13S120	22.0	92.7	13	12	0.78		0.71	0.67
	D13S153	47.5	92.7	10	16	0.81	1.00	0.90	0.90
	D13S279	54.0	92.3	7	9	0.80		0.53	0.50
	D13S170	65.4	92.5	13	15	0.89	0.88	0.85	0.84
	D13S122	73.0	92.3	7	7	0.68		0.78	0.76
	D13S71	73.0	92.3	9	16	0.82		0.71	0.67
	D13S159	81.5	92.0	24	18	0.89	0.84	0.77	0.75
	D13S158	86.9	92.7	8	13	0.81	0.75	0.75	0.72
	D13S280	87.5	92.7	12	16	0.87		0.84	0.83
	D13S173	95.9	92.7	7	9	0.82	0.88	0.70	0.66
	D13S193	112.8	92.7	11	11	0.75		0.72	0.68
14	D14S72	2.1	97.8	7	11	0.82		0.79	0.76
	D14S50	3.5	99.0	8	7	0.77		0.74	0.71
	D14S80	20.6	98.3	11	10	0.83	0.47	0.58	0.54
	D14S49	29.8	99.8	13	13	0.84		0.82	0.81
	D14S70	32.9	99.0	5	6	0.75	0.75	0.70	0.66
	D14S75	36.3	98.1	9	11	0.76		0.84	0.82

Table 1. Continued

Chr	Marker	Position (cM)	Success (%)	NA (CEPH)	NA (Chinese)	Het (CEPH)	Het (Japanese)	Het (Chinese)	PIC (Chinese)
15	D14S63	59.0	98.8	9	14	0.76	0.69	0.71	0.69
	D14S74	76.4	99.5	9	11	0.79	0.91	0.75	0.73
	D14S68	86.3	98.3	16	10	0.91	0.63	0.73	0.70
	D14S51	106.0	99.8	10	9	0.65		0.72	0.69
	D15S128	6.1	92.7	7	12	0.77		0.81	0.78
	D15S118	32.2	92.7	6	7	0.74		0.65	0.61
	D15S117	50.8	92.7	9	9	0.78	0.66	0.70	0.67
	D15S125	63.8	93.0	7	9	0.78		0.76	0.73
	D15S114	72.3	92.5	6	9	0.70		0.59	0.55
	D15S127	84.8	93.0	14	19	0.85	0.78	0.82	0.81
	D15S207	100.8	92.3	9	10	0.76		0.60	0.58
	D15S129	105.0	92.3	9	10	0.84		0.79	0.76
	D15S107	109.3	92.3	8	10	0.77		0.63	0.57
	D16S423	8.4	99.3	10	13	0.73	0.81	0.86	0.85
	D16S407	16.7	98.8	10	16	0.84	0.91	0.86	0.85
16	D16S405	27.0	98.8	9	11	0.76	0.72	0.72	0.70
	D16S287	31.1	98.5	13	12	0.81		0.80	0.78
	D16S420	43.2	99.0	8	10	0.81	0.34	0.74	0.70
	D16S261	56.5	99.8	7	7	0.70		0.58	0.55
	D16S411	57.8	99.5	6	10	0.78	0.66	0.72	0.70
	D16S415	65.6	99.0	8	13	0.72	0.66	0.67	0.63
	D16S265	81.8	98.3	15	13	0.74		0.75	0.72
	D16S503	81.8	99.5	8	10	0.80	0.63	0.61	0.57
	D16S515	90.2	97.1	11	16	0.80	0.72	0.84	0.83
	D16S320	93.9	99.8	13	15	0.86		0.82	0.81
	D16S516	98.3	99.5	7	9	0.73	0.72	0.61	0.58
	D16S289	103.1	99.0	9	9	0.77		0.76	0.73
	D16S422	109.1	97.3	9	13	0.78		0.69	0.67
17	D17S513	18.1	78.2	11	12	0.88		0.73	0.70
	D17S786	18.1	94.8	7	9	0.76		0.81	0.79
	D17S799	32.8	94.3	5	10	0.68	0.69	0.64	0.61
	D17S798	53.9	92.7	7	7	0.80	0.53	0.54	0.46
	D17S250	60.9	93.0	10	12	0.83		0.75	0.72
	D17S934	63.7	93.6	10	12	0.84		0.75	0.72
	HOX2B	?	94.5	11	6	0.74		0.47	0.43
	D17S787	75.7	93.4	9	10	0.81	0.78	0.80	0.77
	D17S807	87.6	92.7	9	14	0.85		0.68	0.65
	D17S789	90.8	92.7	8	9	0.82		0.64	0.61
	D17S939	106.9	94.8	9	12	0.83		0.70	0.67
	D17S784	117.7	93.2	7	9	0.77	0.66	0.63	0.59
18	D18S59	0.1	94.3	8	12	0.81	0.75	0.78	0.76
	D18S476	1.6	93.4	7	6	0.76		0.57	0.48
	D18S1146	1.7	93.4	6	7	0.51		0.58	0.49
	D18S1105	1.8	93.6	12	12	0.52		0.45	0.43
	D18S1098	5.0	93.6	5	9	0.57		0.51	0.46
	D18S481	6.2	93.0	9	9	0.75	0.91	0.80	0.78
	D18S1138	7.6	92.7	7	11	0.78		0.80	0.77
	D18S632	7.9	93.0	9	12	0.79		0.68	0.66
	D18S52	8.3	90.9	9	9	0.76	0.69	0.78	0.75
	D18S62	17.7	93.0	5	6	0.66	0.78	0.67	0.63
	D18S53	40.4	92.0	10	14	0.79	0.78	0.81	0.79
	D18S71	42.8	92.0	12	20	0.81		0.81	0.78
	D18S57	63.2	93.0	11	13	0.86	0.78	0.83	0.81
	D18S64	83.0	92.3	8	12	0.73	0.81	0.79	0.76
	D18S61	102.8	92.7	10	11	0.86	0.88	0.82	0.80
19	D19S247	6.5	99.8	29	12	0.73		0.62	0.60
	D19S221	35.5	88.6	10	11	0.86	0.72	0.80	0.79
	D19S49	49.4	99	9	9	0.67		0.77	0.74
	D19S225	55.9	91	6	6	0.71		0.61	0.57
	D19S220	61.4	91	10	11	0.84	0.81	0.85	0.83
	D19S180	86.3	98.5	13	12	0.75		0.52	0.50
	D19S210	104.9	97.6	6	6	0.73	0.63	0.63	0.59
20	D20S199	6.2	98.1	9	14	0.85		0.88	0.87
	D20S186	33.2	99.5	11	14	0.85	0.91	0.83	0.82
	D20S112	39.3	99.3	10	11	0.81	0.73	0.75	0.72
	D20S200	49.7	98.1	7	7	0.78		0.61	0.58
	D20S106	50.2	99.5	8	8	0.71		0.62	0.60

Table 1. Continued

Chr	Marker	Position (cM)	Success (%)	NA (CEPH)	NA (Chinese)	Het (CEPH)	Het (Japanese)	Het (Chinese)	PIC (Chinese)
21	D20S107	54.9	97.8	10	8	0.79	0.78	0.73	0.70
	D20S120	82.1	99	10	17	0.85		0.81	0.80
	D20S171	94.4	99.3	13	9	0.73	0.88	0.71	0.68
	D21S258	0.0	99.3	12	13	0.87		0.73	0.71
	D21s172	3.0	99.0	16	10	0.65		0.62	0.60
	D21S265	20.5	99.3	8	14	0.84		0.73	0.72
	D21S219	31.3	99.5	5	10	0.75		0.54	0.52
	D21S65	33.0	99.5	12	12	0.83		0.62	0.60
	D21S270	38.1	99.3	11	10	0.85		0.79	0.75
	D21S167	38.7	99.5	14	20	0.56		0.55	0.54
	D21S268	40.5	99.0	12	13	0.85		0.81	0.80
22	D22S264	11.2	99.5	11	13	0.80		0.77	0.75
	D22S156	15.2	98.8	8	11	0.69		0.54	0.52
	CYP2D	40.9	97.8	10	12	0.61		0.65	0.61
	D22S274	45.5	96.8	7	12	0.76	0.88	0.79	0.78

STR, Short tandem repeat; Chr, chromosome; NA, not available; Het, heterozygosity; PIC, polymorphism information content

the markers differed by more than 15%. The figure for the Chinese–Japanese comparison was smaller at 15.5%. For difference in heterozygosity in excess of 20%, 9.3% of markers differed between Chinese and Caucasians, whereas 3.6% differed between Chinese and Japanese. The observation was consistent with the primer set being assembled based on the heterozygosity data for Caucasians because 90.8% of the markers have a heterozygosity level of at least 0.7 for Caucasians compared with only 63.6% for Chinese and 63.7% for Japanese.

Forty-one (14.5%) markers had the same number of alleles for both Caucasians and Chinese, 69 markers (24.4%) differed by only 1 in terms of allele number, and 57 (20.1%) differed by 2. The remaining 116 markers (40%) differed by at least 3. The number of alleles could differ by as many as 11, as in the case of D12S1670 with 27 in Caucasians compared with 16 in Chinese.

Discussion

PCR amplification of hypervariable loci is a useful tool in population genetics for linkage mapping and forensic studies. It is fast to perform, can be multiplexed at both the amplification and the detection steps, and is relatively undemanding in terms of the quantity and quality of genomic DNA needed (Sunnucks 2000). The large number of alleles for each marker also provides discriminatory power beyond the traditional biallelic markers based on RFLP.

Because the large number of alleles makes them useful for nucleic acid-based identification and construction of genetic relationships, there are optimized sets of primers for paternity testing, individual identification, chromosome-specific scanning, or whole genome scanning. However, the mechanism that gives rise to the higher number of alleles in tandem repeats such as microsatellites also makes them more prone to mutation drift because of slippage during replication or genetic recombination. Hence, the patterns of

allele frequency at different microsatellite loci are variable between populations in terms of the number of alleles and the distribution of the alleles within a population. Such variation between ethnic groups means that the discriminatory power of the markers would need to be verified in different populations. One such commercial set of markers designed for whole genome scanning was used on our Chinese population and the data compared with European and Japanese populations.

Heterozygosity for most markers in our study population was lower than the available data for Caucasians. Another genome-wide scan in the Chinese population also reported a similar overall average heterozygosity of 0.72 (Luo et al. 2001). The obvious explanation is that the marker set was chosen based on its discriminatory power for the Caucasian population. A larger number of alleles were found in our population, which could be expected because of our larger sample size of 208 individuals compared with only 38 used for the CEPH data for Caucasians. It has been found that the number of alleles within a population tends to increase with sample size (Nei 1987).

Despite the higher number of alleles overall, the level of heterozygosity was still lower than that for Caucasians. Computer-simulation studies using our data were conducted to investigate how sample size affects heterozygosity. Using sample sizes from 10 to 200 individuals and simulating allele numbers from 2 to 27, the number of individuals genotyped had minimal influence on heterozygosity (data not shown).

Despite the overall lower heterozygosity in this study population, using a subpanel of 192 STR markers with heterozygosities of >0.7 in the Chinese population would provide coverage of 84% of the human genome at the 20-cM interval on either side of each marker. With the average chromosomal interval for these markers between 15 and 20cM, this set of markers clearly needs to be expanded to increase coverage of the genome to 100%.

There is less difference between our results and those for the Japanese population. This finding is in agreement with

classification based on blood group and other protein markers and genetic markers (Ryman et al. 1983; Constans et al. 1985; Bowcock et al. 1994; Nei and Takezaki 1996; Perez-Lezaun et al. 1997; Destro-Bisol et al. 2000; Jorde et al. 2000). Our reported figures for heterozygosity and allele number for the 285 markers will be useful as a reference for future genome-wide screens using Chinese populations.

Microsatellite markers that showed a high allele content difference between populations have been found to be useful in mapping by admixture linkage disequilibrium (MALD). Several markers that showed large differences when Caucasians were compared with Mexican Americans, American Indians, Africans, and African Americans, respectively, have been identified (Smith et al. 2001; Collins-Schramm et al. 2002). The results of this study will provide information on STR markers that could be developed for a MALD approach for use in studies in which one of the founding populations is Chinese. The use of MALD will facilitate population-based association studies in the identification of causative mutations or polymorphisms and fine mapping of regions found to be positive by linkage studies.

Acknowledgments This work is partially supported by Singapore National Medical Research Council grant no. SERI/MG/96-01/0001 and by the Singapore Ministry of Defence. We would like to thank the SAF Medical Classification Centre and the Singapore Eye Research Institute Clinic staff for help in recruitment. We also thank S.J. Chew and D. Tan for their support.

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
- Collaborative Study on the Genetics of Asthma (1997) A genome-wide search for asthma susceptibility loci in ethnically diverse populations. *Nat Genet* 15:389–392
- Collins-Schramm HE, Phillips CM, Operario DJ, Lee JS, Weber JL, Hanson RL, Knowler WC, Cooper R, Li H, Seldin MF (2002) Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. *Am J Hum Genet* 70:737–750
- Constans J, Hazout S, Garruto RM, Gajdusek DC, Spees EK (1985) Population distribution of the human vitamin D binding protein: anthropological considerations. *Am J Phys Anthropol* 68:107–122
- DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW, Wellman N, Loftus J, Nanthakumar B, Razi K, Stewart J, Comazzi M, Vita A, Heffner T, Sherrington R (2002) A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. *Am J Psychiatry* 159:803–812
- Destro-Bisol G, Spedini G, Pascali VL (2000) Application of different genetic distance methods to microsatellite data. *Hum Genet* 106:130–132
- Deterra-Wadleigh SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G, Rollins DY, Moses T, Sanders AR, Karkera JD, Esterling LE, Zeng J, Ferraro TN, Guroff JJ, Kazuba D, Maxwell ME, Nurnberger Jr J, Gershon ES (1999) A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc Natl Acad Sci USA* 96:5604–5609
- Faraoane SV, Matise T, Svrakic D, Pepple J, Malaspina D, Suarez B, Hampe C, Zambuto CT, Schmitt K, Meyer J, Markel P, Lee H, Harkavy Friedman J, Kaufmann C, Cloninger CR, Tsuang MT (1998) Genome scan of European-American schizophrenia pedigrees: results of the NIMH Genetics Initiative and Millennium Consortium. *Am J Med Genet* 81:290–295
- Ghosh S, Watanabe RM, Valle TT, Hauser ER, Magnuson VL, Langefeld CD, Ally DS, Mohlke KL, Silander K, Kohtamaki K, Chines P, Balow J Jr, Birznieks G, Chang J, Eldridge W, Erdos MR, Karanjawala ZE, Knapp JJ, Kudelko K, Martin C, Morales-Mena A, Musick A, Musick T, Pfahl C, Porter R, Rayman JB (2000) The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. *Am J Hum Genet* 67:1174–1185
- Ikari K, Onda H, Furushima K, Maeda S, Harata S, Takeda J (2001) Establishment of an optimized set of 406 microsatellite markers covering the whole genome for the Japanese population. *J Hum Genet* 46:207–210
- Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 66:979–988
- Luo TH, Zhao Y, Li G, Yuan WT, Zhao JJ, Chen JL, Huang W, Luo M (2001) A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. *Diabetologia* 44:501–506
- Mizutani M, Yamamoto T, Torii K, Kawase H, Yoshimoto T, Uchihi R, Tanaka M, Tamaki K, Katsumata Y (2001) Analysis of 168 short tandem repeat loci in the Japanese population, using a screening set for human genetic mapping. *J Hum Genet* 46:448–455
- Mori Y, Otake S, Dina C, Yasuda K, Populaire C, Lecoeur C, Vatin V, Durand E, Hara K, Okada T, Tobe K, Boutin P, Kadowaki T, Froguel P (2002) Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate loci on 7p and 11p. *Diabetes* 51:1247–1255
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Nei M, Takezaki N (1996) The root of the phylogenetic tree of human populations. *Mol Biol Evol* 13:170–177
- Perez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J (1997) Microsatellite variation and the differentiation of modern humans. *Hum Genet* 99:1–7
- Polymeropoulos MH, Schaffer AA (1996) Scanning the genome with 1772 microsatellite markers in search of a bipolar disorder susceptibility gene. *Mol Psychiatry* 1:404–407
- Reed PW, Davies JL, Copeman JB, Bennett ST, Palmer SM, Pritchard LE, Gough SC, Kawaguchi Y, Cordell HJ, Balfour KM, Jenkins SC, Powell EE, Vignal A, Todd JA (1994) Chromosome-specific microsatellite sets for fluorescence-based, semi-automated genome mapping. *Nat Genet* 7:390–395
- Ryman N, Chakraborty R, Nei M (1983) Differences in the relative distribution of human gene diversity between electrophoretic and red and white cell antigen loci. *Hum Hered* 33:93–102
- Smith MW, Lautenberger JA, Shin HD, Chretien JP, Shrestha S, Gilbert DA, O'Brien SJ (2001) Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. *Am J Hum Genet* 69:1080–1094
- Sunnucks P (2000) Efficient genetic markers for population biology. *Trends Ecol Evol* 15:199–203
- Vaessen N, Heutink P, Houwing-Duistermaat JJ, Snijders PJ, Rademaker T, Testers L, Batstra MR, Sandkuijl LA, van Duijn CM, Oostra BA (2002) A genome-wide search for linkage-disequilibrium with type 1 diabetes in a recent genetically isolated population from the Netherlands. *Diabetes* 51:856–859
- Watanabe RM, Ghosh S, Langefeld CD, Valle TT, Hauser ER, Magnuson VL, Mohlke KL, Silander K, Ally DS, Chines P, Blaschak-Harvan J, Douglas JA, Duren WL, Epstein MP, Fingerlin TE, Kaleta HS, Lange EM, Li C, McEachin RC, Stringham HM, Trager E, White PP, Balow J Jr, Birznieks G, Chang J, Eldridge W (2000) The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. II. An autosomal genome scan for diabetes-related quantitative-trait loci. *Am J Hum Genet* 67:1186–1200
- Xu J, Meyers DA, Ober C, Blumenthal MN, Mellen B, Barnes KC, King RA, Lester LA, Howard TD, Solway J, Langefeld CD, Beaty TH, Rich SS, Bleecker ER, Cox NJ (2001) Genome-wide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S. populations: collaborative study on the genetics of asthma. *Am J Hum Genet* 68:1437–1446