

Table S3. The effect of chromosome number on microsatellite sequence properties and measures of variation across individuals

	1 STR region				2 STR regions				3 STR regions			
	Di	Tri	Tetra		Di	Tri	Tetra		Di	Tri	Tetra	
Number of loci	30	133	325		10	15	97		10	15	97	12
G/C content of STR region(s) flanking sequence	0.556	0.064	0.428		0.361	0.650	0.058		0.361	0.650	0.058	0.347
Number of nucleotides separating distinct STR regions	-	-	-		0.403	0.382	0.514		0.403	0.382	0.514	0.515
Heterozygosity	0.864	0.362	0.756		0.347	0.598	0.520		0.347	0.598	0.520	0.625
Number of distinct alleles	0.581	0.668	0.932		0.353	0.860	0.450		0.353	0.860	0.450	0.485
Variance in number of repeats	0.448	0.206	0.921		0.347	0.537	0.678		0.347	0.537	0.678	0.515
Range of number of repeats	0.487	0.895	0.896		0.362	0.801	0.786		0.362	0.801	0.786	0.471
Skewness in number of repeats	0.166	0.853	0.532		0.347	0.710	0.343		0.347	0.710	0.343	0.332
Mean PCR fragment length	0.489	0.323	0.996		0.361	0.508	0.748		0.361	0.508	0.748	0.323
Mean number of repeats	0.564	0.403	0.525		0.347	0.372	0.727		0.347	0.372	0.727	0.550
Maximum number of repeats	0.314	0.731	0.647		0.347	0.230	0.549		0.347	0.230	0.549	0.637
Minimum number of repeats	0.509	0.552	0.784		0.386	0.408	0.389		0.386	0.408	0.389	0.563

P values from Kruskal-Wallis tests for differences in microsatellite heterozygosity, sequence properties, and measures of variation across individuals in the HGDP-CEPH data set when grouped by chromosome number. Microsatellites were classified by the number of separate STR regions embedded in their sequence and by their repeat unit size. Loci with three separate di-nucleotide or tri-nucleotide STR regions were not evaluated due to small sample size (3 and 2, respectively). Hyphens indicate comparisons that were not evaluated.

Table S4. Summary of the properties of the measures of variation across individuals

		1 STR region			2 STR Regions			3 STR regions
		Di	Tri	Tetra	Di	Tri	Tetra	Tetra
Number of loci		30	133	325	10	15	97	12
Number of repeats in RefSeq	Mean	17.73	12.68	11.20	22.10	15.93	16.31	22.42
	Minimum	4	6	4	16	12	10	17
	Maximum	27	17	23	30	21	27	32
Heterozygosity	Mean	0.779	0.749	0.739	0.789	0.721	0.772	0.814
	Minimum	0.606	0.313	0.507	0.614	0.526	0.568	0.675
	Maximum	0.882	0.867	0.907	0.855	0.843	0.908	0.895
Number of distinct alleles	Mean	15.37	11.47	10.86	14.30	10.67	12.59	14.83
	Minimum	9	5	5	6	4	7	9
	Maximum	24	23	29	21	20	35	28
Variance in number of repeats	Mean	8.53	3.57	2.39	5.61	3.06	3.87	4.21
	Minimum	1.41	0.36	0.49	1.87	0.77	0.59	1.17
	Maximum	29.79	13.77	24.62	17.51	6.93	33.92	8.11
Range of number of repeats	Mean	16.13	9.85	9.13	14.40	9.29	10.79	13.13
	Minimum	10	4	4	5	3	5	8
	Maximum	24	19	23	23	12	24	20
Skewness in number of repeats	Mean	0.067	-0.060	-0.177	-0.015	-0.531	-0.057	0.191
	Minimum	-0.921	-2.053	-2.649	-1.182	-1.728	-1.900	-0.459
	Maximum	1.153	2.661	2.425	2.234	0.648	2.804	0.974
Mean PCR fragment size	Mean	158.56	190.21	201.90	181.54	204.68	218.26	241.34
	Minimum	98.30	94.32	102.44	118.28	144.30	115.52	148.27
	Maximum	257.77	319.15	412.49	232.10	273.66	413.16	355.83
Mean number of repeats	Mean	18.16	13.79	12.03	23.72	17.58	16.91	21.63
	Minimum	13.36	9.19	5.18	17.83	12.25	9.46	17.66
	Maximum	23.72	18.94	22.30	31.05	23.00	27.20	32.27
Maximum number of repeats	Mean	26.57	18.80	16.52	31.50	21.84	22.28	28.31
	Minimum	21	12.67	8.75	21	17	13.25	21.50
	Maximum	34	26.67	30.75	43	27	33.50	42.25
Minimum number of repeats	Mean	10.43	8.95	7.39	17.10	12.56	11.49	15.19
	Minimum	6	3	-0.25*	12	6.67	3.50	8
	Maximum	17	14.33	13.25	24	19	20.50	22.25

Microsatellites were classified by the number of separate STR regions embedded in their sequence and by their repeat unit size. For three STR regions, no values are presented for di-nucleotide and tri-nucleotide loci because of small sample size (3 and 2, respectively).

*This negative value was obtained with the smallest allele at the tetra-nucleotide locus AGAT113Z (Table S2), present in only a single individual in our data set. The fragment identified for this locus in RefSeq was 185 bp and contained a single STR region with 7 repeats (Table S1). The smallest allele identified for this locus in our data set was 156 bp. However, the size expected for an allele containing 0 repeats is 157 bp based on the identified RefSeq fragment. There are at least two likely scenarios under which this size anomaly may have occurred: (1) The 156 bp allele contains 0 repeats but because of a -1 bp shift in the PCR fragment size during genotyping it appears to have a negative non-integer number of repeats. (2) A deletion could have occurred in the non-repetitive part of the region between the two primers, creating a PCR fragment smaller than expected for reasons unrelated to the properties of the repetitive element.

Table S5. Spearman's rank correlations between measures of variation across individuals for microsatellites with one or two STR regions embedded in their sequence

	Number of distinct alleles		Variance in number of repeats		Range of number of repeats		Skewness in number of repeats		Mean PCR fragment size		Mean number of repeats		Maximum number of repeats		Minimum number of repeats		
	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P	
Number of distinct alleles	Di	-	0.651	9.88x10⁻⁵	0.735	3.74x10⁻⁶	0.338	0.087	-0.006	0.975	0.286	0.126	0.765	8.37x10⁻⁷	-0.002	0.991	
	Tri	-	0.375	8.94x10⁻⁶	0.696	1.46x10⁻²⁰	-0.131	0.133	0.085	0.329	0.247	0.004	0.400	1.81x10⁻⁶	-0.143	0.101	
	Tetra	-	0.513	3.14x10⁻²³	0.825	3.48x10⁻⁸²	0.027	0.625	0.052	0.348	0.258	2.38x10⁻⁶	0.575	5.22x10⁻³⁰	-0.110	0.047	
Variance in number of repeats	Di	0.494	0.147	-	-	0.657	8.08x10⁻⁵	0.004	0.984	0.004	0.984	0.014	0.940	0.616	2.89x10⁻⁴	-0.071	0.709
	Tri	0.713	0.003	-	-	0.415	6.64x10⁻⁷	-0.142	0.104	0.058	0.506	-0.029	0.740	0.231	0.008	-0.130	0.135
	Tetra	0.674	3.75x10⁻¹⁴	-	-	0.549	4.97x10⁻²⁷	0.127	0.022	0.009	0.873	0.007	0.904	0.300	3.38x10⁻⁸	-0.205	1.99x10⁻⁴
Range of number of repeats	Di	0.985	2.44x10⁻⁷	0.394	0.260	-	-	0.405	0.026	0	0.998	0.067	0.726	0.628	2.00x10⁻⁴	-0.505	0.004
	Tri	0.893	7.59x10⁻⁶	0.679	0.005	-	-	-0.096	0.270	0.033	0.706	0.239	0.006	0.501	7.84x10⁻¹⁰	-0.281	1.06x10⁻³
	Tetra	0.838	9.71x10⁻²⁷	0.771	2.52x10⁻²⁰	-	-	0.108	0.053	0.071	0.203	0.288	1.24x10⁻⁷	0.650	2.23x10⁻⁴⁰	-0.183	8.90x10⁻⁴
Skewness in number of repeats	Di	0.207	0.565	-0.139	0.701	0.172	0.634	-	-	-0.043	0.820	-0.094	0.623	0.153	0.420	-0.432	0.017
	Tri	-0.148	0.599	-0.161	0.567	0.114	0.686	-	-	0.104	0.232	0.002	0.983	-0.041	0.639	0.084	0.334
	Tetra	0.216	0.034	0.163	0.111	0.157	0.124	-	-	0.016	0.779	-0.079	0.156	-0.052	0.347	-0.164	0.003
Mean PCR fragment size	Di	0.439	0.204	-0.103	0.777	0.437	0.207	0.309	0.385	-	-	0.114	0.550	0.020	0.916	-0.002	0.991
	Tri	-0.042	0.882	0.321	0.243	-0.147	0.602	-0.254	0.362	-	-	0.072	0.410	0.062	0.477	0.029	0.742
	Tetra	0.127	0.216	-0.005	0.964	0.120	0.241	0.031	0.765	-	-	0.052	0.348	0.045	0.424	-0.036	0.521
Mean number of repeats	Di	0.866	1.20x10⁻³	0.455	0.187	0.825	0.003	-0.018	0.960	0.394	0.260	-	-	0.656	8.20x10⁻⁵	0.597	5.03x10⁻⁴
	Tri	-0.115	0.683	-0.104	0.713	-0.163	0.562	0.229	0.413	-0.279	0.315	-	-	0.838	3.40x10⁻³⁶	0.737	4.77x10⁻²⁴
	Tetra	0.294	0.004	0.239	0.019	0.391	7.43x10⁻⁵	-0.015	0.884	0.319	1.47x10⁻³	-	-	0.817	2.47x10⁻⁷⁹	0.753	1.17x10⁻⁶⁰
Maximum number of repeats	Di	0.945	3.76x10⁻⁵	0.559	0.093	0.914	2.19x10⁻⁴	-0.024	0.947	0.413	0.235	0.912	2.37x10⁻⁴	-	-	0.322	0.083
	Tri	0.149	0.597	0.185	0.509	0.026	0.926	-0.059	0.834	-0.072	0.799	0.876	1.85x10⁻⁵	-	-	0.642	7.75x10⁻¹⁷
	Tetra	0.617	1.65x10⁻¹¹	0.546	7.11x10⁻⁹	0.725	4.65x10⁻¹⁷	0.100	0.328	0.251	0.013	0.836	1.75x10⁻²⁶	-	-	0.549	5.23x10⁻²⁷
Minimum number of repeats	Di	0.352	0.319	0.280	0.434	0.265	0.459	-0.292	0.413	0.235	0.235	0.596	0.069	0.573	0.083	-	-
	Tri	-0.351	0.200	-0.259	0.352	-0.507	0.054	-0.052	0.854	-0.115	0.683	0.785	5.33x10⁻⁴	0.726	0.002	-	-
	Tetra	-0.169	0.099	-0.140	0.171	-0.199	0.051	-0.093	0.366	0.231	0.023	0.677	2.65x10⁻¹⁴	0.445	4.96x10⁻⁶	-	-

Spearman's rank correlation coefficients (ρ) and their associated P values are shown for comparisons of the measures of variation across individuals in the HGDP-CEPH data set for loci with one (upper triangle) or two (lower triangle) STR regions embedded in their sequence. $|\gamma|$, the absolute value of skewness γ_1 , was used in evaluating correlations with skewness in number of repeats. Microsatellites were classified by the number of separate STR regions embedded in their sequence and by their repeat unit size. Sample sizes for loci with one STR region were 30, 133, and 325 for di-nucleotides, tri-nucleotides, and tetra-nucleotides, respectively, and for loci with two separate STR regions they were 11, 15, and 97, respectively. Correlations with $P < 0.05$ are highlighted in bold.

Table S6. Spearman's rank correlations between measures of variation across individuals for microsatellites with three tetra-nucleotide STR regions embedded in their sequence

	Variance in number of repeats		Range of number of repeats		Skewness in number of repeats		Mean PCR fragment size		Mean number of repeats		Maximum number of repeats		Minimum number of repeats	
	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P	ρ	P
Number of distinct alleles	0.848	4.98×10^{-4}	0.979	3.24×10^{-8}	0.336	0.285	0.308	0.330	0.602	0.038	0.771	0.003	0.312	0.323
Variance in number of repeats	-	-	0.843	5.72×10^{-4}	0.259	0.417	0.203	0.527	0.350	0.265	0.629	0.028	0.070	0.829
Range of number of repeats	-	-	-	-	0.317	0.315	0.275	0.387	0.646	0.023	0.825	9.45×10^{-4}	0.343	0.275
Skewness in number of repeats	-	-	-	-	-	-	-0.357	0.255	0.084	0.795	0.280	0.379	0.042	0.897
Mean PCR fragment size	-	-	-	-	-	-	-	-	0.238	0.457	0.301	0.342	0.175	0.586
Mean number of repeats	-	-	-	-	-	-	-	-	-	-	0.881	1.53×10^{-4}	0.862	3.15×10^{-4}
Maximum number of repeats	-	-	-	-	-	-	-	-	-	-	-	-	0.687	0.014
Minimum number of repeats	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Spearman's rank correlation coefficients (ρ) and their associated P values are shown for comparisons of the measures of variation across individuals in the HGDP-CEPH data set for the 12 loci with three tetra-nucleotide STR regions embedded in their sequence. No comparisons were performed for di-nucleotide and tri-nucleotide loci with three STR regions embedded in their sequence because of small sample size (3 and 2, respectively). Correlations with $P < 0.05$ are highlighted in **bold**.