Supplemental Information

Statistical Detection of Relatives Typed

with Disjoint Forensic and Biomedical Loci

Jaehee Kim, Michael D. Edge, Bridget F.B. Algee-Hewitt, Jun Z. Li, and Noah A. Rosenberg

Δ_{test} / Δ_{true}	Same individual		Parent–offspring		Sib pairs		Match-assignment scenario
	Median	Min, Max	Median	Min, Max	Median	Min, Max	
Same individual	1.000	0.986, 1.000	0.991	0.963, 1.000	1.000	0.991, 1.000	One-to-one
	0.982	0.945, 0.995	0.913	0.867, 0.950	0.972	0.945, 0.991	One-to-many: SNP query
	0.968	0.940, 0.991	0.922	0.885, 0.954	0.977	0.945, 0.995	One-to-many: STR query
	0.766	0.229, 0.927	0.193	0.005, 0.546	0.422	0.060, 0.780	Needle-in-haystack
Parent-offspring	0.248	0.174, 0.349	0.459	0.349, 0.587	0.404	0.248, 0.514	One-to-one
	0.248	0.138, 0.339	0.431	0.339, 0.541	0.376	0.284, 0.486	One-to-many: SNP query
	0.220	0.138, 0.303	0.450	0.321, 0.532	0.385	0.275, 0.477	One-to-many: STR query
	0.009	0.000, 0.064	0.046	0.000, 0.165	0.028	0.000, 0.138	Needle-in-haystack
Sib pairs	0.404	0.303, 0.541	0.404	0.303, 0.550	0.505	0.385, 0.661	One-to-one
	0.394	0.266, 0.532	0.404	0.312, 0.569	0.477	0.367, 0.615	One-to-many: SNP query
	0.376	0.248, 0.505	0.413	0.303, 0.596	0.477	0.385, 0.615	One-to-many: STR query
	0.046	0.000, 0.211	0.055	0.000, 0.165	0.092	0.009, 0.211	Needle-in-haystack

Table S1: **Record-matching accuracies between genome-wide SNP profiles and STR profiles that include the 13 CODIS loci and 4 additional loci from the 2017 CODIS update (D2S441, D10S1248, D19S433, and D22S1045); related to Table 2.** The table design follows Table 2 and was generated in same manner, employing the same 100 partitions of the sample into a training set (75%) and test set (25%), and adding the 4 loci for the match score computation.

Δ_{test} / Δ_{true}	Same individual		Parent–offspring		Sib pairs		Match-assignment scenario
	Median	Min, Max	Median	Min, Max	Median	Min, Max	
Same individual	1.000	1.000, 1.000	1.000	1.000, 1.000	1.000	1.000, 1.000	One-to-one
	0.995	0.977, 1.000	0.972	0.936, 0.991	0.991	0.977, 1.000	One-to-many: SNP query
	0.991	0.972, 1.000	0.977	0.945, 0.995	0.995	0.982, 1.000	One-to-many: STR query
	0.867	0.638, 0.972	0.468	0.073, 0.716	0.638	0.248, 0.835	Needle-in-haystack
Parent-offspring	0.257	0.156, 0.358	0.541	0.431, 0.679	0.450	0.349, 0.569	One-to-one
	0.266	0.202, 0.330	0.486	0.404, 0.606	0.422	0.330, 0.541	One-to-many: SNP query
	0.229	0.147, 0.284	0.505	0.422, 0.624	0.431	0.367, 0.532	One-to-many: STR query
	0.018	0.000, 0.073	0.073	0.009, 0.156	0.028	0.000, 0.073	Needle-in-haystack
Sib pairs	0.385	0.266, 0.514	0.450	0.349, 0.541	0.514	0.404, 0.633	One-to-one
	0.385	0.266, 0.477	0.431	0.376, 0.505	0.486	0.404, 0.578	One-to-many: SNP query
	0.321	0.229, 0.413	0.431	0.349, 0.523	0.486	0.413, 0.578	One-to-many: STR query
	0.064	0.009, 0.128	0.110	0.018, 0.229	0.101	0.018, 0.211	Needle-in-haystack

Table S2: **Record-matching accuracies between genome-wide SNP profiles and STR profiles that include the 13 CODIS loci, 4 additional loci from the 2017 CODIS update (D2S441, D10S1248, D19S433, and D22S1045), and 3 additional randomly chosen loci; related to Table 2.** The table design follows Table 2. We generated 100 sets of 3 additional tetranucleotide markers selected at random from among the 428 tetranucleotide loci not included among the 17 initial or more recent CODIS loci. For median-accuracy partitions into training and test sets obtained from Table S1 for one-to-one matching, the table presents median, minimum, and maximum match accuracies across the 100 locus sets. For each of the three relationship schemes, the median partition was taken from results with $\Delta_{\text{true}} = \Delta_{\text{test}}$.

Δtrue \ Δtest	Same individual		Parent–offspring		Sib pairs		Match-assignment scenario
	Median	Min, Max	Median	Min, Max	Median	Min, Max	
Same individual	1.000	0.931, 1.000	0.982	0.779, 1.000	1.000	0.898, 1.000	One-to-one
	0.954	0.857, 0.995	0.867	0.692, 0.959	0.931	0.807, 0.995	One-to-many: SNP query
	0.959	0.862, 0.995	0.885	0.715, 0.977	0.950	0.834, 0.995	One-to-many: STR query
	0.766	0.224, 0.963	0.440	0.037, 0.771	0.637	0.160, 0.885	Needle-in-haystack
Parent-offspring	0.390	0.055, 0.705	0.602	0.240, 0.935	0.528	0.175, 0.917	One-to-one
	0.362	0.083, 0.649	0.519	0.212, 0.787	0.454	0.185, 0.760	One-to-many: SNP query
	0.352	0.074, 0.649	0.528	0.204, 0.825	0.472	0.212, 0.742	One-to-many: STR query
	0.056	0.000, 0.335	0.139	0.000, 0.492	0.120	0.000, 0.428	Needle-in-haystack
Sib pairs	0.500	0.175, 0.816	0.528	0.167, 0.825	0.565	0.203, 0.917	One-to-one
	0.454	0.166, 0.686	0.454	0.176, 0.705	0.500	0.221, 0.769	One-to-many: SNP query
	0.454	0.184, 0.714	0.472	0.176, 0.741	0.519	0.240, 0.787	One-to-many: STR query
	0.139	0.000, 0.426	0.139	0.000, 0.464	0.168	0.000, 0.510	Needle-in-haystack

Table S3: **Record-matching accuracies between genome-wide SNP profiles and STR profiles, with geographic information taken into account; related to Table 2.** The table design follows Table 2. In all comparisons, only potential matches from the same geographic region were examined. To generate the pedigrees for use in the parent–offspring and sib-pair comparisons, simulated parental pairs considered individuals only within the same geographic region. For each geographic region, match scores were computed using allele frequencies based only on training-set samples from that region (parents in the training set for parent–offspring and sib-pair comparisons). The table reports a weighted average of regional medians, minima, and maxima, weighting regional results by regional sample sizes in the full data set. Otherwise, the analysis follows Table 2, with 100 partitions into training and test sets obtained in the same manner. Regional sample sizes for the 25% test set proportional to those in the full data set were obtained by the Hamilton, Huntington-Hill, and Webster methods of apportionment (Balinski, M.L., and Young, H.P. [2001]; *Fair Representation: Meeting the Ideal of One Man, One Vote*; Brookings Institution Press, Washington DC), all of which produced the same apportionment. For the same-individual comparisons, the numbers of individuals in the test set were: Africa, 19; Europe, 38; Middle East, 39; Central/South Asia, 49; East Asia, 57; Oceania, 6; America, 10. For the parent–offspring and sib-pair comparisons, the numbers of pedigrees in the test set were: Africa, 9; Europe, 19; Middle East, 19; Central/South Asia, 25; East Asia, 28; Oceania, 3; America, 5.