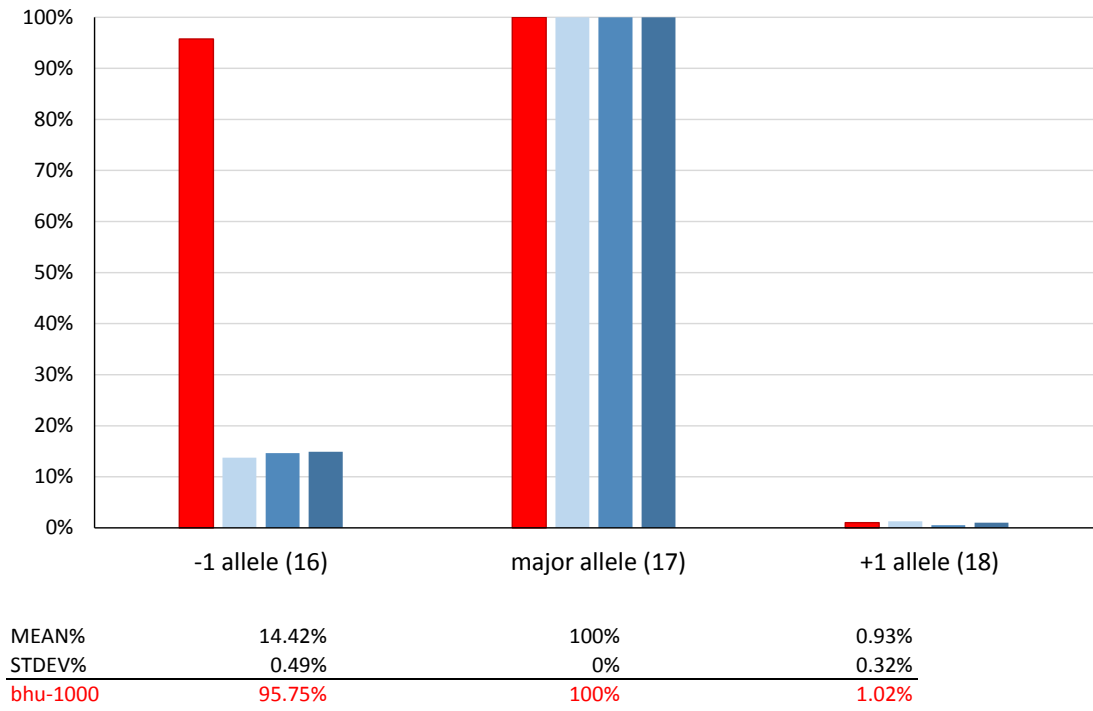
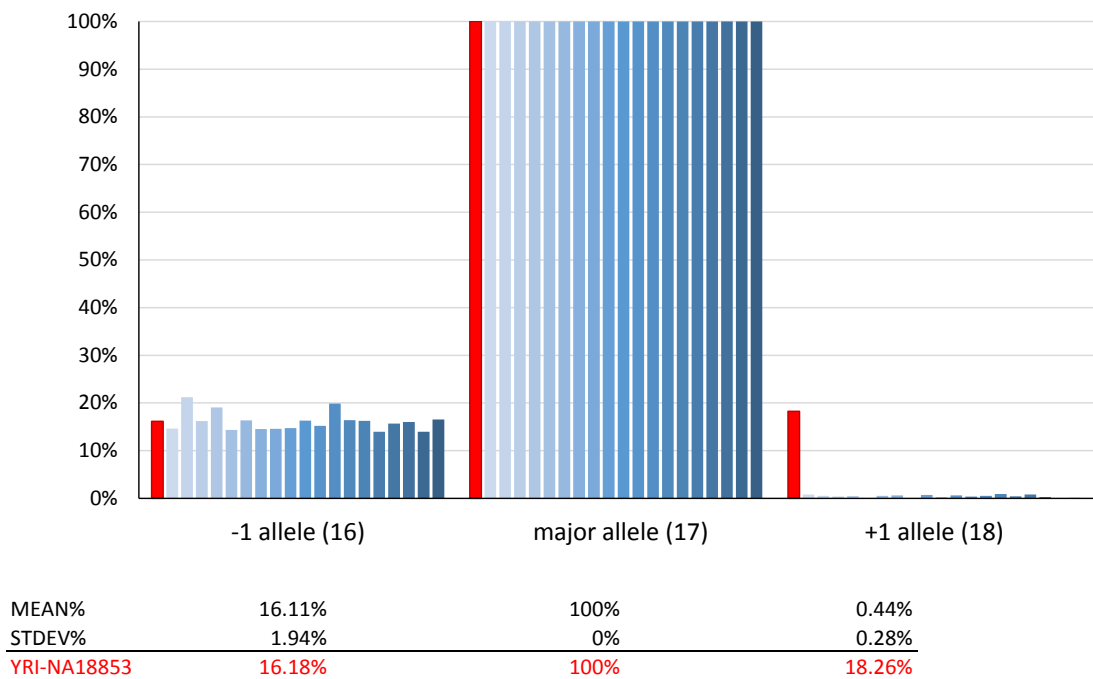


Figure S1: MSY phylogeny showing true branch lengths.

To the left is the schematic phylogeny relating the MSY haplogroups of the 100 samples in this study, as used in Figures 1 and 2, with haplogroup and sample names (as in Table S1) given at the tips. To the right is the same phylogeny, but with branch lengths based on the number of SNPs discovered in resequencing of 3.7 Mb of MSY DNA in each sample (Hallast et al., *Mol. Biol. Evol.* 32 [2015] 661-73). A scale bar representing 50 SNPs is shown.

a**DYS19****b****DYS458****Figure S2: Examples of an allele duplication and a somatic mutation deduced from read-depth observations.**

a) Read-depth of the two DYS19 alleles in sample bhu-1000 (in red), compared to same-size alleles in other samples (blue) and their corresponding stutters (-1 and +1). The sample has two alleles of approximately equal read-depth, consistent with an allele duplication.

b) Read-depth of the two DYS458 alleles in sample YRI-NA18853 (red), compared to same-size alleles in other samples (blue) and their corresponding stutters (-1 and +1). As well as the major allele, the sample has a +1 allele with read-depth much greater than the stutter observed in other samples, consistent with a somatic mutation. Note that stutter at the -1 position is high compared to the +1 position, so we do not call somatic mutant alleles at the -1 position.

Refer to Table S3 for information about the relative read-depth ratio approach to distinguishing allele duplications from somatic mutations.

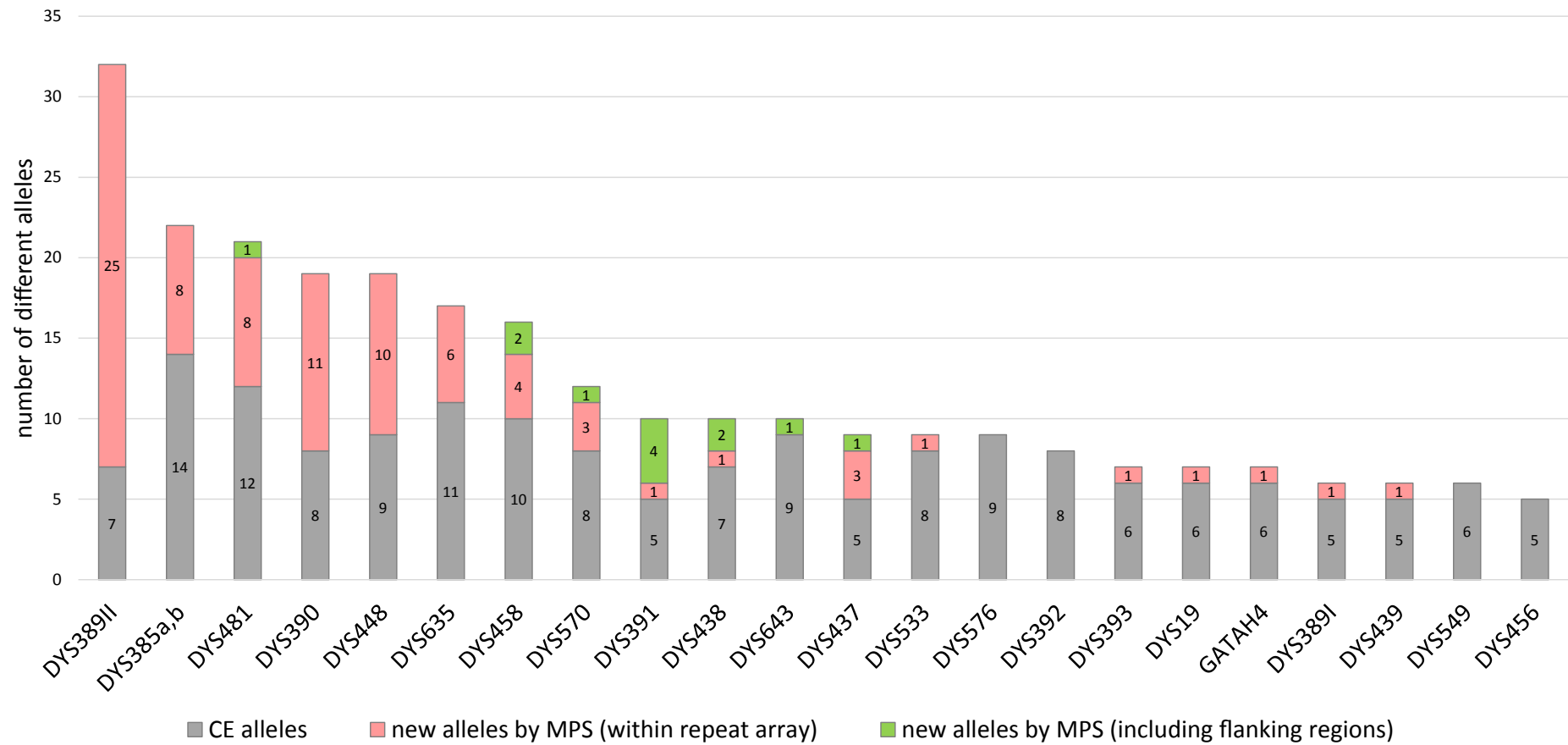


Figure S3. Increase in number of different alleles for each Y-STR using sequence information from MPS analysis.

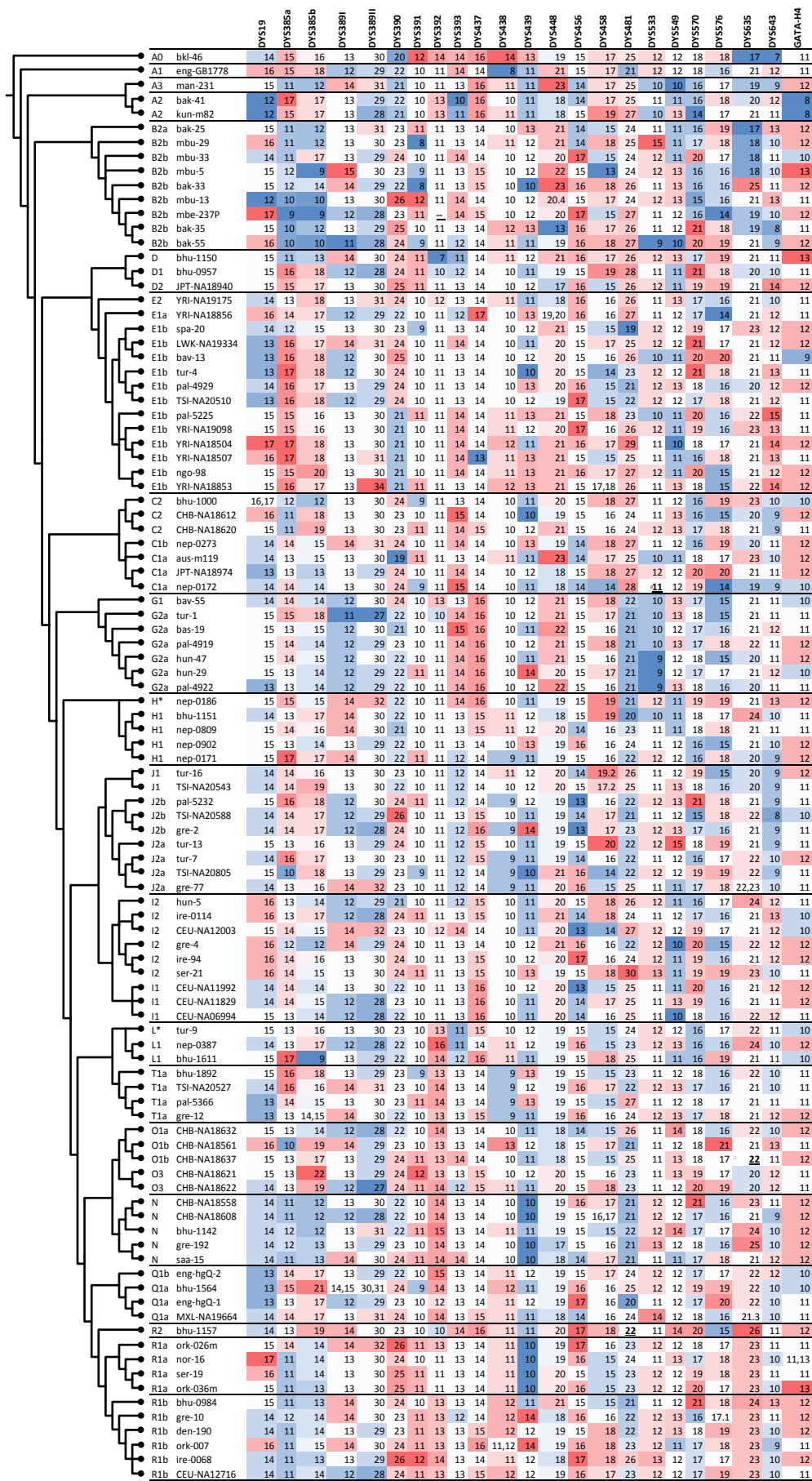


Figure S4: Length-based Y-STR alleles from MPS data in the 100 samples, in their phylogenetic context.

To the left is the schematic phylogeny relating the MSY haplogroups of the 100 samples in this study, with haplogroup and sample names (as in Table S1) given at the tips. To the right are shown the Y-STR allele length-based designations, with Y-STR names given at the top. For each Y-STR, cells containing allele names are heat-map coloured from blue (shortest) to red (longest). The four discordances listed in Table S4 are underlined.