

Database Descriptive Statistics

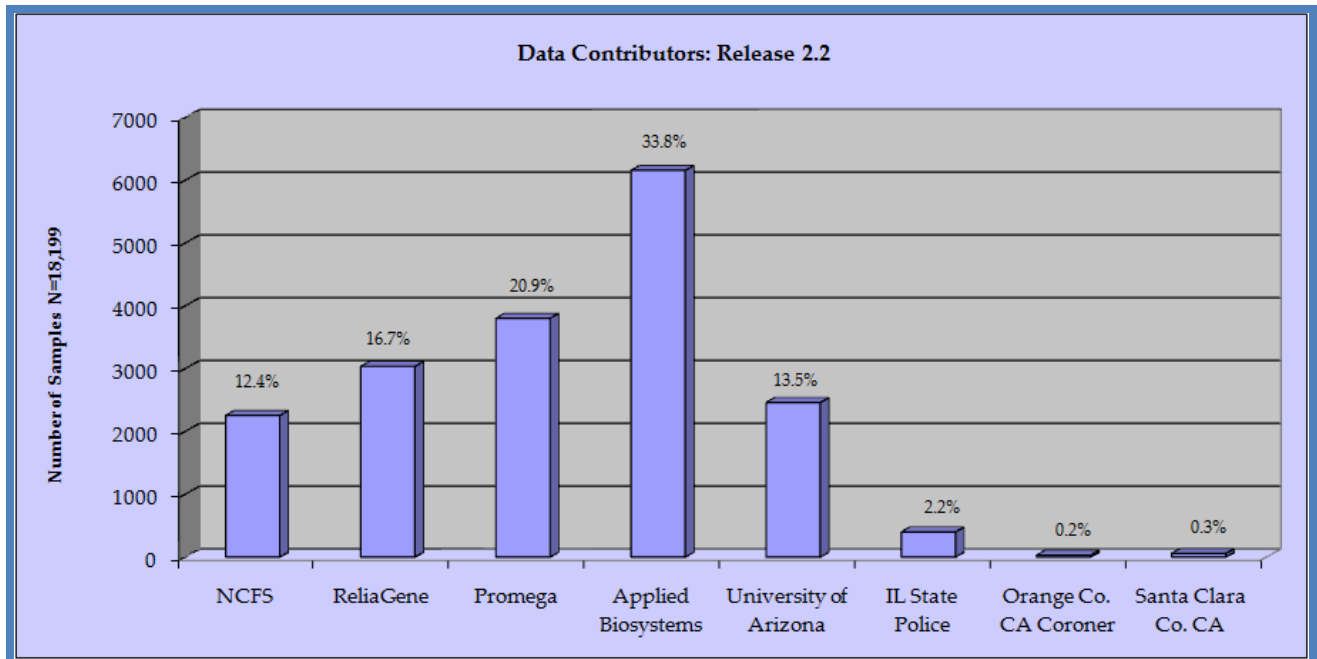
Release 2.2

Last Updated: January 24, 2010

Total Number of Haplotypes (N): 18,199

- [Data Contributors](#)
- [Database Ancestries](#)
- [Populated Loci](#)
- [Discrimination Potential](#)

Data Contributors



NCFs contributed 2,255 samples.

ReliaGene contributed 3,037 samples.

Promega contributed 3,800 samples.

Applied Biosystems contributed 6,159 samples.

The University of Arizona contributed 2,462 samples.

The Illinois State Police contributed 398 samples.

The Orange County CA Coroner contributed 30 samples.

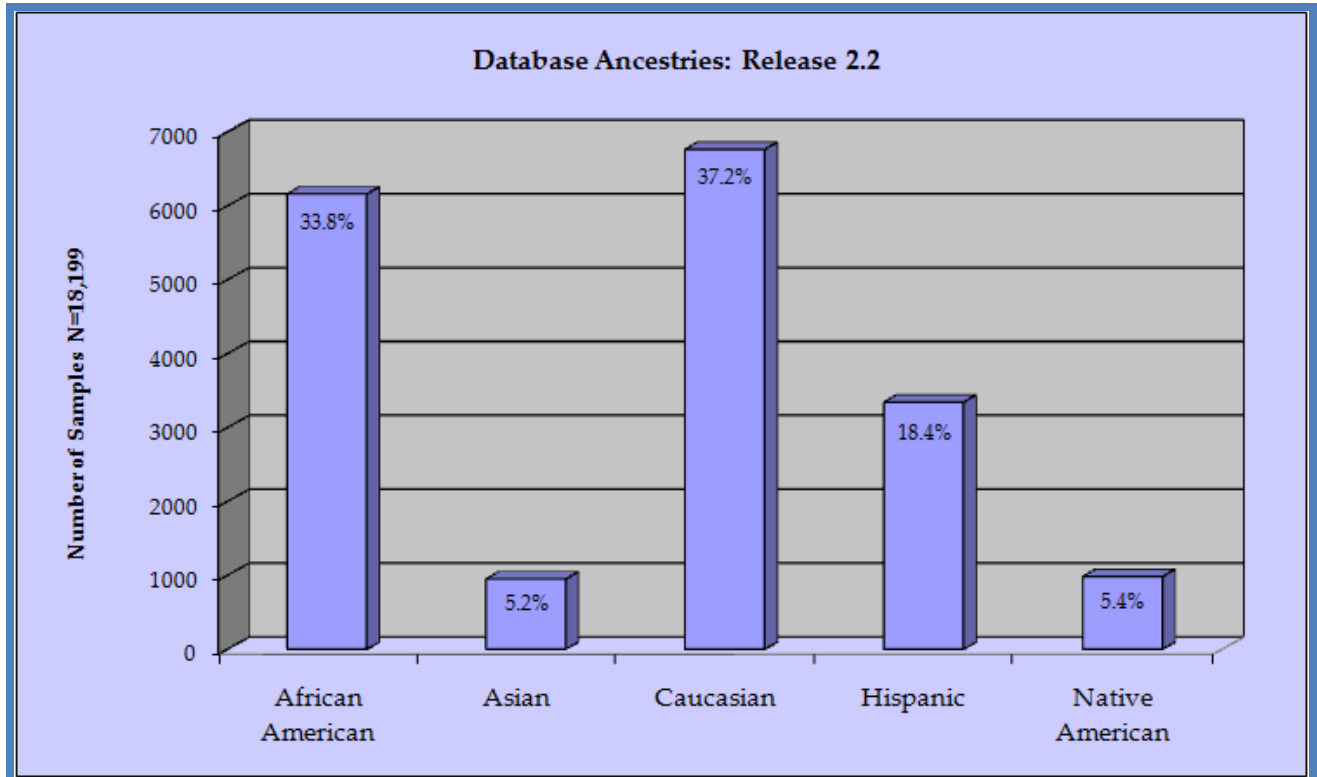
The Santa Clara County CA Crime Laboratory contributed 55 samples.

A number of individual samples were shared among the contributing data sets. All duplicate samples were removed to ensure that each sample in the consolidated database is from a unique individual. Any population group that did not contain at least 100 samples was also removed. These data reconciliation and reorganization steps have resulted in the consolidated US Y-STR Database having slightly different sample numbers than those found in the curated databases currently maintained by the individual contributing institutions.

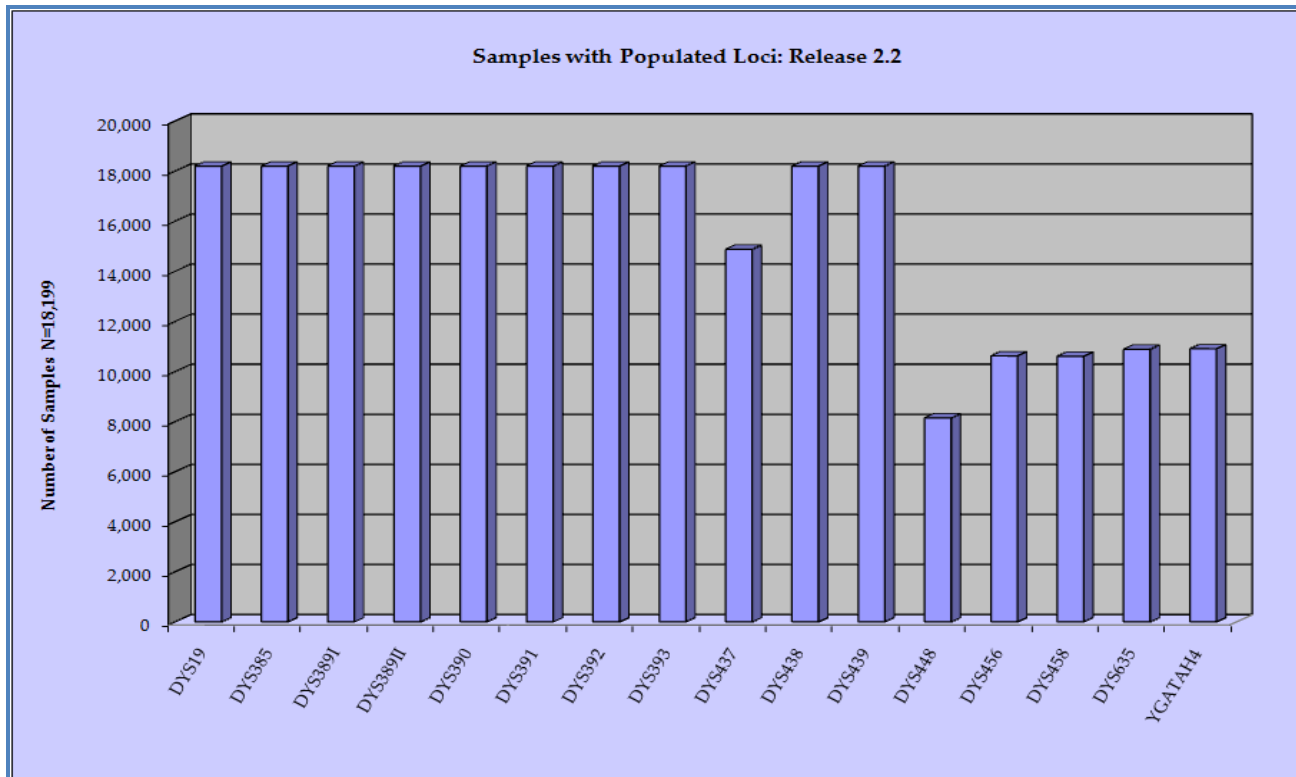
Database Ancestries

Samples are divided into five forensically relevant ancestries: African American, Asian, Caucasian, Hispanic, and Native American.

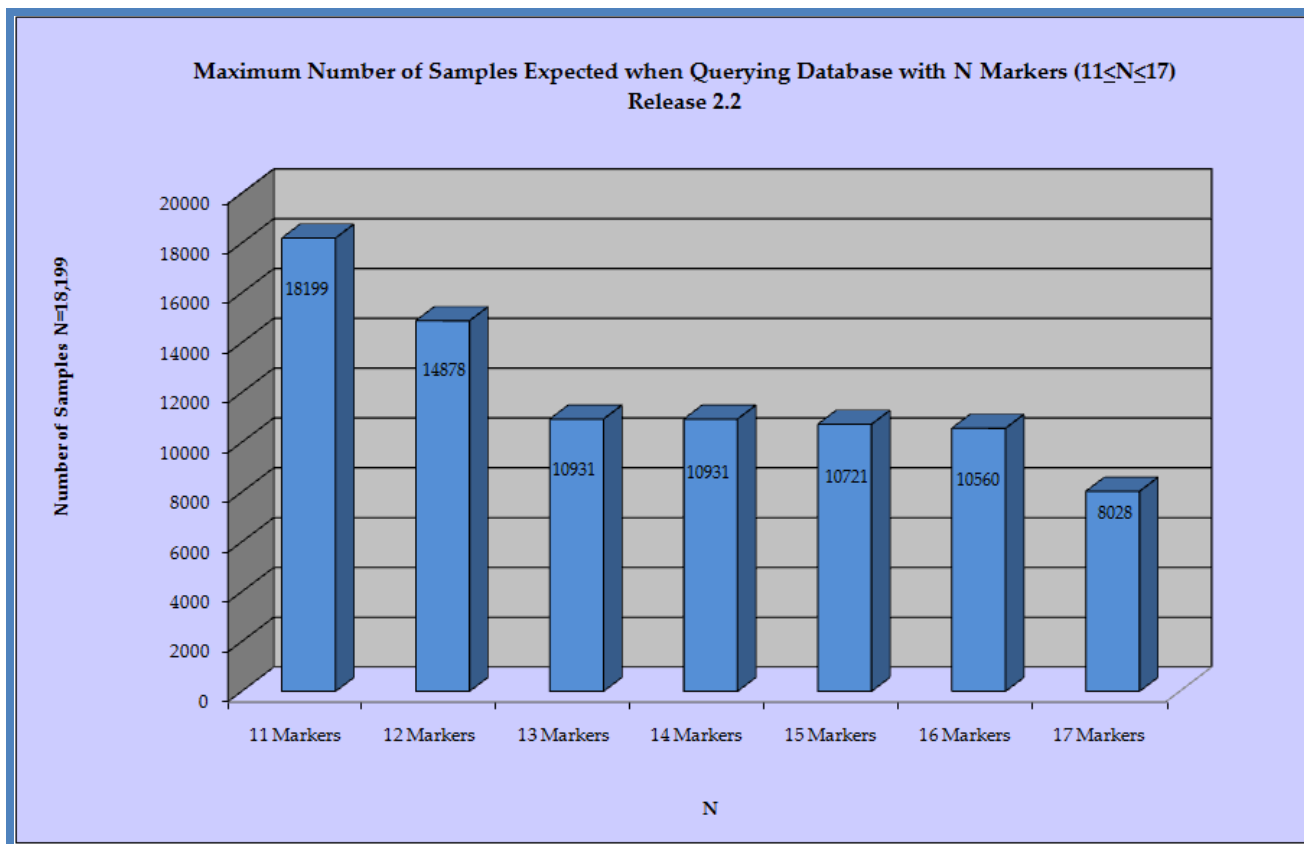
- 6160 African American (All, Undefined, or Select by State)
- 950 Asian (All, Filipino, Vietnamese, Oriental, Southern Indian, Asian)
- 6763 Caucasian (All, US, Canada, Europe, Undefined)
- 3343 Hispanic (All, Undefined, or Select by State)
- 983 Native American (All, Apache, Navajo, Shoshone, Sioux)

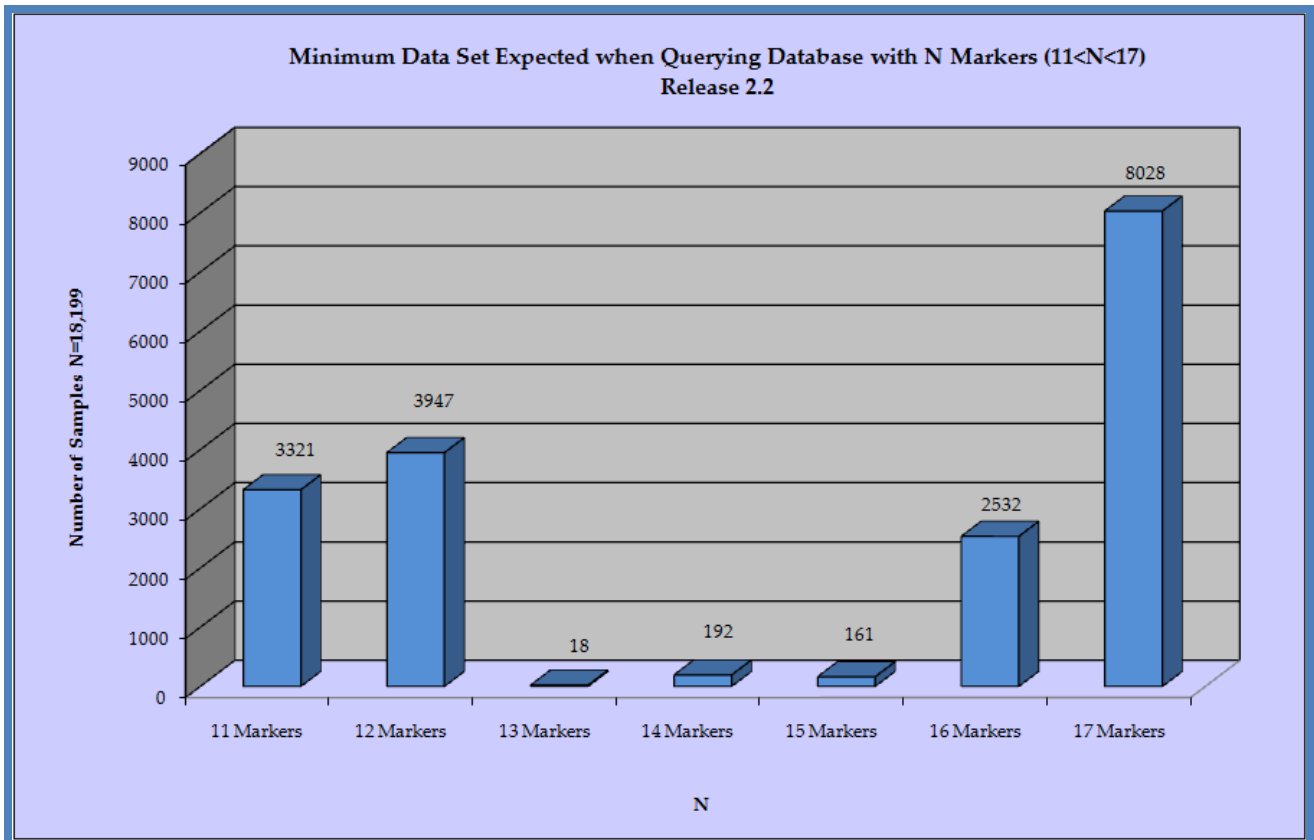


Populated Loci

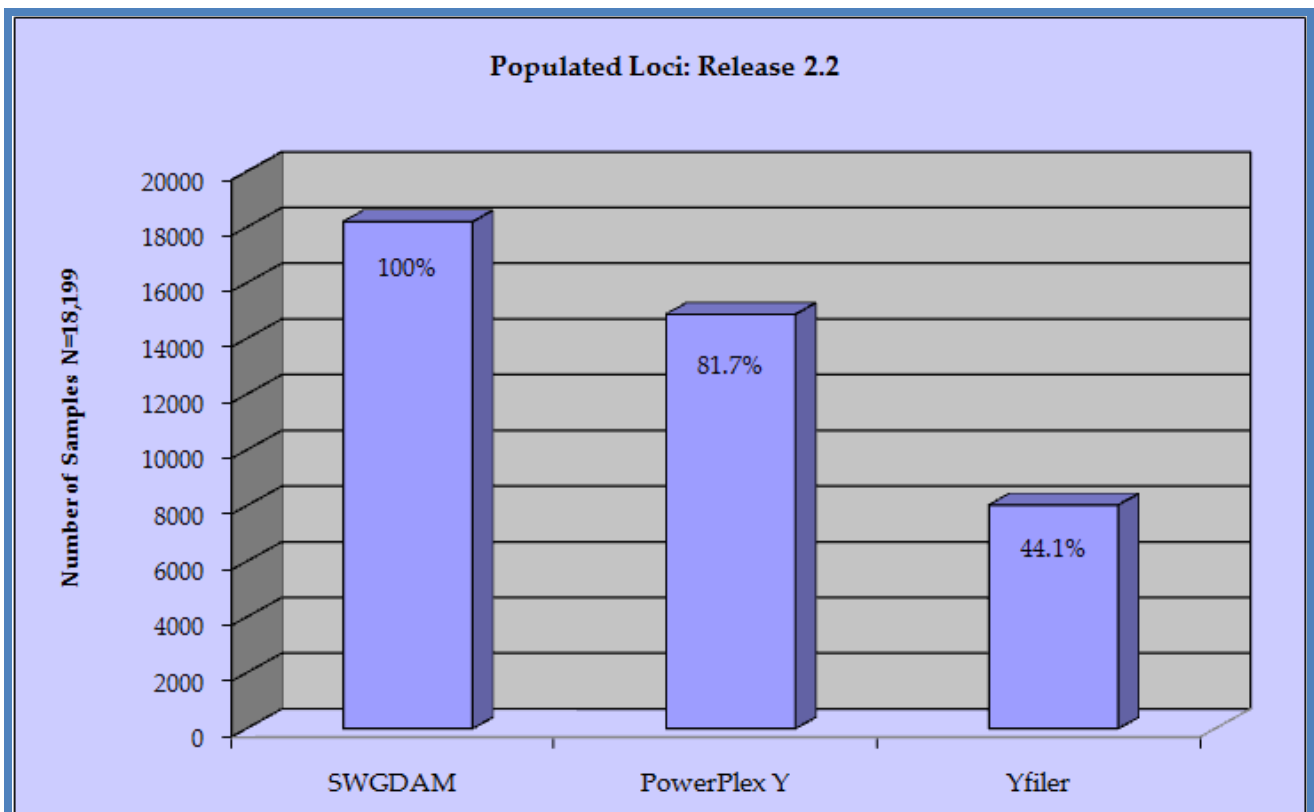


The sample data donated by NCFS and the University of Arizona were generated using Y-STR typing systems developed at these institutions rather than the more commonly used commercial kits. This has resulted in a variation in the number of populated markers within the Database. The Database is designed to query only those samples that possess data at the particular markers chosen by the user, resulting in a data set that varies depending on which markers are selected. The graph below shows the maximum data set that can be expected. The values for 12 – 16 markers can still vary depending on the markers chosen.

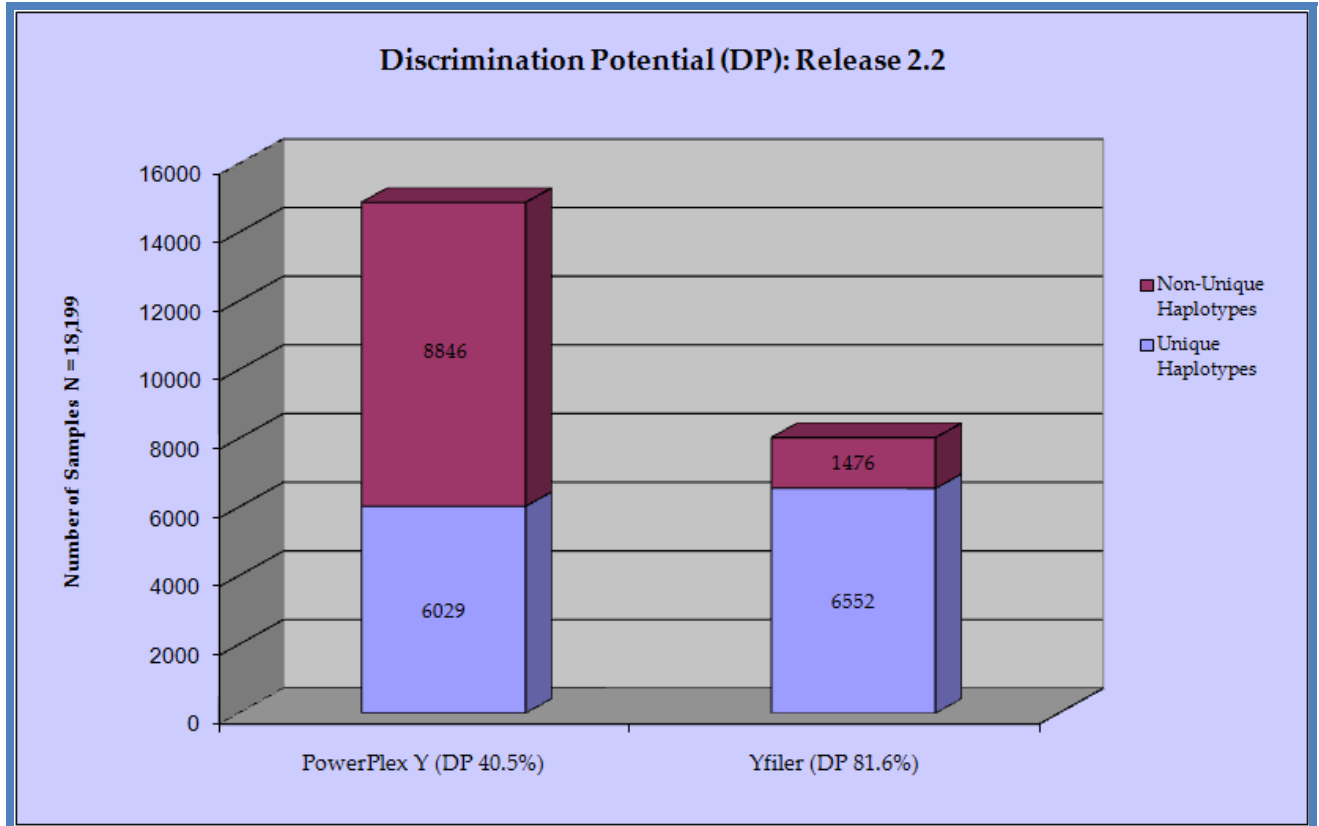




Release 2.2 of the US Y-STR Database contains 18,199 samples with a complete 11-marker SWGDAM core haplotype: DYS19, DYS385a,b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS439; 14,875 samples with a complete 12-marker PowerPlex Y haplotype: SWGDAM + DYS437; and 8,0828 samples with a complete 17-marker Yfiler haplotype: SWGDAM + DYS437, DYS448, DYS456, DYS458, DYS635 (YGATA C4), and YGATAH4



Discrimination Potential



Release 2.2 of the US Y-STR Database contains 14,875 complete PowerPlex Y (12-locus) haplotypes. Of these, 6029 haplotypes are unique (i.e., seen only once in the database) while 8,846 haplotypes are seen more than once, giving a discrimination potential of 40.5%.

The Database contains 8028 complete Yfiler (17-locus) haplotypes. Of these, 6552 haplotypes are unique while 1476 haplotypes are seen more than once, giving a discrimination potential of 81.6%.