

Database Descriptive Statistics

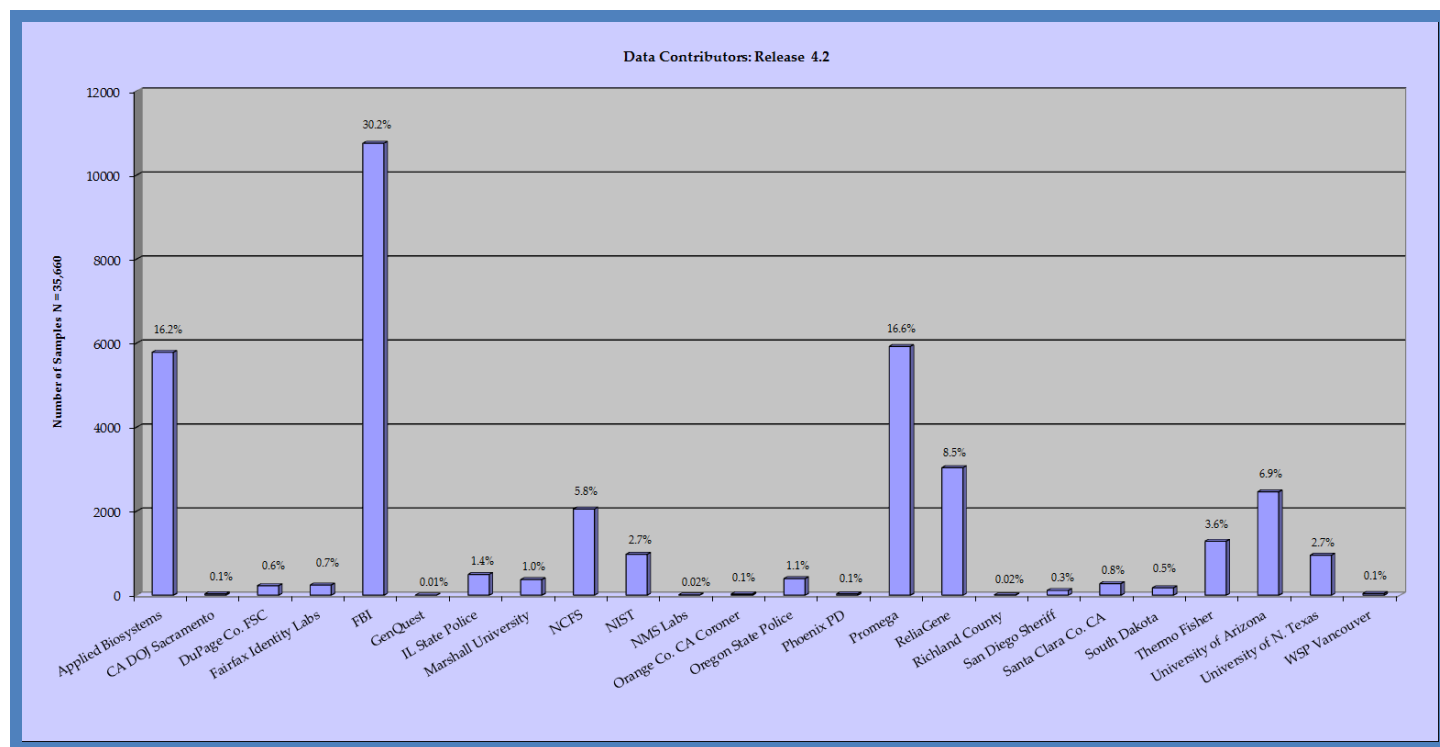
Release 4.2.1

Last Updated: January 26, 2018

Total Number of Haplotypes (N): 35,658

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

Data Contributors



Applied Biosystems: 5781 samples

California Department of Justice, Sacramento: 32 samples

DuPage County Forensic Science Center: 227 samples

Fairfax Identity Laboratories: 241 samples

Federal Bureau of Investigation: 10758 samples

GenQuest DNA Analysis Laboratory: 3 samples

Illinois State Police: 494 samples

Marshall University Forensic Science Center: 371 samples

National Center for Forensic Science: 2052 samples

National Institute of Standards and Technology: 974 samples

NMS Labs: 7 samples

Orange County CA Coroner: 30 samples

Oregon State Police: 394 samples

Phoenix Police Department: 32 samples

Promega: 5921 samples

ReliaGene: 3036 samples

Richland County Sheriff's Department: 7 samples

San Diego Sheriff's Department: 113 samples

Santa Clara County CA Crime Laboratory: 276 samples

South Dakota Forensic Lab: 175 samples

Thermo Fisher Scientific: 1281 samples

University of Arizona: 2462 samples

University of North Texas: 951 samples

Washington State Patrol Crime Lab, Vancouver: 40 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.

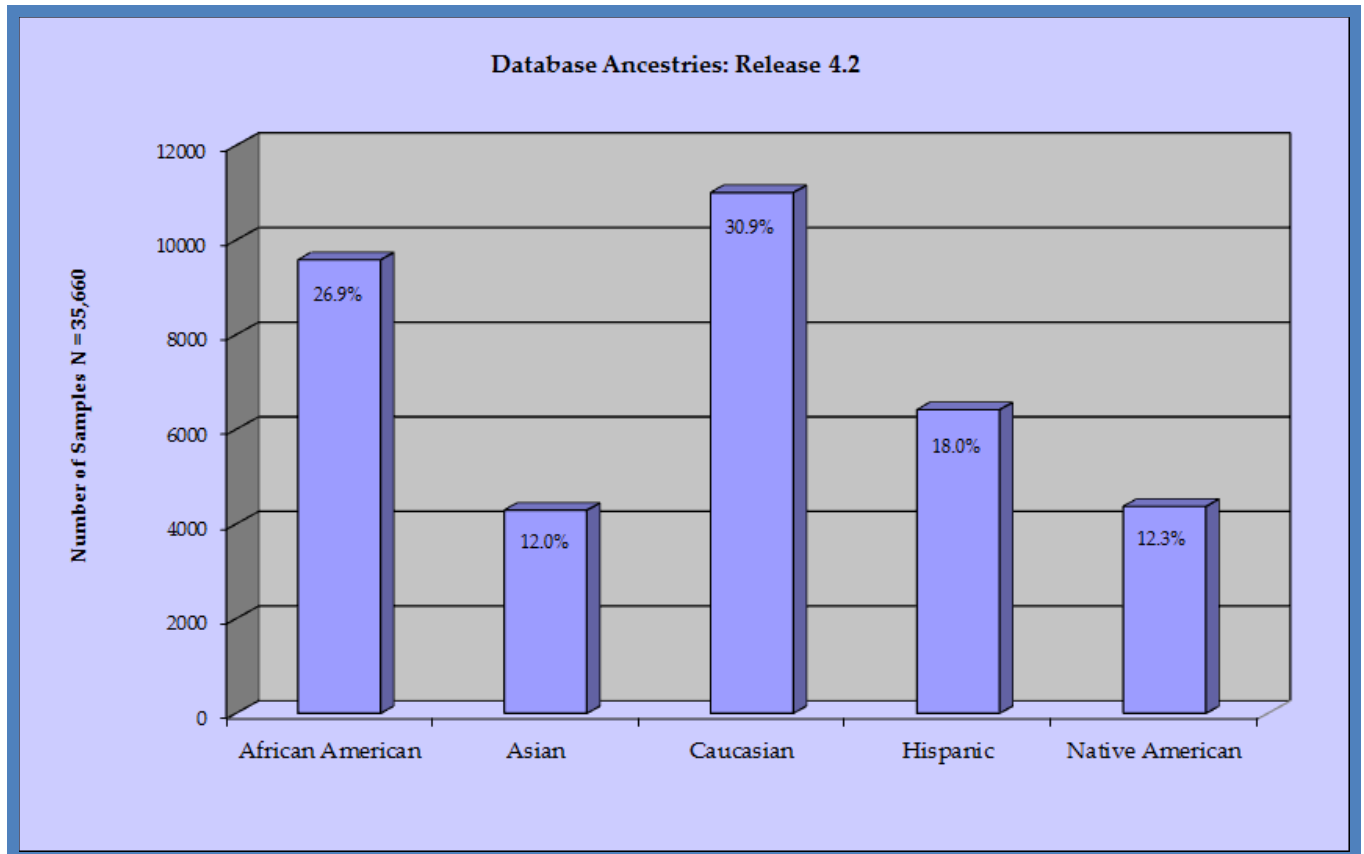
9581 African American (All, Undefined, or Select by State)

4289 Asian (All, Arabic, Asian, Chinese, Filipino, Indian, Jordanian, Middle East, Oriental, Southern Indian, Vietnamese)

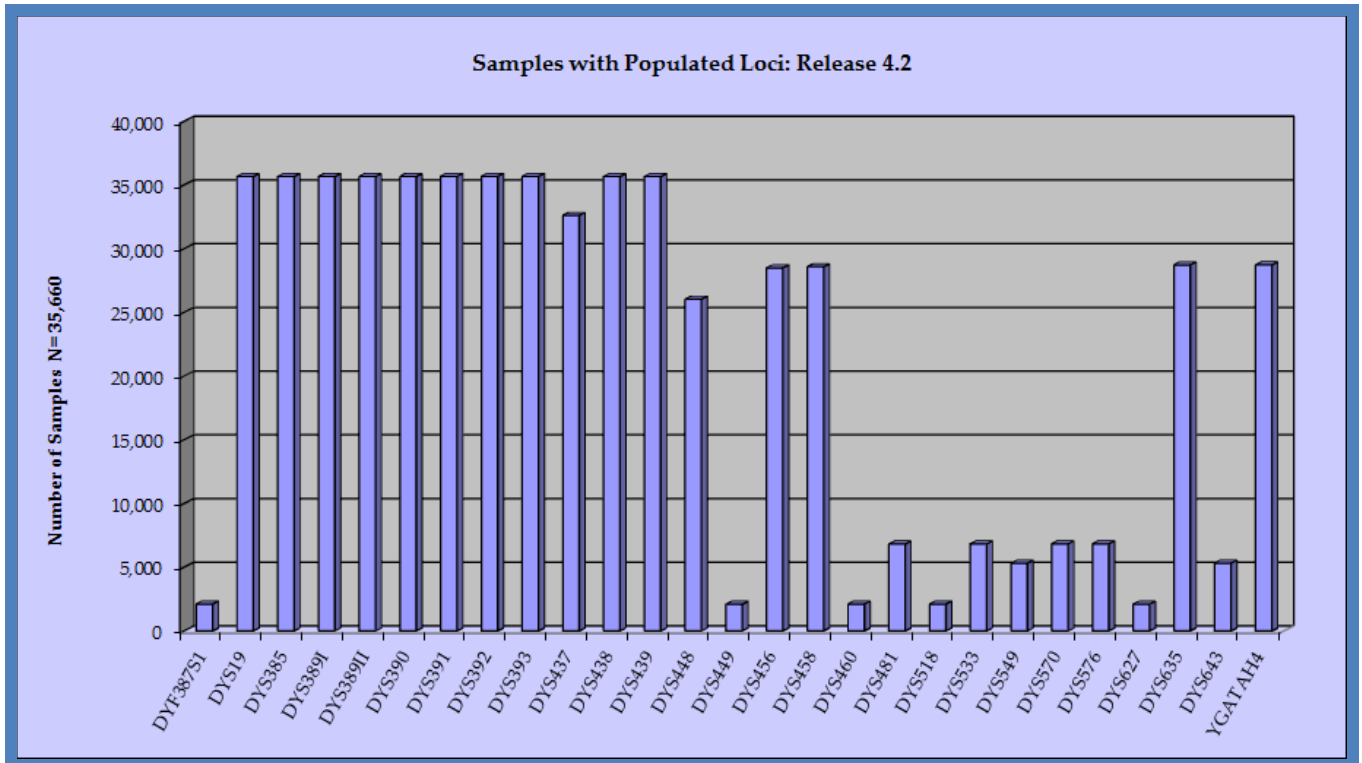
11003 Caucasian (All, US, Canada, Europe, India, Philippines, Select by State, Undefined)

6414 Hispanic (All, Undefined, or Select by State)

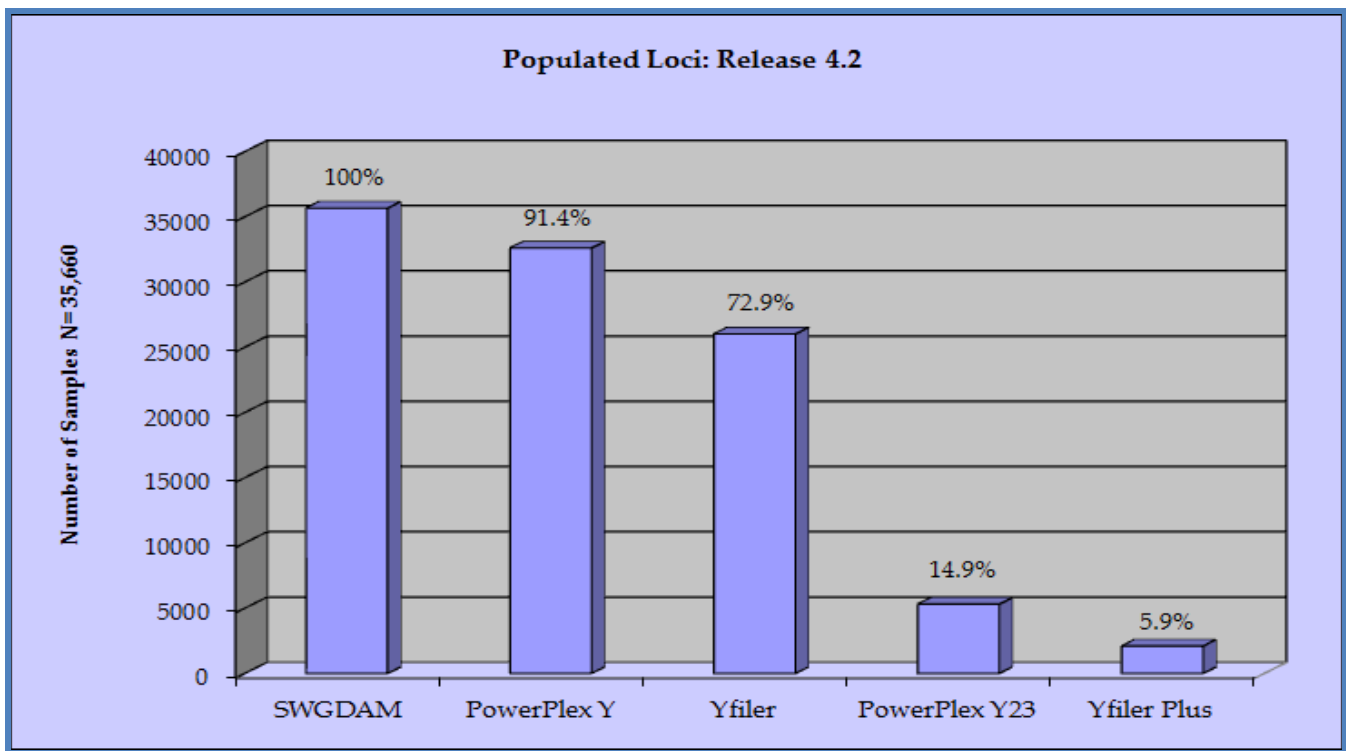
4371 Native American (All, Apache, Athabaskan, Inupiat, Navajo, Shoshone, Sioux, Yupik)



Populated Loci

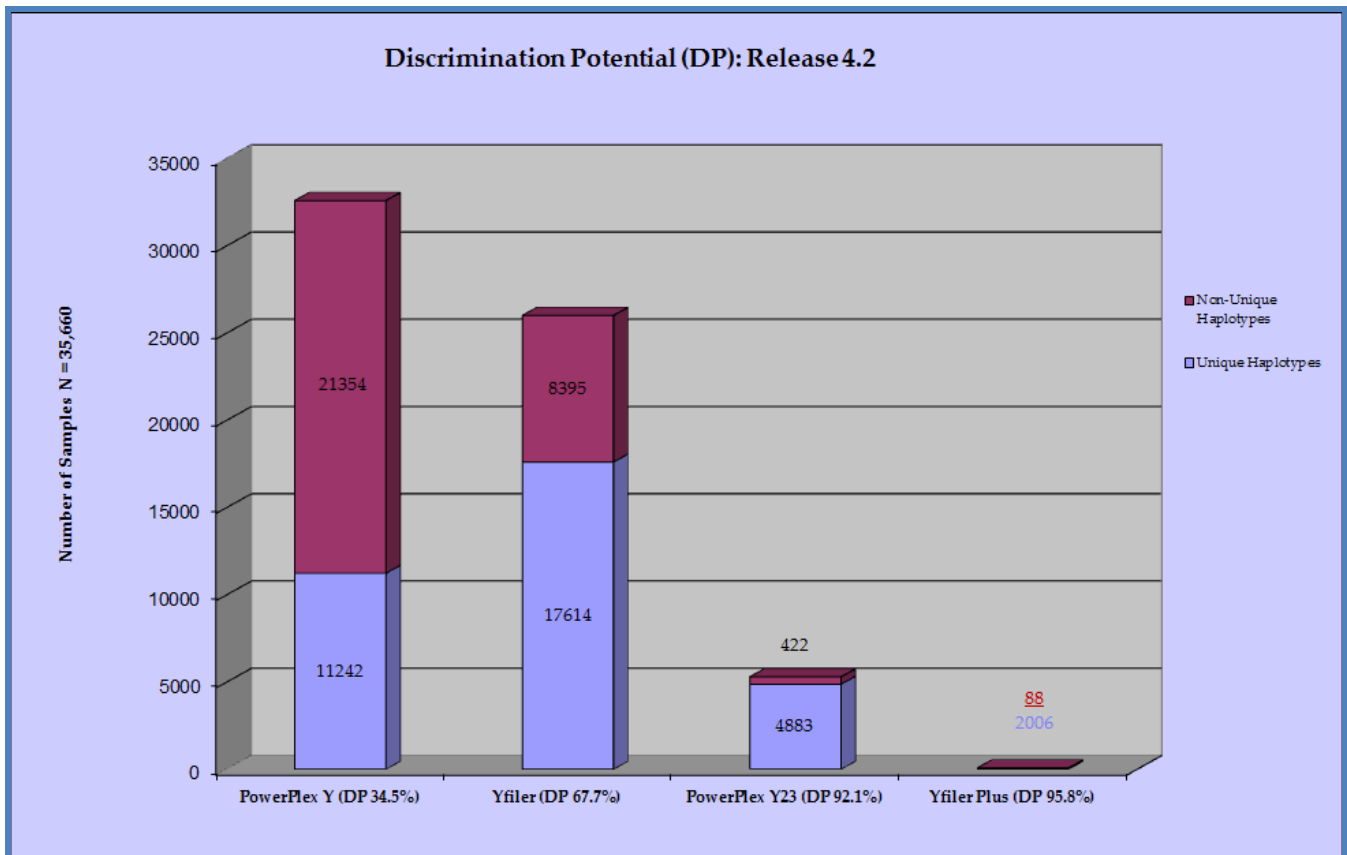


The sample data donated by NCFS and the University of Arizona were generated using Y-STR typing systems developed at these institutions rather than the more commonly used commercial kits. This has resulted in a variation in the number of populated markers within the Database. The Database is designed 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.



Release 4.2 of the US Y-STR Database contains **35,660** samples with a complete 11-marker SWGDAM core haplotype: DYS19, DYS385a,b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, and DYS439; **32,596** samples with a complete 12-marker PowerPlex Y haplotype: SWGDAM + DYS437; **26,009** samples with a complete 17-marker Yfiler haplotype: SWGDAM + DYS437, DYS448, DYS456, DYS458, DYS635 (YGATA C4), and YGATAH4; **5305** samples with a complete 23-marker PowerPlex Y23 haplotype: Yfiler + DYS481, DYS533, DYS549, DYS570, DYS576, and DYS643; and **2094** samples with a complete 27-marker Yfiler Plus haplotype: Yfiler + DYF387S1, DYS449, DYS460, DYS481, DYS518, DYS533, DYS570, DYS576, and DYS627. The database also contains samples with both PowerPlex Y23 and Yfiler Plus data, resulting in **575** haplotypes with a complete 29-marker profile.

Discrimination Potential



Release 4.2 of the US Y-STR Database contains 32,596 complete PowerPlex Y (12-locus) haplotypes. Of these, 11,242 haplotypes are unique (i.e., seen only once in the database) while 21,354 haplotypes are seen more than once, giving a discrimination potential of 34.5%.

The Database contains 26,009 complete Yfiler (17-locus) haplotypes. Of these, 17,614 haplotypes are unique while 8,395 haplotypes are seen more than once, giving a discrimination potential of 67.7%.

The Database contains 5305 complete PowerPlex Y23 (23-locus) haplotypes. Of these, 4,883 haplotypes are unique while 422 haplotypes are seen more than once, giving a discrimination potential of 92.1%.

The Database contains 2094 complete Yfiler Plus (27-locus) haplotypes. Of these, 2006 haplotypes are unique while 88 haplotypes are seen more than once, giving a discrimination potential of 95.8%.