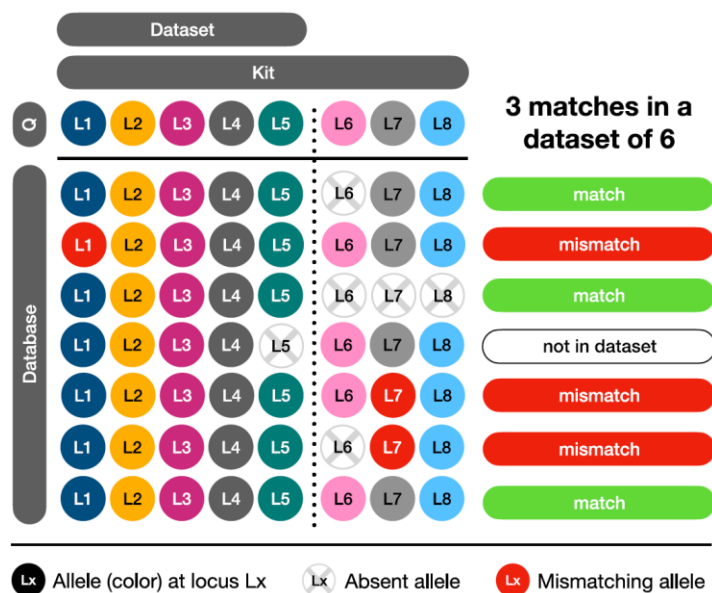YHRD updates for U.S. Laboratories

January 28, 2022

This letter supplants all previous drafts or versions of this letter.

As was announced during the SWGDAM update at the CODIS conference (November 15, 2021) the Y chromosome Haplotype Reference Database (YHRD) is implementing changes to its search capabilities. This notice serves to provide laboratories the most up-to-date information about the approaching YHRD changes and how laboratories can prepare.
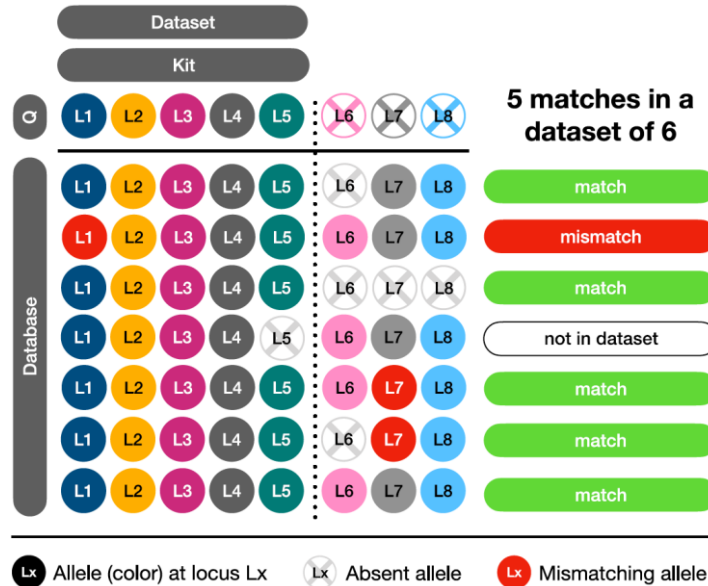
Beginning February 7, 2022, YHRD will separate the dataset (Minimal, Y12, Y17, Y23, Y27, and YMax) and kit (Minimal, PowerPlex Y, Yfiler, PowerPlex Y23, Yfiler Plus, etc.). A dataset is a subset of database samples that minimally includes the selected set of loci (e.g., when selecting dataset Y17, all database samples used for comparison will contain all loci in the Yfiler kit but may have additional loci). A kit refers to a multiplex amplification system (set of loci) and therefore represents the resolution of the query haplotype. Both attributes, dataset and kit, can be selected independently when performing searches at YHRD.

There are effectively three different options when combining dataset and kit regarding their corresponding resolution:
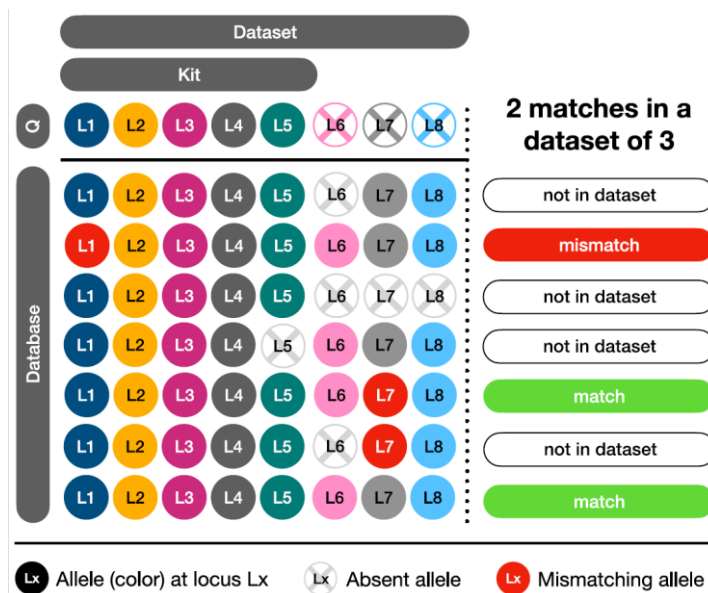
Dataset < Kit ("transient search"): The database is filtered to exclude any haplotypes that do not have full coverage of the dataset loci. This set is then compared with the dataset proportion of the query haplotype. Finally, any loci beyond those of the dataset will be used to exclude non-matches (mismatch between query and database) from that set. The results of that search are in most cases equal to the "reduced locus search" that was previously only possible through an iterative process described in the 2014 SWGDAM Compliant YHRD User's Guide (available at YHRD). An example of a transient search is a Yfiler Plus query against the Y17 dataset. This query eliminates any mismatches at the Y17 loci as well as any additional Yfiler Plus loci that were typed in those database samples.

Dataset = Kit ("masked search"): This search replicates the searches performed prior to the 2022 changes. The database is filtered to exclude any haplotypes that do not have full coverage of the dataset/kit loci. This set is then compared with the dataset/kit proportion of the query haplotype. An example of a masked search is a Yfiler query against the Y17 dataset. This query eliminates any mismatches at the Y17 loci that were typed in those database samples. Any additional loci present in the database samples cannot be used for comparison since only Yfiler loci were entered for the query haplotype.

Dataset > Kit: The database is filtered to exclude any haplotypes that do not have full coverage of the dataset. This set is then compared with the kit proportion of the query haplotype. Loci within the dataset which are absent from the kit are ignored and do not affect matches. An example of a search is a Yfiler query against the Y23 dataset. This query eliminates any mismatches at the Yfiler loci only.

Conducting a reduced locus search by toggling back and forth between kit loci and recording each kit's results will no longer be necessary. However, a laboratory may choose to perform transient searches of the haplotype using datasets with fewer loci than the kit in order to determine the most informative statistic.

Until the changes take full effect, YHRD will allow users to perform searches using both the pre-2022 (current) and 2022 configurations through a beta test site. This will provide laboratories the opportunity to conduct necessary tests. To access the beta test site, visit https://beta.yhrd.org/. For current YHRD users, these changes to the YHRD search algorithms are considered a major software revision and therefore require a validation based on FBI Quality Assurance Standard 8.8.3.3.