User Manual v1.3

Universal Y-SNP Database

2026

Contents

1. Introduction

The Universal Y-SNP Database (UYSD) is an interactive online platform designed to facilitate the study of human Y-chromosome genetic variation. By harmonizing diverse datasets under a unified phylogenetic tree, UYSD provides researchers and the public with a comprehensive resource for exploring Y-SNP and haplogroup distributions across global populations.

Scientific interest in paternal lineages has grown over the decades, but relevant data has often been fragmented across numerous publications, making it difficult to consolidate and analyze effectively. UYSD addresses this challenge by offering a searchable and accessible database that standardizes haplogroup information, ensuring consistency in nomenclature despite historical changes.

The platform enables users to:

- Search and visualize Y-haplogroup distributions on an interactive world map.
- Estimate haplogroup frequencies in regions with sparse data using interpolation.
- Access and explore a detailed phylogenetic tree of Y-chromosomal haplogroups based on the YFull v14 tree (42,041 nodes).
- Submit and integrate new datasets, fostering collaboration and data sharing within the scientific community.
- Download search results as CSV files for offline analysis.

UYSD is compatible with both high- and low-throughput sequencing technologies and supports automated haplogroup prediction via **Yleaf v4** or later. This functionality allows researchers to analyze the deepest branches of the Y-chromosome phylogenetic tree more effectively, including support for T2T (CHM13/hs1) coordinates.

By continuously expanding and refining its dataset, UYSD serves as a valuable tool for researchers in fields such as human population genetics, genetic anthropology, forensic genetics, and ancient DNA analysis. The platform encourages authors of newly published and previously published Y-SNP studies to contribute their data, helping to enhance the collective understanding of human migration patterns and genetic diversity.

This manual provides detailed guidance on how to navigate and utilize UYSD, including data submission, searching, and visualization features. A dark mode can be enabled at any time by clicking the moon/sun icon in the top navigation bar.

2. Y-SNP Database

This section of the website serves as a resource for finding Y-SNPs of interest. It includes a database containing all Y-SNPs incorporated into UYSD, allowing users to search for specific Y-SNPs and access their associated information simply by typing the name of the Y-SNP of interest.

Alternatively, users can search for a specific haplogroup to retrieve all Y-SNPs (including alternative names and equivalent Y-SNPs) that define that haplogroup. It is also possible to search within a specific genomic region by entering a coordinate range in one of the following formats:

- **hg19:start..stop** — GRCh37 / hg19 coordinates
- **hg38:start..stop** — GRCh38 / hg38 coordinates
- **t2t:start..stop** — T2T CHM13 / hs1 coordinates

While the majority of Y-SNPs do not have an associated rs number, users can search for those that are included in the database. The Y-SNP database currently contains over 400,000 entries, with positions available for hg19, hg38, and T2T reference assemblies.

Additionally, this section provides seamless navigation to related features based on the selected haplogroup. With a single click, users can access the haplogroup map, sample list, or locate the haplogroup in the phylogenetic tree.

When a search returns results, a ■ **CSV** button appears in the navigation bar. Clicking it downloads the complete result set (all matching rows, not just the current page) as a CSV file.

3. Sample Database

All samples included in UYSD are available in the Sample Database (without displaying sample IDs). This page provides additional details about each sample, including the date of submission, the contributor's name, and, if analyzed using Yleaf, the corresponding quality control (QC) score.

Additionally, the page includes a feature that allows users to directly contact the individual who submitted the data through a built-in contact form. This facilitates collaboration and enables researchers to seek further information about specific datasets.

The sample database supports multiple search criteria, allowing users to filter results based on haplogroup, author, country, or region. When searching for a country without specifying a region, a colon (:) must be added after the country name to ensure accurate results. For example, to search for all samples from Belgium without regional details, enter **"Belgium:"** in the search bar. To search specifically for East Flanders, type **"Belgium:East Flanders"** instead.

When searching by haplogroup, a checkbox labeled **Include sub-haplogroups** will appear. Enabling this option returns all samples assigned to any descendant of the queried haplogroup in addition to exact matches. A **View as tree** link is also available to visualize the sub-haplogroup distribution in a phylogenetic tree view.

When results are displayed, a ■ **CSV** button appears allowing the full result set to be downloaded as a comma-separated file.

4. Haplogroup Map

This page features one of the core functionalities of UYSD — generating interactive geographic maps that display the frequency of a given haplogroup in populations where it has been analyzed.

4.1 Performing Searches

Users can search for a haplogroup or SNP by simply typing its name into the search bar. Searches are not case-sensitive. The percentages displayed on the map represent the proportion of samples typed for the queried haplogroup. This percentage is calculated as:

(Number of samples with the derived allele) / (Total samples with either the derived or ancestral allele)

Results are generated only if at least one Y-SNP defining the searched haplogroup is present in the analyzed data. Consequently, the number of countries which show frequency data and the number of samples per country may vary depending on the search performed.

In addition to single haplogroup searches, users can refine results using the asterisk (*) notation. This feature helps identify samples that are derived for a primary haplogroup but ancestral for specified subclades. For example, searching for **R1b*(xL51)** will return all samples that are derived for R1b but ancestral for R-L51.

4.2 Display Options

Interpolation feature (Use interpolation checkbox)

This feature estimates haplogroup frequencies in regions without available data by analyzing surrounding areas. For a region to be eligible for interpolation, it must have at least three neighboring regions with known frequencies within 1,000 kilometers. The estimated frequency is calculated using a weighted average where closer regions have a stronger influence. When enabled, hovering over a country shows either 'Interpolation: false' (actual data) or 'Interpolation: true' (estimated).

Country vs. Regional View (Only Show Country checkbox)

When enabled, the map displays haplogroup frequencies for entire countries. When disabled and sub-country regional data is available, the map shows individual regional frequencies.

Dynamic vs. Absolute Scaling (Dynamic Colors checkbox)

Dynamic Scaling adjusts the color gradient to reflect the highest observed frequency in the dataset. If unchecked, a fixed scale ranging from 0% to 100% is used for consistent comparisons.

4.3 Analyzing the Results

Once a search has been performed, a world map is displayed where colors indicate regions for which haplogroup frequency data is available. Hovering over a country or region shows the sample size and observed frequency.

At the top of the page, the phylogenetic pathway leading to the searched haplogroup is shown, including parent haplogroups and direct subclades. Each haplogroup in this pathway is clickable.

In the bottom-right corner of the map, a color scale indicates the frequency range. Clicking a country with non-zero frequency opens a detailed sample table including author information, a publication link, and a contact option. Selecting a country also reveals a **"Show all haplogroups for this country in a tree"** link for deeper phylogenetic exploration.

5. Haplogroup Tree

Phylogenetic Y-haplogroup trees have grown increasingly extensive. To help users keep track of the haplogroups included in UYSD and understand their relationships, the platform provides an interactive tree view. The tree is based on the **YFull v14 phylogenetic tree** and contains over 42,000 haplogroup nodes.

The tree is interactive — clicking on a haplogroup highlights the downstream phylogenetic path in red. After clicking a haplogroup, the panel on the right side of the screen provides hyperlinks to:

- Navigate to its geographic frequency map.
- View a list of associated Y-SNPs.
- Access the samples in UYSD that belong to this haplogroup.

Users can also search for a specific haplogroup using the search box to navigate directly to its position in the tree. The tree supports touch and mouse drag for navigation on all devices.

6. Data Submission

The Data Submission page is accessible only to registered users of UYSD. To gain access, users must be affiliated with an academic institution. Once an account has been created, contributors can view or delete the samples they have provided at any time.

Submitting data requires agreeing to the Terms and Conditions, which are available on the Data Submission page.

6.1 Submission Methods

There are three ways to submit data:

Yleaf Dashboard (recommended)

The easiest and most streamlined method. The **Yleaf v4.1** desktop dashboard (Windows, macOS, and Linux) can submit results directly to UYSD. After analysing your samples, log in to your UYSD account from within the dashboard, then import an existing country file or fill one in using the built-in table, which offers country and region autocomplete and validation. A single button then packages the Yleaf output together with the country file, uploads both to UYSD, and reports whether the submission succeeded. Any sample with a QC score below 0.7 is automatically excluded.

Yleaf v4 Output Files

Upload the ZIP or TAR.GZ output directory produced by **Yleaf v4 or later** on the Data Submission page of this website. Any sample with a QC score below 0.7 is automatically excluded. Haplogroup names are matched against the YFull v14 tree embedded in UYSD.

Manual Submission

Upload a list of Y-SNPs typed in an assay, along with a file specifying the final haplogroup assignment for each sample. Suitable for data generated by means other than Yleaf.

For all of these submission methods, an additional **country file** is required containing:

- Sample IDs — these will not be displayed within UYSD (required).
- Country (required).
- Sub-country region (optional).
- Additional comments about the samples to be displayed on UYSD (optional).
- A link to the original publication where the samples were first described (required).

6.2 Sub-Country Region Information

The Data Submission page offers an option to view all sub-country regions that can be included in UYSD. This information is available in two formats:

- Interactive world map.
- Detailed list view.

7. About

The About page contains a contact form which can be used to reach the administrator of UYSD with questions or concerns. It also provides a link to this manual and to the UYSD publication.

If you found this resource useful, please cite **our paper** and the original publications of the used datasets.