– Directions for Use –

The Y-Chromosome Haplotype Reference Database

– Directions for Use –

Copyright and curation: Sascha Willuweit & Lutz Roewer
Institute of Legal Medicine and Forensic Sciences, Charité – Universitätsmedizin Berlin

Endorsed by International Society of Forensic Genetics (ISFG).

Supported by Promega Corporation.

Based on YHRD Release 61 • September 16th, 2019

Corresponding address
Sascha Willuweit & Lutz Roewer
Charité – Universitätsmedizin Berlin
Institute of Legal Medicine and Forensic Sciences
Department of Forensic Genetics
Augustenburger Platz 1
D-13353 Berlin
Germany

E-Mail
Sascha Willuweit: sascha.willuweit@charite.de
Lutz Roewer: lutz.roewer@charite.de

URL
https://yhrd.org

Contents

1 Introduction

The Y-Chromosome Haplotype Reference Database (YHRD) (Figure 1) aims to help with the interpretation of results from comparisons of evidentiary samples typed with Y-STRs and reference samples, and to formulate conclusions. Since Y-STRs are located on the non-recombining part of the Y chromosome, the profile generated by Y-STR analysis should be considered as one trait coded by one locus (a haplotype). Consequently, the YHRD provides allele and haplotype (>1 locus typed per sample) frequencies for common marker sets consisting of up to 29 loci (Figure 2). The database is explained in detail in the article "The new Y Chromosome Haplotype Reference Database" by Willuweit and Roewer (2015).

Figure 1: YHRD Homepage

Name	Description	Loci
Minimal	YHRD Core Loci	DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385
PowerPlex Y	Promega PowerPlex® Y	DYS391 DYS389I DYS439 DYS389II DYS438 DYS437 DYS19 DYS392 DYS393 DYS390 DYS385
Yfiler	Applied Biosystems AmpFLSTR® Yfiler®	DYS456 DYS389I DYS390 DYS389II DYS458 DYS19 DYS385 DYS393 DYS391 DYS439 DYS635 DYS392 YGATAH4 DYS437 DYS438 DYS448
PowerPlex Y23	Promega PowerPlex® Y23	DYS576 DYS389I DYS448 DYS389II DYS19 DYS391 DYS481 DYS549 DYS533 DYS438 DYS437 DYS570 DYS635 DYS390 DYS439 DYS392 DYS643 DYS393 DYS458 DYS385 DYS456 YGATAH4
Yfiler Plus	Applied Biosystems AmpFLSTR® Yfiler® Plus	DYS576 DYS389I DYS635 DYS389II DYS627 DYS460 DYS458 DYS19 YGATAH4 DYS448 DYS391 DYS456 DYS390 DYS438 DYS392 DYS518 DYS570 DYS437 DYS385 DYS449 DYS393 DYS439 DYS481 DYF387S1 DYS533
Maximal	YHRD Max Loci	DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385 DYS438 DYS439 DYS437 DYS448 DYS456 DYS458 DYS635 YGATAH4 DYS481 DYS533 DYS549 DYS570 DYS576 DYS643 DYF387S1 DYS449 DYS460 DYS518 DYS627

Figure 2: Available marker sets (locus information)

2 Current state of the database

By June 2019, when YHRD release 61 was launched, more than 285,000 haplotypes from 1,308 sampling locations in 135 countries were submitted by more than 450 institutes and laboratories worldwide. In geographic terms, about 47% of the YHRD samples are from Asia, 23% from Europe, 11% from Latin America, 14% from North America, 3% from Africa, 1% from Oceania and less than 1% from the Arctic (Figure 3). The YHRD continuously receives new data from submitters which will be validated and used to create updates about three to four times a year (Figures 4 + 5). Four insertion methods are available for selection: **New** (a completely new dataset), **Append** (a dataset which enlarges a previous submission), **Replace** (a dataset which is inserted to replace a previous dataset) and **Merge** (a dataset which adds new haplotypes to previous haplotypes and/or includes previous haplotypes retyped for additional loci). The selection of the appropriate insertion method, which is selected by authors and approved by curators, is crucial to guarantee that no haplotype is inserted unintentionally more than once. Since new releases replace the previous ones, the release number and date is an important part of the results report (Figure 6). The release notes are included in the footer of all printable documents.

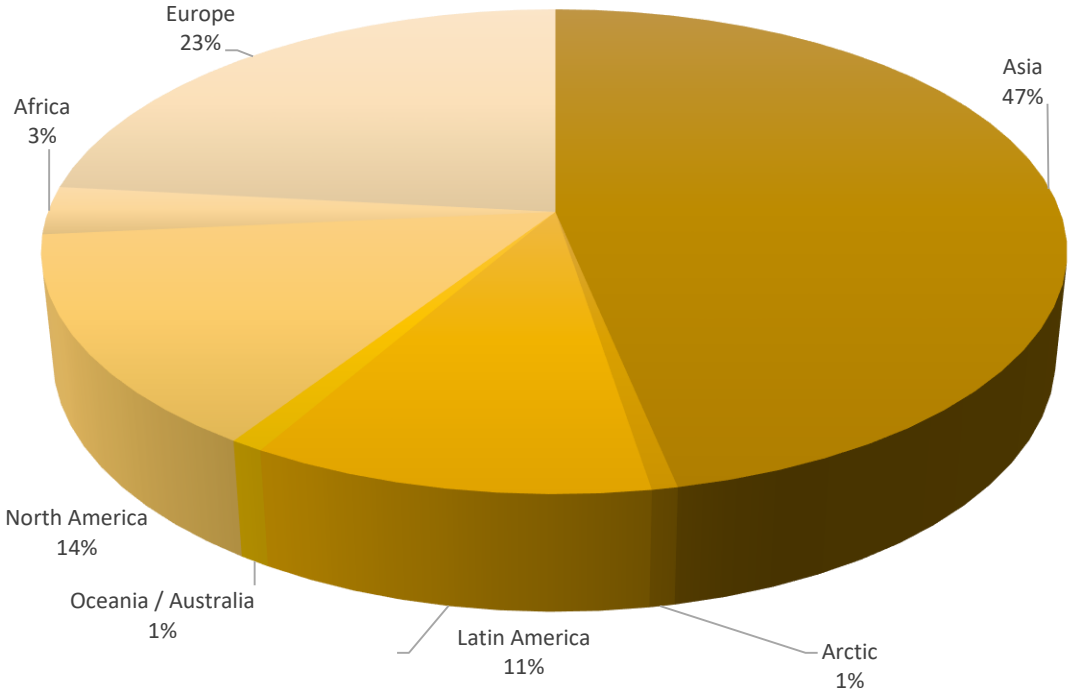


Figure 3: Continental distribution of haplotypes

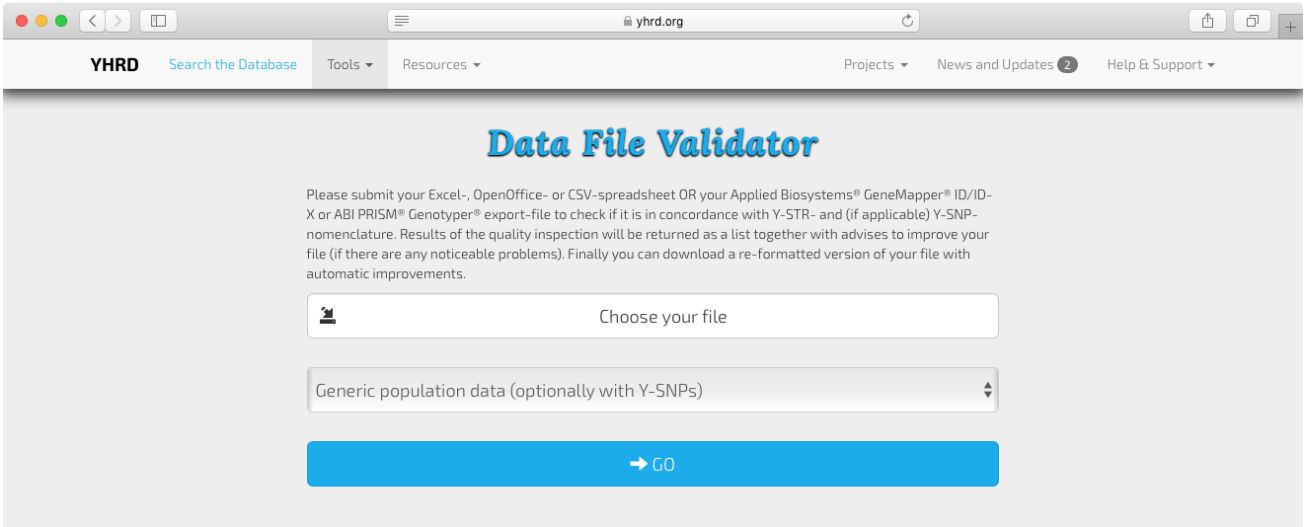


Figure 4: Data File Validator (file upload)

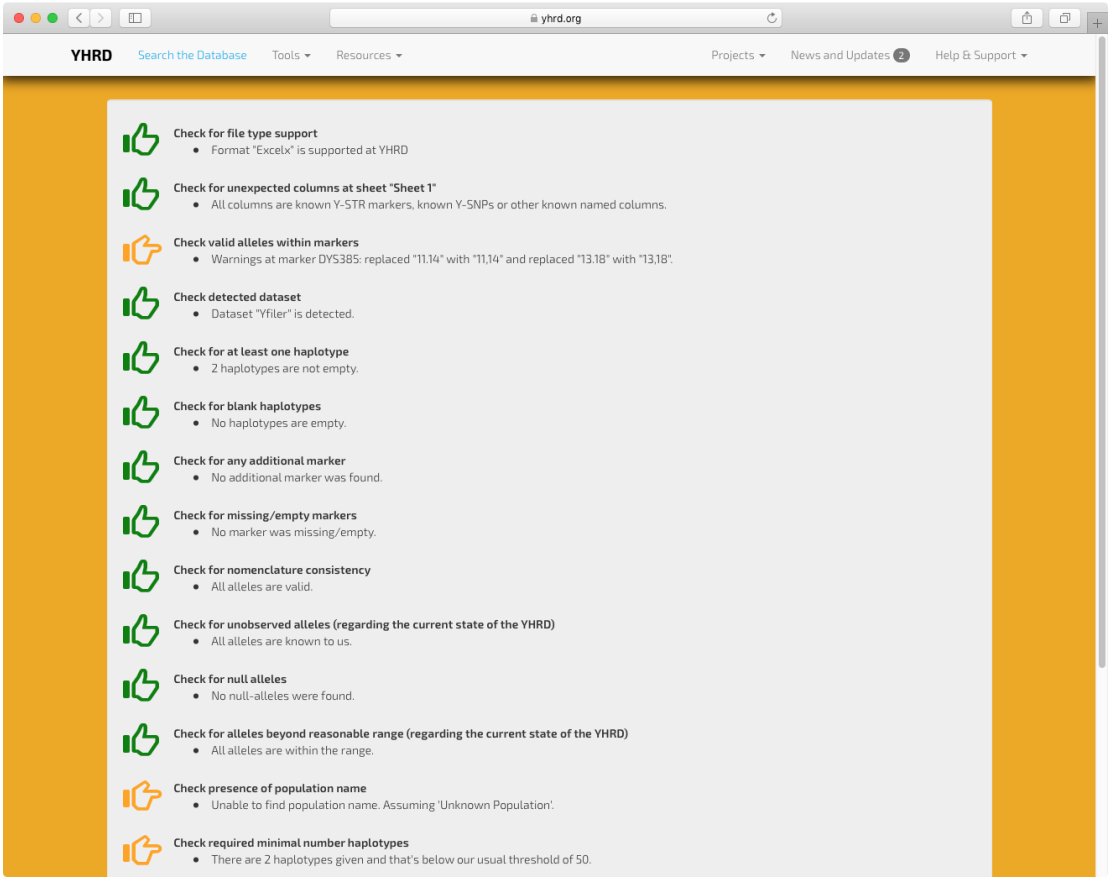


Figure 5: Data File Validator (results)

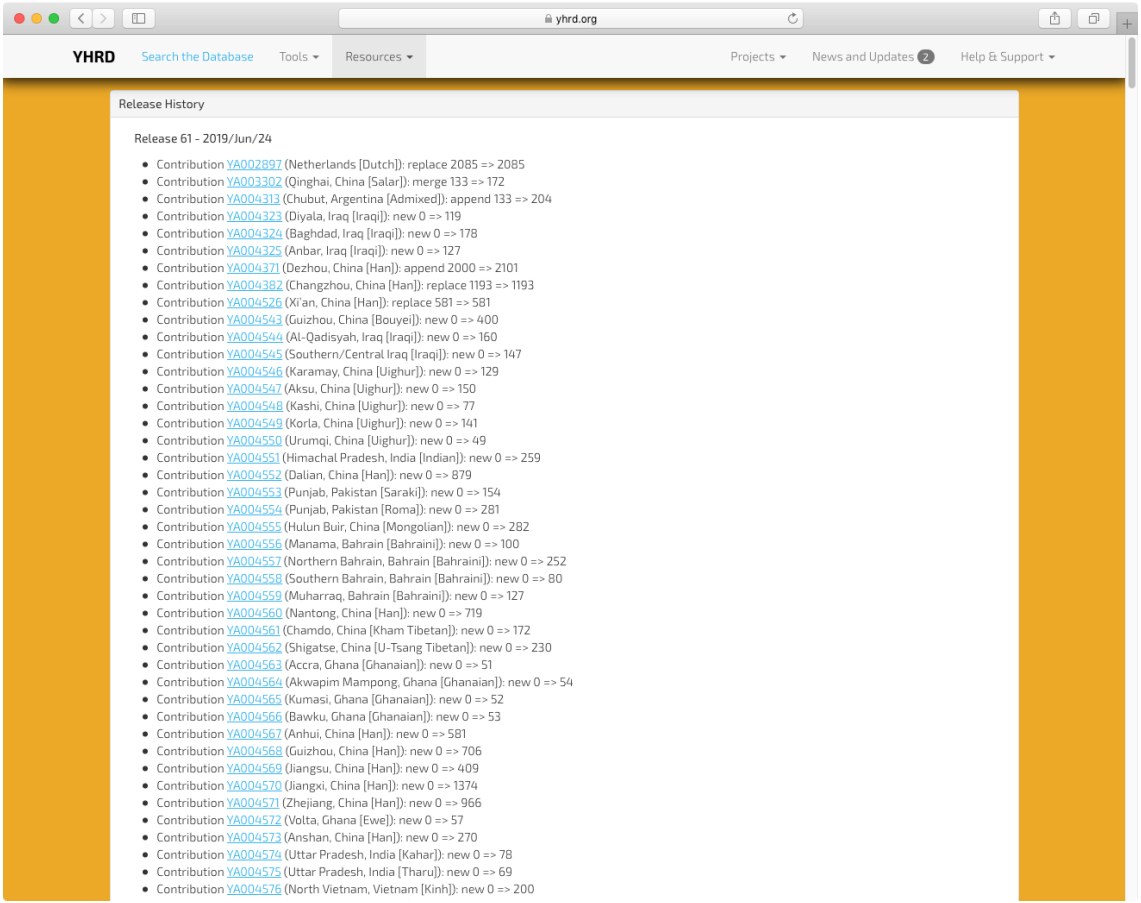


Figure 6: Release History

3 Navigation through the website

3.1 Search the database

The database can be searched either using manual input or by uploading files, including GeneMapper/GenoTyper export files (Figure 7).

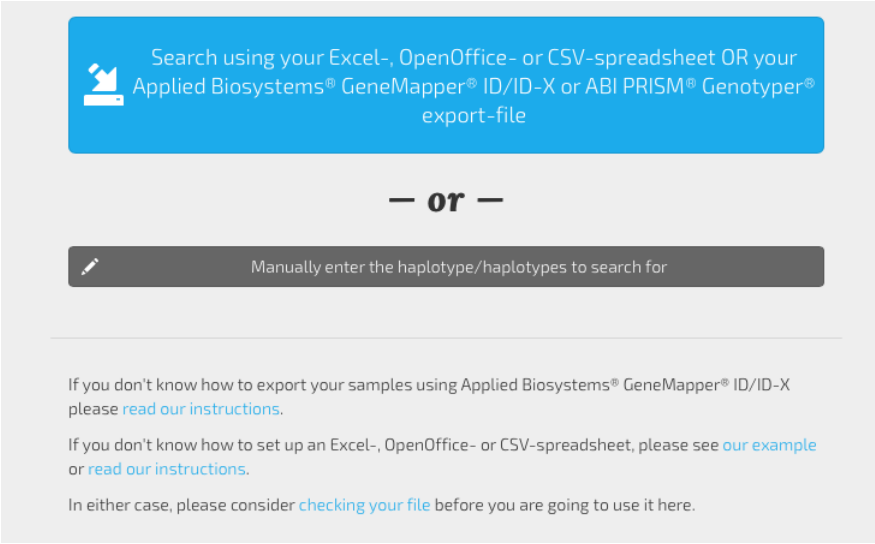


Figure 7: Search database

Without registration and limitations, the YHRD database can be searched for all single alleles, partial haplotypes and full haplotypes in different formats. Nomenclature for all loci follows the ISFG recommendations (Gusmão et al. 2006). The database supports the most frequently used haplotype formats which need to be selected from the "kit" bar at the top, i.e., **Minimal (9 loci)**, **PowerPlex® Y12 (12)**, **YFiler (17)**, **PowerPlex® Y23 (23)**, **YFiler Plus (27)** and **Maximal (29)** (Figure 8). For each of these six panels different-sized databases exist. Note that smaller-sized panels are included in the next larger panels; for example, the YFiler is part of the PPY23 and YFilerPlus. Therefore, a PPY23 analysis in a reference sample will contribute to the three databases with lesser loci. One exception: the two loci DYS549 and DYS643 are not part of the higher resolution panel YFilerPlus, but all loci of PPY23 and YFilerPlus are included in the maximal panel with 29 loci. It is possible to switch from a higher-resolution to a lower-resolution haplotype format, e.g., from PPY23 to YFiler thus enlarging the database in use (~63k to ~225k in release 61) (Figure 11).

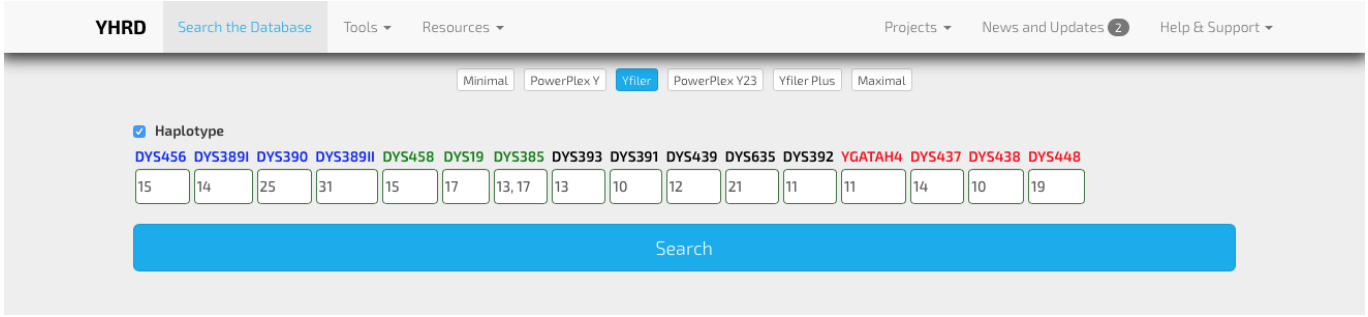


Figure 8: Search database (default panel selection is "YFiler")

At each position of the mask, the respective allele must be entered manually. Alternatively, GeneMapper export files or haplotype lists (Excel, OpenOffice, CSV etc.) can be uploaded for search. The file needs to be prepared according to the example file found at the bottom of the search page. The file is checked using the validator software and issues, e.g., an invalid file format, are highlighted. By using the function "search using the export file" all haplotypes in the file will be displayed as a list. Individual haplotypes need to be selected by clicking the checkbox for search. To avoid clerical errors, wrongly named alleles turn red. For loci which are mandatorily multi-copied (DYS385, DYF387S1) at least two alleles need to be entered. For Null alleles (caused by mutation, not by drop-out due to insufficient DNA quality) a “0” must be entered as a placeholder. Boxes for alleles which dropped or are regarded uncertain due to low DNA amount and/or degradation need to be empty. All alleles need to be called according to their repeat numbers; placeholders for off-ladder alleles like “99” are not allowed. Duplications or higher-order multi-copying events are possible for all Y chromosome sequences. If the sample is not a mixture, all true alleles need to be entered separated by a decimal point. The number of duplications, triplications, quadruplications and other rare intermediate, off-ladder or Null alleles can be viewed at the “Locus Information” page. Note that the allele spectra per locus are based on the full validated YHRD dataset and therefore often include more variants than commercial allelic ladders (Figure 9). You may choose the option “logarithmic axis” and “treat multi-copy alleles (e.g., duplications) as one observation” to get the most information on the variants spectrum.

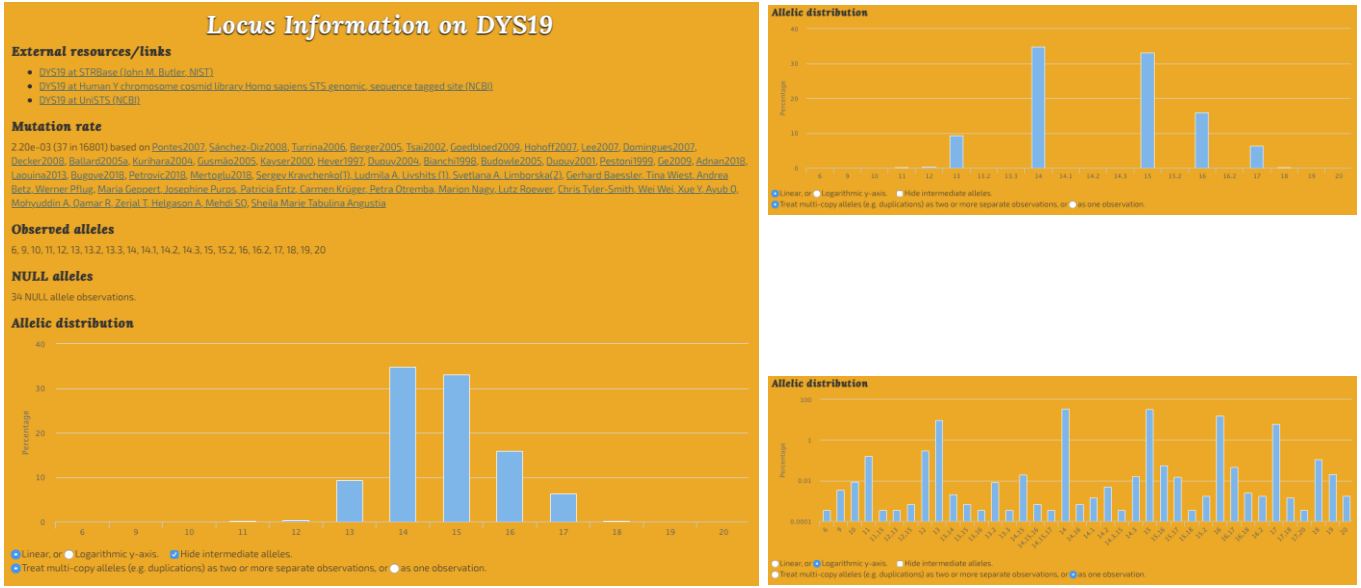


Figure 9: Locus information (DYS 19, multiple graphical options used)

After clicking on the “Search” button, a printable report for the respective sample is generated (Figure 10). It includes two main types of frequency estimates. First is the **observed value** i.e., the number of observations in the worldwide database for the selected panel accompanied by a confidence interval (CI) calculated according to Clopper and Pearson (Clopper and Pearson 1934; see Glossary: Confidence Intervals for more information on different types of CI). Second Is the **expected values** calculated by (1) the Discrete Laplace Method, (2) augmented counting (n+1/N+1) and (3) the Kappa method (see Glossary: Frequency estimation methods for more details on each calculation method).

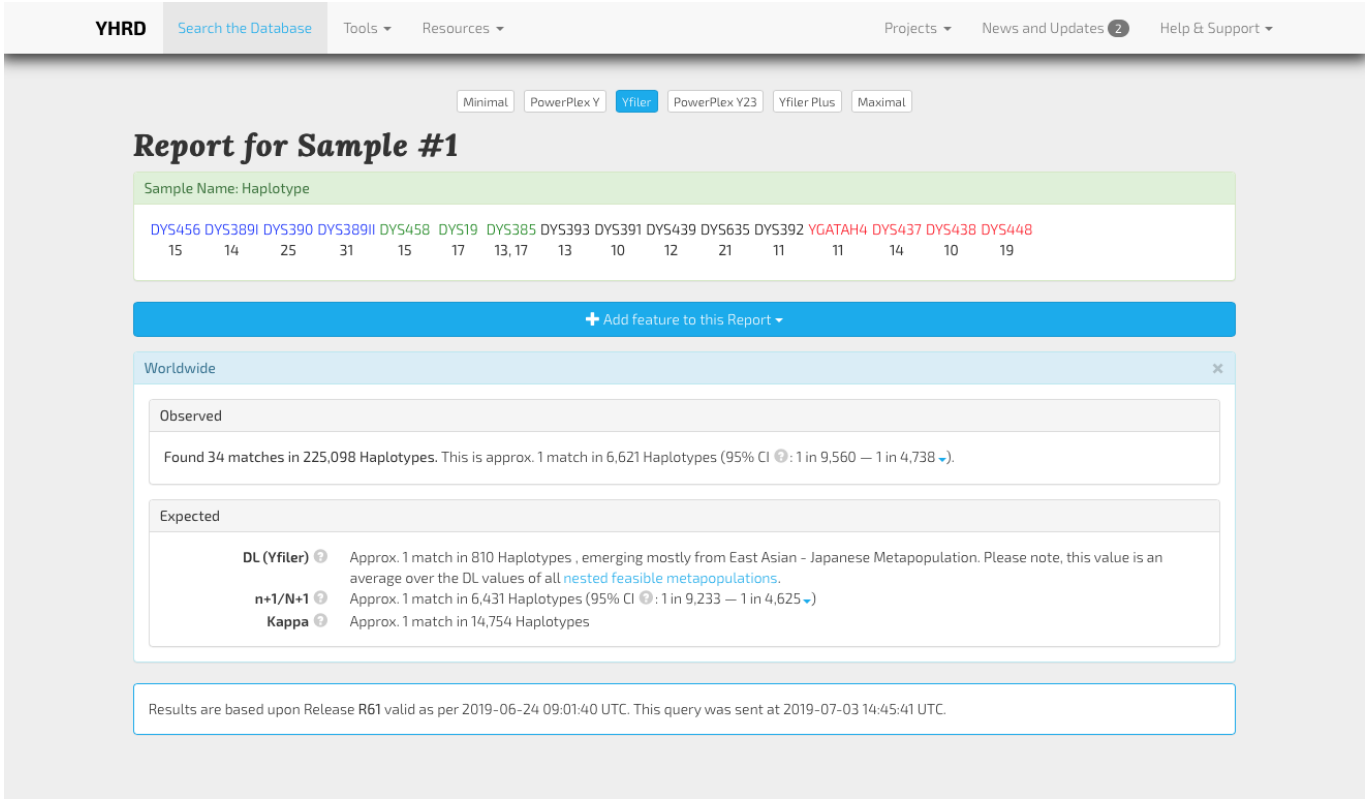


Figure 10: Search result

To switch between different panels click the panel bar at the top (Figure 11).

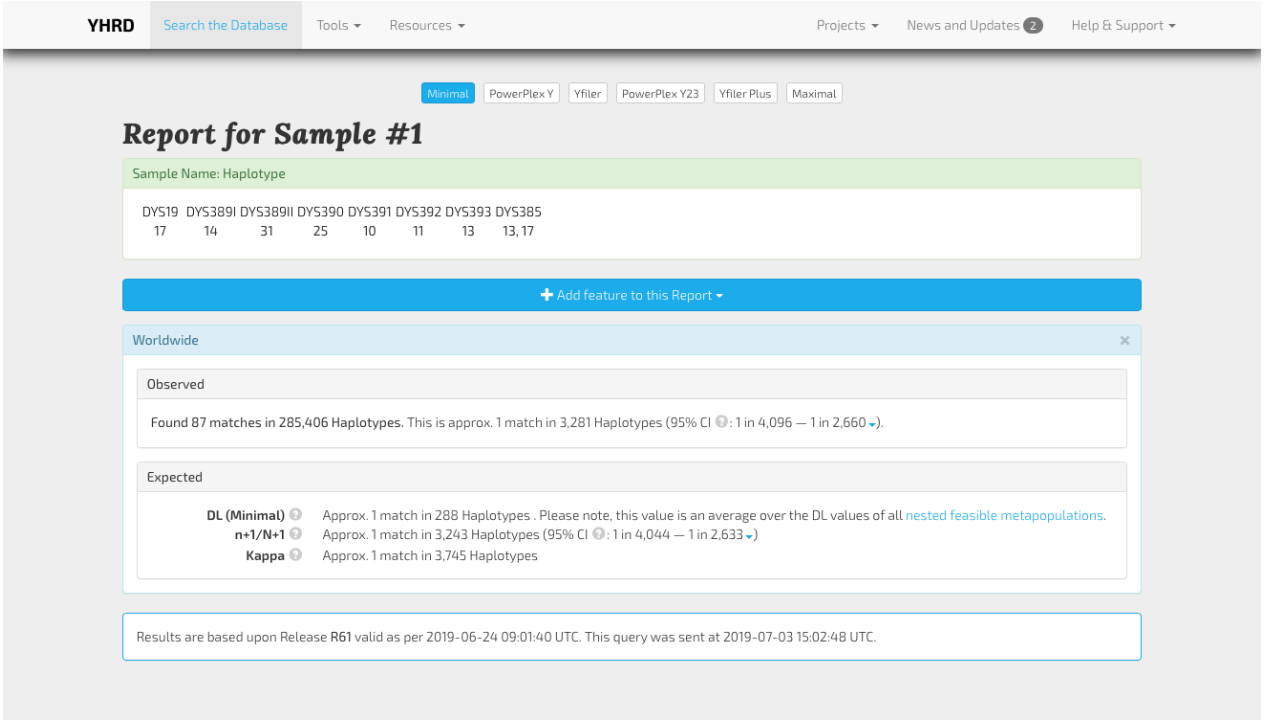


Figure 11: Search result with “kit” switched to “Minimal Haplotype”

Using the button “+ Add feature to this report”, the search result can be further adapted to the needs of the user (Figure 12). The first button **Metapopulation** allows selecting a match statistic within metapopulations (Figure 13), the second button **National** for political entities (countries) (Figure 14), and the third button **National databases with subpopulations** for searches within countries with predefined subpopulations. The next three buttons **Ancestry information (minimal haplotype)**, **Ancestry information (YFiler)** and **Ancestry Information 1-step Neighbors** (haplotype counts with +/- one step-allele per locus) provide access to relevant information on the biogeographical ancestry of the searched haplotype (Figure 15). At the bottom of the "Result page" you' will find a release note with the date and number of the current version of the database, and the time of the query.

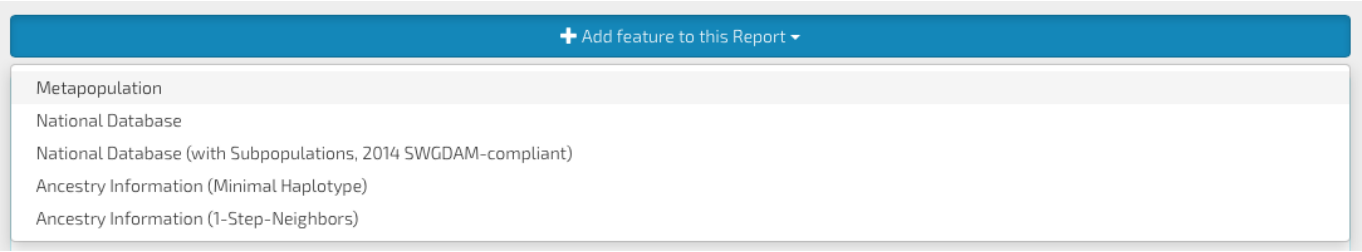


Figure 12: Search result, Add feature to this report

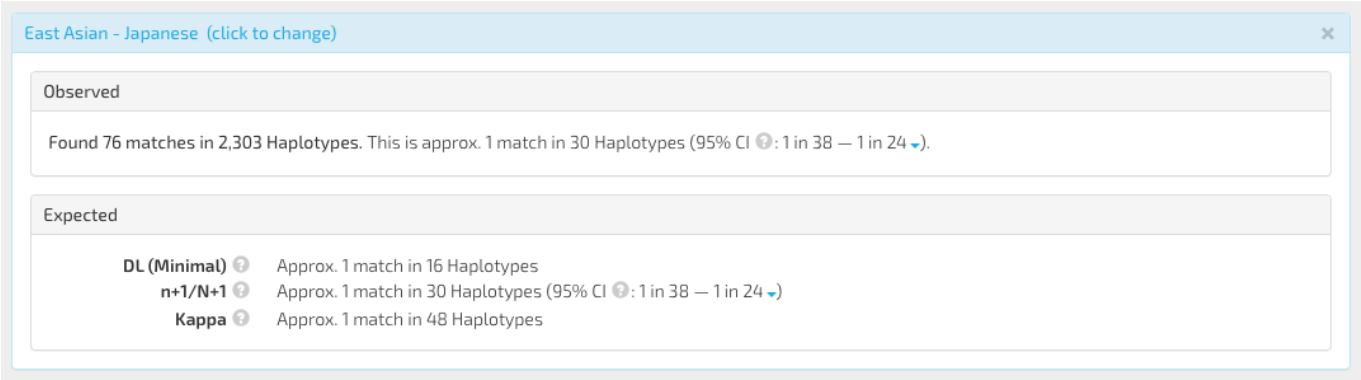


Figure 13: Search result, Add “metapopulation” feature to this report (*East Asian-Japanese*)

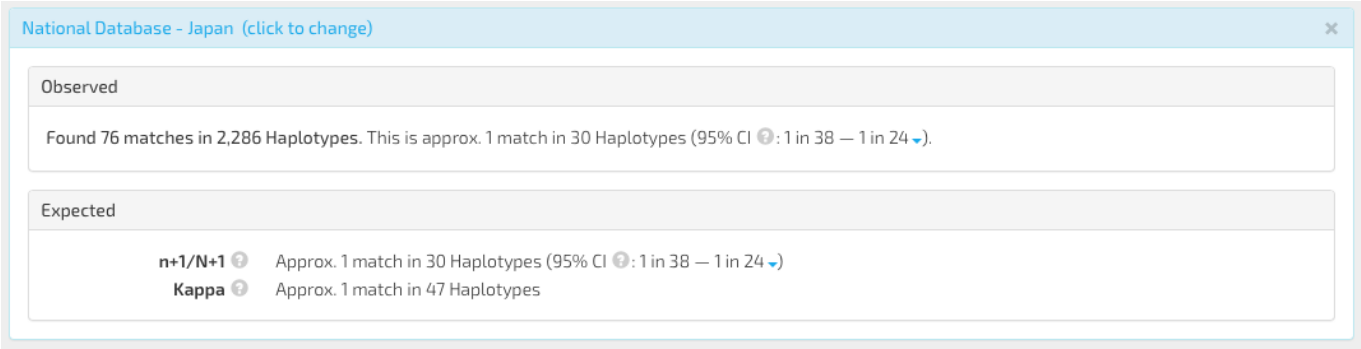


Figure 14: Search result, Add “National Database - Japan” feature to this report

The selected features will be added one after another to the general search report. Note that the computer IP is used to select the default metapopulation and national database, e.g. a user in Japan will see the “East Asian - Japanese metapopulation” and the national database “Japan” as default. The evaluation of the haplotype frequency using the DL method is restricted to the YFiler database which incorporates the 17 loci analyzed using higher-resolution panels. Haplotypes with intermediate, duplicated or missing alleles within the Yfiler panel cannot be evaluated using DL. It is adviseable to calculate the DL value in a pre-selected metapopulation, since the worldwide DL value is an average estimate over the DL values of all nested metapopulations.

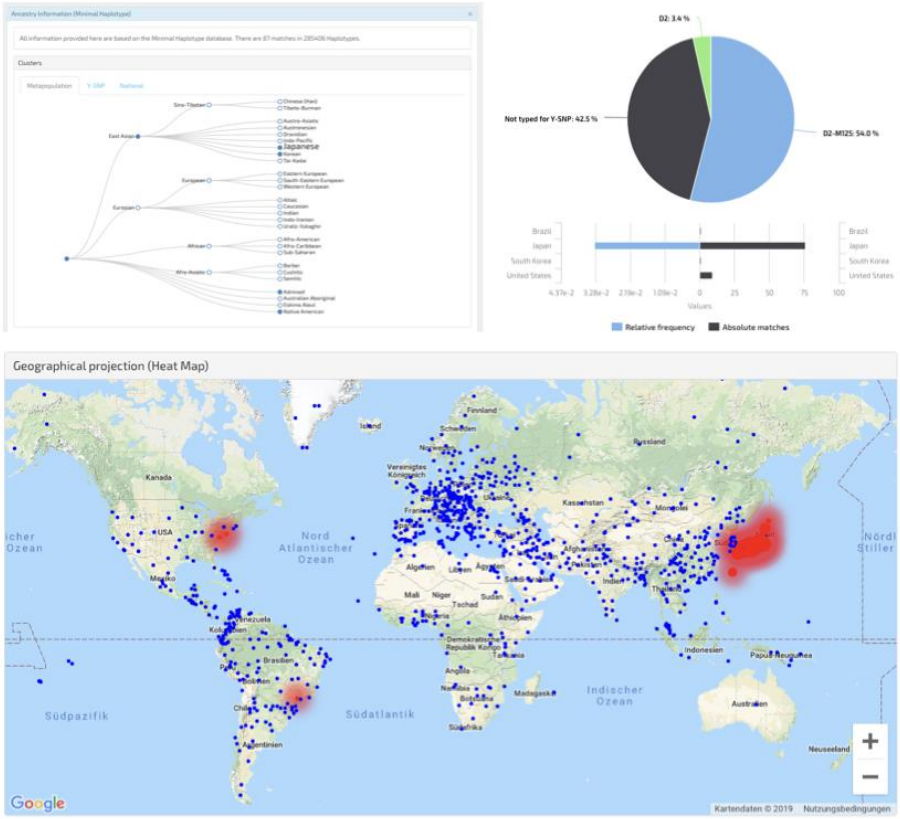


Figure 15: Add “ancestry information (Minimal Haplotype)” to this report (Metapopulation tree, Y-SNP, national frequencies and heat map)

If the feature ancestry information is selected, the database provides a movable and scalable world map with the relative proportion of haplotype matches shown in red/blue for each geoposition. The more matches per population occur, the larger is the red-colored symbol of the population sample. Clicking the population symbol provides information on the number of matches and the population size. Please note that, at a given geoposition, more than one population could have been sampled. Furthermore, the feature presents absolute and relative frequencies of the concerned haplotype in countries and the number of matches with SNP-typed chromosomes (haplogroups) collected in the YHRD (Figure 15).

3.2 Resources

3.2.1 Database details

Enter names of population samples, countries, contributors, accession number, etc. Choose from a list of found items for further information (Figures 16-19).

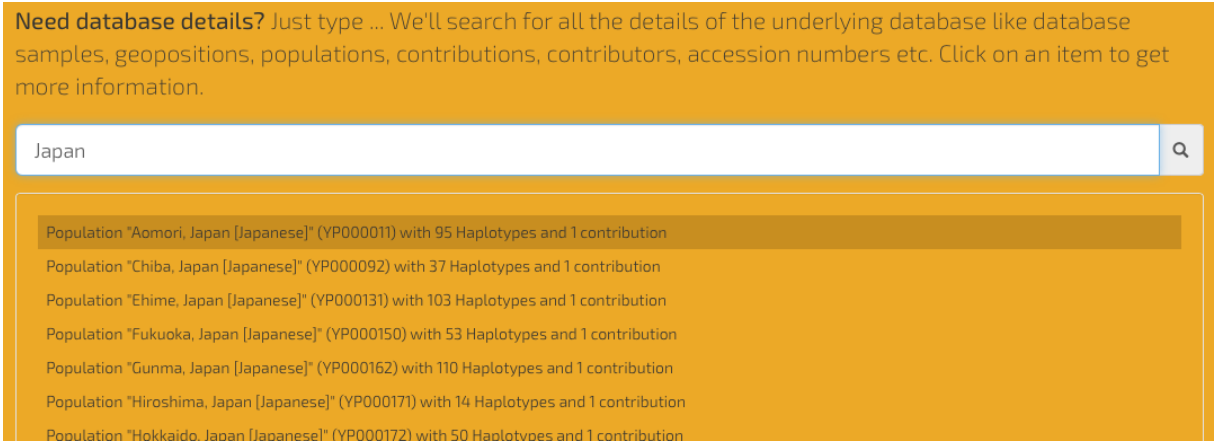


Figure 16: Detail results when typing “Japan”

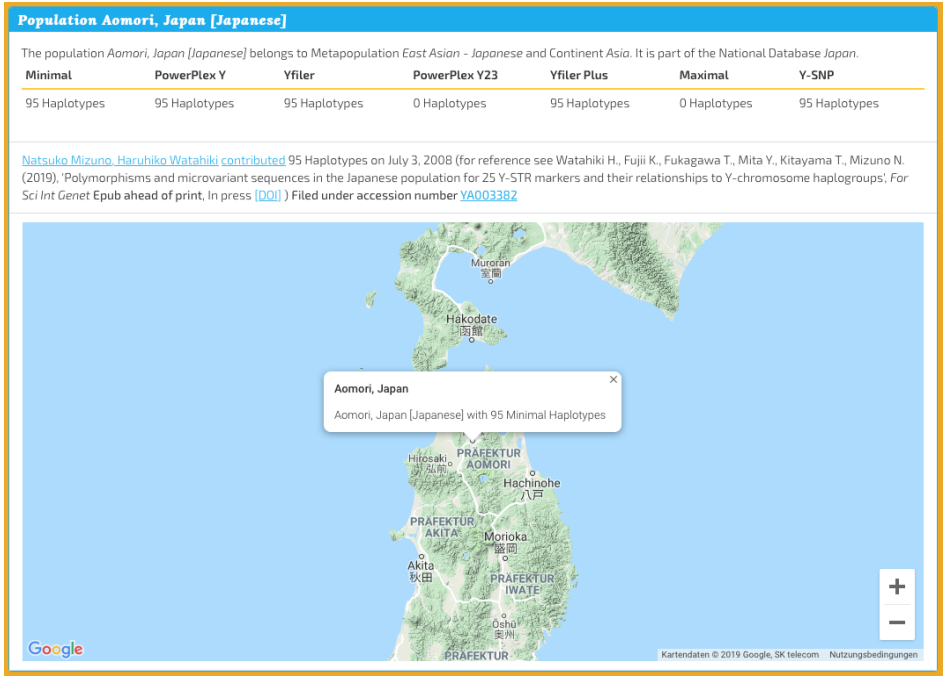
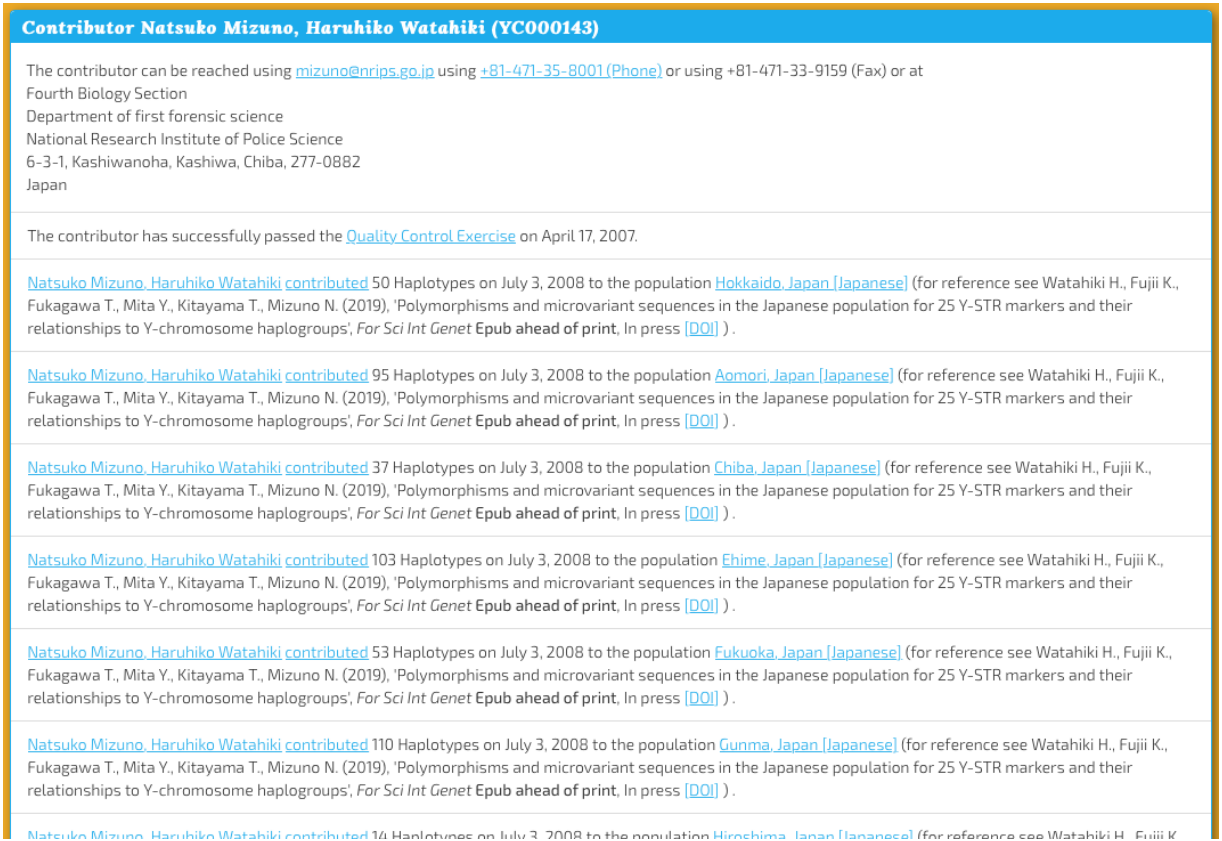


Figure 17: Details on the population sample “Aomori, Japan [Japanese]”



Contributor Natsuko Mizuno, Haruhiko Watahiki (YC000143)

The contributor can be reached using mizuno@nrips.go.jp using +81-471-35-8001 (Phone) or using +81-471-33-9159 (Fax) or at
Fourth Biology Section
Department of first forensic science
National Research Institute of Police Science
6-3-1, Kashiwanoha, Kashiwa, Chiba, 277-0882
Japan

The contributor has successfully passed the Quality Control Exercise on April 17, 2007.

Natsuko Mizuno, Haruhiko Watahiki contributed 50 Haplotypes on July 3, 2008 to the population Hokkaido, Japan [Japanese] (for reference see Watahiki H., Fujii K., Fukagawa T., Mita Y., Kitayama T., Mizuno N. (2019), 'Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups', *For Sci Int Genet* **Epub ahead of print**, In press [DOI]) .

Natsuko Mizuno, Haruhiko Watahiki contributed 95 Haplotypes on July 3, 2008 to the population Aomori, Japan [Japanese] (for reference see Watahiki H., Fujii K., Fukagawa T., Mita Y., Kitayama T., Mizuno N. (2019), 'Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups', *For Sci Int Genet* **Epub ahead of print**, In press [DOI]) .

Natsuko Mizuno, Haruhiko Watahiki contributed 37 Haplotypes on July 3, 2008 to the population Chiba, Japan [Japanese] (for reference see Watahiki H., Fujii K., Fukagawa T., Mita Y., Kitayama T., Mizuno N. (2019), 'Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups', *For Sci Int Genet* **Epub ahead of print**, In press [DOI]) .

Natsuko Mizuno, Haruhiko Watahiki contributed 103 Haplotypes on July 3, 2008 to the population Ehime, Japan [Japanese] (for reference see Watahiki H., Fujii K., Fukagawa T., Mita Y., Kitayama T., Mizuno N. (2019), 'Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups', *For Sci Int Genet* **Epub ahead of print**, In press [DOI]) .

Natsuko Mizuno, Haruhiko Watahiki contributed 53 Haplotypes on July 3, 2008 to the population Fukuoka, Japan [Japanese] (for reference see Watahiki H., Fujii K., Fukagawa T., Mita Y., Kitayama T., Mizuno N. (2019), 'Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups', *For Sci Int Genet* **Epub ahead of print**, In press [DOI]) .

Natsuko Mizuno, Haruhiko Watahiki contributed 110 Haplotypes on July 3, 2008 to the population Gunma, Japan [Japanese] (for reference see Watahiki H., Fujii K., Fukagawa T., Mita Y., Kitayama T., Mizuno N. (2019), 'Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups', *For Sci Int Genet* **Epub ahead of print**, In press [DOI]) .

Natsuko Mizuno, Haruhiko Watahiki contributed 14 Haplotypes on July 3, 2008 to the population Hiroshima, Japan [Japanese] (for reference see Watahiki H., Fujii K.

Figure 18: Details on the Contributor YC000143

Contribution with the Accession Number YA003381

Natsuko Mizuno, Haruhiko Watahiki contributed 50 Haplotypes on July 3, 2008 to the population Hokkaido, Japan [Japanese] (for reference see Watahiki H., Fujii K., Fukagawa T., Mita Y., Kitayama T., Mizuno N. (2019), 'Polymorphisms and microvariant sequences in the Japanese population for 25 Y-STR markers and their relationships to Y-chromosome haplogroups', *For Sci Int Genet* **Epub ahead of print**, In press [DOI]) .

Submission accession number	State	Accepted at	Release	Insertion method	Minimal Haplotypes	PowerPlex Y Haplotypes	Yfiler Haplotypes	PowerPlex Y23 Haplotypes	Yfiler Plus Haplotypes	Maximal Haplotypes	Y-SNP Haplotypes
YA003381-2	active submission	Sat, 28 Jul 2018 12:43:08 +0200	58	merge	50	50	50	0	50	0	50
YA003381-1	former submission	Tue, 25 Oct 2016 10:05:44 +0200	51	new	50	50	50	0	0	0	50

Figure 19: Details on submissions building the population sample YA003381

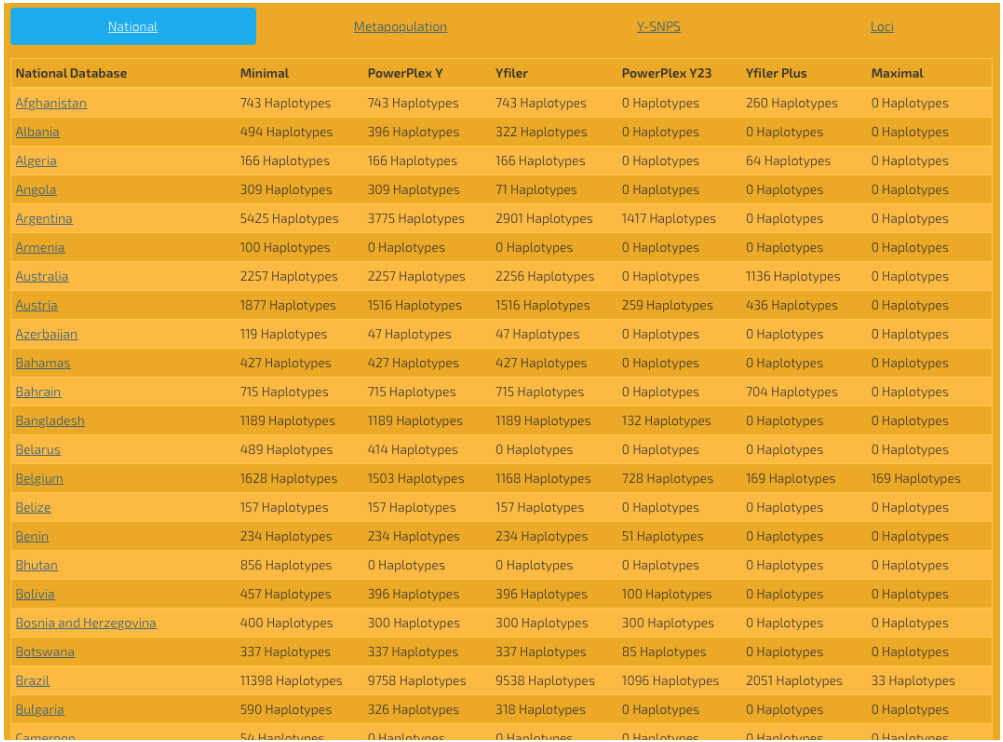
Note that each submitted population sample goes through a validation and receives a unique accession number. A further prerequisite is the laboratory accreditation certificate.

3.2.2 Database statistics

Dataset	Y-STR loci	Number of haplotypes	Number of population samples	Number of national databases	Number of metapopulations
Minimal	DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385	285,406	1308	135	33
PowerPlex Y	DYS391 DYS389I DYS439 DYS389II DYS438 DYS437 DYS19 DYS392 DYS393 DYS390 DYS385	244,777	1084	126	32
Yfiler	DYS456 DYS389I DYS390 DYS389II DYS458 DYS19 DYS385 DYS393 DYS391 DYS439 DYS635 DYS392 YGATAH4 DYS437 DYS438 DYS448	225,098	999	119	32
PowerPlex Y23	DYS576 DYS389I DYS448 DYS389II DYS19 DYS391 DYS481 DYS549 DYS533 DYS438 DYS437 DYS570 DYS635 DYS390 DYS439 DYS392 DYS643 DYS393 DYS458 DYS385 DYS456 YGATAH4	62,737	338	67	28
Yfiler Plus	DYS576 DYS389I DYS635 DYS389II DYS627 DYS460 DYS458 DYS19 YGATAH4 DYS448 DYS391 DYS456 DYS390 DYS438 DYS392 DYS518 DYS570 DYS437 DYS385 DYS449 DYS393 DYS439 DYS481 DYF387S1 DYS533	56,114	250	52	30
Maximal	DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385 DYS438 DYS439 DYS437 DYS448 DYS456 DYS458 DYS635 YGATAH4 DYS481 DYS533 DYS549 DYS570 DYS576 DYS643 DYF387S1 DYS449 DYS460 DYS518 DYS627	15,395	57	15	21

Figure 20: Database statistics

3.2.3 Database composition



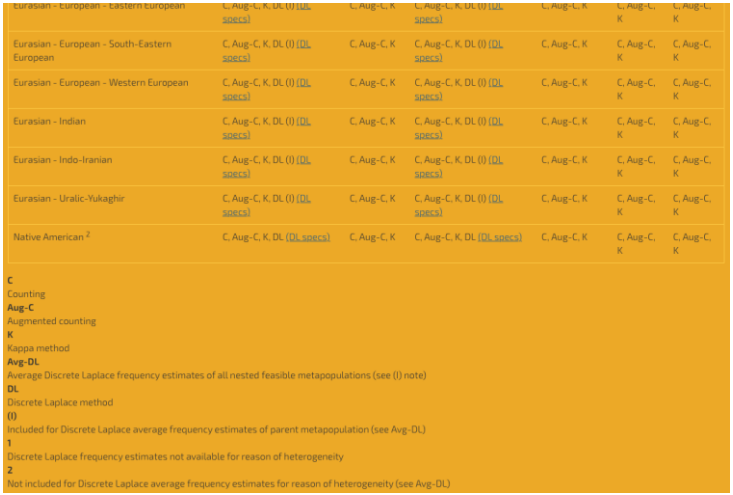
National | Metapopulation | Y-SNPS | Loci

National Database	Minimal	PowerPlex Y	Yfiler	PowerPlex Y23	Yfiler Plus	Maximal
Afghanistan	743 Haplotypes	743 Haplotypes	743 Haplotypes	0 Haplotypes	260 Haplotypes	0 Haplotypes
Albania	494 Haplotypes	396 Haplotypes	322 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Algeria	166 Haplotypes	166 Haplotypes	166 Haplotypes	0 Haplotypes	64 Haplotypes	0 Haplotypes
Angola	309 Haplotypes	309 Haplotypes	71 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Argentina	5425 Haplotypes	3775 Haplotypes	2901 Haplotypes	1417 Haplotypes	0 Haplotypes	0 Haplotypes
Armenia	100 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Australia	2257 Haplotypes	2257 Haplotypes	2256 Haplotypes	0 Haplotypes	1136 Haplotypes	0 Haplotypes
Austria	1877 Haplotypes	1516 Haplotypes	1516 Haplotypes	259 Haplotypes	436 Haplotypes	0 Haplotypes
Azerbaijan	119 Haplotypes	47 Haplotypes	47 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Bahamas	427 Haplotypes	427 Haplotypes	427 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Bahrain	715 Haplotypes	715 Haplotypes	715 Haplotypes	0 Haplotypes	704 Haplotypes	0 Haplotypes
Bangladesh	1189 Haplotypes	1189 Haplotypes	1189 Haplotypes	132 Haplotypes	0 Haplotypes	0 Haplotypes
Belarus	489 Haplotypes	414 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Belgium	1628 Haplotypes	1503 Haplotypes	1168 Haplotypes	728 Haplotypes	169 Haplotypes	169 Haplotypes
Belize	157 Haplotypes	157 Haplotypes	157 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Benin	234 Haplotypes	234 Haplotypes	234 Haplotypes	51 Haplotypes	0 Haplotypes	0 Haplotypes
Bhutan	856 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Bolivia	457 Haplotypes	396 Haplotypes	396 Haplotypes	100 Haplotypes	0 Haplotypes	0 Haplotypes
Bosnia and Herzegovina	400 Haplotypes	300 Haplotypes	300 Haplotypes	300 Haplotypes	0 Haplotypes	0 Haplotypes
Botswana	337 Haplotypes	337 Haplotypes	337 Haplotypes	85 Haplotypes	0 Haplotypes	0 Haplotypes
Brazil	11398 Haplotypes	9758 Haplotypes	9538 Haplotypes	1096 Haplotypes	2051 Haplotypes	33 Haplotypes
Bulgaria	590 Haplotypes	326 Haplotypes	318 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
Cameroon	54 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes

Figure 21: Database composition

3.2.4 Calculation methods

Note that the applicability of calculation methods depends on the size and structure of sampled metapopulations. For example, small-sized or very heterogeneous metapopulations are not suitable for DL calculations. For suitable metapopulations, the DL specs provide information on the central haplotypes and the methods to identify these (Figure 22a). The DL values are recalculated each time the YHRD is regularly updated. Further information on the DL estimation is given for each available dataset and metapopulation (Figure 22b).



Eurasian - European - Eastern European	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K	C, Aug-C, K
Eurasian - European - South-Eastern European	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K	C, Aug-C, K
Eurasian - European - Western European	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K	C, Aug-C, K
Eurasian - Indian	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K	C, Aug-C, K
Eurasian - Indo-Iranian	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K	C, Aug-C, K
Eurasian - Uralic-Yukaghir	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K, DL (I) (DL specs)	C, Aug-C, K	C, Aug-C, K	C, Aug-C, K
Native American [2]	C, Aug-C, K, DL (DL specs)	C, Aug-C, K	C, Aug-C, K, DL (DL specs)	C, Aug-C, K	C, Aug-C, K	C, Aug-C, K

C
Counting
Aug-C
Augmented counting
K
Kappa method
Avg-DL
Average Discrete Laplace frequency estimates of all nested feasible metapopulations (see (I) note)
DL
Discrete Laplace method
(I)
Included for Discrete Laplace average frequency estimates of parent metapopulation (see Avg-DL)
1
Discrete Laplace frequency estimates not available for reason of heterogeneity
2
Not included for Discrete Laplace average frequency estimates for reason of heterogeneity (see Avg-DL)

Figure 22a: Calculation methods

Discrete-Laplace frequency estimation (Yfiler)

Release	R61
Loci included	DYS19, DYS389I, DYS389II*, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, YGATAH4
Central haplotypes	25
EM converged	true
EM iterations	88
GLM method	internal_coef
Init method	pam
Per locus details	DL centers distribution (download PDF)

Table of haplotype centers (click to show/hide)

	DYS19	DYS389I	DYS389II*	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	YGATAH4
1	15	14	16	23	10	11	15	14	10	12	21	15	16	21	11
2	15	14	16	22	10	13	13	14	13	12	18	15	17	20	12
3	14	11	16	23	10	14	12	15	11	12	20	15	18	21	11
4	16	12	17	24	10	13	12	14	10	12	19	15	18	23	12
5	15	14	16	22	10	13	13	14	13	12	18	15	19	20	12
6	16	14	17	25	10	11	13	14	10	12	19	15	15	21	11

Figure 22b: Information on Yfiler DL estimation (Japanese MP)

3.2.5 Metapopulation information

The haplotype frequencies can be sensibly reported in groups of spatially distributed populations (metapopulations) sharing a common ancestry, and thus a similar pool of deep-rooting lineages (Figure 23a). To describe the hierarchy of metapopulations, YHRD uses a knowledge based terminology, which incorporates linguistic, geographic resources and genetic parameters (genetic distance measures). YHRD provides information on each metapopulation including the assigned populations, a map, its common haplotypes (Figure 23b) and applicable calculation methods.

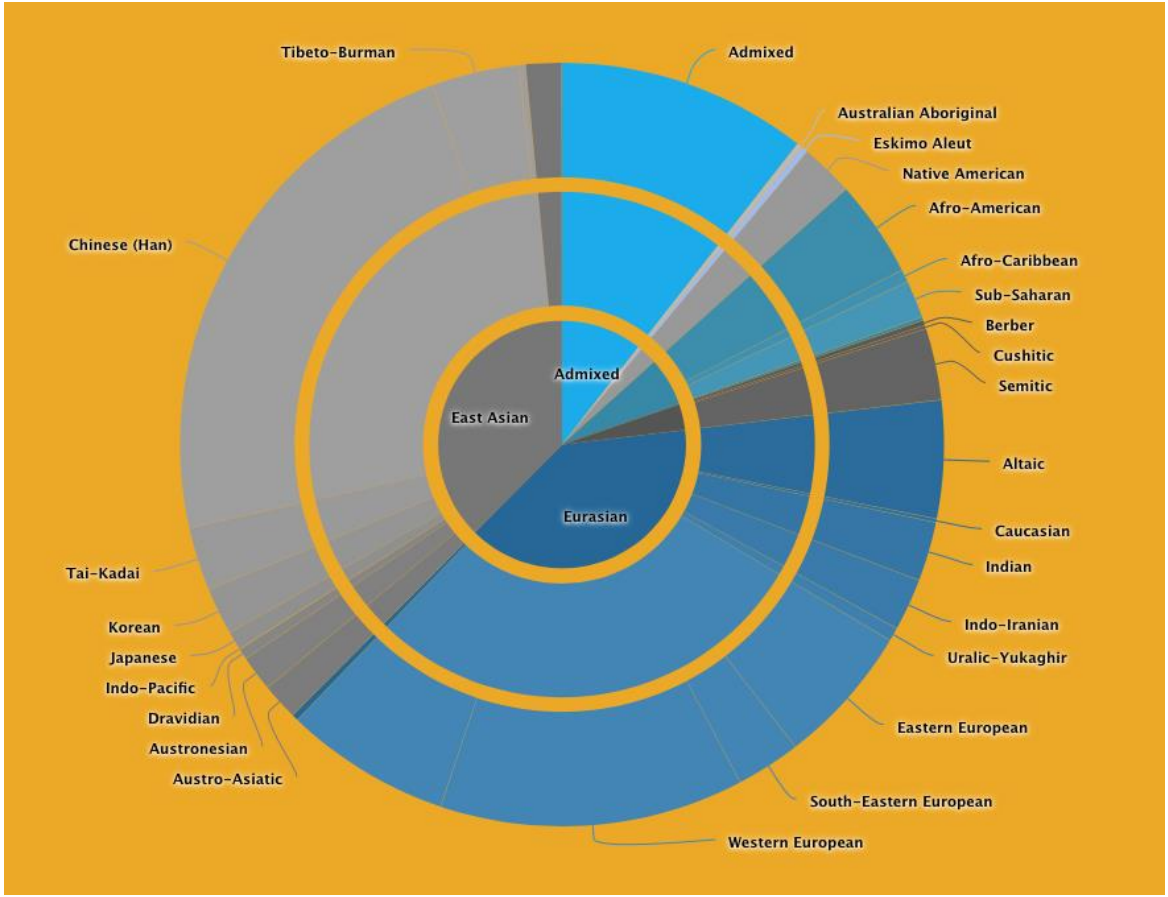


Figure 23a: Metapopulation structure

The Japanese Metapopulation (East Asian - Japanese) consists of 2303 haplotypes (at least minimal) and 26 unique population samples.

Most common haplotypes

Minimal | PowerPlex Y | Yfiler | PowerPlex Y23 | Yfiler Plus | Maximal

Count	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439	DYS635	DYS392	YGATAH4	DYS437	DYS438	DYS448
32	15	14	25	31	15	17	13,17	13	10	12	21	11	11	14	10	19
13	15	14	22	30	17	15	10,20	13	10	12	20	13	12	14	13	18
12	16	13	25	30	15	15	14,17	13	10	12	21	11	12	14	10	17
11	15	14	25	31	15	16	13,17	13	10	12	21	11	11	14	10	19
8	15	14	22	30	17	15	10,19	13	10	12	20	13	12	14	13	18
8	15	14	25	31	15	17	14,17	13	10	12	21	11	11	14	10	19
7	15	14	22	30	18	15	10,19	13	10	12	20	13	12	14	13	18

Figure 23b: Summary of the Japanese Metapopulation

3.2.6 National Databases

Each national database in the YHRD comprises all individuals sampled in a particular country regardless of the ancestry of the individuals (Figure 24). Some national databases are further structured into sub-populations (e.g., United States).

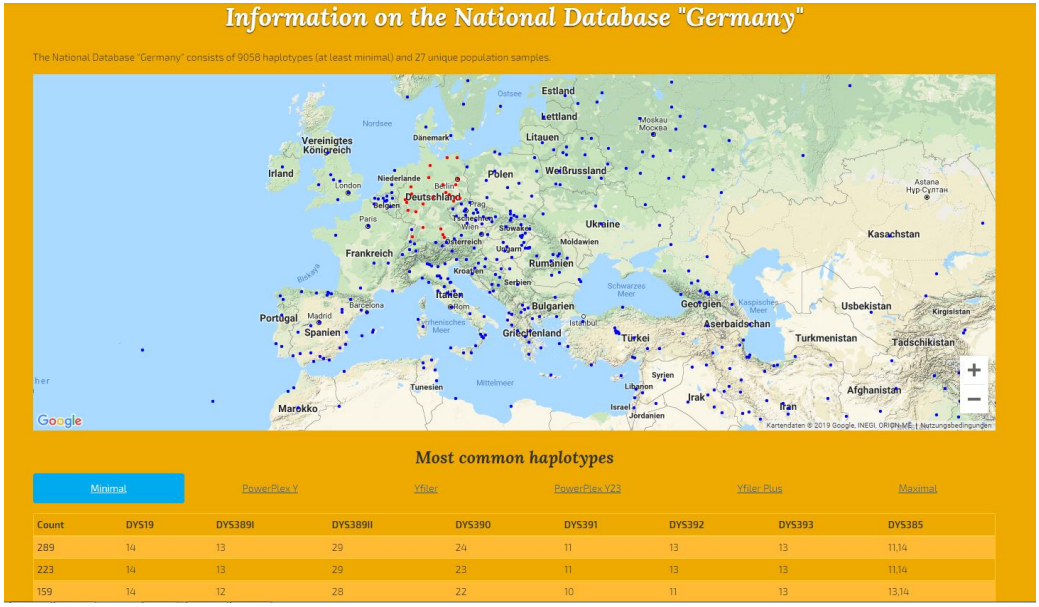


Figure 24: National Database Germany

3.2.7 Y-SNP Information

YHRD uses the resource *Phylotree Y* (Van Oven et al. 2014) to assign submitted SNP-analyzed haplotypes to haplogroups (Figure 25). A click on the respective haplogroup in the tree provides access to a map depicting the frequency of the selected haplogroup and to a table with the number of haplotypes typed in the current release for that haplogroup (Figure 26).

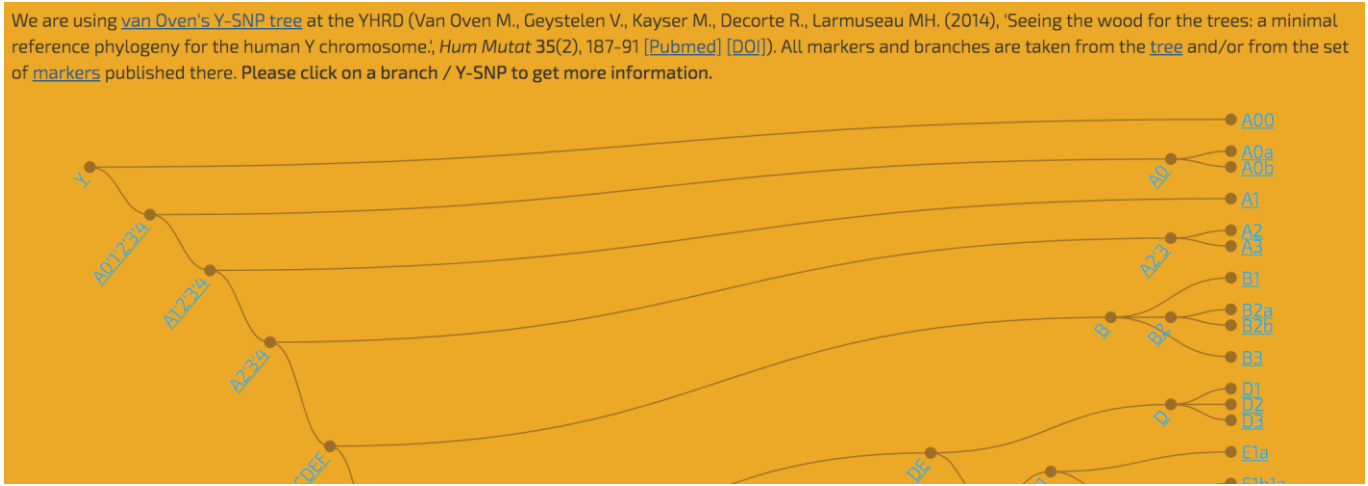
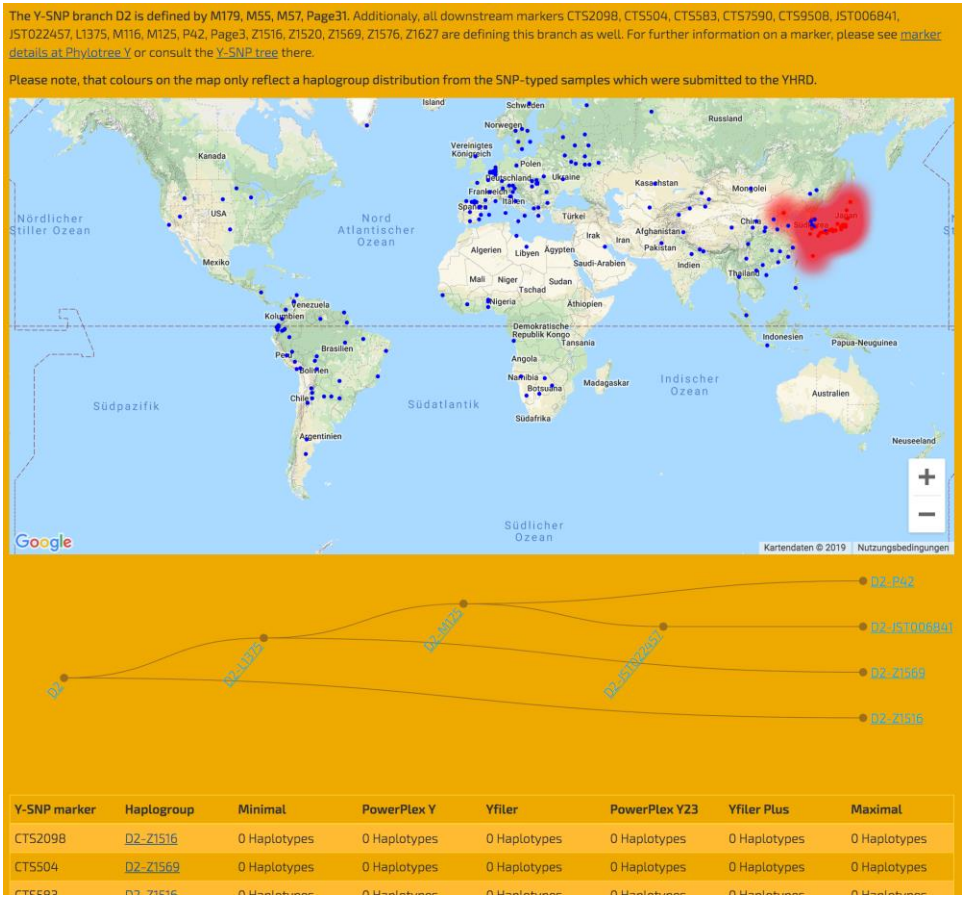


Figure 25: Y-SNP Information

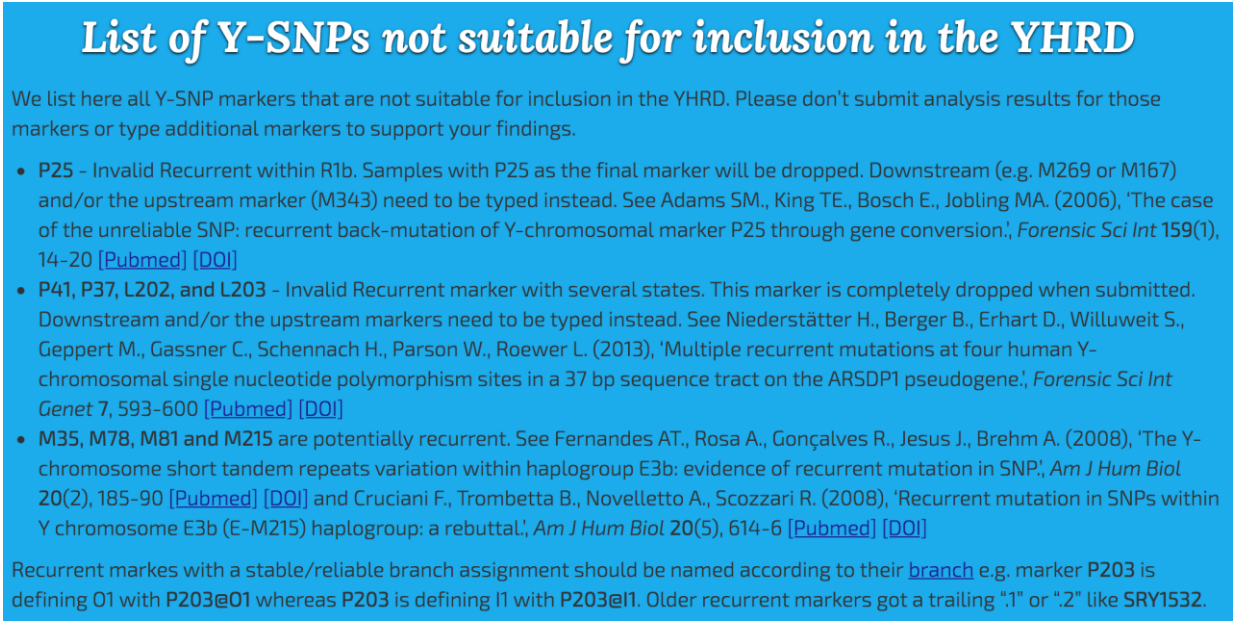


Y-SNP marker	Haplogroup	Minimal	PowerPlex Y	Yfiler	PowerPlex Y23	Yfiler Plus	Maximal
CTS2098	D2-Z1516	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes
CTS504	D2-Z1569	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes	0 Haplotypes

Figure 26: Description of Y-SNP branch D2

3.2.8 List of Invalid Y-SNPs

Invalid Y-SNPs which are not accepted by the YHRD for various reasons, namely recurrence, are listed on a separate page (Figure 27).



List of Y-SNPs not suitable for inclusion in the YHRD

We list here all Y-SNP markers that are not suitable for inclusion in the YHRD. Please don't submit analysis results for those markers or type additional markers to support your findings.

- **P25** - Invalid Recurrent within R1b. Samples with P25 as the final marker will be dropped. Downstream (e.g. M269 or M167) and/or the upstream marker (M343) need to be typed instead. See Adams SM., King TE., Bosch E., Jobling MA. (2006), 'The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion.', *Forensic Sci Int* **159**(1), 14-20 [Pubmed] [DOI]
- **P41, P37, L202, and L203** - Invalid Recurrent marker with several states. This marker is completely dropped when submitted. Downstream and/or the upstream markers need to be typed instead. See Niederstätter H., Berger B., Erhart D., Willuweit S., Geppert M., Gassner C., Schennach H., Parson W., Roewer L. (2013), 'Multiple recurrent mutations at four human Y-chromosomal single nucleotide polymorphism sites in a 37 bp sequence tract on the ARSDP1 pseudogene.', *Forensic Sci Int Genet* **7**, 593-600 [Pubmed] [DOI]
- **M35, M78, M81 and M215** are potentially recurrent. See Fernandes AT., Rosa A., Gonçalves R., Jesus J., Brehm A. (2008), 'The Y-chromosome short tandem repeats variation within haplogroup E3b: evidence of recurrent mutation in SNP.', *Am J Hum Biol* **20**(2), 185-90 [Pubmed] [DOI] and Cruciani F., Trombetta B., Novelletto A., Scozzari R. (2008), 'Recurrent mutation in SNPs within Y chromosome E3b (E-M215) haplogroup: a rebuttal.', *Am J Hum Biol* **20**(5), 614-6 [Pubmed] [DOI]

Recurrent markes with a stable/reliable branch assignment should be named according to their branch e.g. marker **P203** is defining 01 with **P203@01** whereas **P203** is defining I1 with **P203@I1**. Older recurrent markers got a trailing ".1" or ".2" like **SRY1532**.

Figure 27: List of Invalid Y-SNPs

3.2.9 List of publications and References

Most population studies submitted to the YHRD for validation and upload are published in parallel in forensic journals (Forensic Science International: Genetics, International Journal of Legal Medicine and other journals). The YHRD AccessionNumber. connects the article to the database (Figures 18, 19 + 28).

List of Publications and References

Abban2019
Kofi AE., Hakim HM., Khan HO., Ismail SA., Ghansah A., David AA., Mat NFC., Chambers GK., Edinur HA. (2019), 'Population data of 23 Y chromosome STR loci for the five major human subpopulations of Ghana', *Int J Legal Med* **Epub ahead of print**, In press [DOI]

Aboukhalid2010
Aboukhalid R., Bouabdellah M., Abbassi M., Bentayebi K., Elmzibri M., Squalli D., Amzazi S. (2010), 'Haplotype frequencies for 17 Y-STR loci (AmpFlSTRY-filer) in a Moroccan population sample.', *Forensic Sci Int Genet* **4**(3), e73-4 [Pubmed] [DOI]

Achakzai2012
Achakzai NM., Rahman Z., Shahzad MS., Daud S., Zar MS., Israr M., Husnain T., Willuweit S., Roewer L. (2012), 'Y-chromosomal STR analysis in the Pashtun population of Southern Afghanistan.', *Forensic Sci Int Genet* **6**(4), e103-5 [Pubmed] [DOI]

Adams2006
Adams SM., King TE., Bosch E., Jobling MA. (2006), 'The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion.', *Forensic Sci Int* **159**(1), 14-20 [Pubmed] [DOI]

Adnan2017
Adnan A., Rakha A., Noor A., Van M., Ralf A., Kayser M. (2017), 'Population data of 17 Y-STRs (Yfiler) from Punjabis and Kashmiris of Pakistan', *Int J Legal Med* **Epub ahead of Print**, In press [DOI]

Adnan2018
Adnan A., Rakha A., Lao O., Kayser M. (2018), 'Mutation analysis at 17 Y-STR loci (Yfiler) in father-son pairs of male pedigrees from Pakistan.', *For Sci Int Genet* **36**, e17-e18 [DOI]

Adnan2019
Adnan A., Rakha A., Kasim K., Noor A., Nazir S., Hadi S., Pang H. (2019), 'Genetic characterization of Y-chromosomal STRs in Hazara ethnic group of Pakistan and confirmation of DYS448 null allele', *Int J Legal Med* **133**(3), 789-93 [DOI]

Alakoc2010

Figure 28: List of Publications and References

3.2.10 Release History

The release history provides Information about the particular submissions and how these are inserted in the current and former releases (Figures 6 + 29).

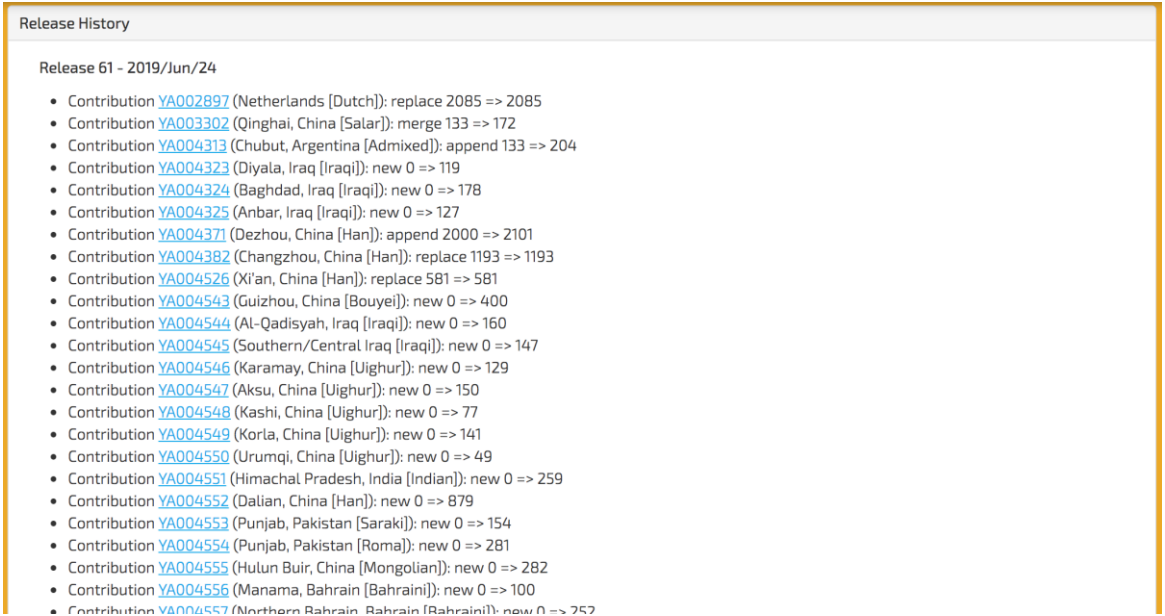


Figure 29: Release History

3.3 Tools

3.3.1 Data file validator

This tool can be used to validate all kinds of YHRD input files like haplotype search lists or population samples (Figures 4 + 30).

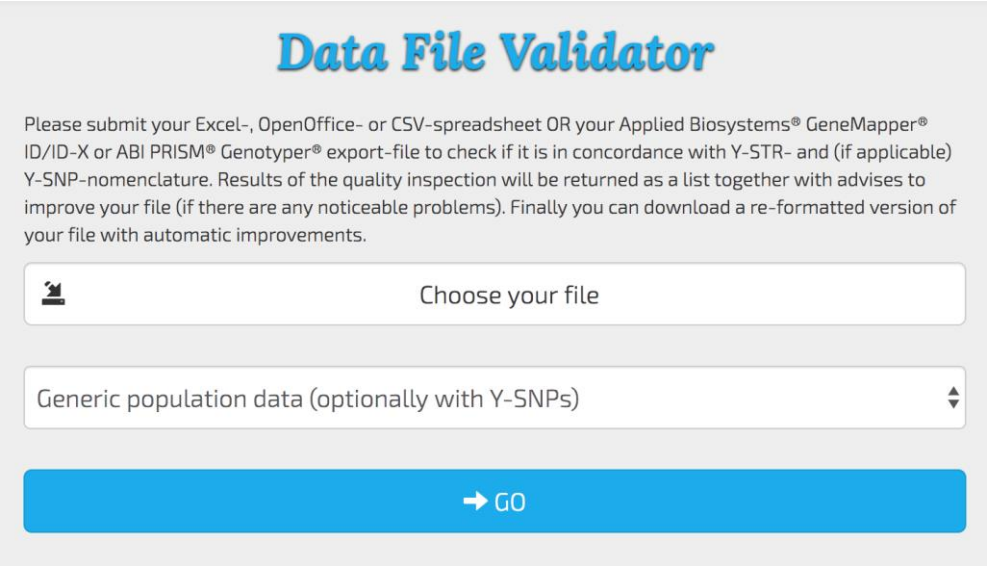


Figure 30: Data File Validator

3.3.2 Y-STR Mixture Calculation

This tool can be applied when a mixed trace (≥2 male donors) and one person (suspect) which is included in the mixed profile should be analyzed. The trace Y-STR profile is compared to the known putative donor as well as to a number of unknown donors. The result will be a likelihood ratio of donorship vs. non-donorship of the suspect to the trace (Wolf et al. 2005) (Figures 31a – c)

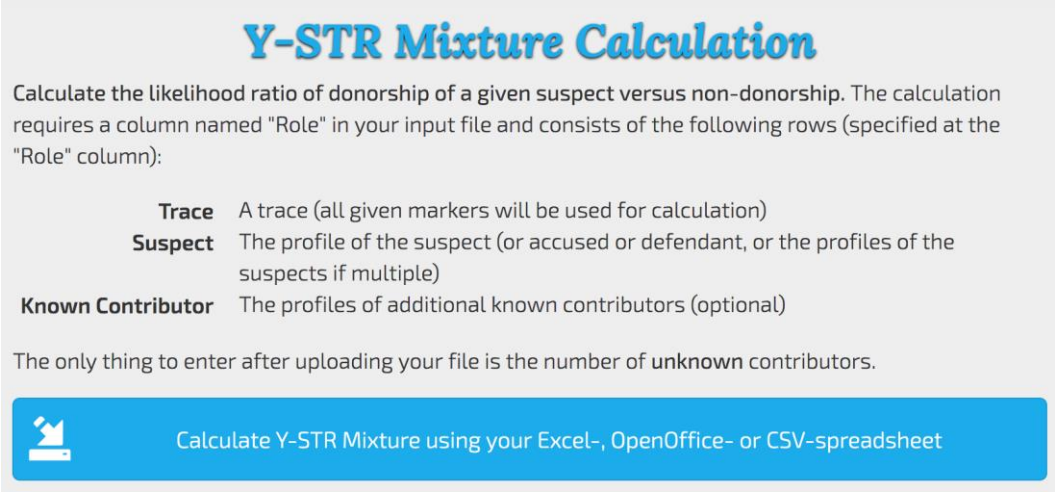


Figure 31a: Mixture analysis

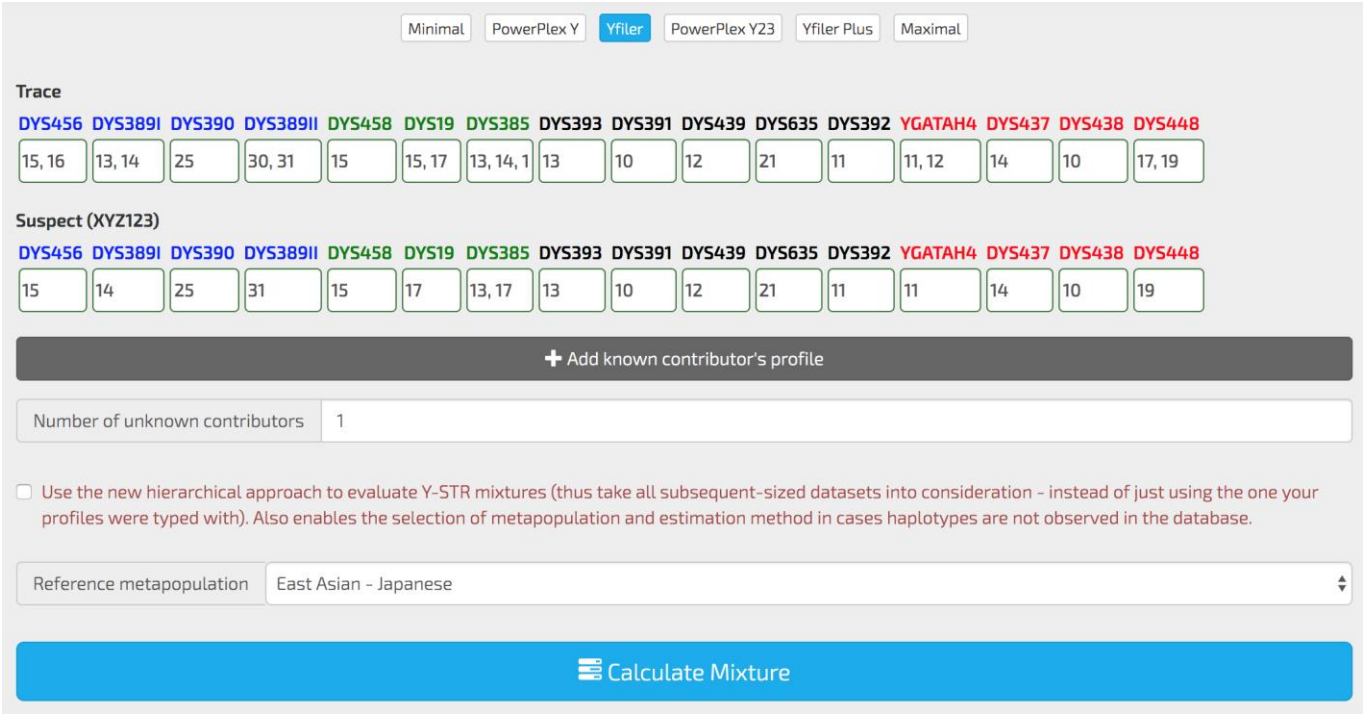


Figure 31b: Mixture analysis – options and confirmation

Mixture Calculation Result

Congratulations! Your job has been successfully processed. It took less than a minute.

Name	Role	DYS456	DYS389I	DYS390	DYS389II	DYS458	DYS19	DYS385	DYS393	DYS391	DYS439
Trace A	Trace	15, 16	13, 14	25	30, 31	15	15, 17	13, 14, 17	13	10	12
XYZ123	Suspect	15	14	25	31	15	17	13, 17	13	10	12

Given the Haplotypes above and that there is 1 additional unknown contributor, the likelihood of the donorship of the suspect is approx. 15 times more likely than the non-donorship. This calculation is based on 1909 haplotypes (Yfiler) in the East Asian - Japanese Metapopulation.

Figure 31c: Mixture analysis –results

3.3.3 Kinship analysis

This tool calculates the likelihood ratio of patrilineal relationship versus non-relationship of male pairs given their Y-STR profiles (Figures 32a – d). The calculation is based on a formula adapted from Rolf et al. (Figure 32a). Calculation details can be reviewed as shown in Figure 32e.

YHRD kinship formula

L : set of loci with typed results for ancestor and offspring

s_l : allelic difference between ancestor and offspring at locus l

$\mu_{l,i}$: i-step mutation rate at loci l

$\frac{\mu_{l,i}}{2}$: probability of exactly one i-step mutation at locus l

$\left(\frac{\mu_{l,i}}{2}\right)^{s_l}$: probability of exactly one s_l i-step mutations at locus l

$1-\mu_l$: probability of one non-mutation at locus l

$(1-\mu_l)^{s_l}$: probability of s_l non-mutations at locus l

H_0 : hypothesis that ancestor and offspring are related

H_1 : hypothesis that ancestor and offspring are unrelated

m : number of transmission events under H_0 and H_1

$\binom{m}{s_l}$: binomial coefficient 'm choose s_l'

$$LR = \frac{P(E|H_0)}{P(E|H_1)}$$

$$LR_{\text{ only 1-steps}} = \frac{\prod_{l \in L}\left[\binom{m}{s_l}\left(\left(\frac{\mu_l}{2}\right)^{s_l}(1-\mu_l)^{m-s_l}\right)\right]}{f(\text{offspring})}$$

Figure 32a: YHRD Kinship formula

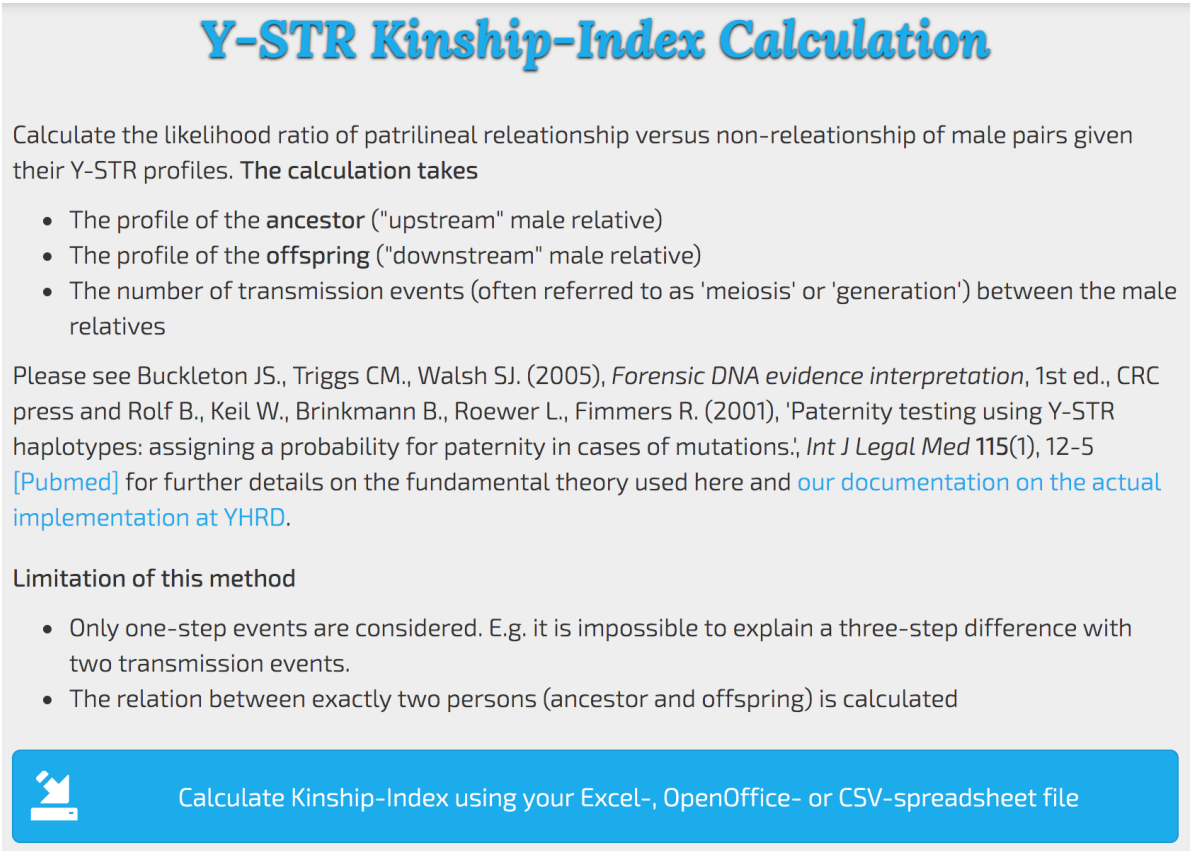


Figure 32b: Kinship analysis

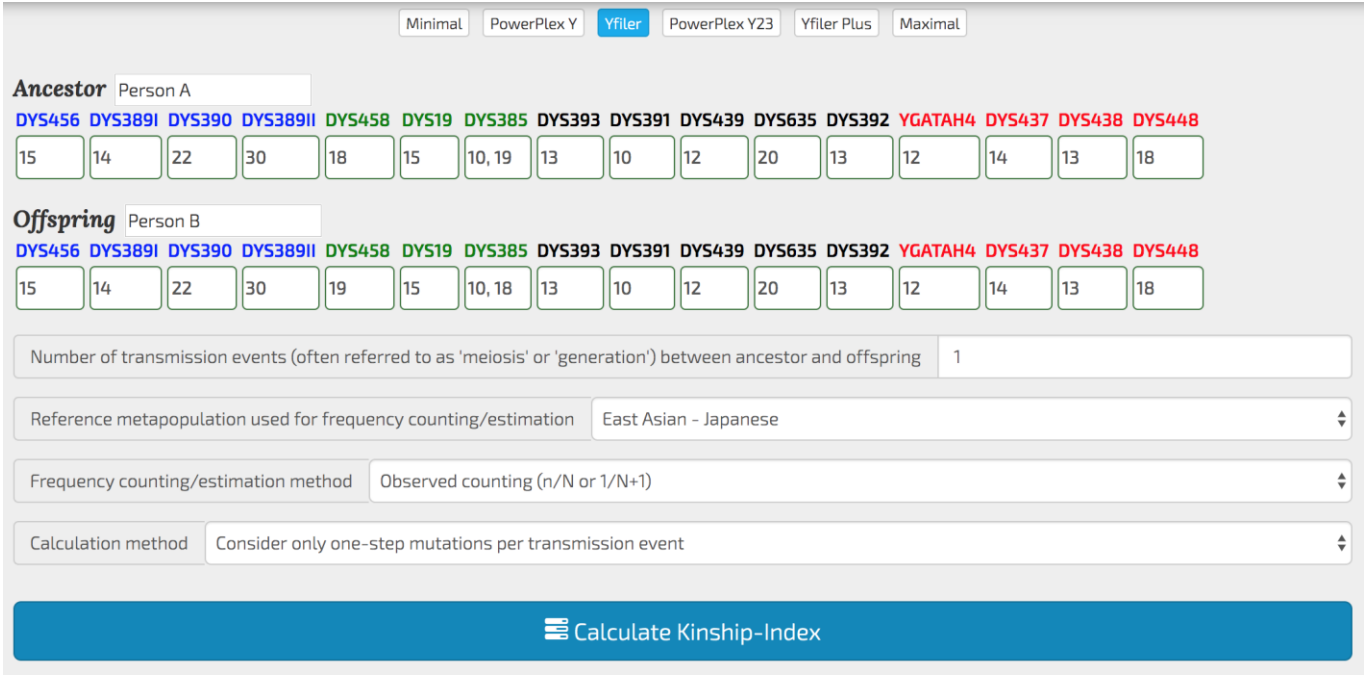


Figure 32c: Kinship analysis – options and confirmation

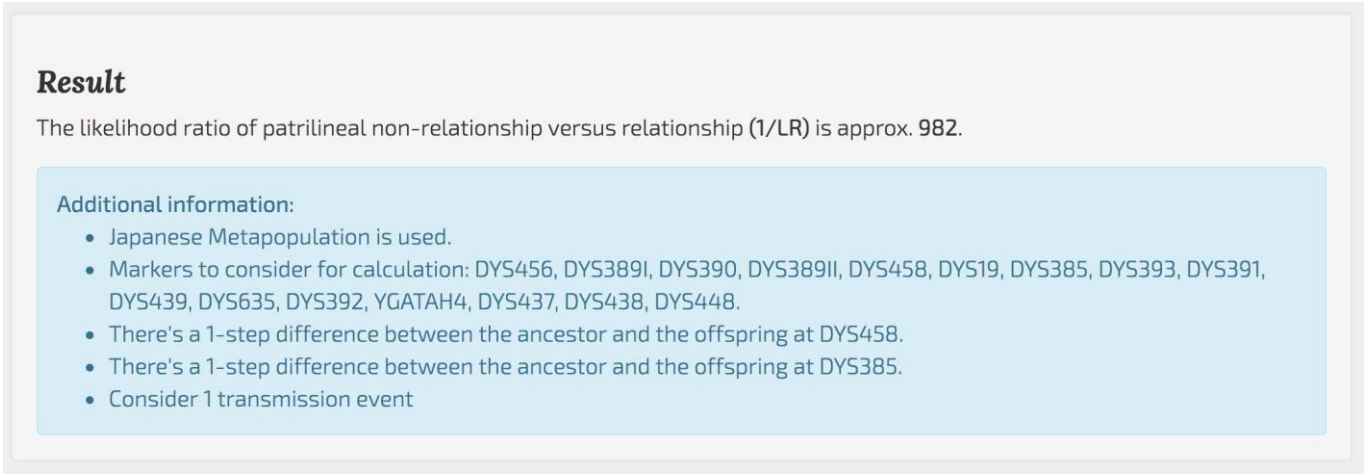


Figure 32d: Kinship analysis – results

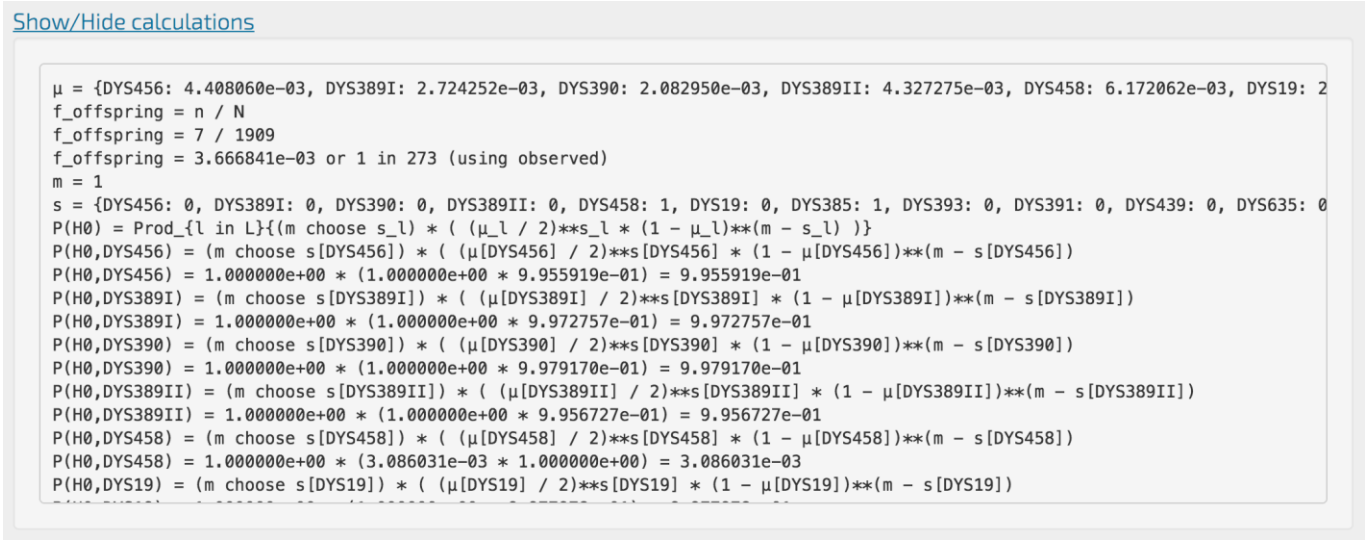


Figure 32e: Kinship analysis – calculation details

3.3.4 AMOVA & MDS

Analysis of Molecular Variance (AMOVA) is a method for analyzing genetic distances between populations using molecular data, e.g., Y-STR haplotypes (Excoffier et al. 1992). Besides the population in question, up to 10,000 haplotypes from reference populations and national databases of the YHRD can be selected. The online calculation returns as a result a CSV file (Figure 33g) with pairwise F_{ST} or R_{ST} values accompanied by p-values as a test for significance (10,000 permutations). In addition, an MDS plot is generated to illustrate the genetic distance between the analyzed populations graphically (Figure 33f). The program shows the references for the selected population studies which facilitates the correct citation. The step-by-step procedure is explained on the index page (Figures 33a – g).

Calculate AMOVA & MDS

Please perform online AMOVA as follows

1. Use your population sample(s) file (Excel) to start AMOVA. Be sure to provide haplotypes only for markers you are going to use in your AMOVA calculation. *E.g. when using a PowerPlex Y23 file, all AMOVA calculations are based on full Y23 haplotypes.*
2. Select populations and/or whole national databases from the YHRD. You can select up to 10,000 haplotypes in sum.
3. Check and adjust the calculation options/preferences as needed.

The result will consist of: a table of calculated pairwise F_{ST} or R_{ST} and p-values (as .csv file), a MDS plot (as .pdf file) and additionally, if "relaxing" was selected, a relaxation file with all the clusters built prior calculation (as .txt file).

Calculate AMOVA & MDS using your Excel-, OpenOffice- or CSV-spreadsheet population sample

Figure 33a: Calculate AMOVA and MDS

Calculate AMOVA & MDS

Please carefully check the data extracted from your file. There is 1 population sample given:

- 128 PowerPlex Y23 Haplotypes in a population sample called **Berlin, Germany [German]**

If the name doesn't match your expectations, please modify the appropriate column in your input file and re-submit it here.

ID	Population	DYS576	DYS389I	DYS448	DYS389II	DYS19	DYS391	DYS481	DYS549	DYS533	DYS4
TRV6L5A	Berlin, Germany [German]	17	12	20	28	15	10	27	12	12	10

➜ Continue with calculate AMOVA & MDS

Figure 33b: Calculate AMOVA and MDS – confirmation

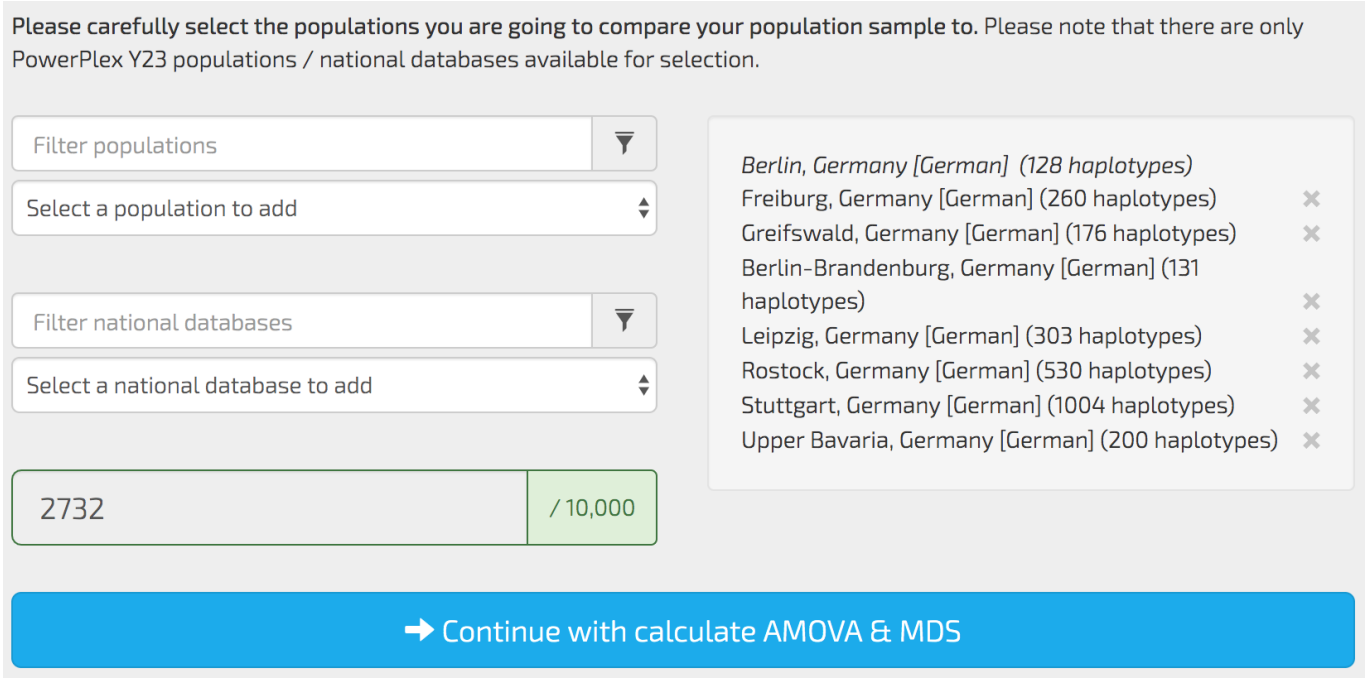


Figure 33c: Calculate AMOVA and MDS – selection

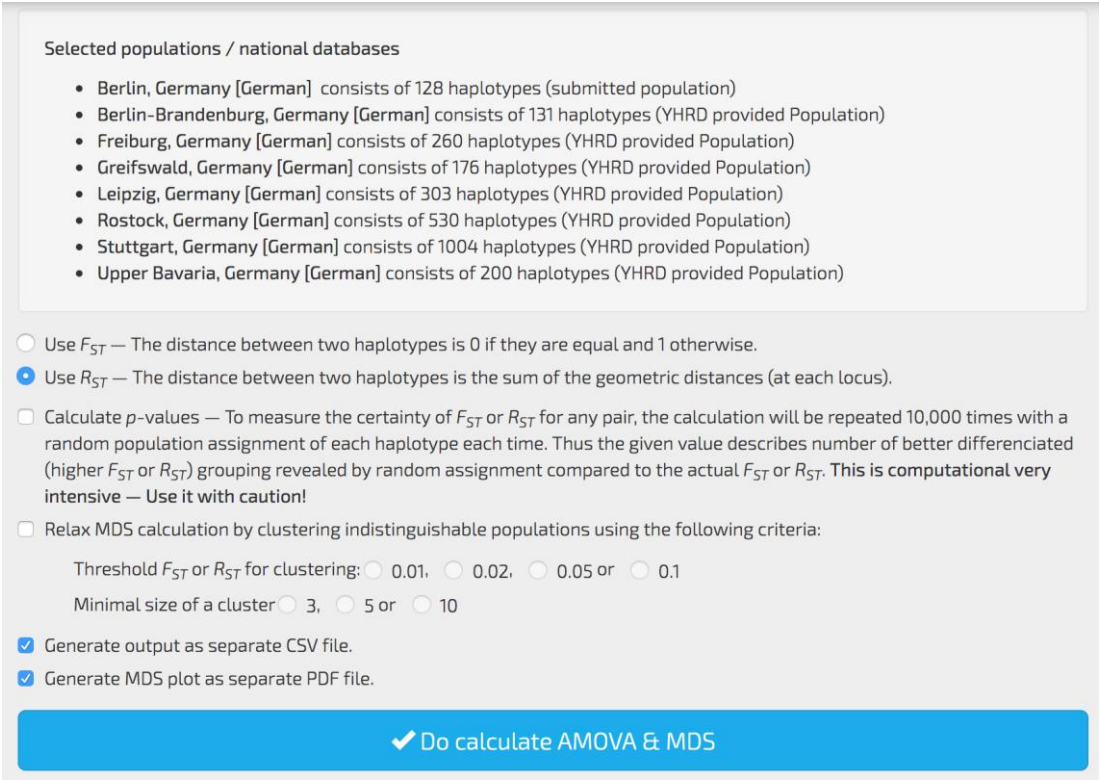


Figure 33d: Calculate AMOVA and MDS – options

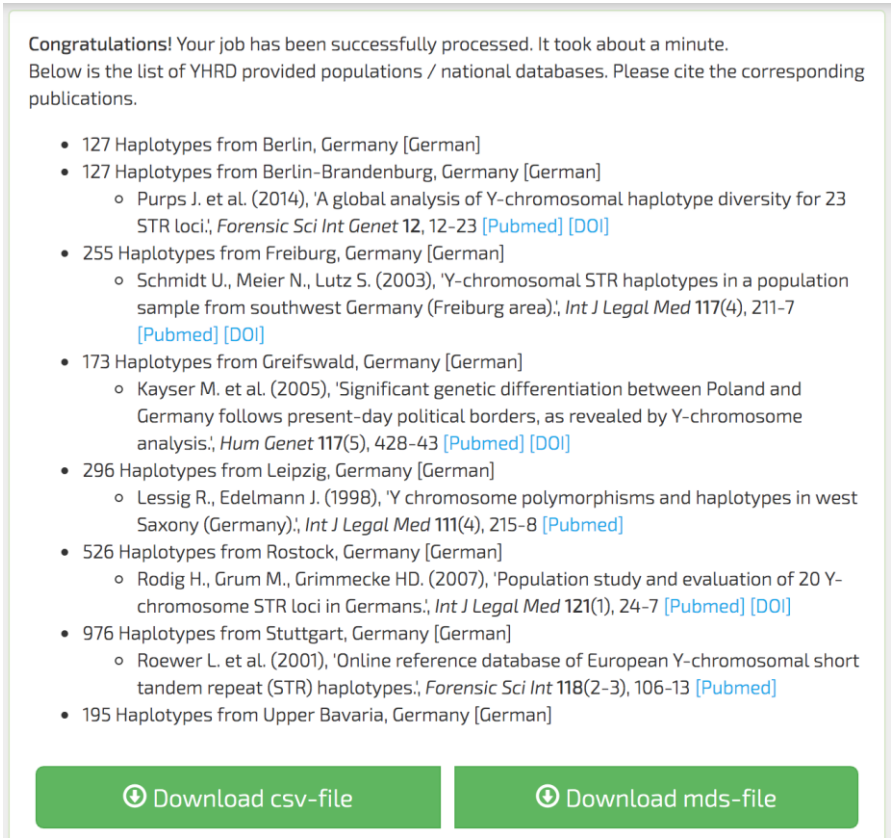


Figure 33e: Calculate AMOVA and MDS – results

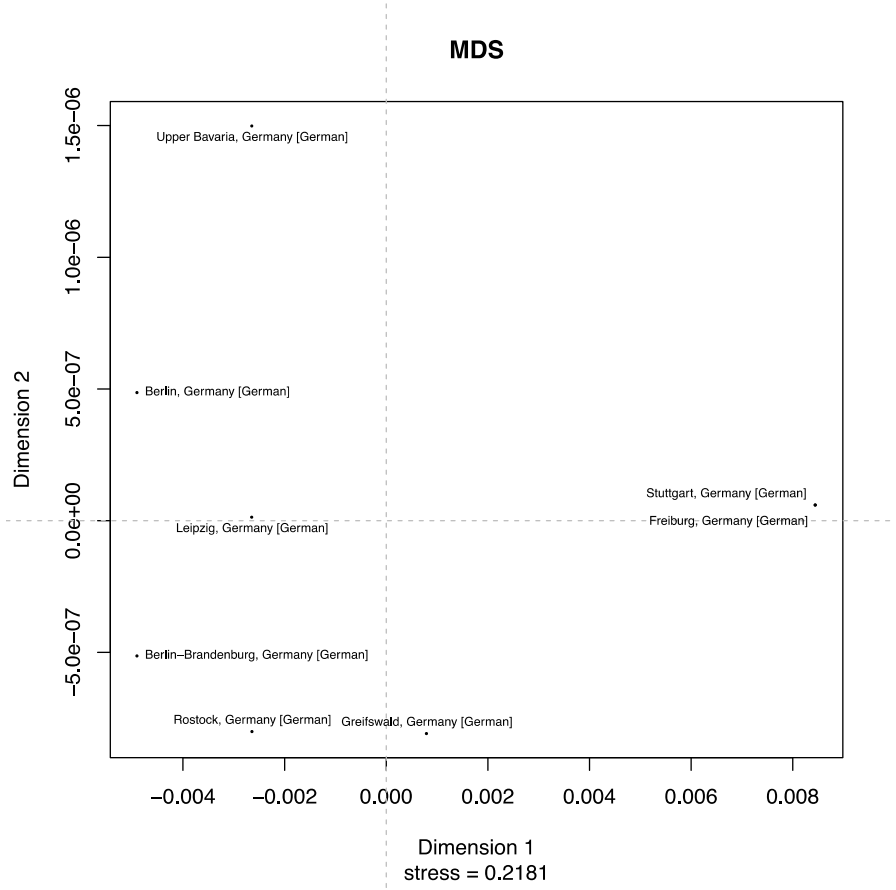


Figure 33f: Calculate AMOVA and MDS – MDS plot

	A	B	C	D	E	F	G	H	I
1	Population	Berlin, Germany [German]	Berlin-Brandenburg, Germany [German]	Freiburg, Germany [German]	Greifswald, Germany [German]	Leipzig, Germany [German]	Rostock, Germany [German]	Stuttgart, Germany [German]	Upper Bavaria, Germany [German]
2	Berlin, Germany [German]	-	10.000	0.0002	0.0175	0.2810	0.1761	0.0000	0.1877
3	Berlin-Brandenburg, Germany [German]	-0.0079	-	0.0002	0.0168	0.2752	0.1722	0.0001	0.1902
4	Freiburg, Germany [German]	0.0230	0.0230	-	0.1012	0.0002	0.0003	0.4041	0.0177
5	Greifswald, Germany [German]	0.0110	0.0110	0.0032	-	0.0625	0.2400	0.0058	0.1046
6	Leipzig, Germany [German]	0.0010	0.0010	0.0128	0.0040	-	0.6452	0.0000	0.6223
7	Rostock, Germany [German]	0.0020	0.0020	0.0116	0.0010	-0.0007	-	0.0000	0.2247
8	Stuttgart, Germany [German]	0.0256	0.0256	-0.0001	0.0073	0.0132	0.0137	-	0.0034
9	Upper Bavaria, Germany [German]	0.0024	0.0024	0.0072	0.0034	-0.0010	0.0010	0.0075	-

Figure 33g: Calculate AMOVA and MDS – R_{ST}- and p-values

3.4 Help & Support

The YHRD offers a number of help and support pages which include this manual, a screencast, FAQs, videos and links to educational webinars.

3.4.1 Contribute

This page explains in detail the submission of population samples to the YHRD and the procedure to receive an accession number (mandatory for peer-reviewed publication).

4 Glossary

4.1 Metapopulations

A metapopulation is generally considered to consist of several distinct populations. In forensic genetics the term "metapopulation" was adapted to describe an assemblage of genetic variants with shared ancestry spread over a territory (Millstein 2010; Willuweit and Roewer 2015). Population genetic analyses on different Y chromosomal marker sets show that metapopulations are stabilized over time by cultural and social factors, including a common language (Quintana-Murci et al. 2001, Baker et al. 2017), patrilocality (Oota et al. 2001) and/or geographical barriers (Rosser et al. 2000). Consequently, the YHRD was reorganized with the activation of YHRD version 4.0 in August 2014 in order to reflect Y-specific metapopulation structures and allow meaningful frequency calculations taking the cladistic structure of Y-STR haplotypes into account. Haplotype frequencies can still be reported country-wise (135 political entities are searchable) but preferably in metapopulations (MP), e.g., in the East Asian MP or the East Asian-Japanese MP (Figures 34 + 35). Frequency estimation methods such as DL use the collected samples of an MP to model the probability distribution of Y-STR haplotypes. Because these distributions differ between metapopulations, the estimated frequency of a certain haplotype also differs considerably between MPs.

Contributors to the YHRD are requested to provide metadata (geographic coordinates, ethnic ancestry and language group) on their population samples in order to assign these sensibly to a metapopulation.

It is important to state that the current metapopulation structure of the YHRD is an a-priori categorization which needs a continuous evaluation and verification by means of statistical methods to quantify the genetic similarity/dissimilarity between the samples. This research will show which groups can be abandoned, further divided or need a new definition.

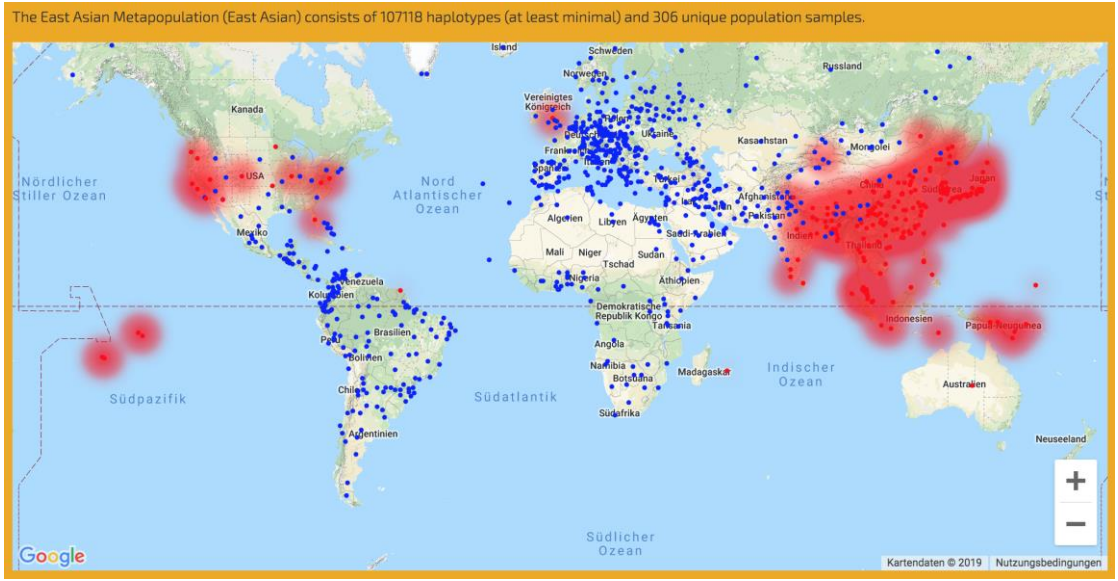


Figure 34: Distribution of the East Asian Metapopulation

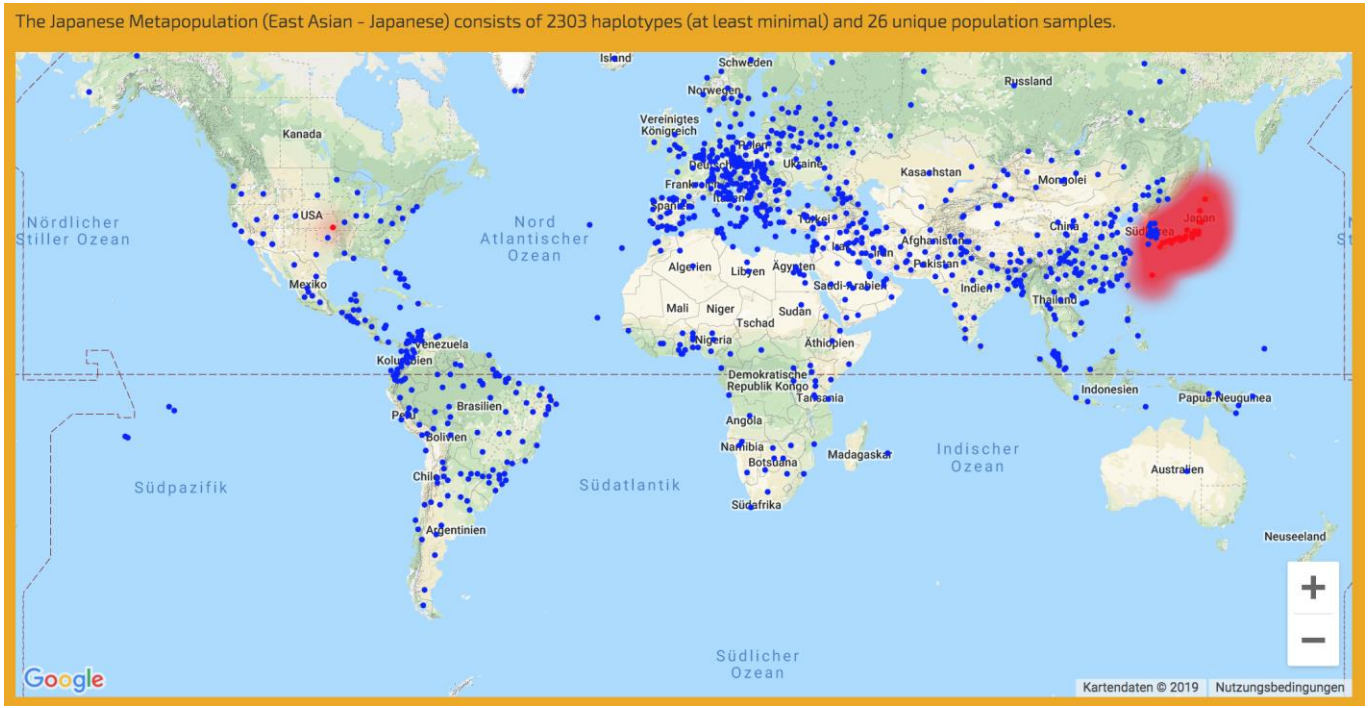


Figure 35: Distribution of the East Asian - Japanese Metapopulation

4.2 Haplogroups

The typing of Y chromosomes submitted to the YHRD is continuously extended for binary Y-SNP polymorphisms. Such "phylogenetic" Y-SNPs define a haplogroup which pertains to a single line of paternal descent. All Y chromosomes sharing a SNP mutation are related by descent, until a further mutation splits the branch. Haplogroups can be related by a single phylogeny using the principle of maximum parsimony (Jobling and Tyler-Smith 2003). The YHRD takes advantage of a consensus phylogeny used in forensic genetics (PhyloTree *Y*) which has a clear structure and nomenclature and captures the main Y-haplogroups of worldwide significance (van Oven et al. 2014). The YHRD submissions are prepared using the PhyloTree *Y* nomenclature defining Y-chromosome haplogroups by letters and the marker name, e.g., R1b-M269. The analysis of worldwide human Y-chromosome diversity has revealed a clear association of haplogroups and related haplotypes with geography and/or language (Underhill et al. 2000, Roewer et al. 2005). The Y-SNP/Y-STR typed reference samples of the YHRD can thus be used to predict the paternal ancestry of a DNA sample.

4.3 Frequency estimation methods

4.3.1 Counting

The observed match count in a database is used as the frequency estimate. The method is limited by the database size.

4.3.2 Augmented counting

The frequency of a haplotype when adding the haplotype in question once to both the database and the observations. The method is limited by the database size.

4.3.3 Discrete Laplace

The Discrete Laplace (DL) method estimates haplotype frequencies within a metapopulation by taking the phylogenetic relationship of haplotypes and its allelic distributions into account (Andersen et al. 2013). Please note that DL is calculated for minimal and YFiler haplotypes only and that the multicopy loci (DYS385ab) is excluded. Also, partial haplotypes or haplotypes with intermediate alleles or Null alleles cannot be calculated with the DL method. The method is not limited by database size but needs representatively sampled metapopulations.

4.3.4 Kappa

The Kappa method is a haplotype frequency estimation method which uses the frequency of singletons within a population sample (Brenner 2010). The method is limited by the database size.

4.3.5 Confidence Intervals

Following the correct Clopper-Pearsons (Clopper & Pearson 1934) method by evaluating the corresponding quantiles (given by α = 5% in case of 95%-CI) using the beta distribution from

lower confidence interval (LCI = 2.5% in case of 95%-CI) to upper confidence interval (UCI = 97.5% in case of 95%-CI):

$$B\left(\frac{\alpha}{2}; x, n-x+1\right) \quad \ldots \quad B\left(1-\frac{\alpha}{2}; x+1, n-x\right)$$

where x is the number of successes (haplotype matches), and n is the number of trials (database size). The formula can be easily evaluated using Excel by utilizing the function "BETA.INV" and the same parameters as shown above.

Note that there are two special cases which cannot be evaluated using the initial formula:

(a) When there are no observations (successes, x = 0). In those cases, the LCI is 0 and the UCI is given by

$$1-\left(\frac{\alpha}{2}\right)^{1/n}$$

(b) When all haplotypes in the database are matching (all trials are successes, x = n). In those cases, the LCI is given by

$$\left(\frac{\alpha}{2}\right)^{1/n}$$

and the UCI is 1.

YHRD provides four different presets for CI calculation: 95%-CI (2.5%-97.5%), 95%-UCI, 99%-CI (0.5%-99.5%) and 99%-UCI. Click the blue triangle next to the CI value to choose your appropriate CI value (see Figure 36).

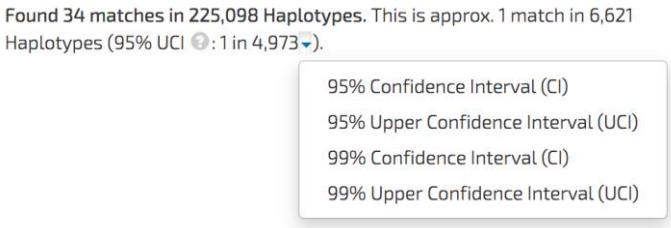


Figure 36: Choose Confidence Interval

4.4 AMOVA (Analysis of Molecular Variance)

4.4.1 *F*-Statistics

The values of *F*-statistics measure the correlation between genes drawn at different levels of a hierarchically subdivided population and allow the characterization of the level of genetic distinctiveness of supposedly inbred or isolated populations, and discrimination even between closely related populations; specifically the degree of (usually) a reduction in heterozygosity when compared to Hardy-Weinberg expectation.

F_{ST} describes the correlation between molecular diversity of random Haplotypes between two populations relative to random pairs of Haplotypes drawn from the whole species.

There are two different ways of dealing with a molecular distance of two Haplotypes: F_{ST} -based, where two Haplotypes are either equal (F_{ST} =0) or have at least one inequality (F_{ST} =1) and R_{ST}-based, where the sum of all squared differences between corresponding haplotype-pairs is used (Excoffier et al. 1992 and Roewer et al. 1996).

4.4.2 MDS (Multidimensional scaling)

This analysis is used to assign pair-wise similarities (or dissimilarities) to points in an N-dimensional space representing those similarities as distances between points. There are two major principles of MDS:

Metric MDS (M-MDS), a distance matrix D (similarities) into a set of coordinates such that the Euclidean distances derived from these coordinates fitting D as well as possible. The basic idea

of M-MDS is to transform the distance matrix into a cross-product matrix and then to find its Eigen-decomposition which gives a principal component. This requires linearity assumptions to be met.

Non-Metric MDS (N-MDS), on the other hand, uses the rank of a distance matrix delta (dissimilarities) to iteratively assign locations to monotonic parts of delta. In every iteration, the configuration of assigned locations is evaluated with respect to a stress criterion (how well the configuration approximates the original input dissimilarities).

5 References

Andersen MM, Eriksen PS, Morling N (2013) The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. J Theor Biol 329:39-51

Baker JL, Rotimi CN, Shriner D (2017) Human ancestry correlates with language and reveals that race is not an objective genomic classifier. Sci Rep. 7(1):1572

Brenner CH (2010) Fundamental problem of forensic mathematics – the evidential value of a rare haplotype. Forensic Sci Int Genet 4:281–91

Clopper CJ, Pearson ES (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26:404–13

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131(2):479–91.

Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM and International Society of Forensic Genetics (2006) DNA commission of the international society of forensic genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Int J Legal Med 120(4):191–200.

Jobling MA and Tyler-Smith C (2003) The human y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4(8):598–612.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human y chromosomal haplogroup tree. Genome Res 18(5):830–8.

Millstein RL (2010) The concepts of population and metapopulation in evolutionary biology and ecology. In M. A. Bell, D. J. Futuyma, W. F. Eanes, & J. S. Levinton (Eds.), Evolution since Darwin: The first 150 years. Sunderland, MA: Sinauer.

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29(1):20-1.

Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K (2001) Y chromosome lineages trace diffusion of people and languages in southwestern Asia. Am J Hum Genet. 68(2):537-42.

Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1996) Analysis of molecular variance (AMOVA) of y-chromosome-specific microsatellites in two closely related human populations. Hum Mol Genet 5(7):1029–33.

Roewer L, Croucher PJP, Willuweit S, Lu TL, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M on behalf of the Forensic Y-Chromosome Research Group (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279-91.

Roewer L, Kayser M, de Knijff P, Anslinger K, Betz A, Caglià A, Corach D, Füredi S, Henke L, Hidding M, Kärgel HJ, Lessig R, Nagy M, Pascali VL, Parson W, Rolf B, Schmitt C, Szibor R, Teifel-Greding J, Krawczak M (2000) A new method for the evaluation of matches in non-recombining genomes: application to y-chromosomal short tandem repeat (STR) haplotypes in european males. Forensic Sci Int 114(1):31–43.

Roewer L (2009) Y chromosome STR typing in crime casework. Forensic Sci Med Pathol 5(2):77–84.

Roewer L (2019) Y-chromosome STRs in forensics – sexing, profiling and matching male DNA. WIREs Forensic Sci 1(4):e1336.

Rolf B, Keil W, Brinkmann B, Roewer L, Fimmers R (2001) Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations. Int J Legal Med 115(1): 12-5.

Rosser ZH, Zerjal T, Hurles ME et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67(6):1526-43.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26(3):358–61.

van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MH (2014) Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Hum Mutat 35(2):187-91.

Willuweit S, Roewer L (2015) The new Y Chromosome Haplotype Reference Database. Forensic Sci Int Genet 15:43-8.

Wolf A, Caliebe A, Junge O, Krawczak M (2005) Forensic interpretation of y-chromosomal DNA mixtures. Forensic Sci Int 152(2-3):209–13.