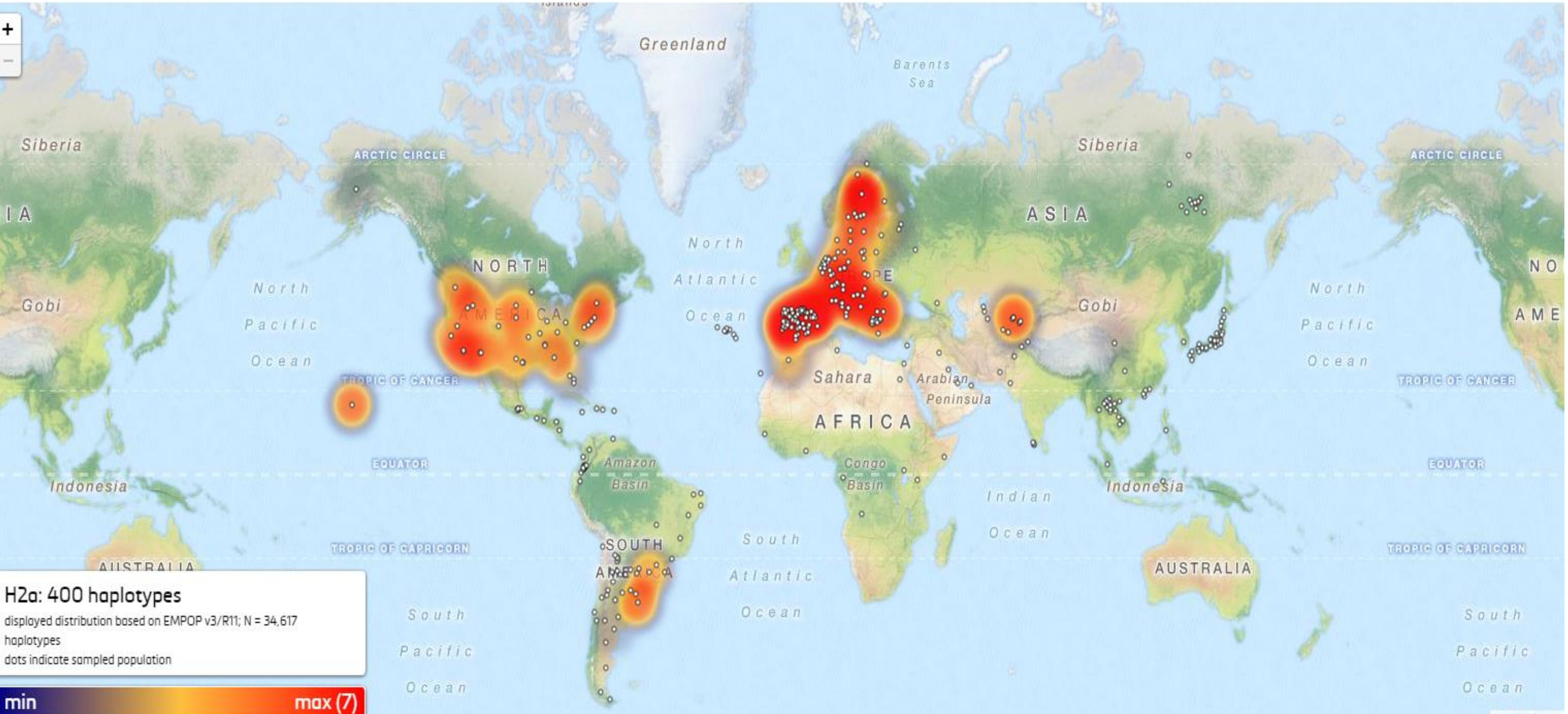
DIRECTIONS FOR USE

Version v3/R11 July 2015

EMPOP mtDNA Database



EMPOP mtDNA Database – Directions for Use

Revision Overview

Version V3/R11

- December 2015
    - Section 4.3.2. – When no matches are found was updated
        - Tolerance value of EMMA was changed from „0.3” to „0.1”
- November 2015
    - Section 4.1.8. EMPOP haplogroup estimation – EMMA was added
- October 2015
    - Information about special positions was added in Section 4.1.2. Ranges
- July 2015
    - Revision Overview was added
    - Section 4.3.3. Ambiguous haplogroup estimates was added
- May2015
    - Initial Release of EMPOP mtDNA Database – Directions for Use

Table of Contents

How to use EMPOP

1. Introduction

The high copy number per cell, the stability against degradation and the maternal mode of inheritance make the mitochondrial (mt) genome particularly suitable for palaeo-, medical- and forensic genetic investigations. Its increased evolutionary rate led to sequence variation that has been generated by sequential accumulation of new mutations along radiating maternal lineages during human dispersal into different parts of the world.

Forensic molecular biology takes advantage of this variation for human identity testing by sequence analysis of the (hypervariable segments within the) mtDNA control region. New developments in Massively Parallel Sequencing demonstrated that also full mtGenome information can be obtained from even degraded forensic samples. MtDNA analysis is a powerful tool to exclude samples as originating from the same individual/matriline. If two samples cannot be excluded the significance of the mtDNA match is assessed by making reference to the abundance of that particular mtDNA sequence (= mtDNA haplotype) in a relevant population.

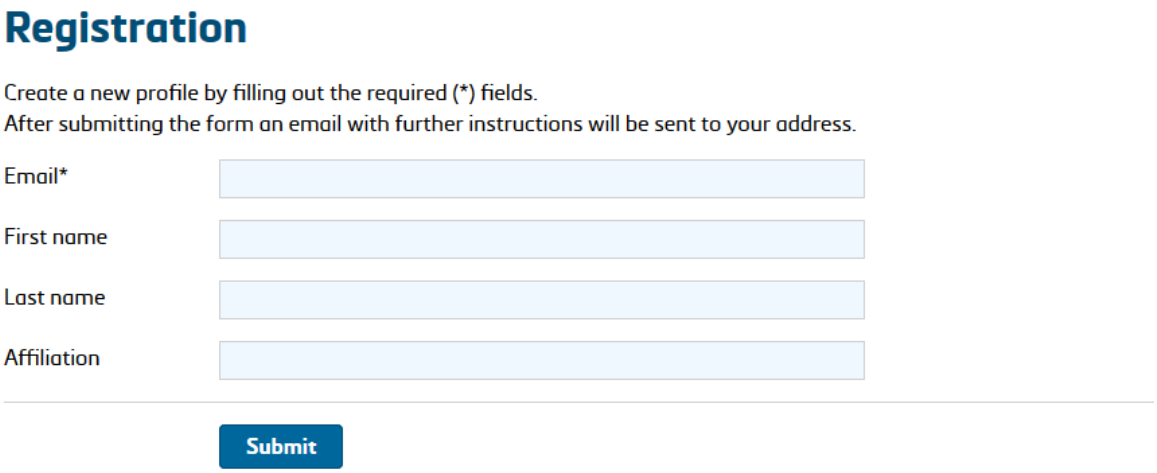
2. Concept

The EMPOP database aims at the collection, quality control and searchable presentation of mtDNA haplotypes from all over the world. EMPOP has carefully been envisioned and designed as high quality mtDNA database, where available primary sequence lane data are permanently linked to the database entries. The scientific concept and the quality control measures using logical and phylogenetic tools were found suitable for forensic purposes, e.g. by a declaration of the German Supreme Court of Justice (2010), the SWGDAM mtDNA interpretation guidelines (2013), and the updated ISFG guidelines for mtDNA analysis and interpretation (2014).

The scientific contents presented in EMPOP were developed by the Institute of Legal Medicine (GMI), Medical University of Innsbruck and the Institute of Mathematics, University of Innsbruck.
The haplotypes stored in EMPOP are not considered for partial or full download. The concept of data quality management requires a centralized supervision of the data. Necessary updates (e.g. haplogroup status) will be introduced by the database holders to ensure continuous data quality and are made publicly available (see Release history).

3. Register/login

An EMPOP user is identified by the Email address to which account information and search history are connected. Follow the instructions for registration. An Email will be sent with a link that completes registration.



Registration

Create a new profile by filling out the required (*) fields.
After submitting the form an email with further instructions will be sent to your address.

Email*

First name

Last name

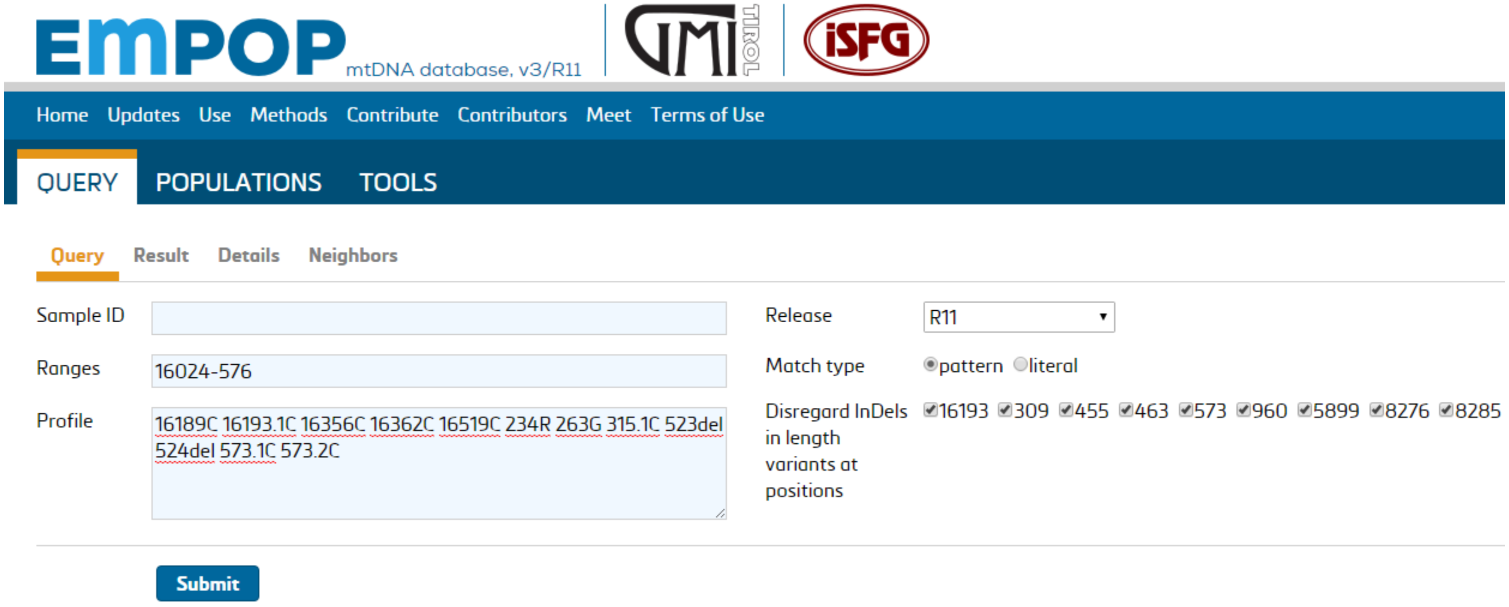
Affiliation

Figure 1 – User Registration

4. Using EMPOP to perform mtDNA haplotype frequency estimates

EMPOP follows the revised and extended guidelines for mitochondrial DNA typing issued by the DNA commission of the ISFG (Parson et al. 2014). See document for further details.

4.1. Query options



EMPOP mtDNA database, v3/R11

Home Updates Use Methods Contribute Contributors Meet Terms of Use

QUERY POPULATIONS TOOLS

Query Result Details Neighbors

Sample ID:

Ranges: 16024-576

Profile: 16189C 16193.1C 16356C 16362C 16519C 234R 263G 315.1C 523del 524del 573.1C 573.2C

Release: R11

Match type: ● pattern ○ literal

Disregard InDels in length variants at positions: ☑16193 ☑309 ☑455 ☑463 ☑573 ☑960 ☑5899 ☑8276 ☑8285

Submit

Figure 2 – Query Input

4.1.1. Sample ID

Use this field to enter the ID of an mtDNA haplotype. Search results are linked to this information and also provided on printouts. Sample IDs are used to identify queries in the search history of each individual user.

4.1.2. Ranges

Database queries require specification of the sequence range(s). EMPOP depend on the information captured by the submitting contributors. This is why also individual sequence ranges can be found, e.g. 16024-16390 73-300. Some data are represented by HVS-I and HVS-II (73-340) or by the entire control region (16024-576). With the emergence of Massively Parallel Sequencing techniques full mitochondrial genomes (1-16569) will be increasingly developed in forensic science. EMPOP3 holds full mitochondrial genome sequences for database queries.

Examples:

16024-16356 73-340	represents a standard range for a query in HVS-I and HVS-II.
16024-576	represents the control region range
16024-16365 489 3010	represents a query range including HVS-I and the two SNPs 489 and 3010. Note that an insertion between 489 and 490 is not included in that query range.

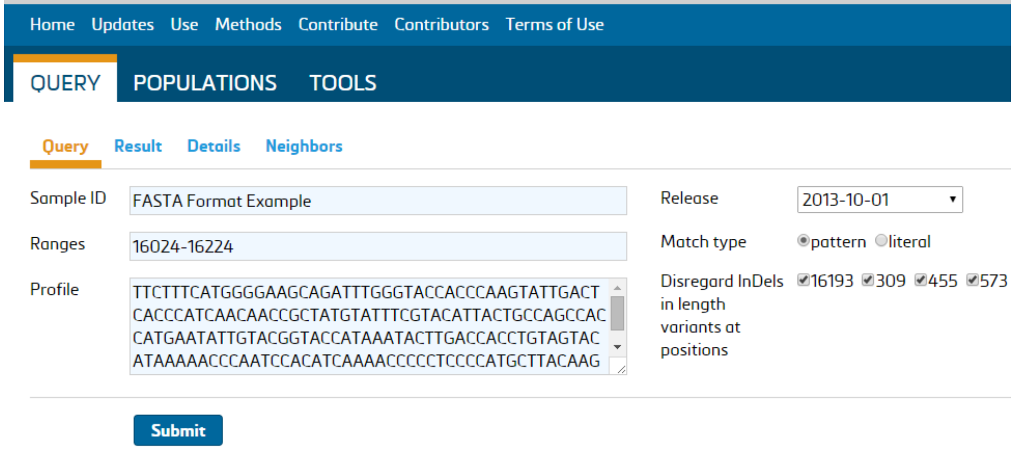
Note that there are special positions where a query does not make sense. For example, position 3107: A deletion at this position can not be queried as it serves only as a place holder to keep the original numbering system. EMPOP will display an appropriate error message in such cases.

4.1.3. Profile

MtDNA haplotypes can be entered as alignment-free sequence strings or aligned relative to the rCRS.

Query sequence strings:

Copy&paste a sequence string from a text file or a consensus from sequence analysis software. Do not enter header information like in FASTA format, enter nucleotides only. Mixtures (e.g point heteroplasmy) can be entered using IUPAC format.



Sample ID: FASTA Format Example

Ranges: 16024-16224

Profile:

```
TTCTTTCATGGGGAAGCAGATTTGGGTACCACCCAAGTATTGACT
CACCCATCAACAACCGCTATGTATTTCGTACATTACTGCCAGCCAC
CATGAATATTGTACGGTACCATAAATACTTGACCACCTGTAGTAC
ATAAAAACCCAATCCACATCAAAACCCCCTCCCCATGCTTACAAG
```

Release: 2013-10-01 ▾

Match type: ● pattern ○ literal

Disregard InDels in length variants at positions: ☑16193 ☑309 ☑455 ☑573

Figure 3 – Query Input in FASTA Format

Query rCRS aligned haplotypes:

Differences to the revised Cambridge Reference Sequence (rCRS, Andrews et al 1999) are entered as haplotypes.

Table 1 – Notation Guidelines:

Type	Possible annotations	Comment
Base changes	73G, A73G	If preceding bases are used they must match rCRS base at the given position
Insertions	315.1C -315.1C 315+C 309.1C 309.2C 309+CC	For multiple insertions all preceding insertions need to be stated, i.e. annotating 309.2C is not possible without annotating 309.1C
Deletions	249- A249- 249delA 249del	'del' is treated case insensitive, e.g. Del, DEL, dEL, deL etc is accepted. Please note that the single character 'D' is considered a mixture of A, G, and T (IUB code). The single character 'd' is considered a mixture of A, G, T, and deletion (see Röck et al. 2013 for details).

Note that the EMPOP query discerns capital letters (A,G, C, T, Y, ...) from uncapitalized letters (a, g, c, t, y, ...). Uncapizalized letters stand for a mixture of a deletion and a non-deleted variant. E.g. T152c represents two variants, T152C and T152del.

4.1.4. Release

EMPOP 3 offers release-specific queries. The most recent database release is selected by default. Earlier database releases can be selected.

4.1.5. Match type

This is relevant for the consideration of point heteroplasmy in both the query sequence as well as the database sequences.

Pattern match: mixture designations match its individual components (Y={C,T,Y}). Example: 152Y matches 152T and 152C.

Literal match: mixture designations are considered exclusive to all other nucleotide designations (Y={Y}). Example: 152Y matches only 152Y.

Pattern match is default setting for forensic frequency estimates.

4.1.6. Disregard InDels in length variants at positions

Length variants that are known hotspots for insertion/deletions (indels) should be ignored in a forensic database query. This involves the C-runs around positions 16193, 309, 463 and 573 and the T-run around position 455 relative to the rCRS in the control region. In the coding region length variants around positions 960, 5899, 8276 and 8285 should be ignored for a forensic query.

Standard query settings disregard discrepancies in hot spot length variant regions between query and database sequences.

Note that costs of disregarded InDels do not contribute to the final costs which influences the ranking of results. See section 4.4. Neighbors.

4.1.7. EMPOP query engine - SAM

MtDNA sequences are traditionally reported relative to the human reference sequence (rCRS). This format is short and convenient, however nucleotide sequences strings can be translated into more than one rCRS-coded haplotype and are therefore ambiguous. As a consequence, database searches may suffer from biased results when query and database haplotypes are aligned differently. In the forensic context that could lead to an underestimation of absolute and relative frequencies and thus to an overestimation of the statistical power of the evidence.

EMPOP uses SAM, a string-based search algorithm that converts query and database sequences into alignment-free nucleotide strings and thus eliminates the possibility that identical sequences will be missed in a database query.

EMPOP 3 introduces an updated query engine that considers block insertions and block deletions (indels) as a single phylogenetic event. In the CA-repeat of the control region (positions 513 and 524) only tandem indels are observed, e.g. 523del 524del. While this tandem deletion nominally constitutes two individual differences to the rCRS it is considered as single event by the new version of SAM. This better reflects the phylogenetic nature of the mitochondrial molecule.

The following events are considered by SAM-E:

Table 2 - Considered Events:

Position	Type	Length of Ins/Del	Inserted/Deleted Block	Since SAM Version
16193	Length Variation	-	-	20
309	Length Variation	-	-	20
455	Length Variation	-	-	20
463	Length Variation	-	-	20
573	Length Variation	-	-	20
960	Length Variation	-	-	20
5899	Length Variation	-	-	20
8276	Length Variation	-	-	20
8285	Length Variation	-	-	20
16032	Exceptional Insertion	15	TCTCTGTTCTTTCAT	20
104	Exceptional Insertion	6	CGGAGC	20
105	Exceptional Insertion	6	GGAGCA	20
209	Exceptional Insertion	7	GTGTGTT	20
241	Exceptional Insertion	3	TAA	20

Position	Type	Length of Ins/Del	Inserted/Deleted Block	Since SAM Version
286	Exceptional Insertion	5	TAACA	20
291	Exceptional Insertion	16	ACATCATAACAAAAAA	20
398	Exceptional Insertion	14	ACCAGATTTCAAAT	20
470	Exceptional Insertion	8	TACTACTA	20
514	Exceptional Insertion	2	AC	20
516	Exceptional Insertion	2	AC	20
518	Exceptional Insertion	2	AC	20
520	Exceptional Insertion	2	AC	20
522	Exceptional Insertion	2	AC	20
524.2 – 524.8	Exceptional Insertion	2-8	AC	20
563	Exceptional Insertion	204	AACAAAGAAC…AAA	20
8271	Exceptional Insertion	9	CCCCCTCTA	20
8280	Exceptional Insertion	9	CCCCCTCTA	20
8289	Exceptional Insertion	9	CCCCCTCTA	20
16032	Exceptional Deletion	15	TCTCTGTTCTTTCAT	20
110	Exceptional Deletion	6	CGGAGC	20
111	Exceptional Deletion	6	GGAGCA	20
209.7	Exceptional Deletion	7	GTGTGTT	20

Position	Type	Length of Ins/Del	Inserted/Deleted Block	Since SAM Version
241.3	Exceptional Deletion	3	TAA	20
286.5	Exceptional Deletion	5	TAACA	20
291.16	Exceptional Deletion	16	ACATCATAACAAAAAA	20
398.14	Exceptional Deletion	14	ACCAGATTTCAAAT	20
478	Exceptional Deletion	8	TACTACTA	20
516	Exceptional Deletion	2	AC	20
518	Exceptional Deletion	2	AC	20
520	Exceptional Deletion	2	AC	20
522	Exceptional Deletion	2	AC	20
524.2 – 524.10	Exceptional Deletion	2-10	AC	20
563.204	Exceptional Deletion	204	AACAAAGAAC…AAA	20
8280	Exceptional Deletion	9	CCCCCTCTA	20
8289	Exceptional Deletion	9	CCCCCTCTA	20
8289.9	Exceptional Deletion	9	CCCCCTCTA	20

4.1.8. EMPOP haplogroup estimation – EMMA

The assignment of haplogroups to mitochondrial DNA haplotypes contributes substantial value for quality control, not only in forensic genetics but also in population and medical genetics. The availability of Phylotree, a widely accepted phylogenetic tree of human mitochondrial DNA lineages, led to the development of several (semi-)automated software solutions for haplogrouping. However, the currently existing tools only make use of haplogroup-defining mutations, whereas private mutations (beyond the haplogroup level) can be additionally informative allowing for enhanced haplogroup assignment.

EMPOP uses EMMA, an algorithm for estimating the haplogroup of mtDNA sequence based on 14,990 full mtGenomes from GenBank and 3925 virtual haplotypes from Phylotree. Further, 19,171 full control region haplotypes are used to perform a maximum likelihood estimation of the stability of mutations which is expressed as fluctuation rates.

Assuming independent positions fluctuation rates estimated by

$$r_{\alpha\beta} = \frac{\sum_\gamma min(n(\alpha,\gamma), n(\beta,\gamma))}{\sum_\gamma n(\gamma)}$$

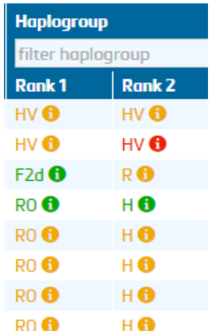
here α, β are elements of the set A, C, G, T, – with α not equal to β, γ runs over all CR-HGs where α or β are dominant, n(x,γ) denotes the number of samples in CR-HG γ with symbol x and n(γ) denotes the total number of samples in CR-HG γ.
The algorithm compares a test profile to every database profile with an appropriate reading frame. Resulting differences are determined and assigned with appropriate costs. By ranking the total costs of the compared profiles, the algorithm is able to cluster optimal and suboptimal profiles. Note that in the output of the algorithm only base profiles with the lowest and second lowest costs are displayed. Further Information and details can be found in Röck et al. 2013.

Note that SAM and EMMA are based on different data sets and thus the calculated costs for specific mutations may differ. Whilst SAM is based on the mtDNA haplotypes which are stored in the EMPOP database, EMMA is based on full mitochondrial genomes from GenBank and on virtual haplotypes from Phylotree.

On the website, color codes indicate the quality of the haplogroup estimation

- green: costs below 1
- yellow: costs between 1 and 2
- red: costs exceeding 2

Individual haplogroup estimates can be found under the ⓘ symbol.



Haplogroup

filter haplogroup

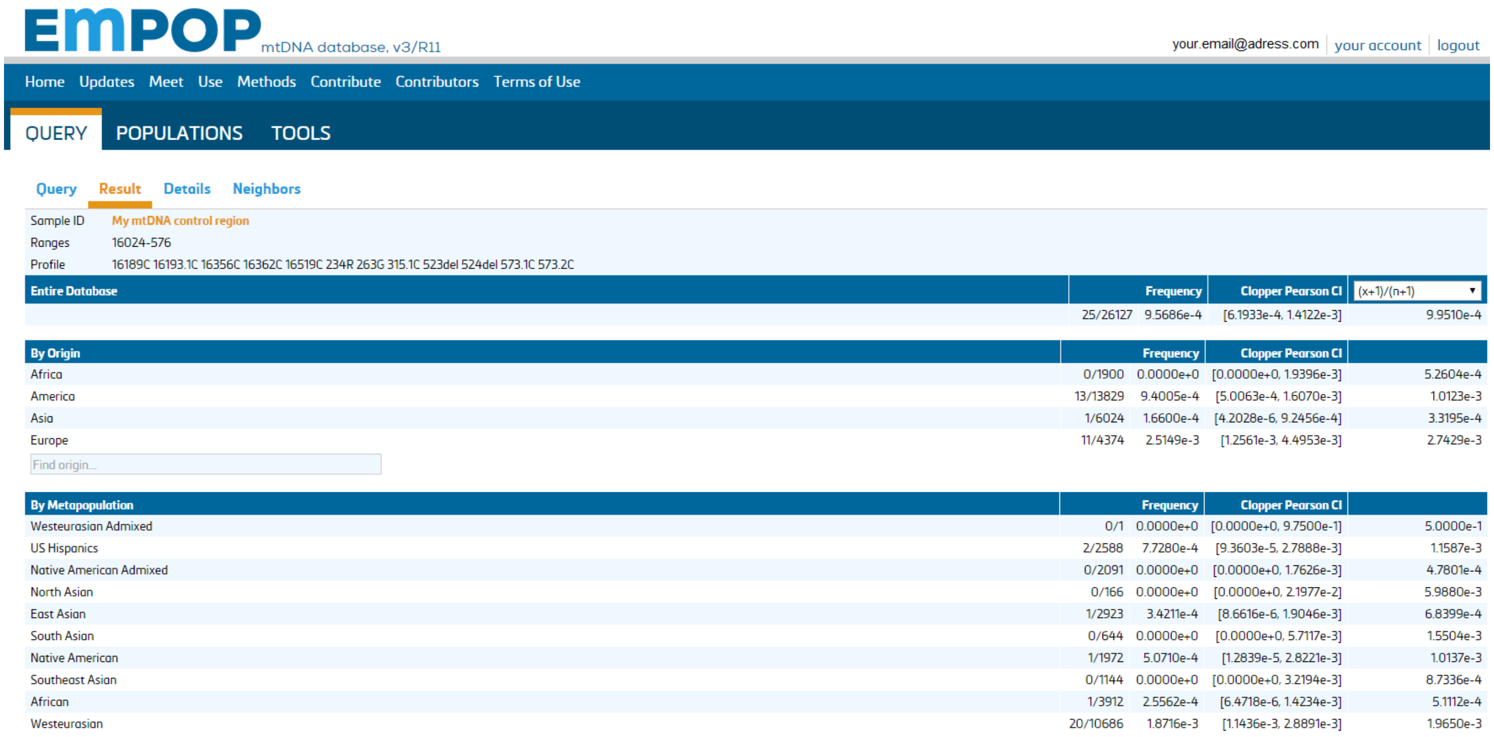
Rank 1	Rank 2
HV ⓘ	HV ⓘ
HV ⓘ	HV ⓘ
F2d ⓘ	R ⓘ
R0 ⓘ	H ⓘ
R0 ⓘ	H ⓘ
R0 ⓘ	H ⓘ
R0 ⓘ	H ⓘ
R0 ⓘ	H ⓘ

Note that the haplogroup estimates presented in this table are based on the full sequence information available for the database haplotypes (which is not necessarily identical to the sequence range of the query haplotype).

4.2. Result

The execution of a database query automatically directs the user to the **Results** tab. Sample ID, query range(s) and haplotype are indicated in the top lines. Following information is listed in the results table:

1. number of observed matches in the entire database
2. number of observed matches sorted by geographic origin and
3. number of observed matches by metapopulation affiliation


EMPOP mtDNA database, v3/R11
your.email@adress.com | your account | logout
Home Updates Meet Use Methods Contribute Contributors Terms of Use
QUERY POPULATIONS TOOLS
Query Result Details Neighbors

Sample ID	My mtDNA control region
Ranges	16024-576
Profile	16189C 16193.1C 16356C 16362C 16519C 234R 263G 315.1C 523del 524del 573.1C 573.2C

Entire Database		Frequency	Clopper Pearson CI	(x+1)/(n+1) ▾
	25/26127	9.5686e-4	[6.1933e-4, 1.4122e-3]	9.9510e-4

By Origin		Frequency	Clopper Pearson CI	
Africa	0/1900	0.0000e+0	[0.0000e+0, 1.9396e-3]	5.2604e-4
America	13/13829	9.4005e-4	[5.0063e-4, 1.6070e-3]	1.0123e-3
Asia	1/6024	1.6600e-4	[4.2028e-6, 9.2456e-4]	3.3195e-4
Europe	11/4374	2.5149e-3	[1.2561e-3, 4.4953e-3]	2.7429e-3

By Metapopulation		Frequency	Clopper Pearson CI	
Westeurasian Admixed	0/1	0.0000e+0	[0.0000e+0, 9.7500e-1]	5.0000e-1
US Hispanics	2/2588	7.7280e-4	[9.3603e-5, 2.7888e-3]	1.1587e-3
Native American Admixed	0/2091	0.0000e+0	[0.0000e+0, 1.7626e-3]	4.7801e-4
North Asian	0/166	0.0000e+0	[0.0000e+0, 2.1977e-2]	5.9880e-3
East Asian	1/2923	3.4211e-4	[8.6616e-6, 1.9046e-3]	6.8399e-4
South Asian	0/644	0.0000e+0	[0.0000e+0, 5.7117e-3]	1.5504e-3
Native American	1/1972	5.0710e-4	[1.2839e-5, 2.8221e-3]	1.0137e-3
Southeast Asian	0/1144	0.0000e+0	[0.0000e+0, 3.2194e-3]	8.7336e-4
African	1/3912	2.5562e-4	[6.4718e-6, 1.4234e-3]	5.1112e-4
Westeurasian	20/10686	1.8716e-3	[1.1436e-3, 2.8891e-3]	1.9650e-3

Figure 4 – Query Result

An uncorrected frequency estimate is provided included a two-tailed Clopper Pearson confidence interval. Correction for sampling bias is provided and alternative methods to calculate probability are provided in the drop-down box to the right. P value can be estimated based on following formulas:

1. (x+1)/(n+1)
2. (x+2)/(n+2)
3. CI from zero pop

Where x… number of database hits and n… database size

Free text searches are possible for origin and metapopulation to address the relevant subset of the database. This depends on the formulation of the hypothesis, e.g. the reduction of the dataset to the country of Spain.

Note that the haplotypes included in a query result depends on the indicated sequence range. Only haplotypes with overlapping sequence ranges to the query sequence are considered. E.g. the query range 16024-576 includes all database sequences that were typed for the entire control region. HVS-I/II data (16024-16365 73-340) are not included in such a query. It may therefore be conservative to also perform a query with standard HVS-I/II sequence ranges.

Below the tabular representation of the database query an interactive map can be found that depicts the sampled populations within the query range (blue) and the matches in the sampled populations (pink).

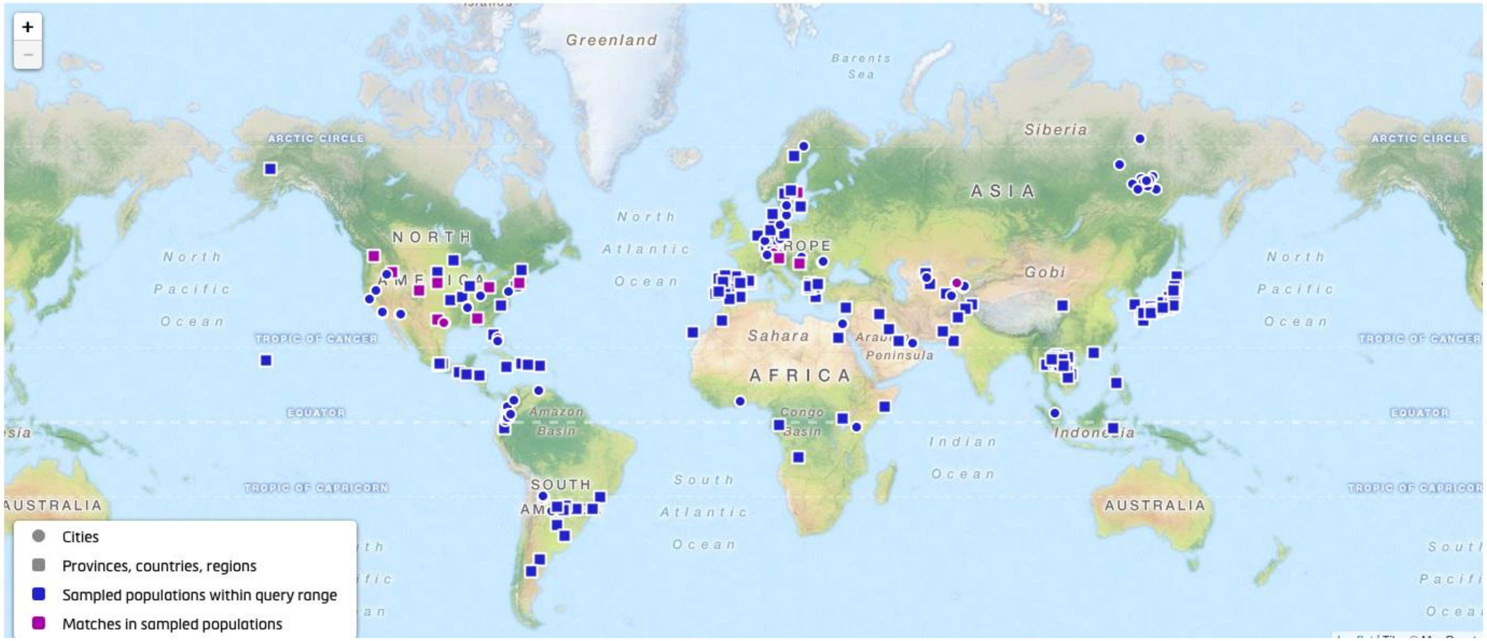
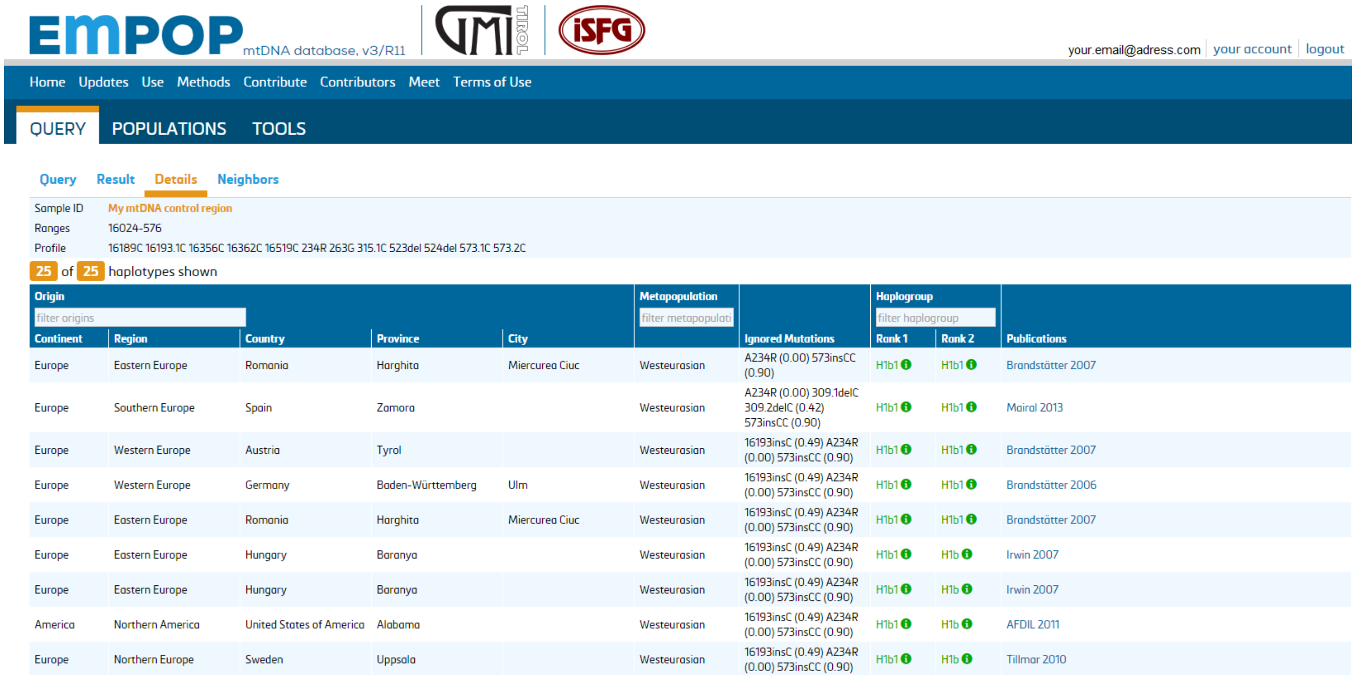


Figure 5 – Result Map

4.3. Details

The Details tab provides a more detailed presentation of the queried profile.

4.3.1. When matches are found



Sample ID My mtDNA control region

Ranges 16024-576

Profile 16189C 16193.1C 16356C 16362C 16519C 234R 263G 315.1C 523del 524del 573.1C 573.2C

25 of 25 haplotypes shown

Origin					Metapopulation		Haplogroup		
Continent	Region	Country	Province	City		Ignored Mutations	Rank 1	Rank 2	Publications
Europe	Eastern Europe	Romania	Harghita	Miercurea Ciuc	Westeurasian	A234R (0.00) 573insCC (0.90)	H1b1 ⓘ	H1b1 ⓘ	Brandstätter 2007
Europe	Southern Europe	Spain	Zamora		Westeurasian	A234R (0.00) 309.1delC 309.2delC (0.42) 573insCC (0.90)	H1b1 ⓘ	H1b1 ⓘ	Mairal 2013
Europe	Western Europe	Austria	Tyrol		Westeurasian	16193insC (0.49) A234R (0.00) 573insCC (0.90)	H1b1 ⓘ	H1b1 ⓘ	Brandstätter 2007
Europe	Western Europe	Germany	Baden-Württemberg	Ulm	Westeurasian	16193insC (0.49) A234R (0.00) 573insCC (0.90)	H1b1 ⓘ	H1b1 ⓘ	Brandstätter 2006
Europe	Eastern Europe	Romania	Harghita	Miercurea Ciuc	Westeurasian	16193insC (0.49) A234R (0.00) 573insCC (0.90)	H1b1 ⓘ	H1b1 ⓘ	Brandstätter 2007
Europe	Eastern Europe	Hungary	Baranya		Westeurasian	16193insC (0.49) A234R (0.00) 573insCC (0.90)	H1b1 ⓘ	H1b ⓘ	Irwin 2007
Europe	Eastern Europe	Hungary	Baranya		Westeurasian	16193insC (0.49) A234R (0.00) 573insCC (0.90)	H1b1 ⓘ	H1b ⓘ	Irwin 2007
America	Northern America	United States of America	Alabama		Westeurasian	16193insC (0.49) A234R (0.00) 573insCC (0.90)	H1b1 ⓘ	H1b ⓘ	AFDIL 2011
Europe	Northern Europe	Sweden	Uppsala		Westeurasian	16193insC (0.49) A234R (0.00) 573insCC (0.90)	H1b1 ⓘ	H1b ⓘ	Tillmar 2010

Figure 6 - Details

EMPOP provides a summary table of all matching haplotypes that meet the queried sequence range. Columns can be sorted by clicking on the column headers.

Geographic and **metapopulation** origins can be filtered using the text boxes.

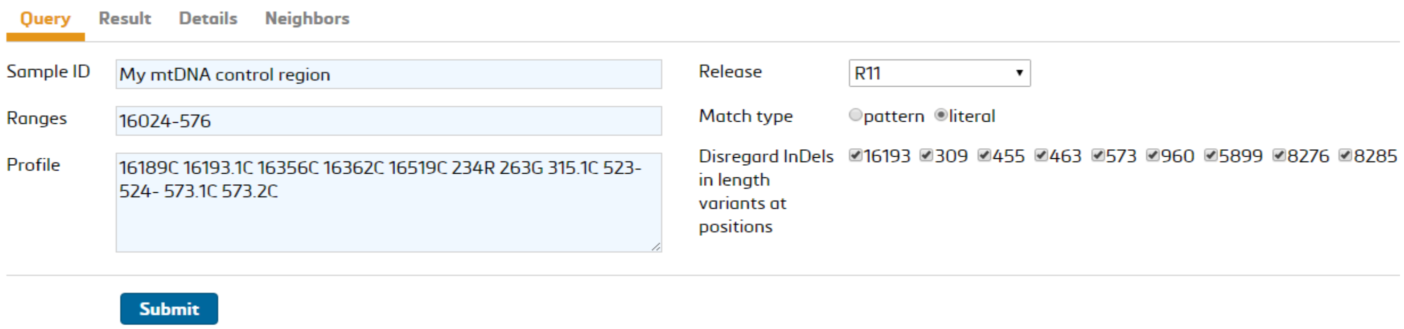
Ignored mutations list the differences between database and query sequences that were disregarded for the search (see 4.1.6. Disregard InDels in length variants at positions). The values in brackets display the costs of the listed mutation.

Haplogroup indicates the samples' haplogroup assignment. In case of a database match, there is no need to estimate the haplogroup why this column simply indicates the haplogroup of the matching samples. Rank 1 displays the haplogroup estimate with lowest costs (including a tolerance of 0.1) and Rank 2 displays the haplogroup estimate with the next lowest costs (including a tolerance of 0.1).

4.3.2. When no matches are found

MtDNA sequence queries often do not result in database matches. Besides statistical parameters (Results tab), EMPOP provides ad hoc haplogroup estimates that can be found under the Details tab. Haplogroups are estimated based on Phylotree and a curated database of full mtGenomes (currently approx. 20k haplotypes) using EMMA. The scientific background is detailed in Röck et al (2013).

Example:



Sample ID: My mtDNA control region

Ranges: 16024-576

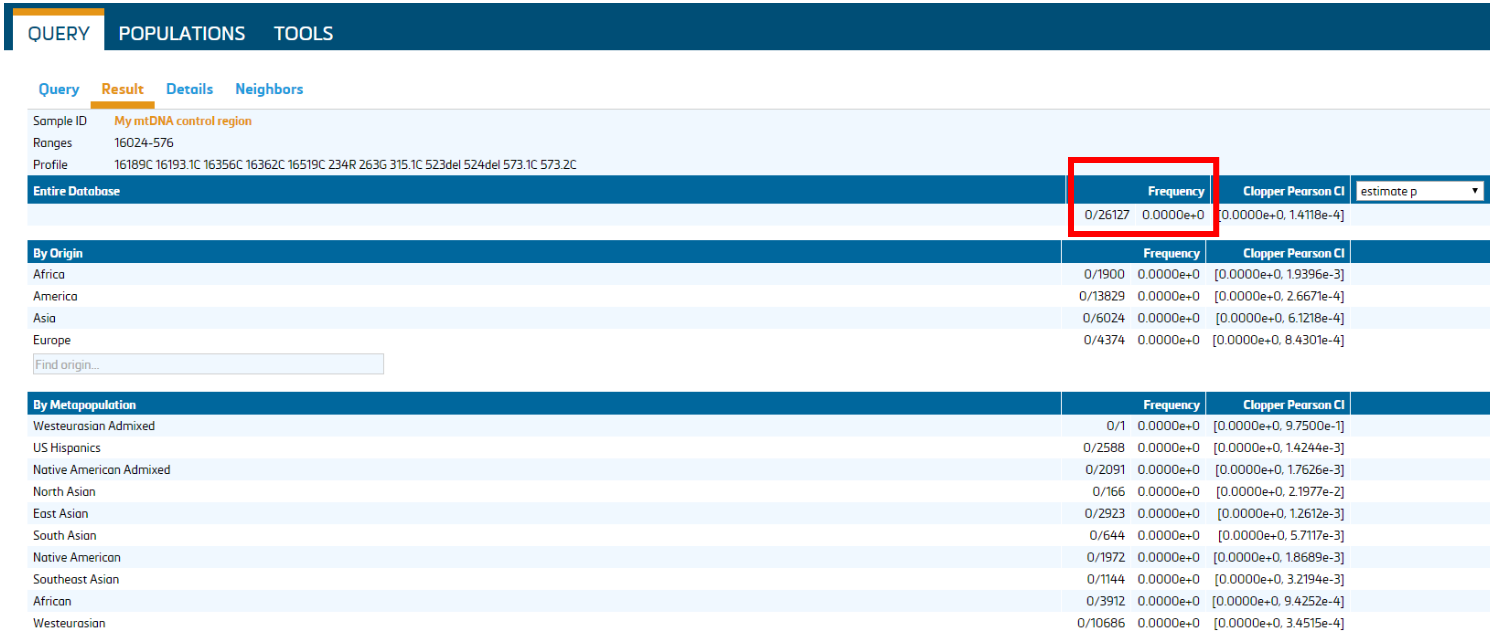
Profile: 16189C 16193.1C 16356C 16362C 16519C 234R 263G 315.1C 523- 524- 573.1C 573.2C

Release: R11

Match type: ○pattern ●literal

Disregard InDels in length variants at positions: ☑16193 ☑309 ☑455 ☑463 ☑573 ☑960 ☑5899 ☑8276 ☑8285

As a result no matching haplotypes were found which is indicated by a frequency value of “0”:



Sample ID My mtDNA control region
Ranges 16024-576
Profile 16189C 16193.1C 16356C 16362C 16519C 234R 263G 315.1C 523del 524del 573.1C 573.2C

Entire Database	Frequency		Clopper Pearson CI	estimate p ▾
	0/26127	0.0000e+0	[0.0000e+0, 1.4118e-4]	

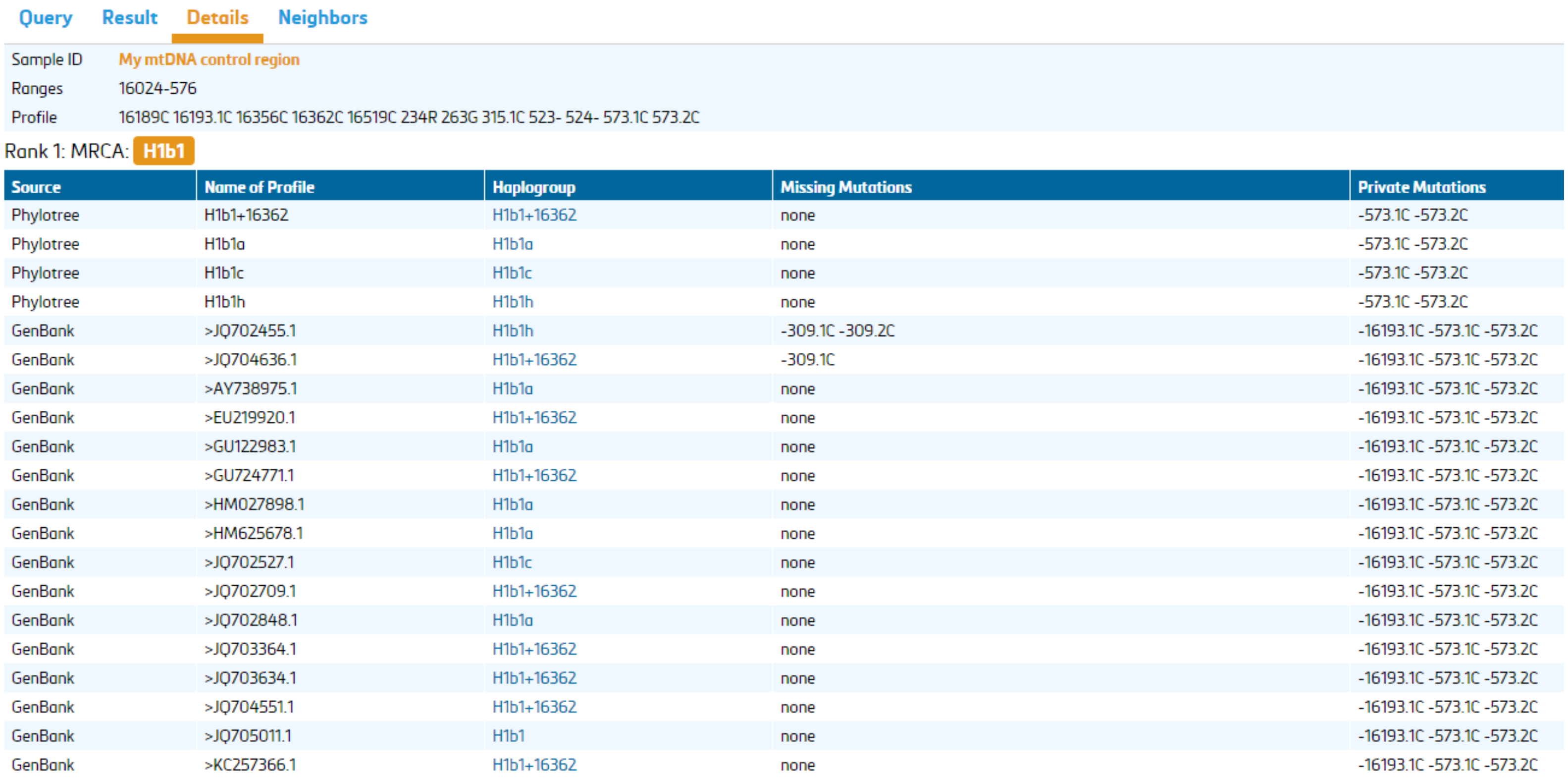
By Origin	Frequency		Clopper Pearson CI
Africa	0/1900	0.0000e+0	[0.0000e+0, 1.9396e-3]
America	0/13829	0.0000e+0	[0.0000e+0, 2.6671e-4]
Asia	0/6024	0.0000e+0	[0.0000e+0, 6.1218e-4]
Europe	0/4374	0.0000e+0	[0.0000e+0, 8.4301e-4]

Find origin...

By Metapopulation	Frequency		Clopper Pearson CI
Westeurasian Admixed	0/1	0.0000e+0	[0.0000e+0, 9.7500e-1]
US Hispanics	0/2588	0.0000e+0	[0.0000e+0, 1.4244e-3]
Native American Admixed	0/2091	0.0000e+0	[0.0000e+0, 1.7626e-3]
North Asian	0/166	0.0000e+0	[0.0000e+0, 2.1977e-2]
East Asian	0/2923	0.0000e+0	[0.0000e+0, 1.2612e-3]
South Asian	0/644	0.0000e+0	[0.0000e+0, 5.7117e-3]
Native American	0/1972	0.0000e+0	[0.0000e+0, 1.8689e-3]
Southeast Asian	0/1144	0.0000e+0	[0.0000e+0, 3.2194e-3]
African	0/3912	0.0000e+0	[0.0000e+0, 9.4252e-4]
Westeurasian	0/10686	0.0000e+0	[0.0000e+0, 3.4515e-4]

Figure 7 - Results view when no matches were found

In **Details** haplogroup results are displayed:

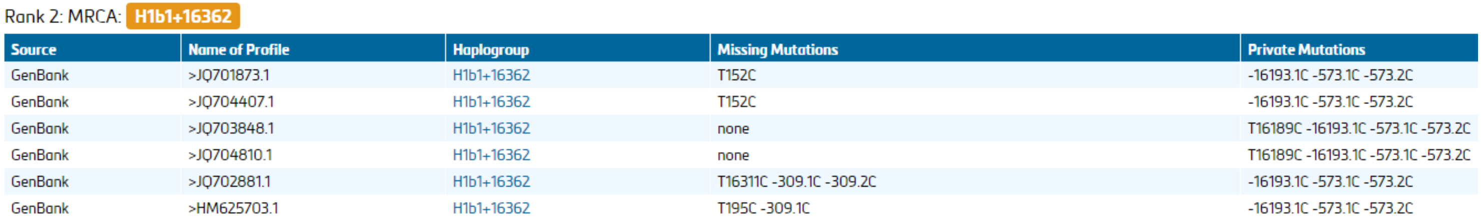


Sample ID	My mtDNA control region
Ranges	16024-576
Profile	16189C 16193.1C 16356C 16362C 16519C 234R 263G 315.1C 523- 524- 573.1C 573.2C

Rank 1: MRCA: H1b1

Source	Name of Profile	Haplogroup	Missing Mutations	Private Mutations
Phylotree	H1b1+16362	H1b1+16362	none	-573.1C -573.2C
Phylotree	H1b1a	H1b1a	none	-573.1C -573.2C
Phylotree	H1b1c	H1b1c	none	-573.1C -573.2C
Phylotree	H1b1h	H1b1h	none	-573.1C -573.2C
GenBank	>JQ702455.1	H1b1h	-309.1C -309.2C	-16193.1C -573.1C -573.2C
GenBank	>JQ704636.1	H1b1+16362	-309.1C	-16193.1C -573.1C -573.2C
GenBank	>AY738975.1	H1b1a	none	-16193.1C -573.1C -573.2C
GenBank	>EU219920.1	H1b1+16362	none	-16193.1C -573.1C -573.2C
GenBank	>GU122983.1	H1b1a	none	-16193.1C -573.1C -573.2C
GenBank	>GU724771.1	H1b1+16362	none	-16193.1C -573.1C -573.2C
GenBank	>HM027898.1	H1b1a	none	-16193.1C -573.1C -573.2C
GenBank	>HM625678.1	H1b1a	none	-16193.1C -573.1C -573.2C
GenBank	>JQ702527.1	H1b1c	none	-16193.1C -573.1C -573.2C
GenBank	>JQ702709.1	H1b1+16362	none	-16193.1C -573.1C -573.2C
GenBank	>JQ702848.1	H1b1a	none	-16193.1C -573.1C -573.2C
GenBank	>JQ703364.1	H1b1+16362	none	-16193.1C -573.1C -573.2C
GenBank	>JQ703634.1	H1b1+16362	none	-16193.1C -573.1C -573.2C
GenBank	>JQ704551.1	H1b1+16362	none	-16193.1C -573.1C -573.2C
GenBank	>JQ705011.1	H1b1	none	-16193.1C -573.1C -573.2C
GenBank	>KC257366.1	H1b1+16362	none	-16193.1C -573.1C -573.2C

Figure 8 – Haplogroup Estimation



Rank 2: MRCA: H1b1+16362

Source	Name of Profile	Haplogroup	Missing Mutations	Private Mutations
GenBank	>JQ701873.1	H1b1+16362	T152C	-16193.1C -573.1C -573.2C
GenBank	>JQ704407.1	H1b1+16362	T152C	-16193.1C -573.1C -573.2C
GenBank	>JQ703848.1	H1b1+16362	none	T16189C -16193.1C -573.1C -573.2C
GenBank	>JQ704810.1	H1b1+16362	none	T16189C -16193.1C -573.1C -573.2C
GenBank	>JQ702881.1	H1b1+16362	T16311C -309.1C -309.2C	-16193.1C -573.1C -573.2C
GenBank	>HM625703.1	H1b1+16362	T195C -309.1C	-16193.1C -573.1C -573.2C

Figure 9 – Haplogroup Estimation (continued)

In this view Rank 1 and Rank 2 MRCAs and candidates are listed in separate tables indicating the source, haplotype information, haplogroup affiliation, **missing mutations** (not present in query haplotype) and **private mutations** (not present in database haplotypes).

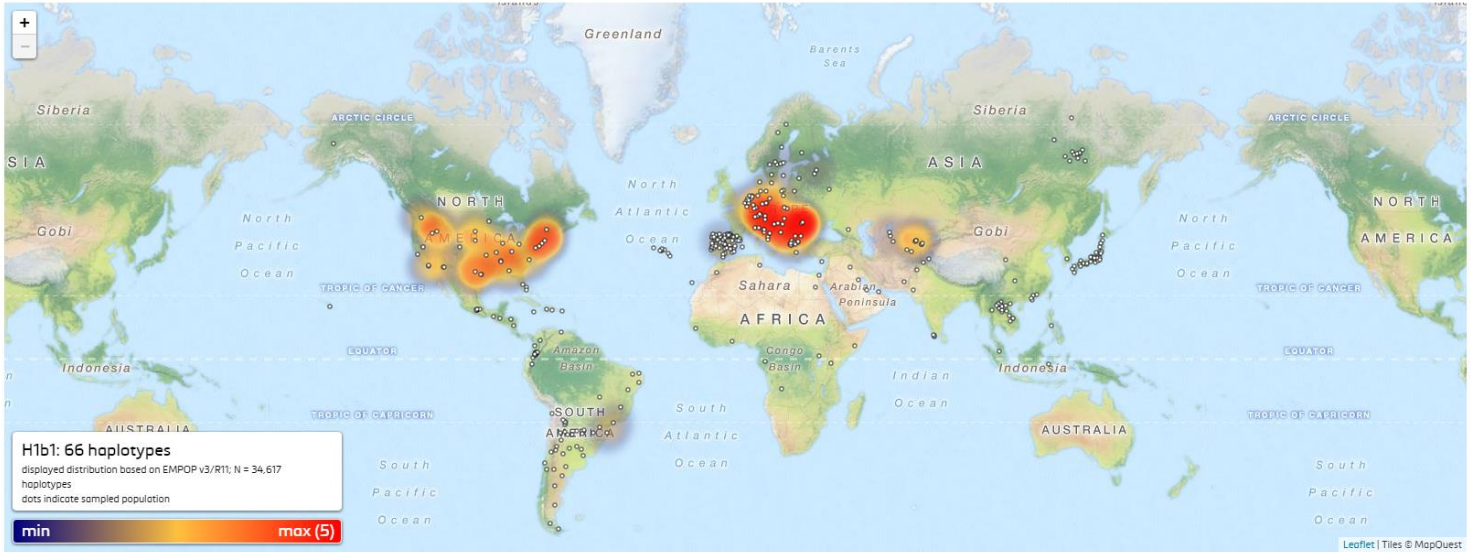
Haplotype names either correspond to a haplogroup designation from Phylotree (e.g. H1b1+16362) and then represents a Phylotree branch ("virtual haplotype", see Röck et al 2013) or relates to a GenBank entry (e.g. >JQ702455.1) which is indicated in column **Source**.

Rank 1 lists the Most Recent Common Ancestor (MRCA) of all haplogroup estimates with the minimum cost to the query haplotype (including a tolerance of 0.1).

Rank 2 presents the MRCA of all haplogroup estimates with the next lowest costs (including a tolerance of 0.1). Note that cost estimates are based on observed fluctuation rates in the haplotypes of the newest EMPOP release and the findings are based on the range of the query haplotype.

Mouse-clicks on the haplogroup affiliations lead to the depiction of the **geographical distribution** of the haplogroup based on EMPOP mtDNA sequences. Note that the distribution is depending on the coverage of the particular haplogroup in EMPOP (e.g. H1b1).

Note that further investigations on haplogroups and their distributions can be performed with Haplogroup Browser (see 6.1. Haplogroup Browser).



H1b1: 66 haplotypes

displayed distribution based on EMPOP v3/R11; N = 34,617 haplotypes

dots indicate sampled population

Figure 10 – Haplogroup Browser

4.3.3. Ambiguous haplogroup estimates

Partial mtDNA sequences that are often the result of highly degraded mtDNA encountered in forensic specimens may give ambiguous haplogroup estimates due to the lack of haplogroup-informative mutations. Also, control region sequences (or parts thereof) usually do not contain all information required to assign a single correct haplogroup. In these (and other) cases the assignment of a single haplogroup estimate may be biased. Instead the Most Recent Common Ancestor (MRCA) of all haplogroup estimates within a defined range (ranks) is reported. This proved to be a more conservative and thus stable approach to assign haplogroups in forensic genetics.

The quality of the MRCA estimate needs to be judged by the haplogroup diversity of the candidates that fall within the two ranks. Homogeneous haplogroup distribution within the rank candidates suggests that the MRCA estimate is unambiguous (see e.g. Figure 8 and Figure 9).
The following counter-example shows a broad MRCA estimate due to the ambiguity of the control region mutations.

Example:

The mtGenome (1-16569)

16184T 16298C 16519C 73G 263G 309.1C 315.1C 466C 750G 1438G 2706G 4580A 4639C 4769G 5263T 7028T 8860G 8869G 15326G 15904T

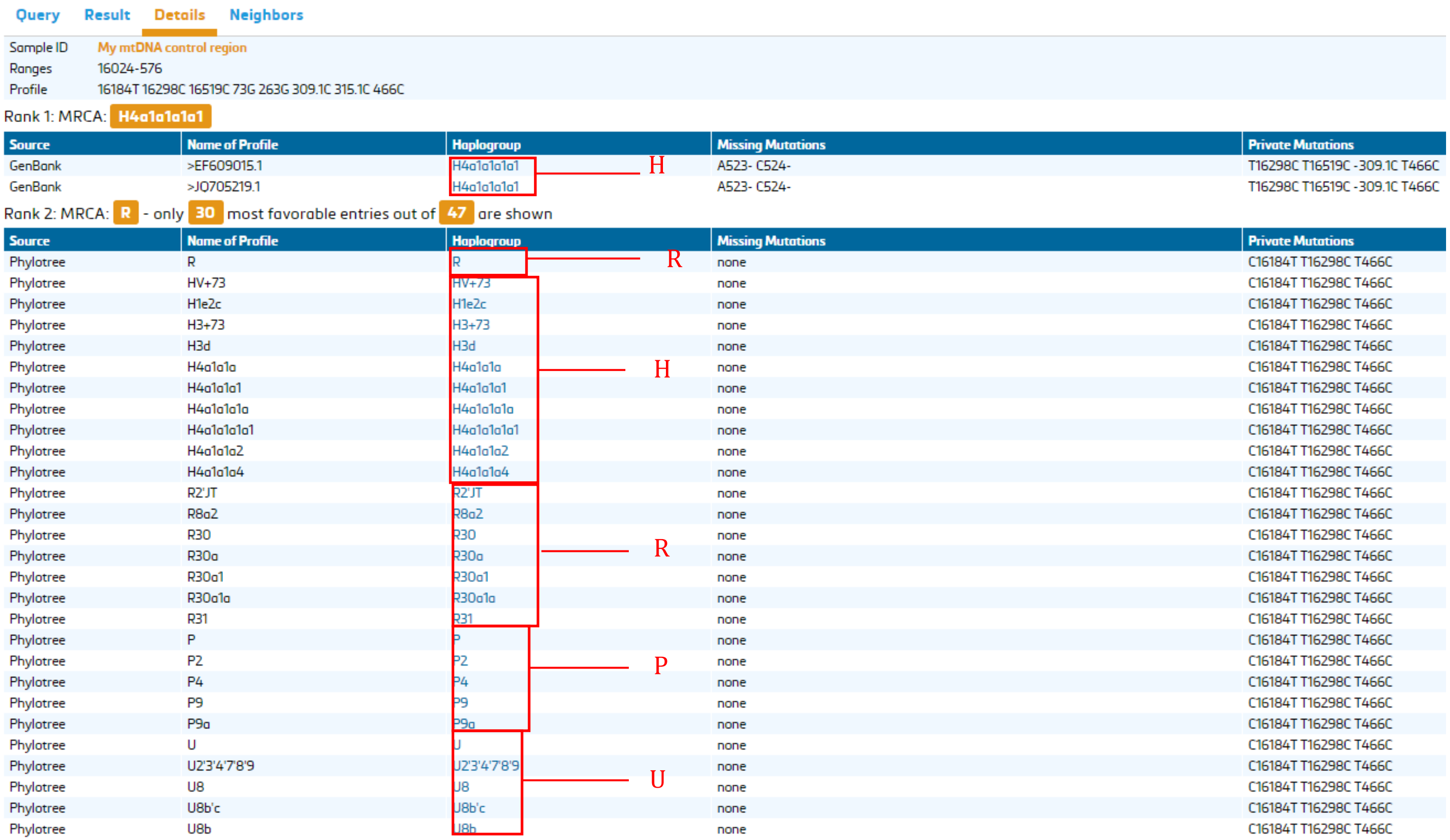
falls within haplogroup V1a1 as can be easily confirmed by querying this haplotype in EMPOP).

The control region (16024-576) of that example

16184T 16298C 16519C 73G 263G 309.1C 315.1C 466C

leads to MCRA estimates R for both ranks with candidates matching various haplogroups including HV, R, P and U (see Figure 11). The reason for this result is that this sequence is not included in the data basis for haplogroup estimation and that the closest neighbor to this sequence is relatively distant.

The HVS-1 segment (16024-16265) of this example leads to the even broader MRCA estimate L3 as more candidates fall within the defined ranks and thus contribute to the final estimate.



Sample ID	My mtDNA control region
Ranges	16024-576
Profile	16184T 16298C 16519C 73G 263G 309.1C 315.1C 466C

Rank 1: MRCA: H4a1a1a1a1

Source	Name of Profile	Haplogroup	Missing Mutations	Private Mutations
GenBank	>EF609015.1	H4a1a1a1a1	A523- C524-	T16298C T16519C -309.1C T466C
GenBank	>JQ705219.1	H4a1a1a1a1	A523- C524-	T16298C T16519C -309.1C T466C

Rank 2: MRCA: R - only 30 most favorable entries out of 47 are shown

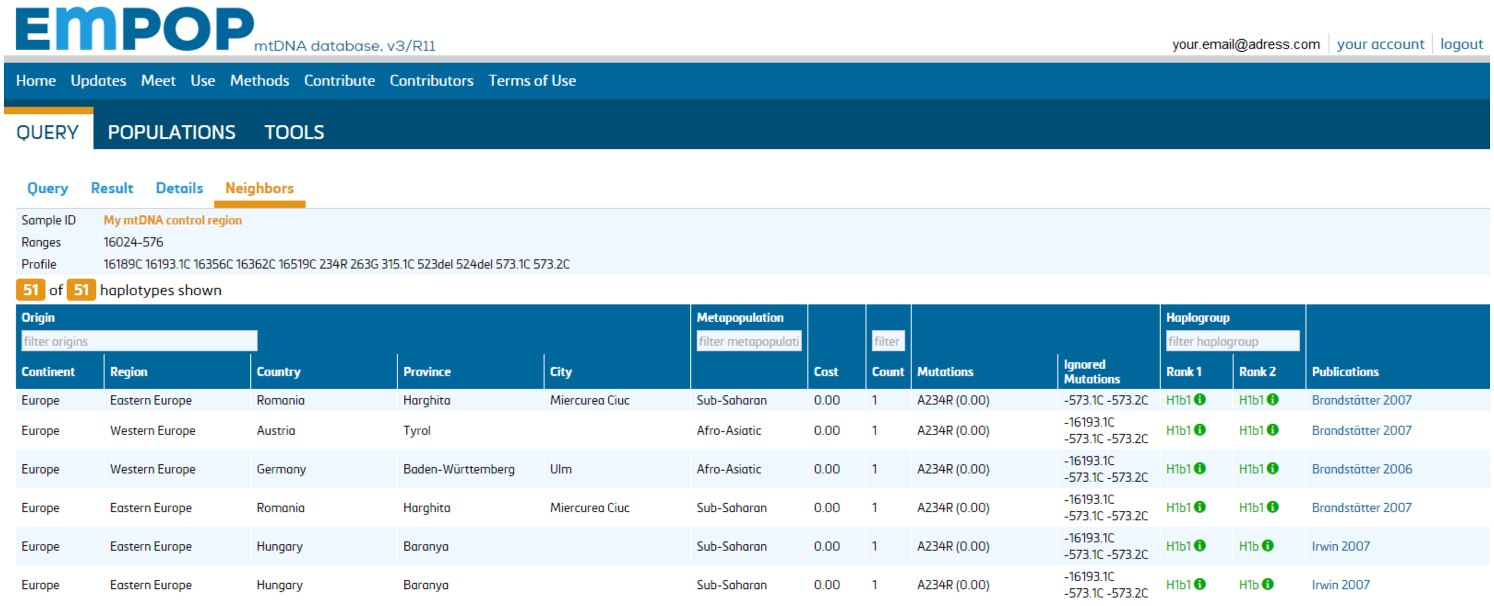
Source	Name of Profile	Haplogroup	Missing Mutations	Private Mutations
Phylotree	R	R	none	C16184T T16298C T466C
Phylotree	HV+73	HV+73	none	C16184T T16298C T466C
Phylotree	H1e2c	H1e2c	none	C16184T T16298C T466C
Phylotree	H3+73	H3+73	none	C16184T T16298C T466C
Phylotree	H3d	H3d	none	C16184T T16298C T466C
Phylotree	H4a1a1a	H4a1a1a	none	C16184T T16298C T466C
Phylotree	H4a1a1a1	H4a1a1a1	none	C16184T T16298C T466C
Phylotree	H4a1a1a1a	H4a1a1a1a	none	C16184T T16298C T466C
Phylotree	H4a1a1a1a1	H4a1a1a1a1	none	C16184T T16298C T466C
Phylotree	H4a1a1a2	H4a1a1a2	none	C16184T T16298C T466C
Phylotree	H4a1a1a4	H4a1a1a4	none	C16184T T16298C T466C
Phylotree	R2'JT	R2'JT	none	C16184T T16298C T466C
Phylotree	R8a2	R8a2	none	C16184T T16298C T466C
Phylotree	R30	R30	none	C16184T T16298C T466C
Phylotree	R30a	R30a	none	C16184T T16298C T466C
Phylotree	R30a1	R30a1	none	C16184T T16298C T466C
Phylotree	R30a1a	R30a1a	none	C16184T T16298C T466C
Phylotree	R31	R31	none	C16184T T16298C T466C
Phylotree	P	P	none	C16184T T16298C T466C
Phylotree	P2	P2	none	C16184T T16298C T466C
Phylotree	P4	P4	none	C16184T T16298C T466C
Phylotree	P9	P9	none	C16184T T16298C T466C
Phylotree	P9a	P9a	none	C16184T T16298C T466C
Phylotree	U	U	none	C16184T T16298C T466C
Phylotree	U2'3'4'7'8'9	U2'3'4'7'8'9	none	C16184T T16298C T466C
Phylotree	U8	U8	none	C16184T T16298C T466C
Phylotree	U8b'c	U8b'c	none	C16184T T16298C T466C
Phylotree	U8b	U8b	none	C16184T T16298C T466C

Figure 11 – Ambiguous haplogroup estimates for Rank 1

It is important to check the haplogroup distribution of the rank candidates in order to evaluate the quality of the MRCA estimate.

4.4. Neighbors

Similar sequences with a low number of differences are displayed here.



Sample ID My mtDNA control region

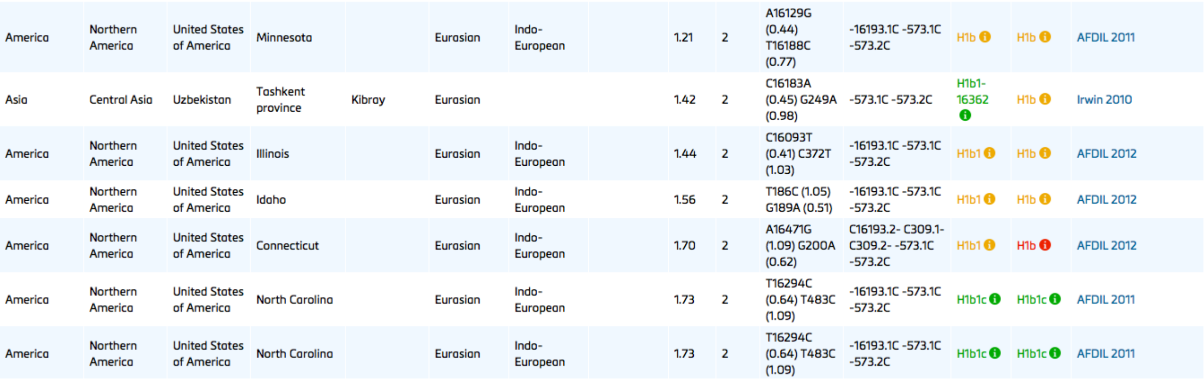
Ranges 16024-576

Profile 16189C 16193.1C 16356C 16362C 16519C 234R 263G 315.1C 523del 524del 573.1C 573.2C

51 of 51 haplotypes shown

Origin					Metapopulation					Haplogroup		
Continent	Region	Country	Province	City		Cost	Count	Mutations	Ignored Mutations	Rank 1	Rank 2	Publications
Europe	Eastern Europe	Romania	Harghita	Miercurea Ciuc	Sub-Saharan	0.00	1	A234R (0.00)	-573.1C -573.2C	H1b1 ℹ	H1b1 ℹ	Brandstätter 2007
Europe	Western Europe	Austria	Tyrol		Afro-Asiatic	0.00	1	A234R (0.00)	-16193.1C -573.1C -573.2C	H1b1 ℹ	H1b1 ℹ	Brandstätter 2007
Europe	Western Europe	Germany	Baden-Württemberg	Ulm	Afro-Asiatic	0.00	1	A234R (0.00)	-16193.1C -573.1C -573.2C	H1b1 ℹ	H1b1 ℹ	Brandstätter 2006
Europe	Eastern Europe	Romania	Harghita	Miercurea Ciuc	Sub-Saharan	0.00	1	A234R (0.00)	-16193.1C -573.1C -573.2C	H1b1 ℹ	H1b1 ℹ	Brandstätter 2007
Europe	Eastern Europe	Hungary	Baranya		Sub-Saharan	0.00	1	A234R (0.00)	-16193.1C -573.1C -573.2C	H1b1 ℹ	H1b ℹ	Irwin 2007
Europe	Eastern Europe	Hungary	Baranya		Sub-Saharan	0.00	1	A234R (0.00)	-16193.1C -573.1C -573.2C	H1b1 ℹ	H1b ℹ	Irwin 2007

…



America	Northern America	United States of America	Minnesota		Eurasian	Indo-European		1.21	2	A16129G (0.44) T16188C (0.77)	-16193.1C -573.1C -573.2C	H1b ⓘ	H1b ⓘ	AFDIL 2011
Asia	Central Asia	Uzbekistan	Tashkent province	Kibray	Eurasian			1.42	2	C16183A (0.45) G249A (0.98)	-573.1C -573.2C	H1b1-16362 ⓘ	H1b ⓘ	Irwin 2010
America	Northern America	United States of America	Illinois		Eurasian	Indo-European		1.44	2	C16093T (0.41) C372T (1.03)	-16193.1C -573.1C -573.2C	H1b1 ⓘ	H1b ⓘ	AFDIL 2012
America	Northern America	United States of America	Idaho		Eurasian	Indo-European		1.56	2	T186C (1.05) G189A (0.51)	-16193.1C -573.1C -573.2C	H1b1 ⓘ	H1b ⓘ	AFDIL 2012
America	Northern America	United States of America	Connecticut		Eurasian	Indo-European		1.70	2	A16471G (1.09) G200A (0.62)	C16193.2- C309.1- C309.2- -573.1C -573.2C	H1b1 ⓘ	H1b ⓘ	AFDIL 2012
America	Northern America	United States of America	North Carolina		Eurasian	Indo-European		1.73	2	T16294C (0.64) T483C (1.09)	-16193.1C -573.1C -573.2C	H1b1c ⓘ	H1b1c ⓘ	AFDIL 2011
America	Northern America	United States of America	North Carolina		Eurasian	Indo-European		1.73	2	T16294C (0.64) T483C (1.09)	-16193.1C -573.1C -573.2C	H1b1c ⓘ	H1b1c ⓘ	AFDIL 2011

Figure 12 - Neighbors

The display of match neighbors follows the same concept as the summary of matches (see 4.3. Details) and includes all haplotypes that are at a distance to the query sequence of **one and two differences (“events”)**.

An “Event” refers to the biological meaning of any difference but not the absolute number of differing nucleotides. As such, a tandem deletion (or insertion) in the AC-repeat region between 514 and 524 is regarded as one event, and therefore one difference between otherwise matching haplotypes. The same rationale applies to the 6 bp Chibcha deletion between 105 and 110 or 106 and 111, the 9 bp deletion between 8281 and 8290, as well as other (less abundant) block indels in the mitochondrial genome.

Additional information is provided with regard to differences between query haplotype and neighbors. These are listed in the columns **cost**, **count** and **mutations**.

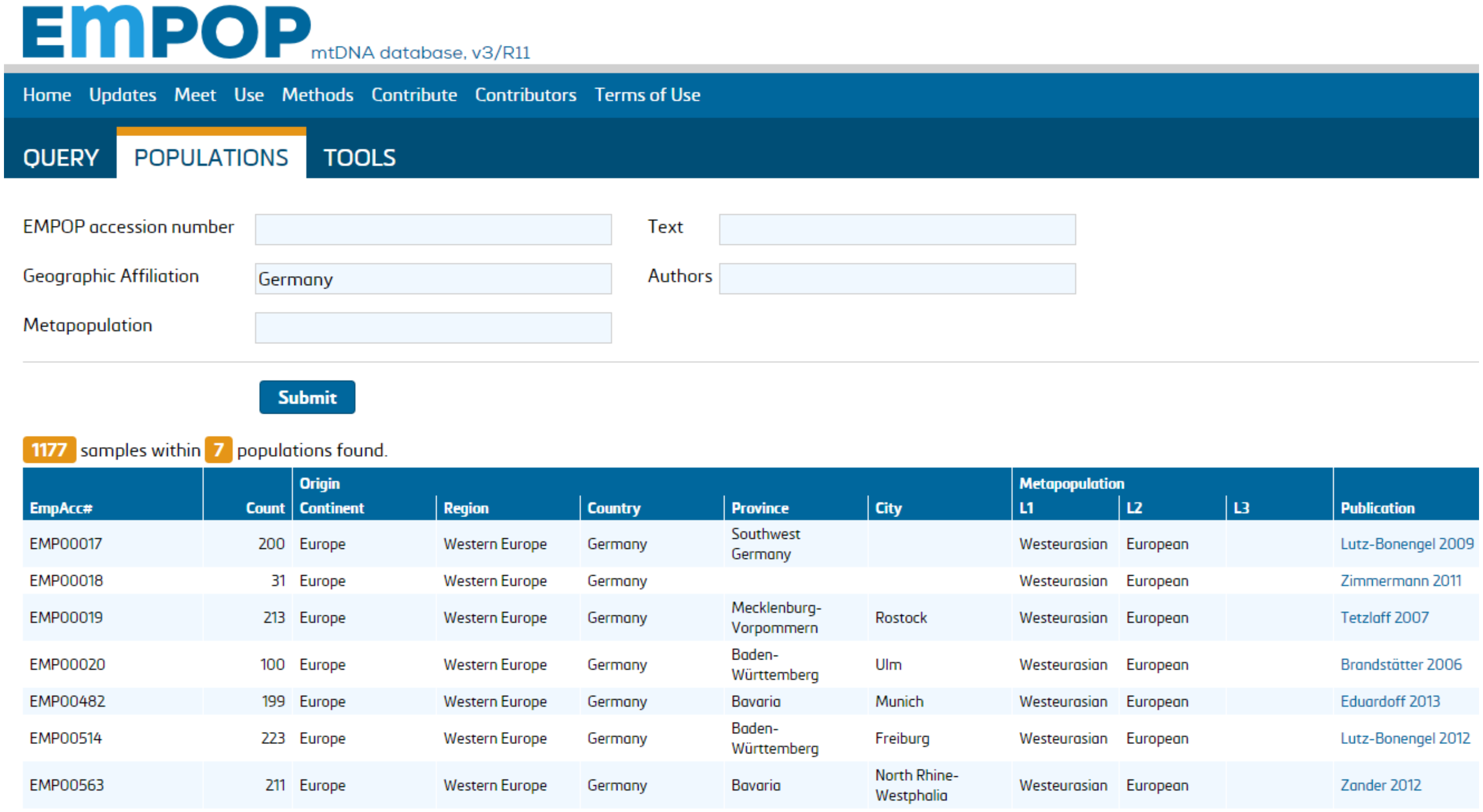
Costs are determined by the change from the base profile symbol to the test profile symbol (approximately 1.0 for an average mutation). See Röck et al. 2013 for further details.

Count lists the number of mutational events between query and database haplotypes.

Note that some combined mutations are single events, e.g. 523del 524del or 106-111del and treated as such in EMPOP.

Mutations specifies differences between query and database haplotypes which are listed with the individual costs. InDels which were disregarded do not contribute to the final costs.

5. Browsing EMPOP for populations



EMPOP mtDNA database, v3/R11

Home Updates Meet Use Methods Contribute Contributors Terms of Use
QUERY POPULATIONS TOOLS

EMPOP accession number:

Text:

Geographic Affiliation: Germany

Authors:

Metapopulation:

1177 samples within 7 populations found.

		Origin					Metapopulation			
EmpAcc#	Count	Continent	Region	Country	Province	City	L1	L2	L3	Publication
EMP00017	200	Europe	Western Europe	Germany	Southwest Germany		Westeurasian	European		Lutz-Bonengel 2009
EMP00018	31	Europe	Western Europe	Germany			Westeurasian	European		Zimmermann 2011
EMP00019	213	Europe	Western Europe	Germany	Mecklenburg-Vorpommern	Rostock	Westeurasian	European		Tetzlaff 2007
EMP00020	100	Europe	Western Europe	Germany	Baden-Württemberg	Ulm	Westeurasian	European		Brandstätter 2006
EMP00482	199	Europe	Western Europe	Germany	Bavaria	Munich	Westeurasian	European		Eduardoff 2013
EMP00514	223	Europe	Western Europe	Germany	Baden-Württemberg	Freiburg	Westeurasian	European		Lutz-Bonengel 2012
EMP00563	211	Europe	Western Europe	Germany	Bavaria	North Rhine-Westphalia	Westeurasian	European		Zander 2012

Figure 13 - Populations

Under the tab POPULATIONS the individual datasets contained in EMPOP can be found by using the accession number (if known), geographic or metapopulation affiliations. Published datasets can be searched by Text (Title) and Authors.

6. EMPOP Tools

The EMPOP tools section provides a suite of software to support the analysis and interpretation of mitochondrial DNA sequence variation.

6.1. Haplogroup Browser

Represents the established most recent Phylotree haplogroups in convenient searchable format and provides the number of EMPOP sequences assigned to the respective haplogroups by EMMA. Note that EMPOP provides the MRCA haplogroup if multiple haplogroup assignments are feasible.

Individual haplogroups can also be found by querying differences to the rCRS in a database of > 20.000 mtGenome sequences.

6.2. EMPcheck

EMPcheck is a tool to perform plausibility checks on a rCRS-coded data table.

The file format must meet the requirements described below and in Carracedo et al 2014.

6.2.1. Structure of the emp-file

Lines starting with "#!" indicate the sequence range of the haplotype. Note that a given sequence range is applied to all mtDNA haplotypes following this range until a new range is defined. Thus, multiple haplotypes with different sequence ranges can be handled in one file.
The file lists the haplotypes in columns with the following contents.

Column A: Sequence name: don't use blank space or special characters (allowed characters are letters (except umlauts ä, ö, ü), numbers, "-", "_", "/")

Column B: Haplogroup (hg) status: indicate hg, if unknown, use "?"

Column C: Frequency of haplotype (0 - 9999). Typically, this value is "1", as individual haplotypes should be presented. If it is set to 0 the sample is not considered for the analysis.

Column D: Annotation of the haplotype relative to the rCRS. Separate differences by tabs (or use individual cells in MS Excel). Use forensic notation of sequences as outlined in the revised and updated ISFG recommendations for mtDNA typing (Parson et al (2014)).

Text lines can be included everywhere in the file for comments or description. They need to be marked with "#".
Avoid blank lines (except when marked with "#").

The structure of an EMP file is illustrated below and the file can also be downloaded from: Downloads section in EMPOP.

Example

```
# Population data of 250 individuals from Austria; Walther Parson (walther.parson@i-med.ac.at)
# 100 samples from Innsbruck, 100 samples from Salzburg, 50 samples from Vienna
#! 16024-576
haplotype1                H1c   1          16519C 263G 523DEL 524DEL 477C
haplotype2                R0    2
#! 16024-16365 73-340
haplotype4                T2b   1          16126C 16294T 16296T 16304C 73G 263G 315.1C
haplotype5                ?     1          16223T 73G 263G 315.1C
```

6.3. NETWORK

This tool can be used to calculate and draw quasi-median networks. They are useful to examine the quality of an mtDNA dataset.

MtDNA data tables can be depicted as quasi-median networks to enhance the understanding of the data in regard to homoplasmy and potential artifacts. Highly recurrent mutations are removed from the dataset (filtering) to help detect data idiosyncrasies that pinpoint sequencing and data interpretation problems. A detailed discussion of the method can be found in Bandelt and Dür (2007) and its application in Parson and Dür (2007).

The following section leads you through

- the input and parameter selection of a network analysis
- the output generated by NETWORK
- network drawing and
- interpretation of the results.

6.3.1. Input

Sample Info

The sample-specific information identifies a search. This is also the reference under which the query is reported. The history of NETWORK searches can be found under YOUR ACCOUNT.

Input file (=emp file)

The input file contains the annotated population data. The emp-file is a tab delimited text file that can be created using standard text software or MS Excel (then, safe file under .txt format and rename "txt" by "emp"). Its format needs to meet the following criteria:

Ambiguous symbols

The software accepts the IUB-code. However, ambiguous symbols (e.g. sequence heteroplasmy Y ~ C/T) can cause artificial nodes and links in the network. Therefore it may be necessary to specify a non-ambiguous symbol either by calling the dominant type or by using the phylogenetic background of the sample. You will be notified on the presence of ambiguous symbols on the screen and in the network analysis report. For your information you also get a list of new insertions in your data set that are not known to the current EMPOP database.

New insertions

We collect positions with observed insertions in an EMPOP datafile to which new data are compared. New insertions that have not been recorded in EMPOP yet are displayed to draw the attention on them. This however does not impact the performance of NETWORK.

Filtering

Highly recurrent mutations are removed from the data set (filtering) that would otherwise increase the complexity of the network. You can choose between different filters depending on the application. The contents of the filters can be viewed by clicking on the symbol next to the dropdown box.

Available filters:

- EMPOPspeedy: This filter removes highly recurrent mutations based on the lists provided in Bandelt et al (2002 and 2006). This filter is typically used for the analysis of mtDNA population data within the hypervariable segments - HVS-I (16024 - 16569) and HVS-II (1 - 576).
- EMPOPspeedyWE: This filter removes highly recurrent mutations as presented in Zimmermann et al (2010). This filter is typically used for the analysis of west Eurasian mtDNA population data within the hypervariable segments - HVS-I (16024 - 16569) and HVS-II (1 - 576).
- EMPOPall_R11: This is a superfine filter that contains all mutations observed in EMPOP. This filter provides a very quick check on the data by highlighting only yet unobserved mutations. We update the EMPOPall filter periodically.
- Unfiltered: None of the mutations are removed from your dataset. This is useful for the analysis of very short sequence stretches in the mtDNA CR (see below). The complexity of the network will increase rapidly if no filter is applied to the analysis of larger sequence regions.

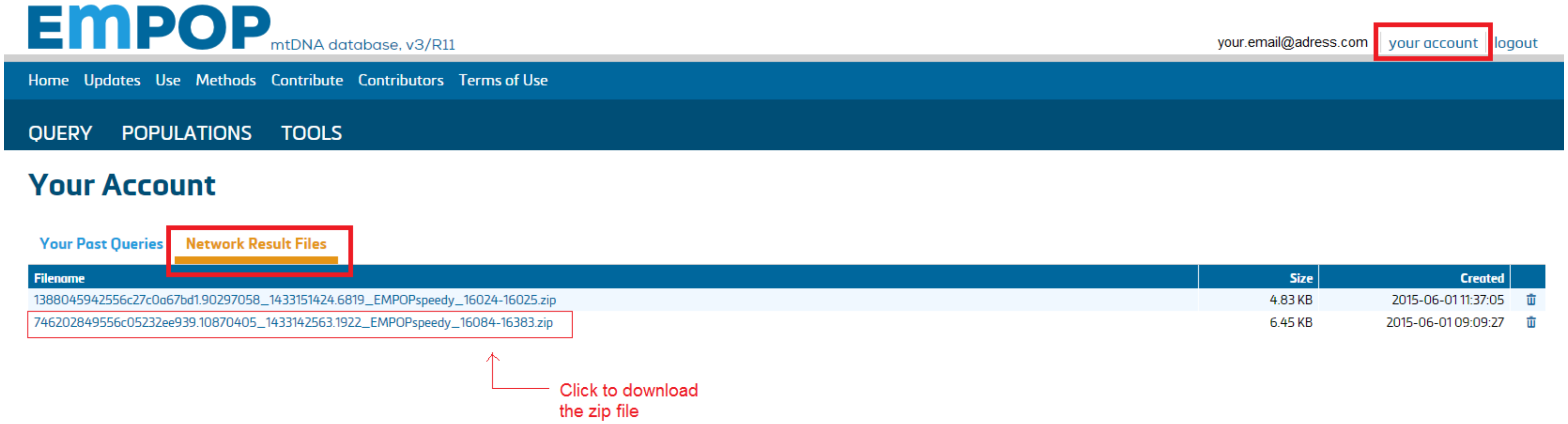
Range

The range determines the region for which the network is computed. Any range within 16024-16569 and 1-576 can be queried. In some data very small regions may be interesting for detailed network analysis (e.g. 450-460).

Submit starts the execution.

6.3.2. Output

After clicking on the **Submit** Button, the network calculation is initiated. Depending on the size of the file and the used filter options this process may take some time. When finished, result files will be listed in "Network Result Files" on the "your account" page.



Your Account

Filename	Size	Created	
1388045942556c27c0a67bd1.90297058_1433151424.6819_EMPOPspeedy_16024-16025.zip	4.83 KB	2015-06-01 11:37:05	🗑
746202849556c05232ee939.10870405_1433142563.1922_EMPOPspeedy_16084-16383.zip	6.45 KB	2015-06-01 09:09:27	🗑

Click to download the zip file

Figure 14 - Download a created Network Result File

Download the file and unzip it to obtain the folder *[RID_FILTERNAME_REGION]*, which contains the following files:

Results file [FILENAME_report.txt]: This file summarizes the settings and the results of the network analysis - for details see chapter Interpretation.

File for drawing the network [FILENAME_network.dnw]: This file can be used to draw the entire network of the mtDNA datafile by dnw.exe.

File for drawing the torso [FILENAME_torso.dnw]: This file can be used to draw the torso of the network of the mtDNA datafile by dnw.exe.

Difference table of the network [FILENAME_network.txt]: This file contains the filtered and reduced haplotypes of the entire network, displayed in dot table format.

Difference table of the torso [FILENAME_torso.txt]: This file contains the filtered and reduced haplotypes of the torso of the network, displayed in dot table format.

EMP-File [EMPFileName.emp]: The emp file which was uploaded

Info-file [FILENAME_info.txt]: Contains the sample identification (which was defined in "Sample info", see 6.3.1 Input) and the title of the emp file.

Drawing

1. Download the software for drawing the network (DrawNetWorkSetup.exe) from the EMPOP download page.
2. Execute the file and follow the instructions given by the software. Choose a destination folder where the software is to be installed.
3. Once the installation is finished you can find a folder called DrawNetWork containing the software and an uninstaller in the start menu. Files having ".dnw" as file ending are automatically linked to the software. Double-clicking a dnw file opens the network in a separate window. The help menu contains a legend of keys to edit the network (e.g. t ... for drawing a draft of the network, l ... for adding

labels, etc.). During execution the current drawing can be exported in SVG (Scalable Vector Graphic), EPS (Encapsulated PostScript) or GIF format for printing or editing.

Interpretation

The Report.txt file summarizes relevant information of the network analysis. The network is described in a table by the number of samples (n), the number of polymorphic positions (p), the number of partitions or condensed characters (p'), the number of haplotypes (h), the number of nodes in the network (q), the number of nodes in the torso (t) and the number of nodes of the peeled torso (t'). These values are indicative for the quality of a network. However, they depend on the size and composition of the population data set in question. Generally, small t'-values (ideally 1) describe a star-like structure of the network, which is in agreement with the expected evolutionary pattern.

A more suggestive representation of the data is the graph of the quasi-median network. The nodes of this graph are given by the haplotypes or the quasi-medians generated from the haplotypes. In the drawing the frequencies of the haplotypes or quasi-medians are also shown. The root node is drawn with a bold circle and contains the filtered and reduced Anderson sequence (In the rare case that no haplotype contains the filtered and reduced Anderson sequence, the first haplotype is chosen instead and a warning is included in the report). The links are single or combined mutations specified by the syntax for single mutations or / for combined mutations, where the orientation is from the root node outwards. Links with the same mutation are drawn parallel and are labeled only once. The torso is obtained from the quasi-median network by collapsing all pendant subtrees into their base nodes. Thus the analysis of homoplasmy can be restricted to the torso which contains all the reticulation of the network. For each base node the coinciding haplotypes are listed in the report to make it easy to find all corresponding samples.

Further reading

- Bandelt HJ et al (2002) The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71:1150-1160
- Bandelt HJ et al (2006) Estimation of mutation rates and coalescence times: some caveats. In: Human mitochondrial DNA and the evolution of Homo sapiens. Springer-Verlag eds. Hans-Jürgen Bandelt, Vincent Macaulay, Martin Richards
- Bandelt and Dür (2007) Translating DNA data tables into quasi-median networks for parsimony analysis and error detection. Mol Phylogenet Evol 42:256-271
- Parson and Dür (2007) EMPOP - A forensic mtDNA database. FSI:Genetics 1:88-92
- Schwarz and Dür (2011) Visualization of quasi-median networks. Discrete Applied Mathematics 159(15):1608-1616
- Zimmermann et al (2014) Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software. Croat Med J 55(2): 115-120.