mitoBench & mitoDB

A Peltzer & J Neukamm

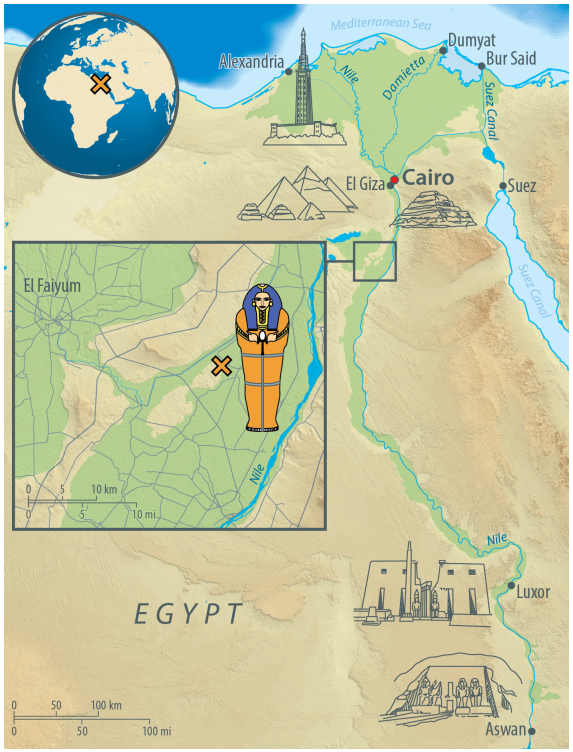
Integrative Transcriptomics, University of Tuebingen
INA, WG: Archaeo- and Paleogenetics, University of Tuebingen
Max-Planck Institute for the Science of Human History, Jena

February 23^{rd}, 2017

Overview

- Introduction & Motivation
- mitoBench
- mitoDB
- Outlook

Introduction: Overview



Intent: Study population history of Egypt

Source: National Geographic, June 2016

Introduction: Overview

- After mitochondrial (bead) capture: 90 individuals, authenticated (damage $10 - 20\%$), endogenous DNA between $14 - 60\%$
- Low Contamination (threshold: $< 3\%$)

Introduction: Overview

- Sequence based analysis: Calculate population differentiation (e.g. F_{st})
- Haplotype based analysis: Determine maternal Haplogroups (e.g. with HaploGrep2[1])

[1] Weissensteiner, Hansi, et al. *"HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing."* Nucleic acids research (2016)

Introduction: Overview

- What do we need?

Introduction: Overview

- What do we need?
  - Comparison data: Both sequence information and haplogroups

Introduction: Overview

- What do we need?
  - Comparison data: Both sequence information and haplogroups
  - Several file formats for downstream tools: ARP, XLSX, FastA, ...

Mitochondrial Genomics: Current status

- Not so much a problem of getting the data

Mitochondrial Genomics: Current status

- Not so much a problem of getting the data
  - Lots of people work on this (dozens of labs)

Mitochondrial Genomics: Current status

- Not so much a problem of getting the data
  - Lots of people work on this (dozens of labs)
  - Fairly cheap per genome - generate more data if you need more (if samples are available)

Mitochondrial Genomics: Current status

region	country		ID	region
Africa, central	Cammeroon		0985	Africa, central
	Central African Republic		0572	Africa, central
	Chad		0575	Africa, central
	Democratic Republic of the Congo		0594	Africa, central
	Equatorial Guinea		0573	Africa, central
	Gabon		0574	Africa, central
	Niger		0577	Africa, central
	São Tomé and Príncipe		0576	Africa, central
Africa, east	Egypt		0578	Africa, central
	Ethiopia		0579	Africa, central
	Kenya		0582	Africa, central
	Rwanda		0583	Africa, central
	Sudan		0584	Africa, central
Africa, north	Algerien		0580	Africa, central
	Lybia		0581	Africa, central

Source: Previous internal data source, credits to A Szécsényi-Nagy, G Brandt, W Haak and others

Mitochondrial Genomics: Current status

- More a problem of what to do with it ...

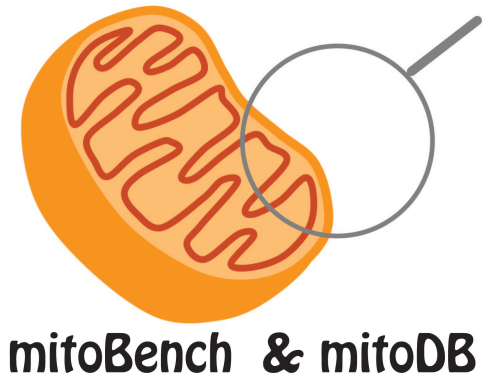
Mitochondrial Genomics: Current status

- More a problem of what to do with it ...
  - Manual conversions (e.g. from FastA to ARP, ...)

Mitochondrial Genomics: Current status

- More a problem of what to do with it ...
  - Manual conversions (e.g. from FastA to ARP, ...)
  - Collections are kept private (though publicly available data) - room for collaborative improvements
  - Lots of potential for error: Copy & Paste errors, inconsistent data sources, citing papers is difficult (e.g. where was this sample published again?)

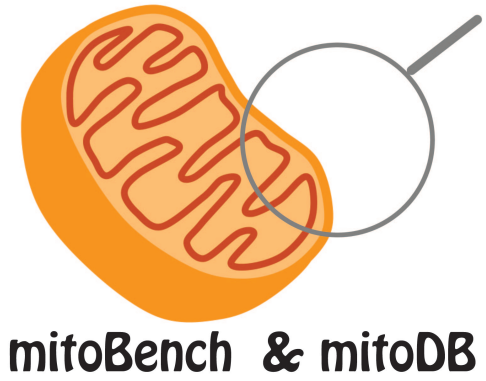
mitoBench: Introduction



Idea: Mitochondrial Workbench

- Collect mitochondrial data

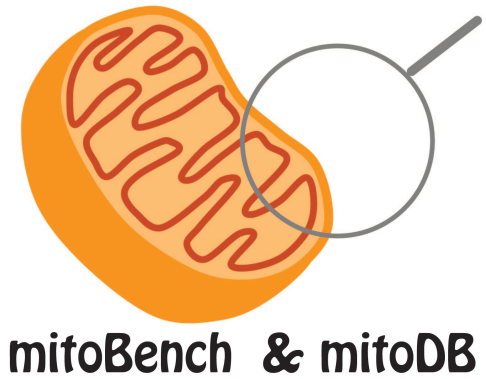
mitoBench: Introduction



Idea: Mitochondrial Workbench

- Collect mitochondrial data
- Ease file handling and file conversion

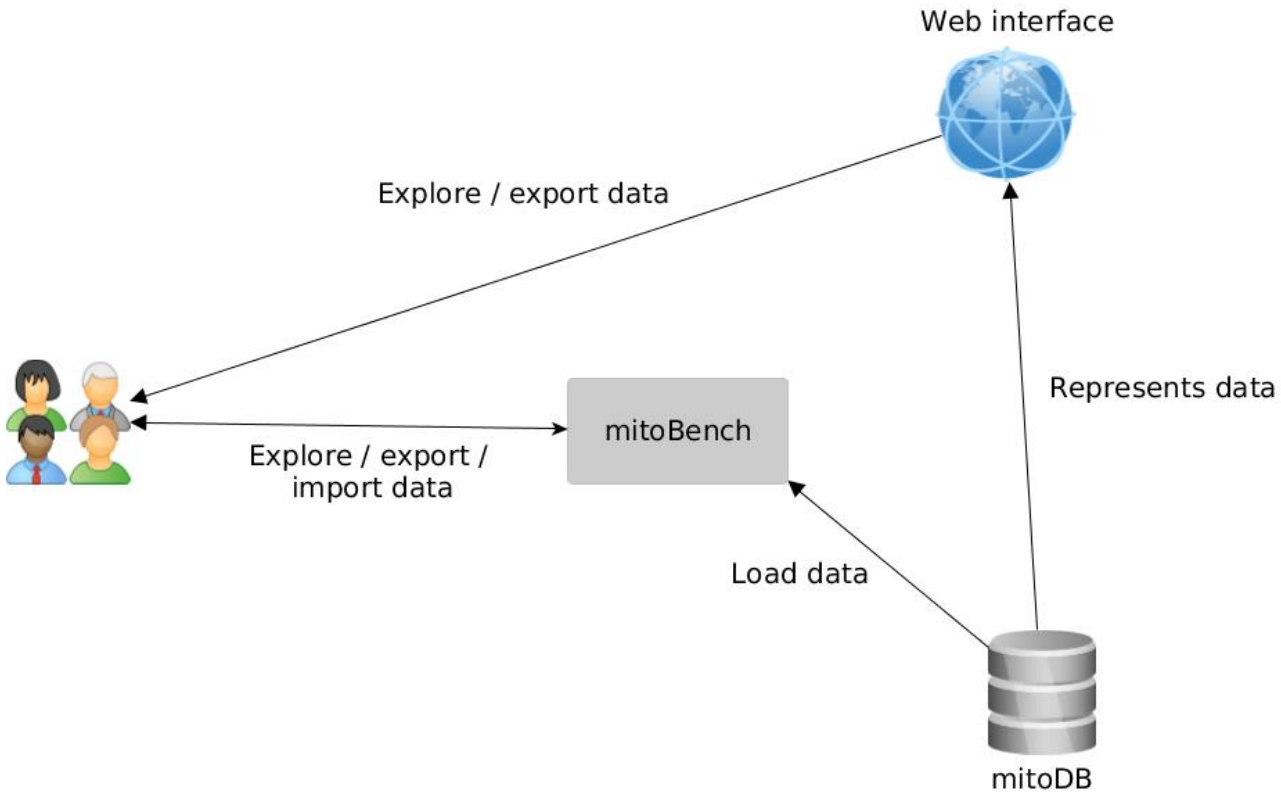
mitoBench: Introduction



Idea: Mitochondrial Workbench

- Collect mitochondrial data
- Ease file handling and file conversion
- Combine, manipulate and visualize MT data

mitoBench: Workflow

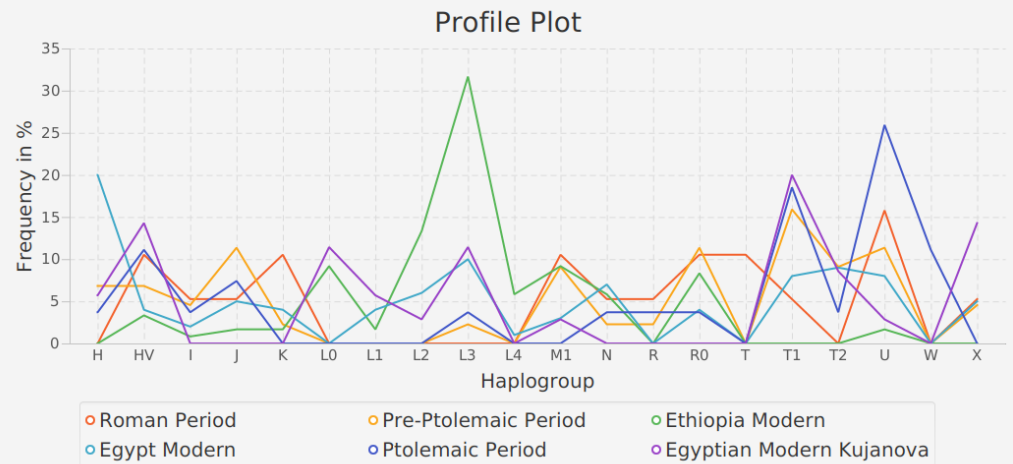


mitoBench: Main view

File Edit Grouping Statistics Table Graphics Help

▼ Haplo tree

Profile Plot ×



Count statistics ×

Groups	Total Number	H	HV	I	J	K	L0	L1	L2	L3	L4	M1	N	R	R0	T	T1
Roman Period	19	0	2	1	1	2	0	0	0	0	0	2	1	1	2	2	1
Pre-Ptolemaic Period	44	3	3	2	5	1	0	0	0	1	0	4	1	1	5	0	7
Ethiopia Modern	120	0	4	1	2	2	11	2	16	38	7	11	7	0	10	0	0
Egypt Modern	100	20	4	2	5	4	0	4	6	10	1	3	7	0	4	0	8
Ptolemaic Period	27	1	3	1	2	0	0	0	0	1	0	0	1	1	1	0	5
Egyptian Modern Kujanova	35	2	5	0	0	0	4	2	1	4	0	1	0	0	0	0	7

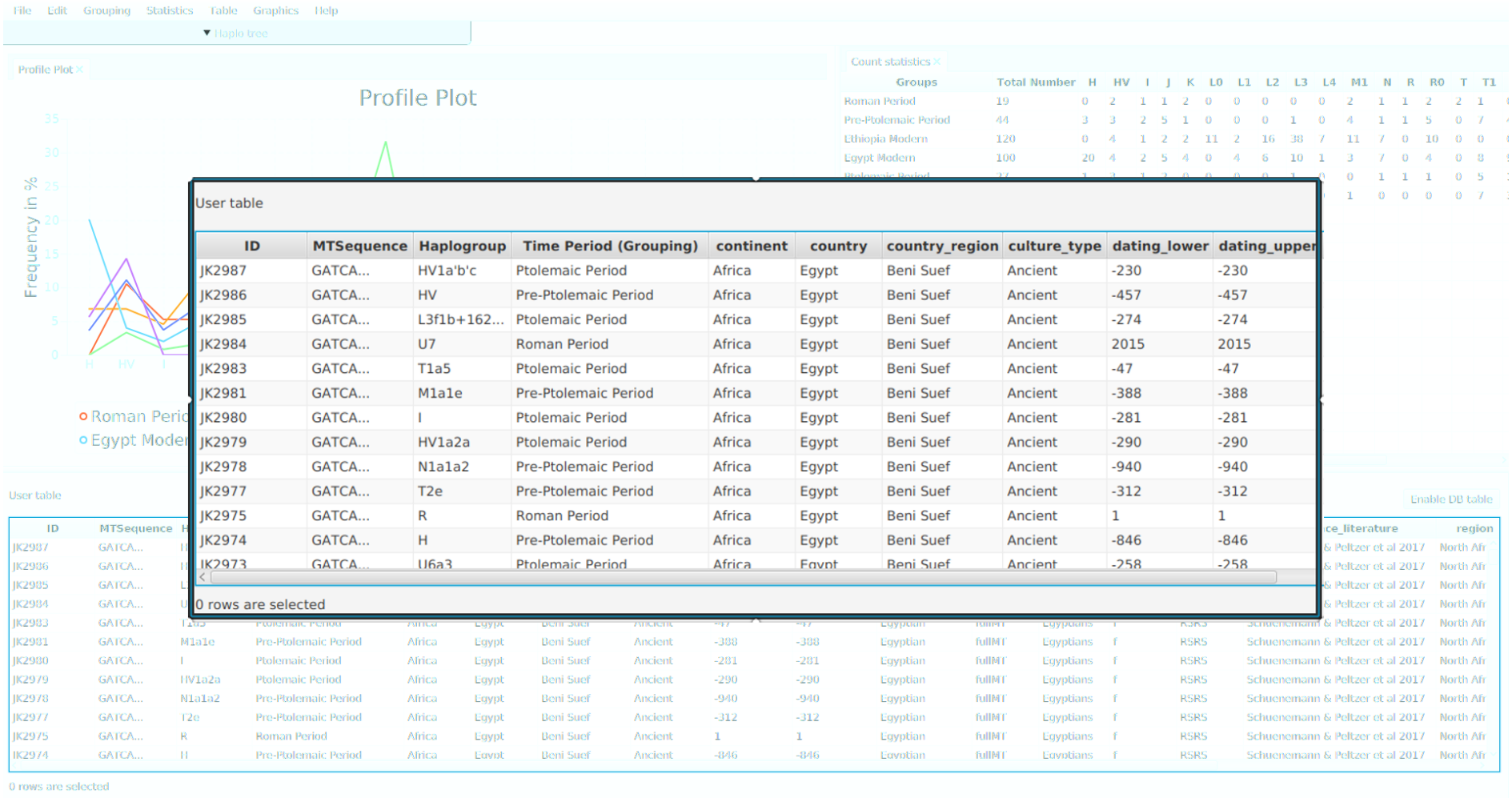
User table

Enable DB table

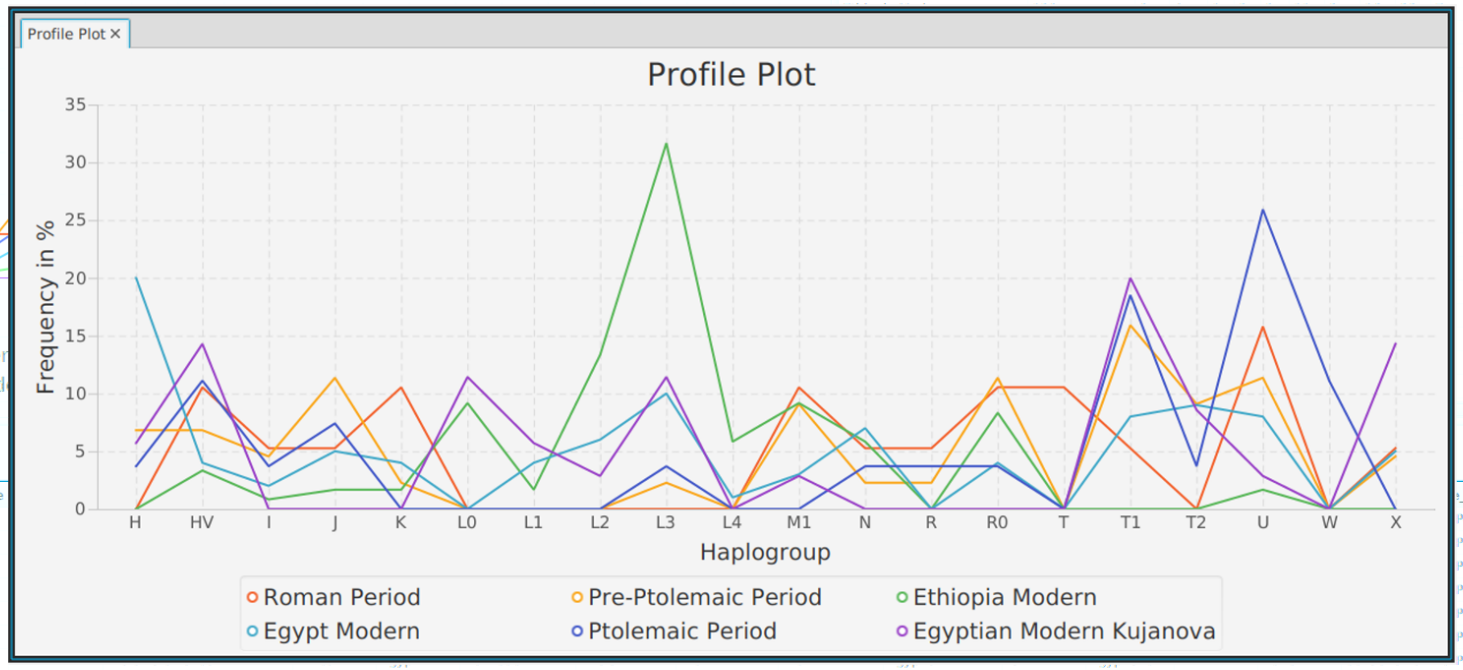
ID	MTSequence	Haplogroup	Time Period (Grouping)	continent	country	country_region	culture_type	dating_lower	dating_upper	language	mito_type	population	published	reference	reference_literature	region
JK2987	GATCA...	HV1a'b'c	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-230	-230	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2986	GATCA...	HV	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-457	-457	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2985	GATCA...	L3f1b+162...	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-274	-274	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2984	GATCA...	U7	Roman Period	Africa	Egypt	Beni Suef	Ancient	2015	2015	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2983	GATCA...	T1a5	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-47	-47	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2981	GATCA...	M1a1e	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-388	-388	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2980	GATCA...	I	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-281	-281	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2979	GATCA...	HV1a2a	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-290	-290	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2978	GATCA...	N1a1a2	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-940	-940	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2977	GATCA...	T2e	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-312	-312	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2975	GATCA...	R	Roman Period	Africa	Egypt	Beni Suef	Ancient	1	1	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr
JK2974	GATCA...	H	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-846	-846	Egyptian	fullMT	Egyptians	f	RSRS	Schuenemann & Peltzer et al 2017	North Afr

0 rows are selected

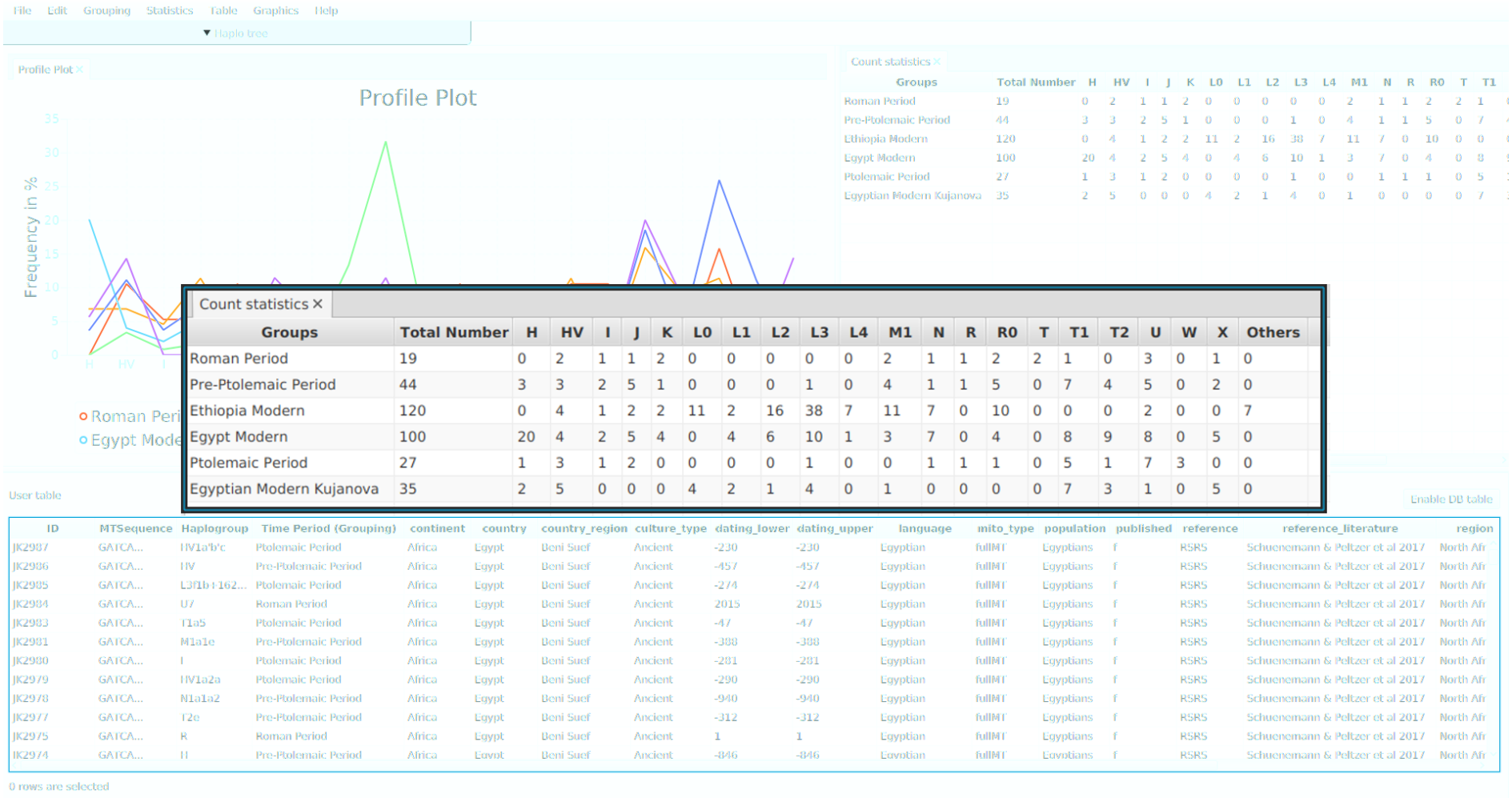
mitoBench: Main view



mitoBench: Main view



mitoBench: Main view



mitoBench: Data import and conversion

Accepted file formats

- (Multi-) FastA

[2] Excoffier, Laurent, and Heidi EL Lischer. *"Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows."* Molecular ecology resources 10.3 (2010): 564-567.

mitoBench: Data import and conversion

Accepted file formats

- (Multi-) FastA
- ARP (Arlequin format)[2]

[2] Excoffier, Laurent, and Heidi EL Lischer. *"Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows."* Molecular ecology resources 10.3 (2010): 564-567.

mitoBench: Data import and conversion

Accepted file formats

- (Multi-) FastA
- ARP (Arlequin format)[2]
- HSD (Haplogrep 2 format)

[2] Excoffier, Laurent, and Heidi EL Lischer. *"Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows."* Molecular ecology resources 10.3 (2010): 564-567.

mitoBench: Data import and conversion

Accepted file formats

- (Multi-) FastA
- ARP (Arlequin format)[2]
- HSD (Haplogrep 2 format)
- Excel file

[2] Excoffier, Laurent, and Heidi EL Lischer. *"Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows."* Molecular ecology resources 10.3 (2010): 564-567.

mitoBench: Data import and conversion

Accepted file formats

- (Multi-) FastA
- ARP (Arlequin format)[2]
- HSD (Haplogrep 2 format)
- Excel file
- Generic formats (e.g. dating information, tsv/csv format)

[2] Excoffier, Laurent, and Heidi EL Lischer. *"Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows."* Molecular ecology resources 10.3 (2010): 564-567.

mitoBench: Data import and conversion

Import from database

- Connect to mitoDB via mitoBench

mitoBench: Data import and conversion

Import from database

- Connect to mitoDB via mitoBench
- Specify query

mitoBench: Data import and conversion

Import from database

- Connect to mitoDB via mitoBench
- Specify query
- Add data to user table

mitoBench: Data import and conversion (Mummies)

- Load sequences (via fasta file)

User table

ID	MTSequence
JK2987	GATCA...
JK2986	GATCA...
JK2985	GATCA...
JK2984	GATCA...
JK2983	GATCA...
JK2981	GATCA...
JK2980	GATCA...
JK2979	GATCA...
JK2978	GATCA...
JK2977	GATCA...
JK2975	GATCA...
JK2974	GATCA...
JK2973	GATCA...
JK2972	GATCA...
JK2970	GATCA...
JK2966	GATCA...
JK2965	GATCA...
JK2963	GATCA...
JK2962	GATCA...
JK2961	GATCA...
JK2960	GATCA...

0 rows are selected

mitoBench: Data import and conversion (Mummies)

- Load Haplogroups (via hsd file)

User table

ID	MTSequence	Haplogroup
JK2880	GATCA...	T1a2
egypt.5AJ129	GATCA...	X1c
egypt.6AJ68	GATCA...	H3w
egpg5305866	GATCA...	L3f2a1
egypt.6AJ63	GATCA...	R0a3
egpg5305863	GATCA...	M1a5
egpg5305862	GATCA...	R0a2
egpg5306038	GATCA...	R0a2g
egpg5306032	GATCA...	L0f2b
egpg5306031	GATCA...	L3i2
egpg5306030	GATCA...	L2a1d1
egypt.7AJ139	GATCA...	I6b
JK2879	GATCA...	U3b
JK2878	GATCA...	T1a7
egypt.8AL34	GATCA...	H20c
JK2876	GATCA...	T1a8a
JK2875	GATCA...	N
JK2874	GATCA...	U
JK2873	GATCA...	T2
JK2872	GATCA...	HV1a2a
egypt.7AJ136	GATCA...	L1c3a1b

0 rows are selected

mitoBench: Data import and conversion (Mummies)

- Load generic file (via tsv file)

User table

ID	MTSequence	Haplogroup	Time Period	continent	country	country_region	culture_type	dating_lower
JK2880	GATCA...	T1a2	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-669
egypt.5AJ129	GATCA...	X1c	Egypt Modern	Africa	Egypt	Other	Modern	2015
egypt.6AJ68	GATCA...	H3w	Egypt Modern	Africa	Egypt	Other	Modern	2015
egpg5305866	GATCA...	L3f2a1	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egypt.6AJ63	GATCA...	R0a3	Egypt Modern	Africa	Egypt	Other	Modern	2015
egpg5305863	GATCA...	M1a5	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg5305862	GATCA...	R0a2	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg5306038	GATCA...	R0a2g	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg5306032	GATCA...	L0f2b	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg5306031	GATCA...	L3i2	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egpg5306030	GATCA...	L2a1d1	Ethiopia Modern	Africa	Ethiopia	Other	Modern	2015
egypt.7AJ139	GATCA...	I6b	Egypt Modern	Africa	Egypt	Other	Modern	2015
JK2879	GATCA...	U3b	Roman Period	Africa	Egypt	Beni Suef	Ancient	2015
JK2878	GATCA...	T1a7	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-235
egypt.8AL34	GATCA...	H20c	Egypt Modern	Africa	Egypt	Other	Modern	2015

mitoBench: Data representation

- Data represented in table format

mitoBench: Data representation

- Data represented in table format
- Merging data in a single table

mitoBench: Data grouping

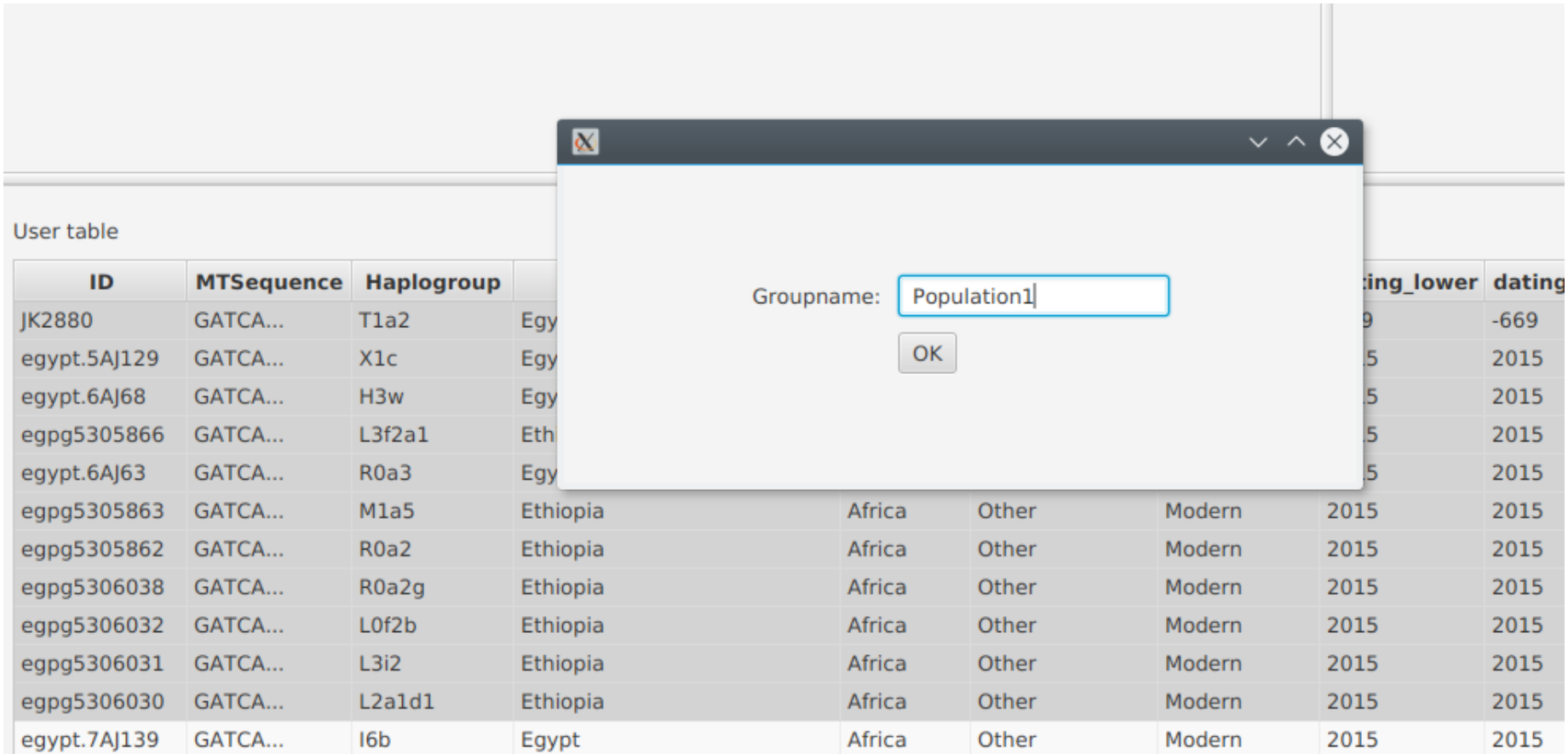
- interested in behavior of different groups
  $\rightarrow$ Group data by feature

mitoBench: Data grouping

- interested in behavior of different groups
  $\rightarrow$ Group data by feature
- internal grouping (columns not sorted)

mitoBench: Data grouping

User-specified groups



ID	MTSequence	Haplogroup					ing_lower	dating
JK2880	GATCA...	T1a2	Egy				9	-669
egypt.5AJ129	GATCA...	X1c	Egy				5	2015
egypt.6AJ68	GATCA...	H3w	Egy				5	2015
egpg5305866	GATCA...	L3f2a1	Eth				5	2015
egypt.6AJ63	GATCA...	R0a3	Egy				5	2015
egpg5305863	GATCA...	M1a5	Ethiopia	Africa	Other	Modern	2015	2015
egpg5305862	GATCA...	R0a2	Ethiopia	Africa	Other	Modern	2015	2015
egpg5306038	GATCA...	R0a2g	Ethiopia	Africa	Other	Modern	2015	2015
egpg5306032	GATCA...	L0f2b	Ethiopia	Africa	Other	Modern	2015	2015
egpg5306031	GATCA...	L3i2	Ethiopia	Africa	Other	Modern	2015	2015
egpg5306030	GATCA...	L2a1d1	Ethiopia	Africa	Other	Modern	2015	2015
egypt.7AJ139	GATCA...	I6b	Egypt	Africa	Other	Modern	2015	2015

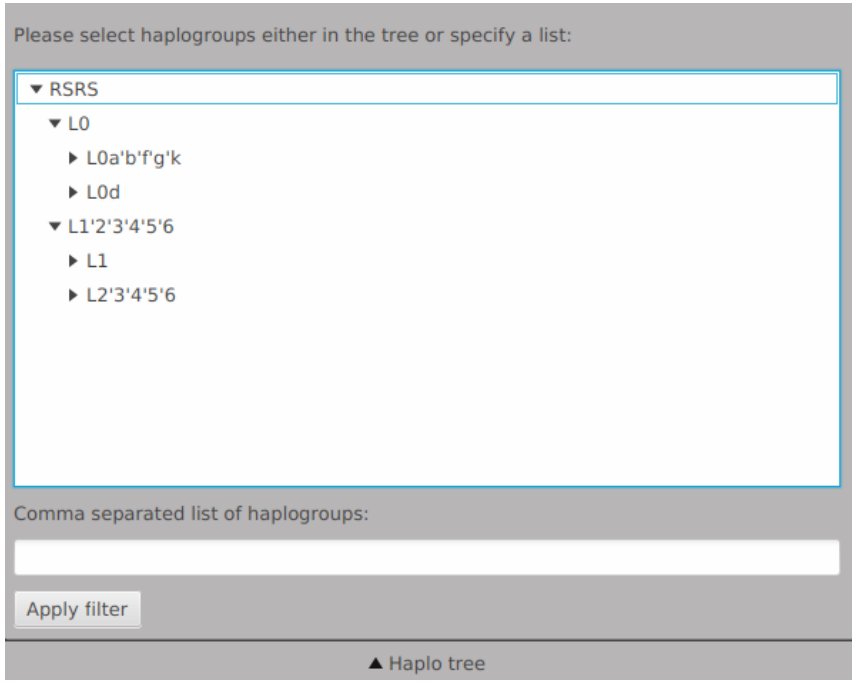
mitoBench: Data grouping

Groups by column

User table

ID	MTSequence	Haplogroup	Time Period (Grouping)	continent	country	language
JK2987	GATCA...	HV1a'b'c	Ptolemaic Period	Africa	Egypt	Egyptian
JK2986	GATCA...	HV	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2985	GATCA...	L3f1b+162...	Ptolemaic Period	Africa	Egypt	Egyptian
JK2984	GATCA...	U7	Roman Period	Africa	Egypt	Egyptian
JK2983	GATCA...	T1a5	Ptolemaic Period	Africa	Egypt	Egyptian
JK2981	GATCA...	M1a1e	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2980	GATCA...	I	Ptolemaic Period	Africa	Egypt	Egyptian
JK2979	GATCA...	HV1a2a	Ptolemaic Period	Africa	Egypt	Egyptian
JK2978	GATCA...	N1a1a2	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2977	GATCA...	T2e	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2975	GATCA...	R	Roman Period	Africa	Egypt	Egyptian
JK2974	GATCA...	H	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2973	GATCA...	U6a3	Ptolemaic Period	Africa	Egypt	Egyptian
JK2972	GATCA...	T1a5	Ptolemaic Period	Africa	Egypt	Egyptian
JK2970	GATCA...	U1a1	Ptolemaic Period	Africa	Egypt	Egyptian
JK2966	GATCA...	T1a7	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2965	GATCA...	T2c1c	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2963	GATCA...	M1a1i	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2962	GATCA...	H13c1	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2961	GATCA...	T1a7	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2960	GATCA...	N1'5	Ptolemaic Period	Africa	Egypt	Egyptian

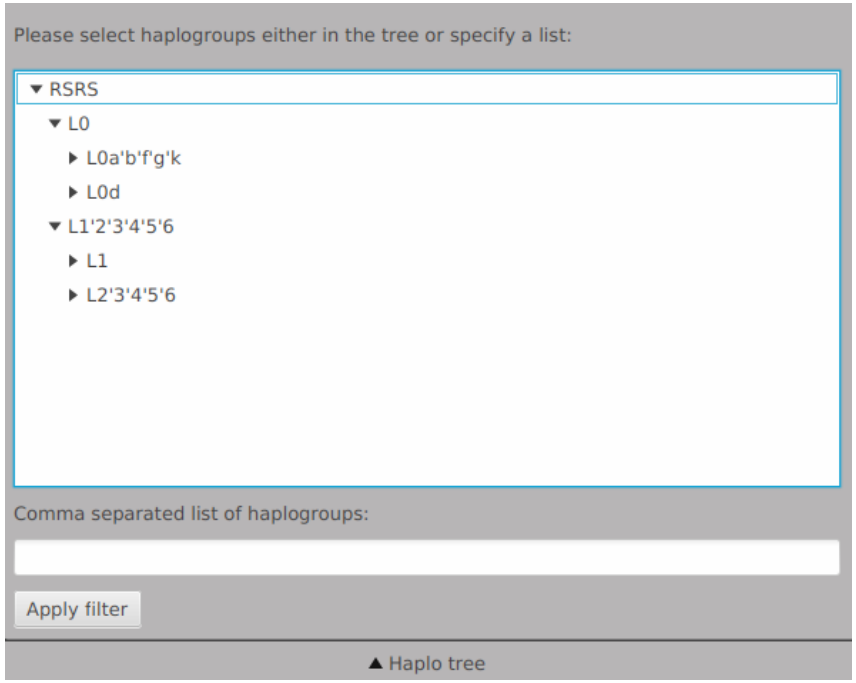
mitoBench: Haplotype filtering



- Phylotree[3] embedded into mitoBench

[3] van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2):E386-E394. `http://www.phylotree.org`.doi:10.1002/humu.20921

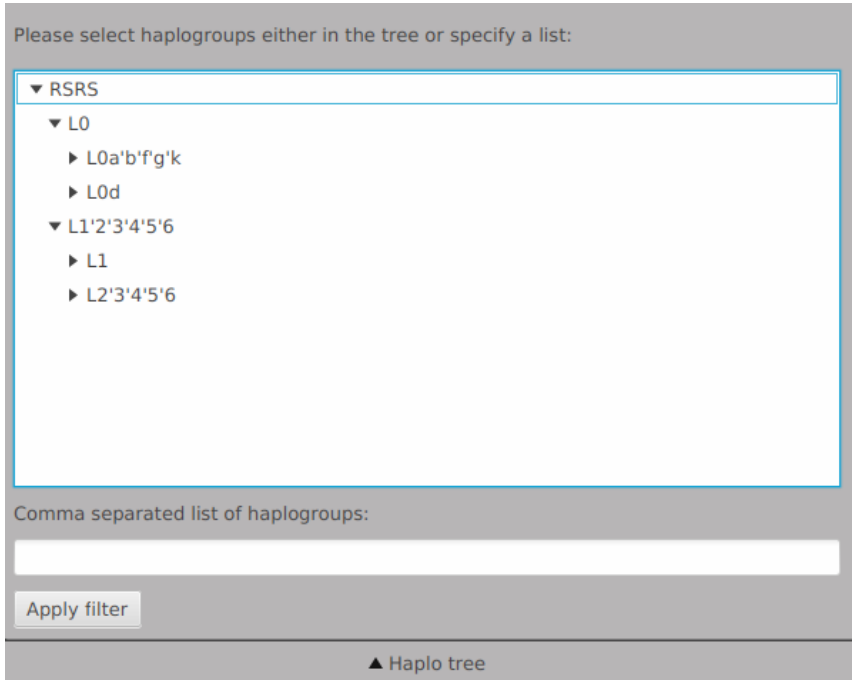
mitoBench: Haplotype filtering



- Phylotree[3] embedded into mitoBench
- Multiple selection possible

[3] van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2):E386-E394. `http://www.phylotree.org`. doi:10.1002/humu.20921

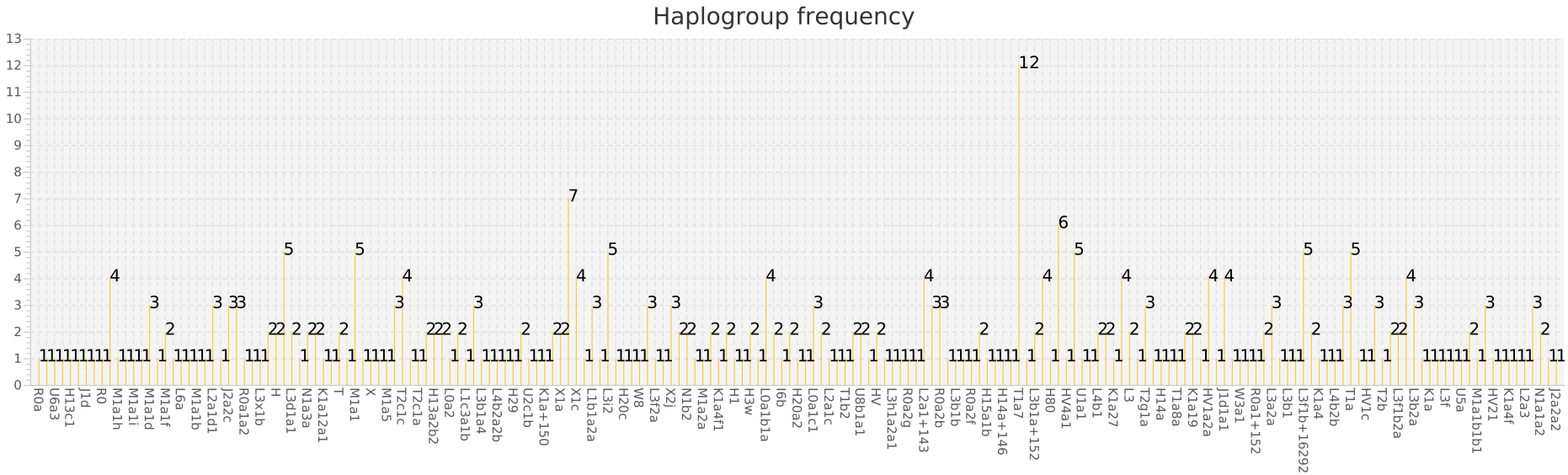
mitoBench: Haplotype filtering



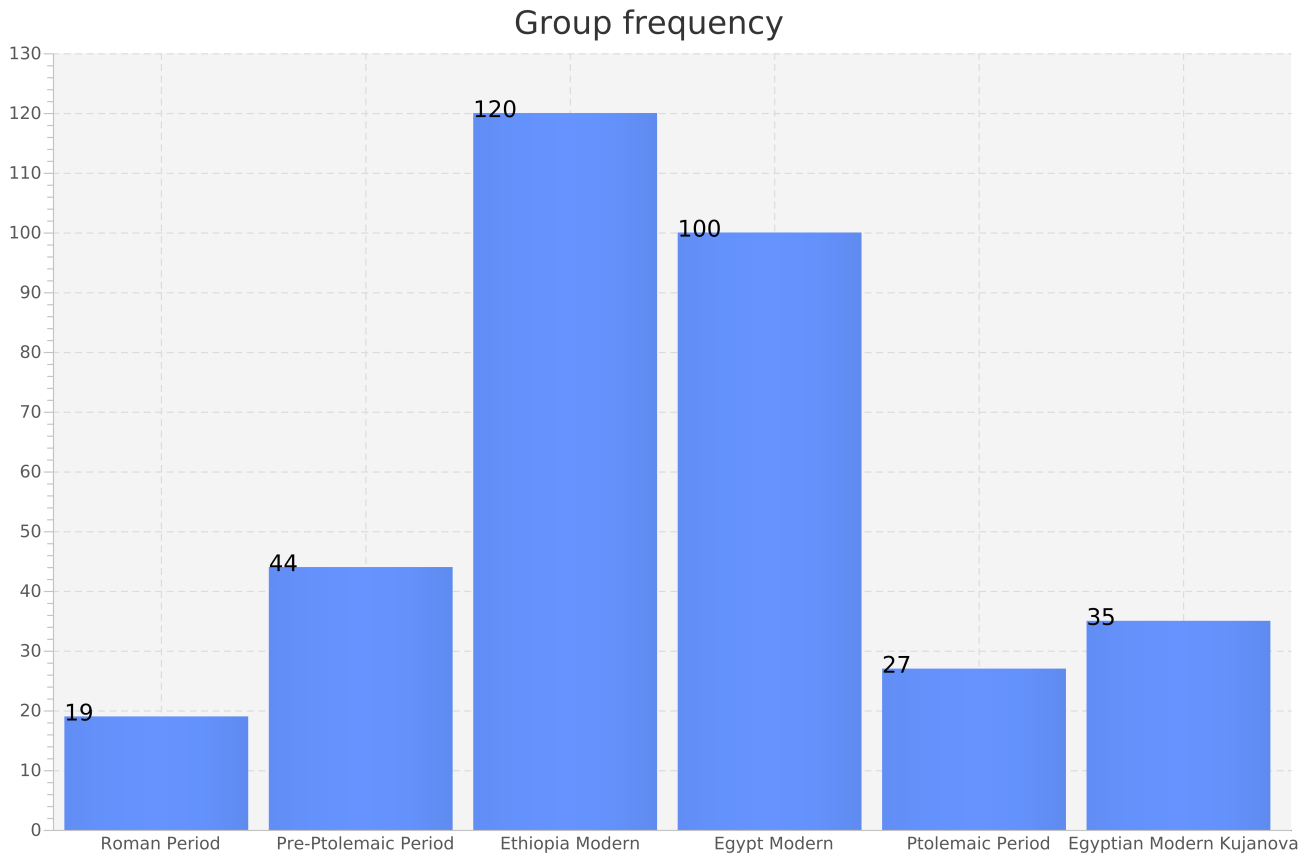
- Phylotree[3] embedded into mitoBench
- Multiple selection possible
- Hierarchical tree-based Haplotype filtering

[3] van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 30(2):E386-E394. `http://www.phylotree.org`.doi:10.1002/humu.20921

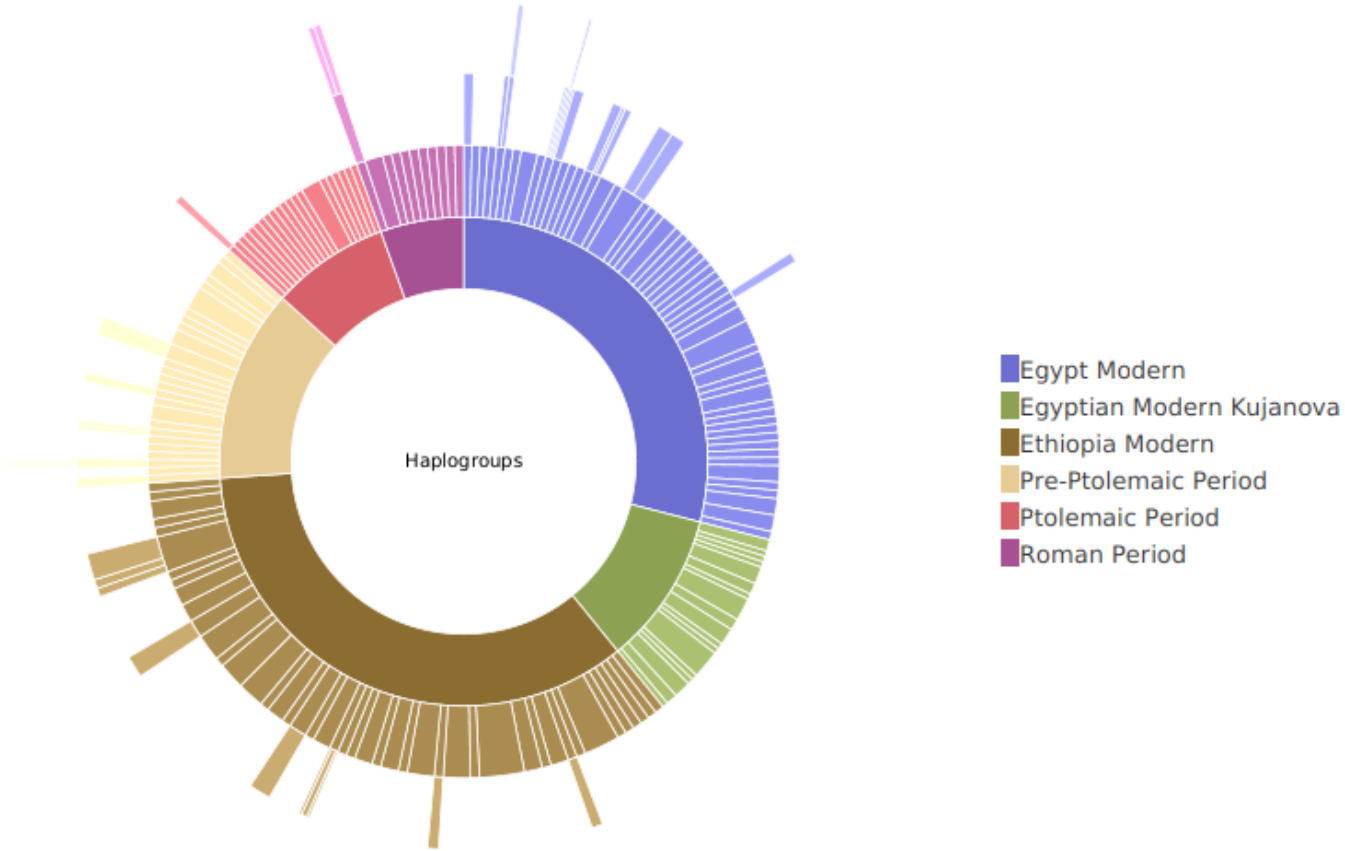
mitoBench: Current visualization methods



mitoBench: Current visualization methods

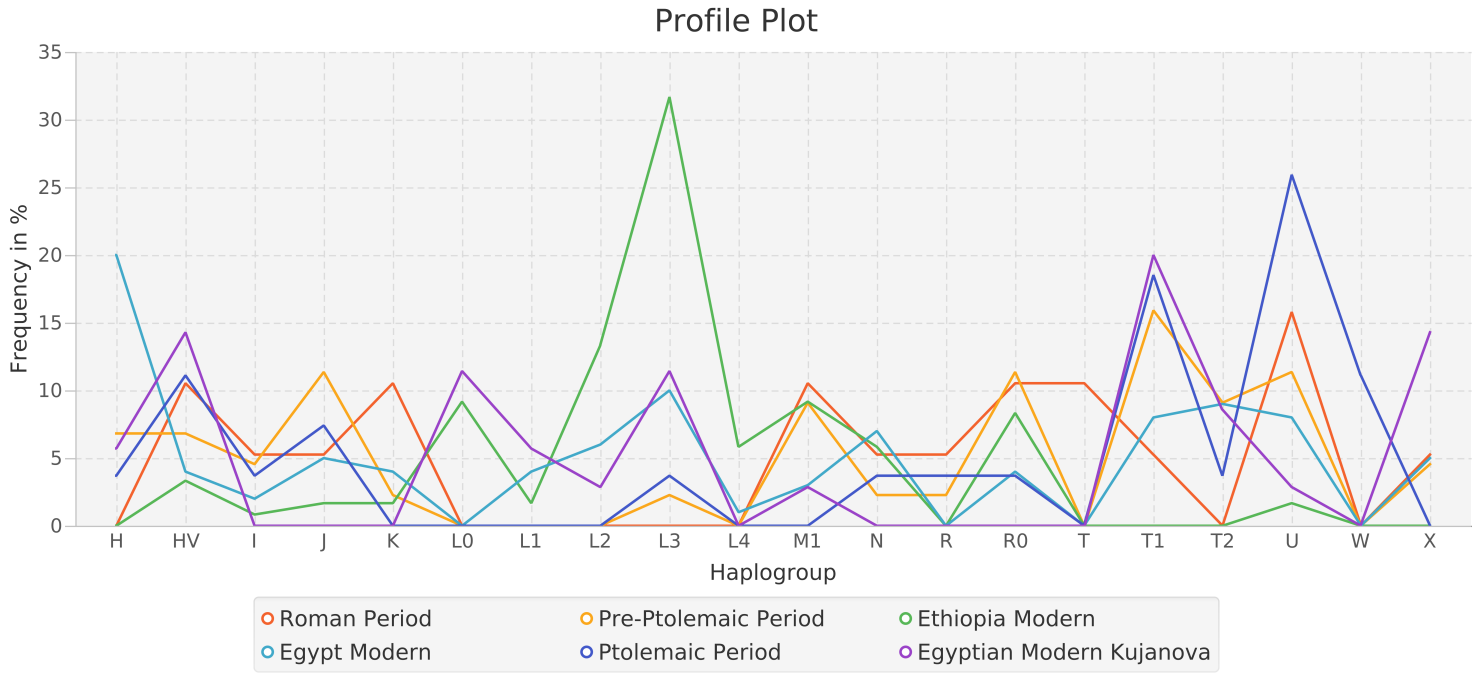


mitobench: Current visualization methods



mitobench: Current visualization methods

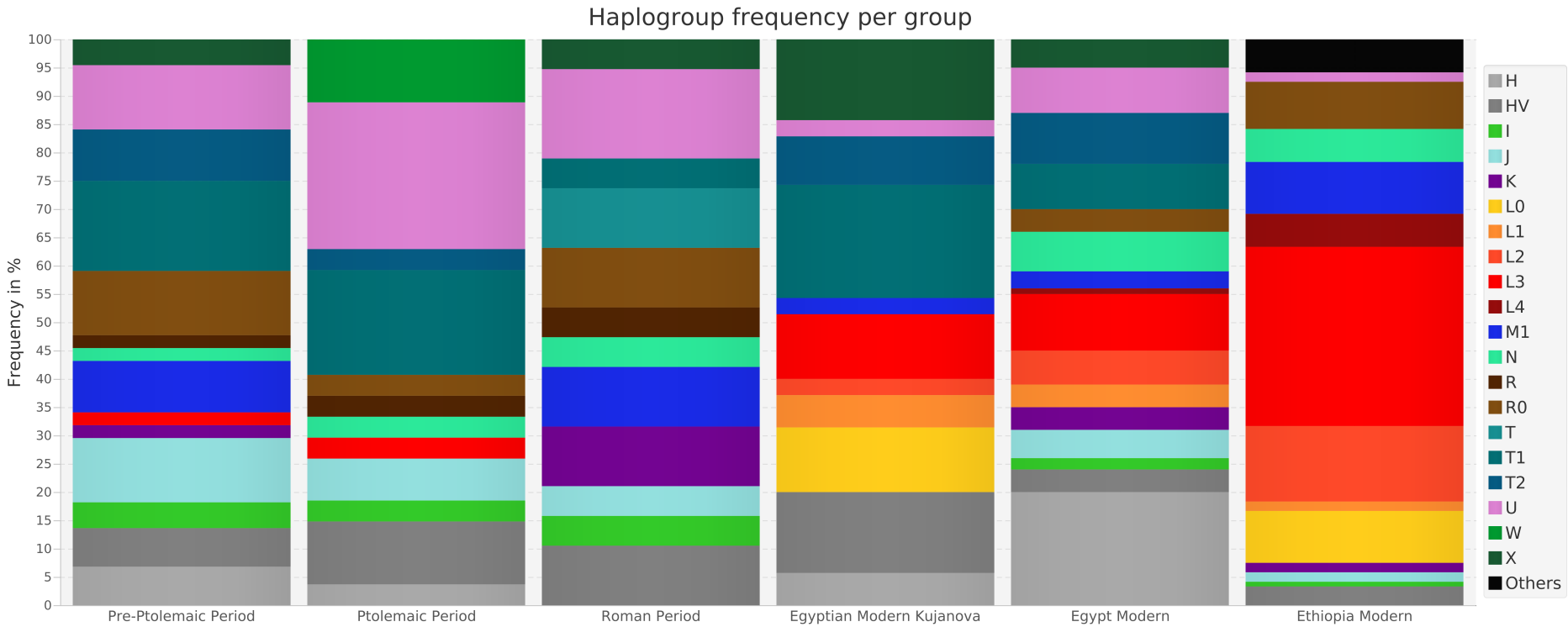
- Groups are represented as line
- Table rows are linked to corresponding lines in profile plot



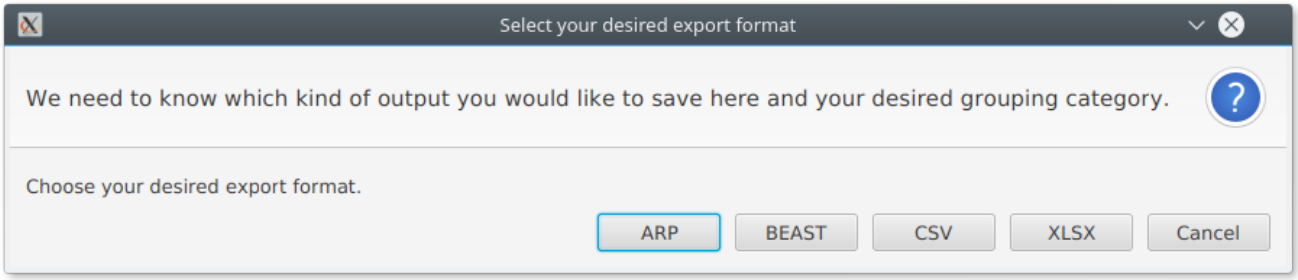
Count statistics ×

Groups	Total Number	H	HV	I	J	K	L0	L1	L2	L3	L4	M1	N	R	R0	T	T1	T2	U	W	X
Roman Period	19	0	2	1	1	2	0	0	0	0	0	2	1	1	2	2	1	0	3	0	1
Pre-Ptolemaic Period	44	3	3	2	5	1	0	0	0	1	0	4	1	1	5	0	7	4	5	0	2
Ethiopia Modern	120	0	4	1	2	2	11	2	16	38	7	11	7	0	10	0	0	0	2	0	0
Egypt Modern	100	20	4	2	5	4	0	4	6	10	1	3	7	0	4	0	8	9	8	0	5
Ptolemaic Period	27	1	3	1	2	0	0	0	0	1	0	0	1	1	1	0	5	1	7	3	0
Egyptian Modern Kujanova	35	2	5	0	0	0	4	2	1	4	0	1	0	0	0	0	7	3	1	0	5

mitoBench: Current visualization methods



mitoBench: Export Data and Figures



- Export data into different file formats
  $\rightarrow$ input for downstream analysis tools
- Export figures

mitoBench: Export Project

- Export whole project
  - Export all table entries and grouping information

mitoBench: Export Project

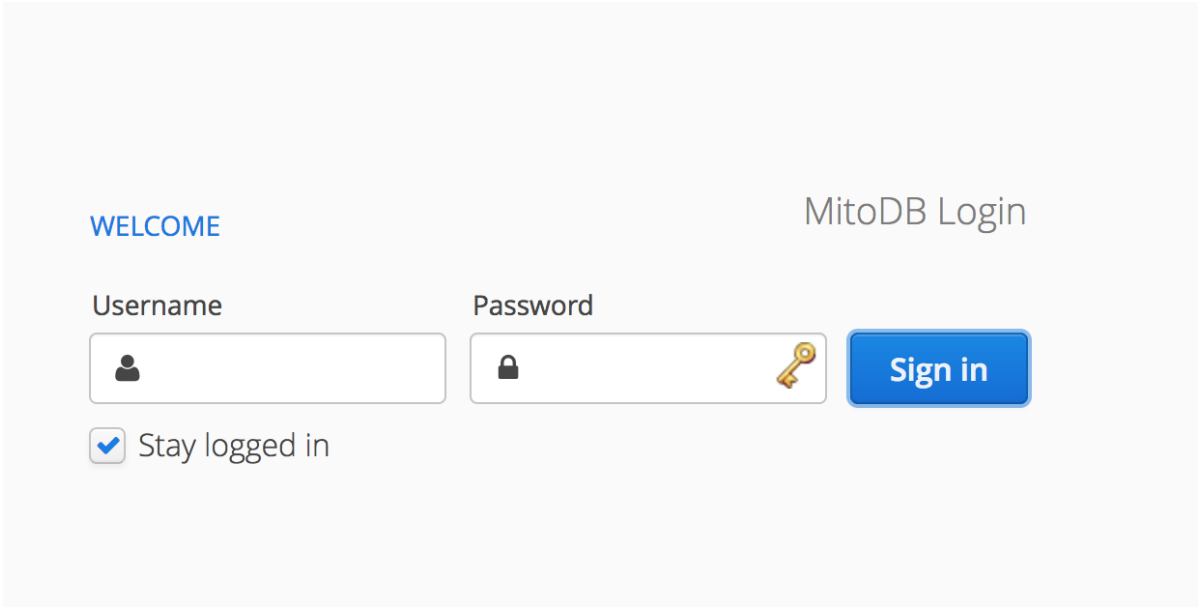
- Export whole project
  - Export all table entries and grouping information
  - Restore a previously stored session with one click

mitoDB

- Idea: Consolidate available data collections in a concise way
  - Web-Frontend with Vaadin Java Framework
  - Backend with PostgreSQL, providing sequence information and meta-data (SQL)
  - Curated data upload with metadata
  - Retrieval possible through WebUI and mitoBench

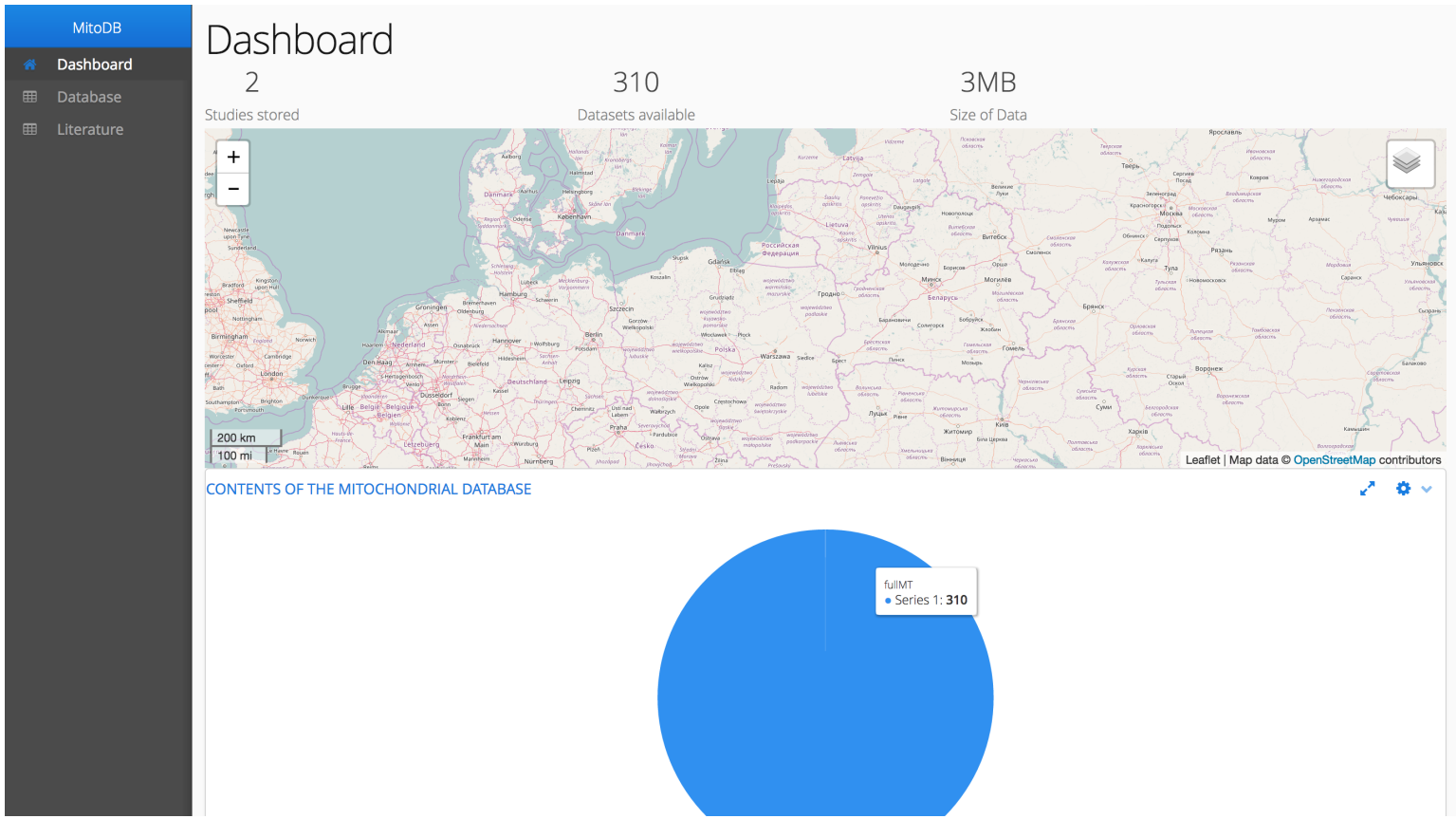
vaadin }>

mitoDB: Login



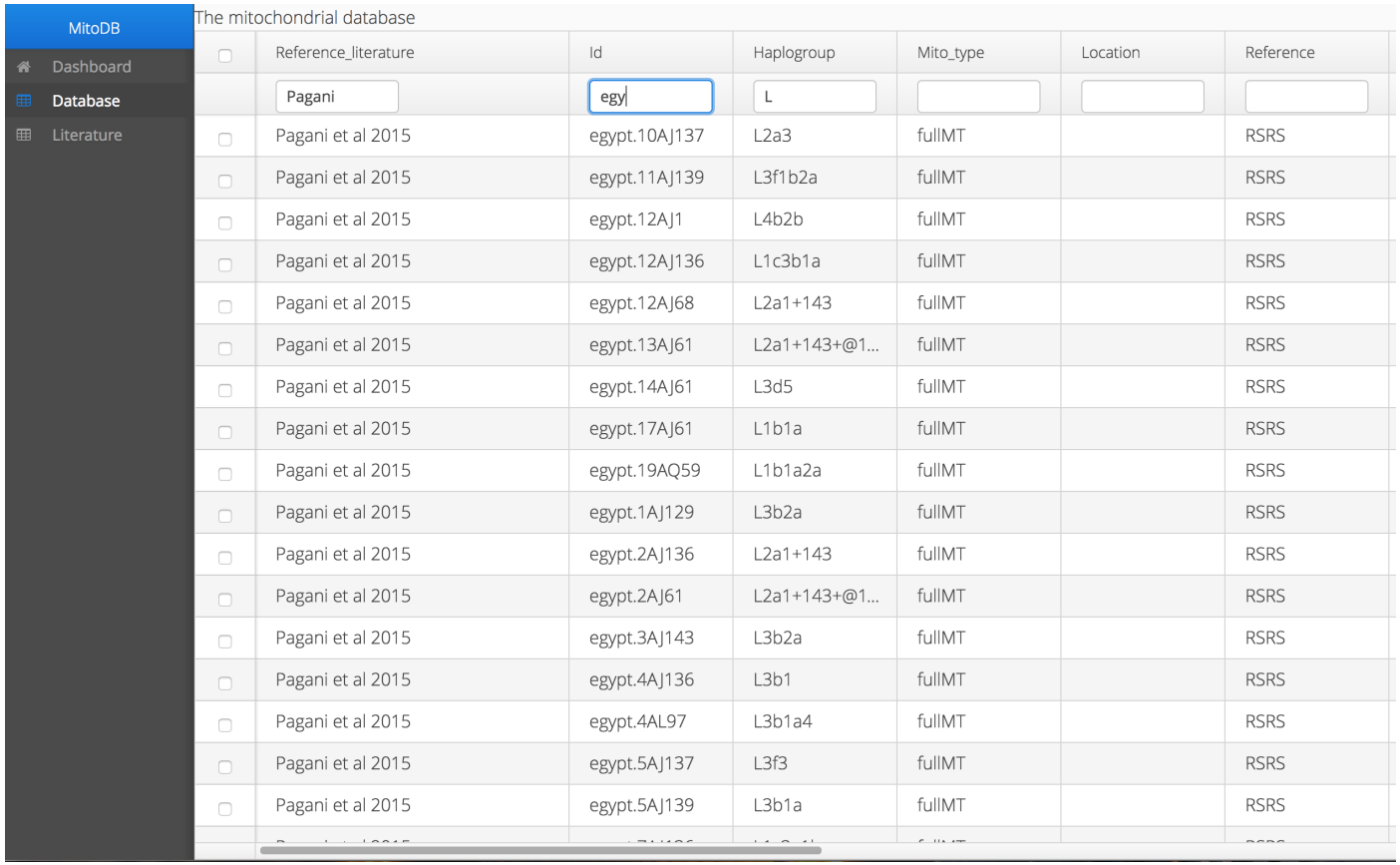
- Secure, encrypted login - registration will be available, too

mitoDB: Dashboard



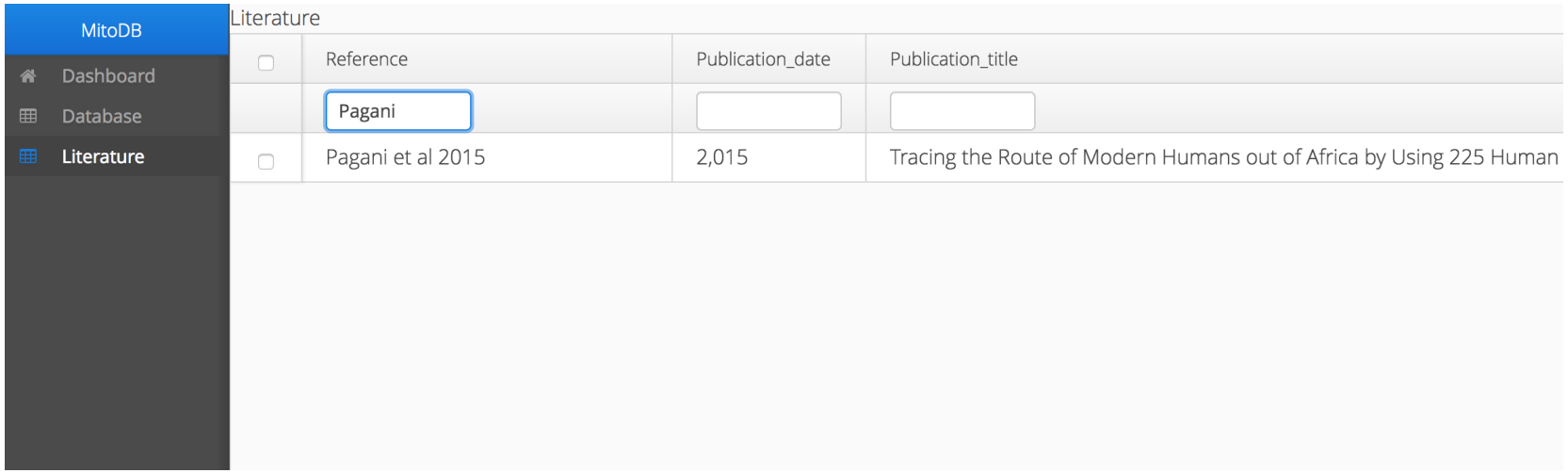
- Central ‘one stop shop‘ for database access
- Left: Navigation, User Menu
- Center: Map with geographic data points
- Bottom: General information on the database (concept, not complete yet)

mitoDB: DataTable



- Data Table: Query functions, multiple selections, handle table data
- Upcoming: Export View: Drag and Drop selected data to different view

mitoDB: References



- Literature database $\leftrightarrow$ Locked to samples (no inconsistencies)
- Simple checking rules: No orphans - consistent updates

mitoDB: Prospects

- Export Functionality: XLSX, ARP, BEAST, FastA, ...

mitoDB: Prospects

- Export Functionality: XLSX, ARP, BEAST, FastA, ...
- Import Functionality, with curation (upload, review mode)

mitoDB: Prospects

- Export Functionality: XLSX, ARP, BEAST, FastA, ...
- Import Functionality, with curation (upload, review mode)
- Data Access: Who can do what? Groups, Access Control etc.

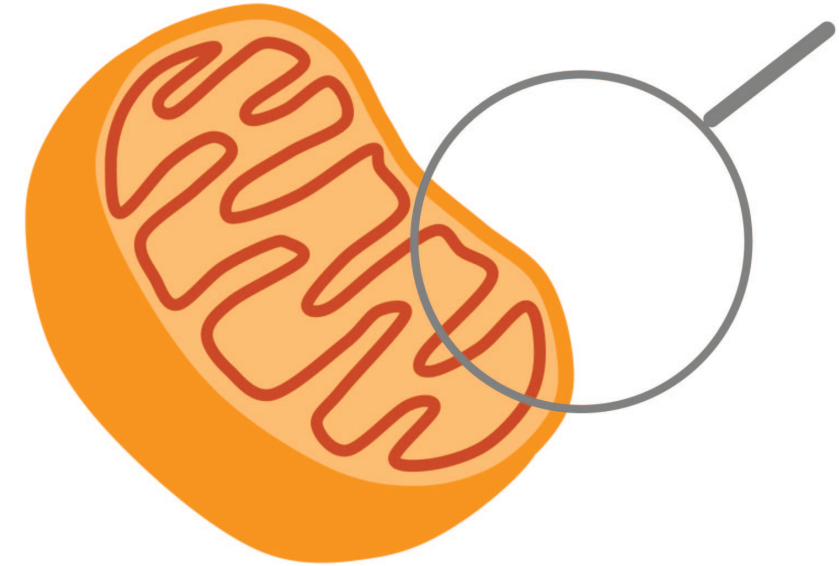
mitoDB: Prospects

- Export Functionality: XLSX, ARP, BEAST, FastA, ...
- Import Functionality, with curation (upload, review mode)
- Data Access: Who can do what? Groups, Access Control etc.
- Workspace: Potentially some analysis inside the WebUI

mitoBench: Prospects

- Looking for early feedback/ideas:
  - Data Manipulation: What can we do to improve further?
  - Data Regulation: Curation, Access rights, Publication, . . .

Thank you



mitoBench & mitoDB