

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 23rd, 2017



MAX-PLANCK-GESELLSCHAFT



ITX
Integrative
Transcriptomics

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 HaploGrep2¹)

¹Weissensteiner, 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	Cameroon	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
	Egypt	0578	Africa, central
	Ethiopia	0579	Africa, central
Africa, east	Kenya	0582	Africa, central
	Rwanda	0583	Africa, central
	Sudan	0584	Africa, central
	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 ...

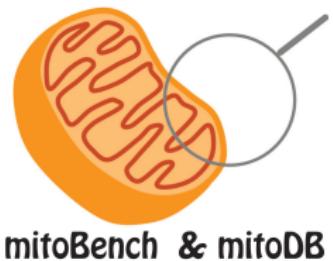
Mitochondrial Genomics: Current status

- 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

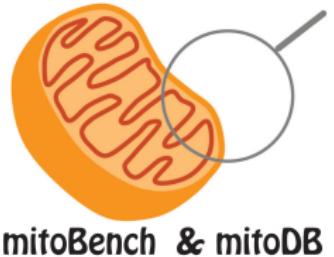


mitoBench & mitoDB

Idea: Mitochondrial Workbench

- Collect mitochondrial data

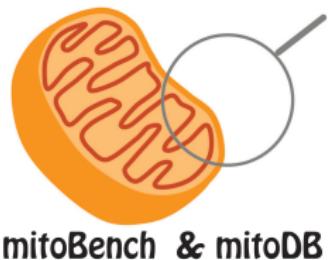
mitoBench: Introduction



Idea: Mitochondrial Workbench

- Collect mitochondrial data
- Ease file handling and file conversion

mitoBench: Introduction

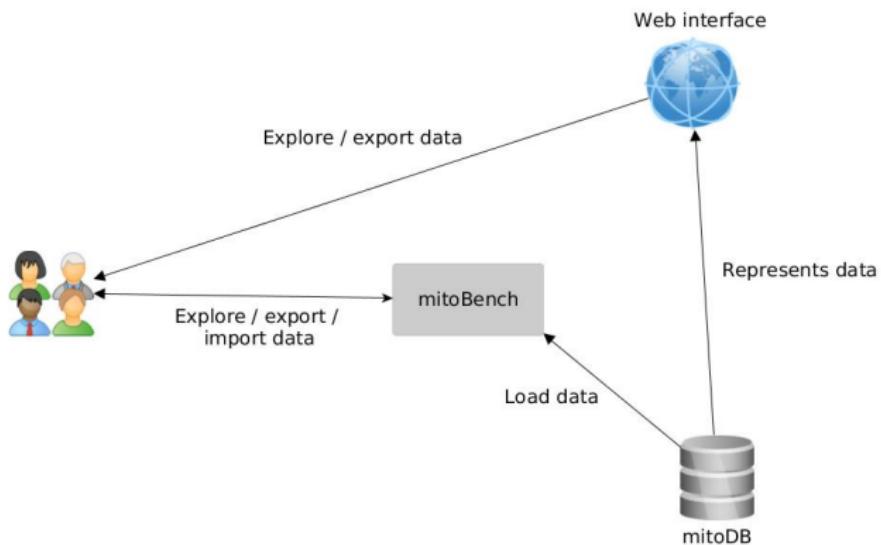


mitoBench & mitoDB

Idea: Mitochondrial Workbench

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

mitoBench: Workflow



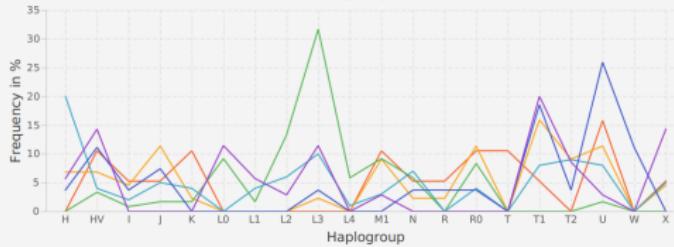
mitoBench: Main view

File Edit Grouping Statistics Table Graphics Help

▼ Haplo tree

Profile Plot X

Profile Plot



Count statistics X

Groups	Total Number	H	HV	I	J	K	L0	L1	L2	L3	L4	M1	N	R	RD	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	1
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...	HV1a1b	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...	L3I1b+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...	N1a1a2a	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

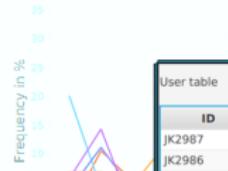
mitoBench: Main view

File Edit Grouping Statistics Table Graphics Help

▼ Log in from

Profile Plot ▾

Profile Plot



Profile Plot

Count statistics ▾

Groups	Total Number	H	HV	I	J	K	L0	L1	L2	L3	L4	M1	N	R	T	T1
Roman Period	29	0	2	1	1	2	0	0	0	0	2	1	1	2	2	1
Pre-Ptolemaic Period	44	3	3	2	5	1	0	0	1	0	4	1	1	5	0	7
Egypt Modern	120	0	4	1	2	2	11	2	16	20	7	11	7	0	10	0
Copt. Modern	100	20	4	2	5	4	0	4	8	10	1	3	7	0	4	0
Modern Egypt	72	3	1	1	0	0	0	0	3	10	0	0	1	1	0	5

User table

ID	MTSequence	Haplogroup	Time Period (Grouping)	continent	country	country_region	culture_type	dating_lower	dating_upper
JK2987	GATCA...	HV1a'b'c	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-230	-230
JK2986	GATCA...	HV	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-457	-457
JK2985	GATCA...	L3f1b+162...	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-274	-274
JK2984	GATCA...	U7	Roman Period	Africa	Egypt	Beni Suef	Ancient	2015	2015
JK2983	GATCA...	T1a5	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-47	-47
JK2981	GATCA...	M1aE	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-388	-388
JK2980	GATCA...	I	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-281	-281
JK2979	GATCA...	HV1a2a	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-290	-290
JK2978	GATCA...	N1a1a2	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-940	-940
JK2977	GATCA...	T2e	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-312	-312
JK2975	GATCA...	R	Roman Period	Africa	Egypt	Beni Suef	Ancient	1	1
JK2974	GATCA...	H	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-846	-846
JK2973	GATCA...	U1a6a3	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-258	-258

Unstable DB table

ice_literature

region

Peltz et al 2017 North Afr

Schoenemann & Peltz et al 2017 North Afr

0 rows are selected

0 rows are selected

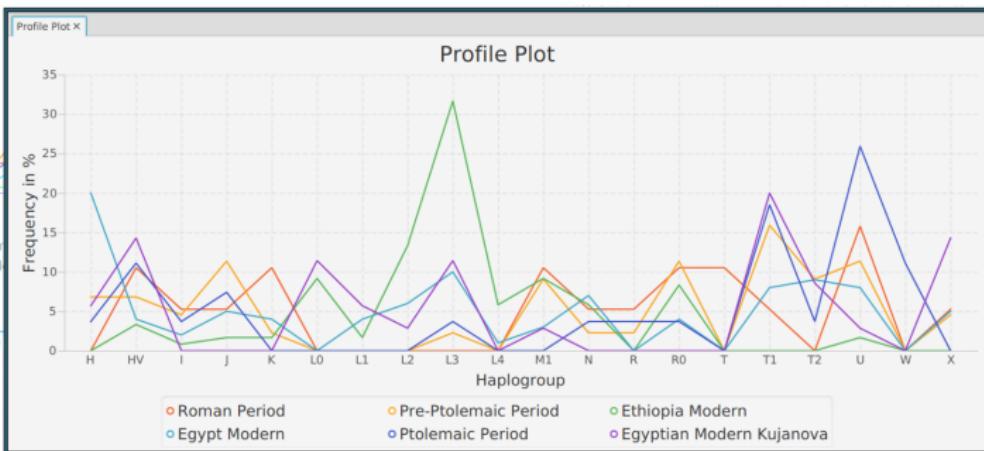
mitoBench: Main view

File Edit Grouping Statistics Table Graphics Help

▼ Log in from

Profile Plot x

Profile Plot



User table

ID	MTSequence
K2997	GATCA...
K2998	GATCA...
K2995	GATCA...
K2994	GATCA...
K2992	GATCA...
K2991	GATCA...
K2990	GATCA...
K2979	GATCA... HV1a1a2
K2978	GATCA... HV1a1a2
K2977	R2b
K2975	GATCA... R
K2974	GATCA... H

0 rows are selected

Count statistics x

Groups	Total Number	H	HV	I	J	K	L0	L1	L2	L3	L4	M1	N	R	R0	T	T1
Roman Period	29	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	1	0	4	1	1	5	0	7	1

11	7	0	10	0	0	1
3	7	0	4	0	0	1
0	1	1	1	0	5	1
1	0	0	0	0	7	1

Unable DB table

literature	region
Pulitzer et al 2017	North Afr
Pulitzer et al 2017	North Afr
Pulitzer et al 2017	North Afr
Pulitzer et al 2017	North Afr
Pulitzer et al 2017	North Afr
Pulitzer et al 2017	North Afr
Pulitzer et al 2017	North Afr
Pulitzer et al 2017	North Afr
Schoenemann & Pulitzer et al 2017	North Afr
Schoenemann & Pulitzer et al 2017	North Afr
Schoenemann & Pulitzer et al 2017	North Afr
Schoenemann & Pulitzer et al 2017	North Afr
Schoenemann & Pulitzer et al 2017	North Afr
Schoenemann & Pulitzer et al 2017	North Afr

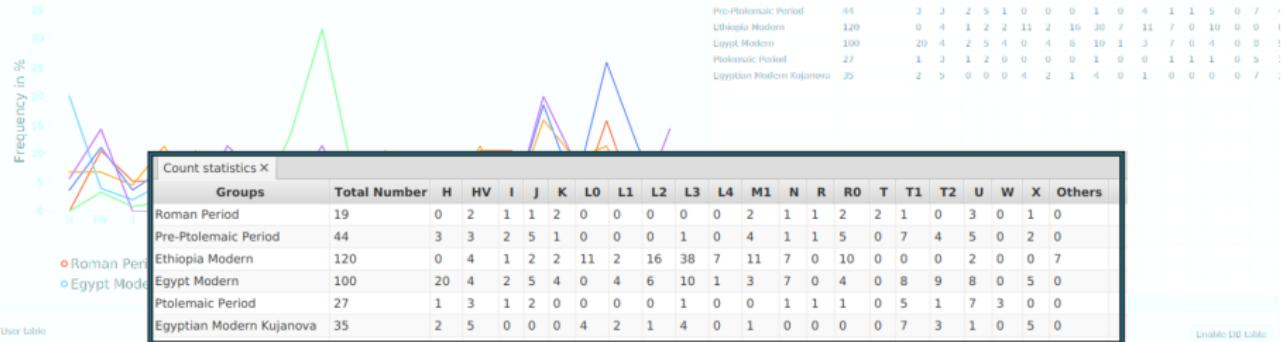
mitoBench: Main view

File Edit Grouping Statistics Table Graphics Help

▼ Log in here

Profile Plot ▾

Profile Plot



User table

Unstable DB table

ID	HTSequence	Haplotype	Time_Period (Grouping)	continent	country	country_region	culture_type	dating_lower	dating_upper	language	mito_type	population	published	reference	reference_literature	region
K2997	GATCA...	IV2a1rc	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-230	-230	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2998	GATCA...	IV	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-157	-157	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2995	GATCA...	LR1b1 162...	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-274	-274	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2994	GATCA...	U7	Roman Period	Africa	Egypt	Beni Suef	Ancient	2012	2012	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2993	GATCA...	Y119	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-47	-47	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2991	GATCA...	M1a1e	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-380	-380	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2998	GATCA...	I	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-281	-281	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2979	GATCA...	IV2a2a	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-299	-299	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2978	GATCA...	RI1a1a2	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-940	-940	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2977	GATCA...	F26	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient	-312	-312	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2975	GATCA...	R	Roman Period	Africa	Egypt	Beni Suef	Ancient	1	1	Egyptian	fullMT	Egyptians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.
K2974	GATCA...	II	Pre-Ptolemaic Period	Africa	Lowest	Beni Suef	Ancient	-846	-846	Croatian	fullMT	Croatians	F	RSRS	Schoenemann & Peltzer et al 2017	North Afr.

0 rows are selected

mitoBench: Data import and conversion

Accepted file formats

- (Multi-) FastA

²Excoffier, 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

²Excoffier, 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)

²Excoffier, 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

²Excoffier, 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)

²Excoffier, 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
→ Group data by feature

mitoBench: Data grouping

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

mitoBench: Data grouping

User-specified groups

User table

ID	MTSequence	Haplogroup		ing_lower	dating			
JK2880	GATCA...	T1a2	Egypt	9	-669			
egypt.5AJ129	GATCA...	X1c	Egypt	5	2015			
egypt.6AJ68	GATCA...	H3w	Egypt	5	2015			
egpg5305866	GATCA...	L3f2a1	Ethiopia	5	2015			
egypt.6AJ63	GATCA...	R0a3	Egypt	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

Groupname: OK

39 / 62

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...	M1ale	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2980	GATCA...	I	Ptolemaic Period	Africa	Egypt	Egyptian
JK2979	GATCA...	HV1a2a	Ptolemaic Period	Africa	Egypt	Egyptian
JK2978	GATCA...	N1ala2	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...	T2clc	Pre-Ptolemaic Period	Africa	Egypt	Egyptian
JK2963	GATCA...	M1ali	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's	Ptolemaic Period	Africa	Egypt	Egyptian

mitoBench: Haplotype filtering

Please select haplogroups either in the tree or specify a list:

- ▼ RSRS
- ▼ L0
 - L0a'b'f'g'k
 - L0d
- ▼ L1'2'3'4'5'6
 - L1
 - L2'3'4'5'6

Comma separated list of haplogroups:

▲ Haplo tree

- PhyloTree³ embedded into mitoBench

³van 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

Please select haplogroups either in the tree or specify a list:

- ▼ RSRS
- ▼ L0
 - L0a'b'f'g'k
 - L0d
- ▼ L1'2'3'4'5'6
 - L1
 - L2'3'4'5'6

Comma separated list of haplogroups:

▲ Haplo tree

- Phylogenetic tree³ embedded into mitoBench
- Multiple selection possible

³van 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

Please select haplogroups either in the tree or specify a list:

▼ RSRS

▼ L0

► L0a'b'f'g'k

► L0d

▼ L1'2'3'4'5'6

► L1

► L2'3'4'5'6

Comma separated list of haplogroups:
L1,L2

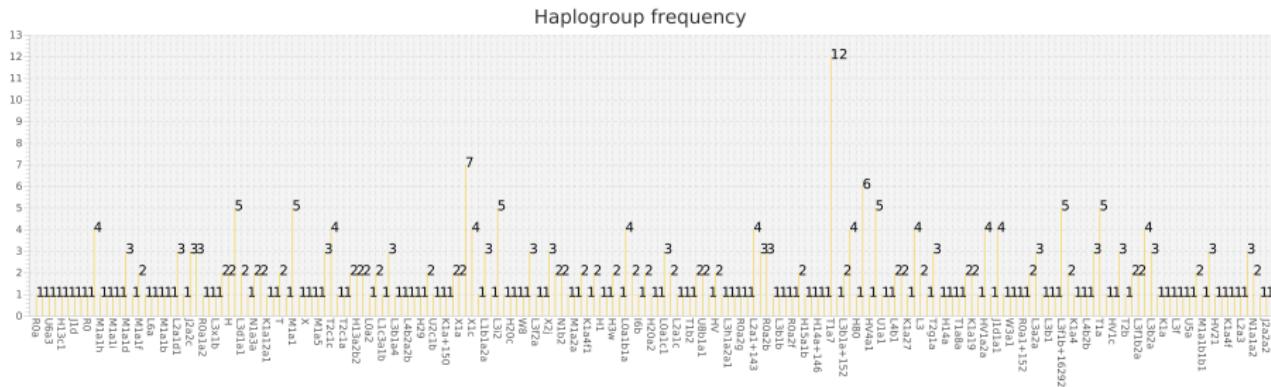
Apply filter

▲ Haplo tree

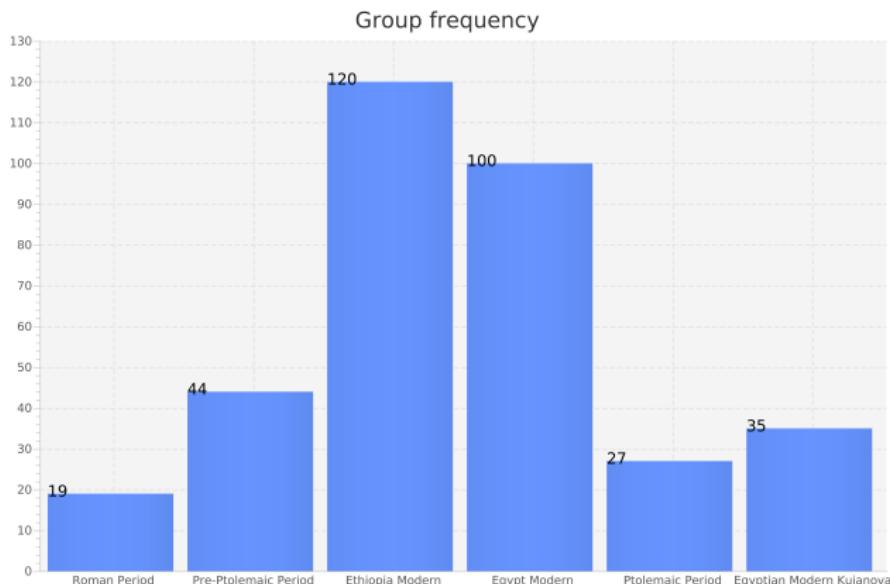
- Phylogenetic tree³ embedded into mitoBench
- Multiple selection possible
- Hierarchical tree-based Haplotype filtering

³van 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

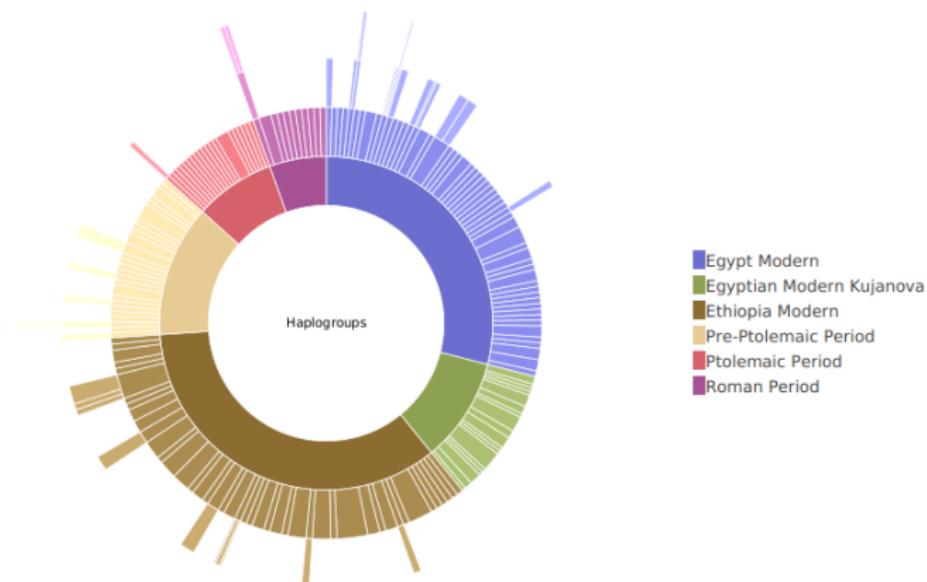
mitoBench: Current visualization methods



mitoBench: Current visualization methods

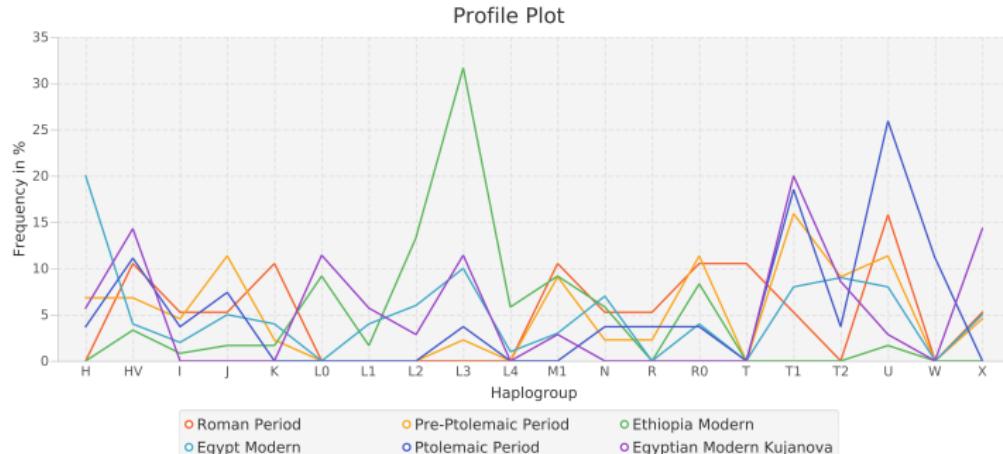


mitobench: Current visualization methods



mitobench: Current visualization methods

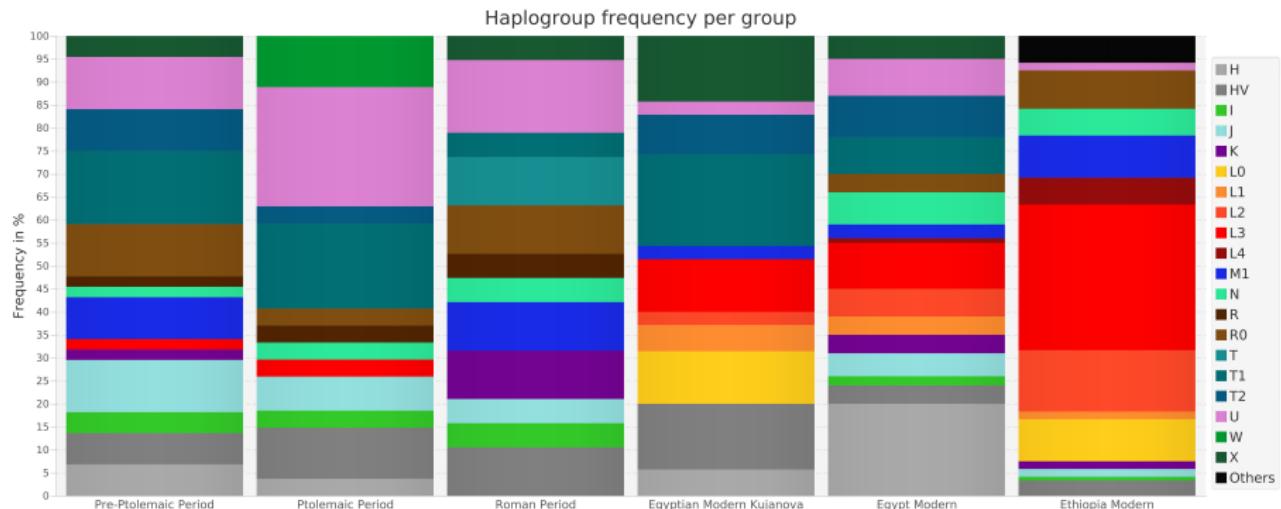
- Groups are represented as line
- Table rows are linked to corresponding lines in profile plot



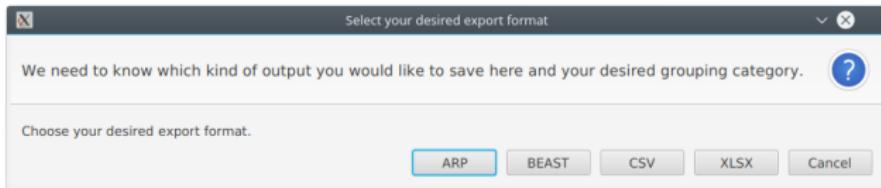
Count statistics X

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	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
→ 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

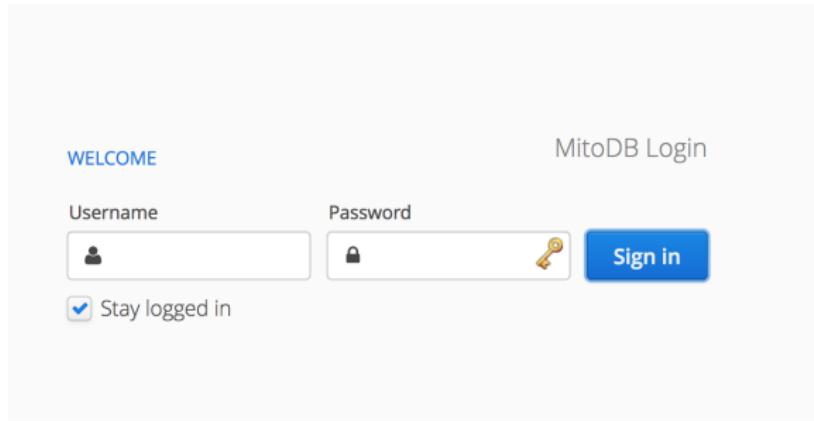
mitoDB

- 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



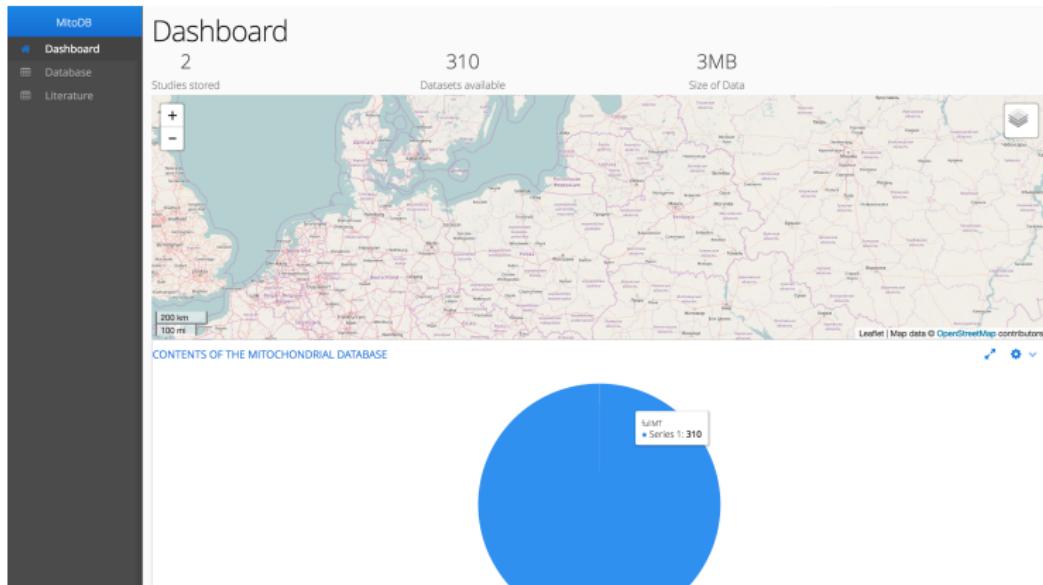
vaadin }>

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

The screenshot shows the mitoDB interface with a sidebar on the left containing navigation links: MitoDB, Dashboard, Database (selected), and Literature. The main area displays a table titled "The mitochondrial database". The table has columns: Reference_literature, id, Haplogroup, Mito_type, Location, and Reference. A search bar at the top of the table allows filtering by "Reference_literature". The "id" column contains entries such as egypt.10AJ137, egypt.11AJ139, etc. The "Haplogroup" column includes L, L2a3, L3f1b2a, L4b2b, L1c3b1a, L2a1+143, L2a1+143+@..., L3d5, L1b1a, L1b1a2a, L3b2a, L2a1+143, L2a1+143+@1..., L3b2a, L3b1, L3b1a4, L3f3, and L3b1a. The "Mito_type" column is mostly "fullMT". The "Location" and "Reference" columns are mostly blank or show "RSRS". The "Reference_literature" column shows "Pagani et al 2015" repeated multiple times. The table has a footer with navigation icons.

Reference_literature	id	Haplogroup	Mito_type	Location	Reference
Pagani	egy	L	fullMT		
Pagani et al 2015	egypt.10AJ137	L2a3	fullMT	RSRS	
Pagani et al 2015	egypt.11AJ139	L3f1b2a	fullMT	RSRS	
Pagani et al 2015	egypt.12AJ1	L4b2b	fullMT	RSRS	
Pagani et al 2015	egypt.12AJ136	L1c3b1a	fullMT	RSRS	
Pagani et al 2015	egypt.12AJ68	L2a1+143	fullMT	RSRS	
Pagani et al 2015	egypt.13AJ61	L2a1+143+@...	fullMT	RSRS	
Pagani et al 2015	egypt.14AJ61	L3d5	fullMT	RSRS	
Pagani et al 2015	egypt.17AJ61	L1b1a	fullMT	RSRS	
Pagani et al 2015	egypt.19AQ59	L1b1a2a	fullMT	RSRS	
Pagani et al 2015	egypt.1AJ129	L3b2a	fullMT	RSRS	
Pagani et al 2015	egypt.2AJ136	L2a1+143	fullMT	RSRS	
Pagani et al 2015	egypt.2AJ61	L2a1+143+@1...	fullMT	RSRS	
Pagani et al 2015	egypt.3AJ143	L3b2a	fullMT	RSRS	
Pagani et al 2015	egypt.4AJ136	L3b1	fullMT	RSRS	
Pagani et al 2015	egypt.4AJ97	L3b1a4	fullMT	RSRS	
Pagani et al 2015	egypt.5AJ137	L3f3	fullMT	RSRS	
Pagani et al 2015	egypt.5AJ139	L3b1a	fullMT	RSRS	

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

mitoDB: References

The screenshot shows the mitoDB interface with a sidebar on the left containing 'MitoDB' and three menu items: 'Dashboard', 'Database', and 'Literature'. The 'Literature' item is selected and highlighted with a blue border. To its right is a table titled 'Literature' with four columns: 'Reference', 'Publication_date', and 'Publication_title'. There is one visible row in the table. The first column contains a checkbox and the value 'Pagani'. The second column is empty. The third column contains the year '2,015'. The fourth column contains the title 'Tracing the Route of Modern Humans out of Africa by Using 225 Human'. The 'Pagani' entry is highlighted with a blue border.

	Reference	Publication_date	Publication_title
<input type="checkbox"/>	Pagani		
<input type="checkbox"/>	Pagani et al 2015	2,015	Tracing the Route of Modern Humans out of Africa by Using 225 Human

- Literature database ↔ 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.

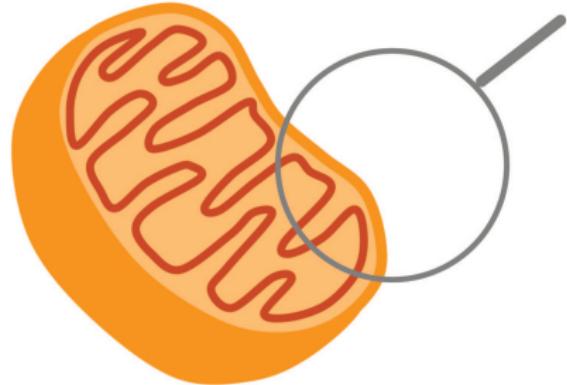
mitoDB: Prospects

- 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