

# mitoBench & mitoDB: Novel interactive methods for population genetics on mitochondrial DNA



Alexander Peltzer<sup>1,2\*</sup>, Judith Neukamm<sup>1,3\*</sup>, The mitoDB project consortium<sup>†</sup>, Kay Nieselt<sup>1</sup> and Wolfgang Haak<sup>2</sup>

<sup>1</sup> Integrative Transcriptomics, Center for Bioinformatics (ZBIT), University of Tuebingen, Germany.

<sup>2</sup> Max Planck Institute for the Science of Human History, Jena, Germany.

<sup>3</sup> Institute for Archaeological Sciences, Archaeo- and Paleogenetics, University of Tuebingen, Germany.

† See <http://bit.ly/2vB8wKx> for full author-list.

\*These authors contributed equally to the work.

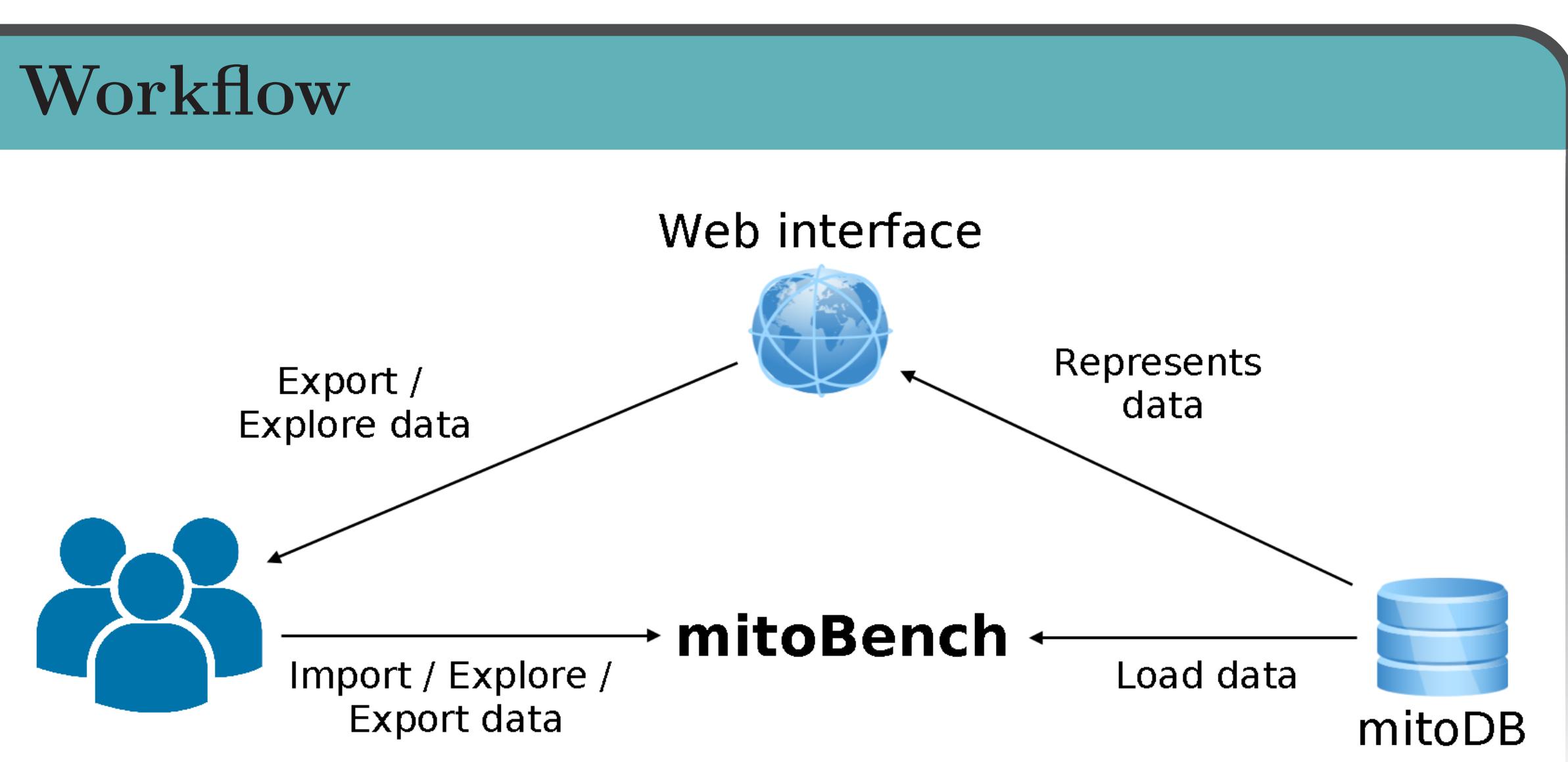


Contact: peltzer@shh.mpg.de, judith.neukamm@uni-tuebingen.de

## Motivation & Goals

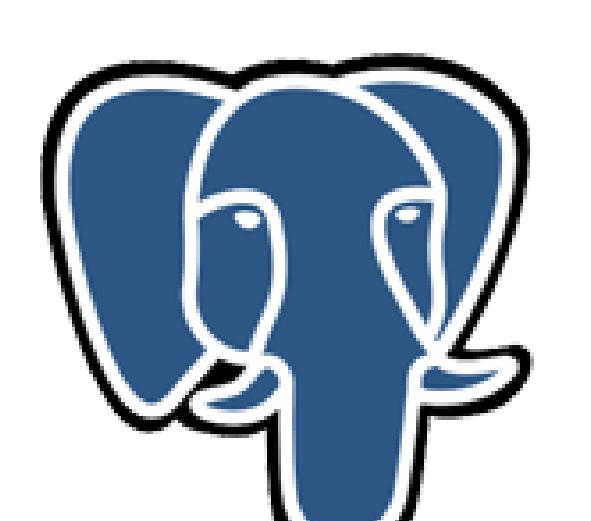
- Mitochondrial DNA (mtDNA) is often the only proxy available to study extinct populations and their relationship with modern populations
- Tools for the analysis typically rely on different file formats → requiring manual interaction with the data for downstream analysis
- mitoBench: workbench to ease file conversions, methods to interactively analyze and visualize human mitochondrial data
- mitoDB: database for human mitochondrial genomes to provide a central reference that can be easily accessed via the workbench and a web-frontend

## Workflow



## mitoDB

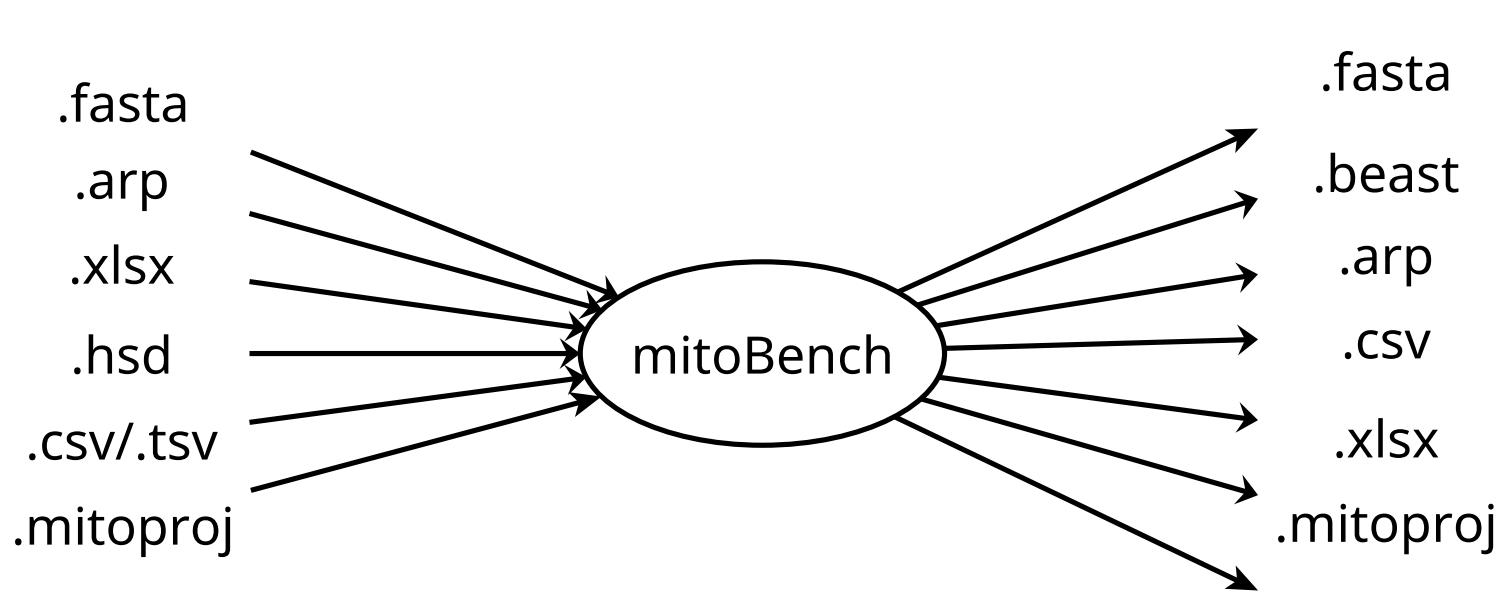
- Database: Providing meta-information (Location, language, Sequence data, ...)
- Web-Frontend: Browsing data, searching for locations, quick look at database contents
- Data Curation: Users can rate samples in the web based on their experiences with data
- Access: Web-Frontend (exporting) and mitoBench (analysis)



vaadin >

## mitoBench

### File conversions



⇒ connect the workbench with existing analysis methods/resources such as BEAST<sup>1</sup>, Arlequin<sup>2</sup>, PhyloTree<sup>3</sup> and others

### Data representation as Table

ID	MTSequence	Haplogroup	Time Period	continent	country	country_region	culture_type
JK2975	GATCA...	R	Roman Period	Africa	Egypt	Beni Suef	Ancient
JK2974	GATCA...	U	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2972	GATCA...	T1a5	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2916	NNCCA...	R0	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2972	GATCA...	IV2a2a	Roman Period	Africa	Egypt	Beni Suef	Ancient
JK2918	GATCA...	J2a2a2	Roman Period	Africa	Egypt	Beni Suef	Ancient
JK2988	GATCA...	T1a5	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2987	GATCA...	Z2a1	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2937	GATCA...	Z2a2	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2963	GATCA...	M1a1	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2985	GATCA...	R2J1T	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2987	GATCA...	J2a1a1	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2553	GATCA...	R0a1a	Roman Period	Africa	Egypt	Beni Suef	Ancient
JK2132	GATCA...	T	Roman Period	Africa	Egypt	Beni Suef	Ancient
JK2128	GATCA...	HV21	Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2130	GATCA...	M1a1	Roman Period	Africa	Egypt	Beni Suef	Ancient
JK2962	GATCA...	M1a1	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2969	GATCA...	Z2a2	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2996	GATCA...	HV1b2	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2883	GATCA...	T1a	Pre-Ptolemaic Period	Africa	Egypt	Beni Suef	Ancient
JK2979	GATCA...	U3b	Roman Period	Africa	Egypt	Beni Suef	Ancient

### Data filtering / Statistics

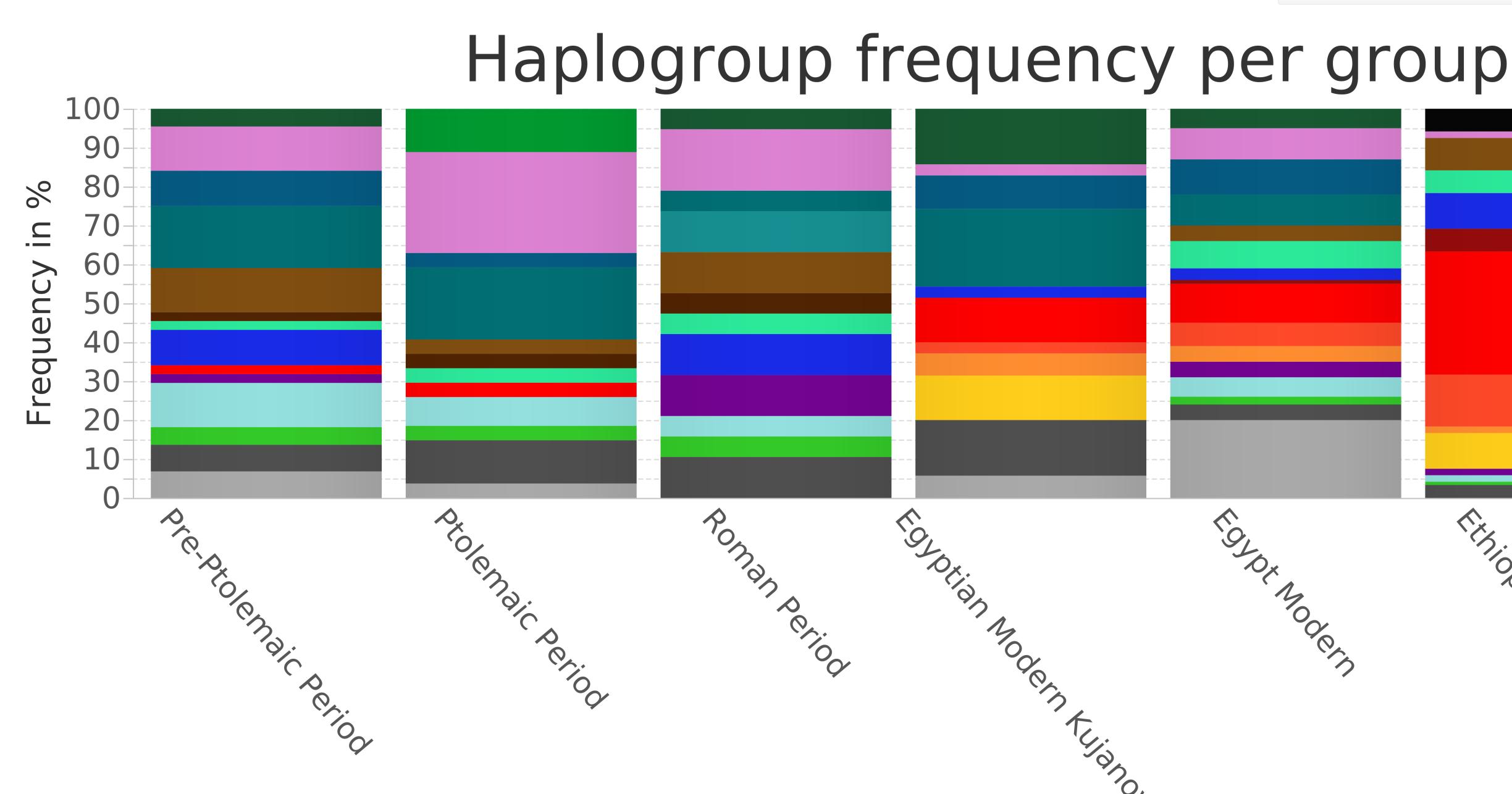
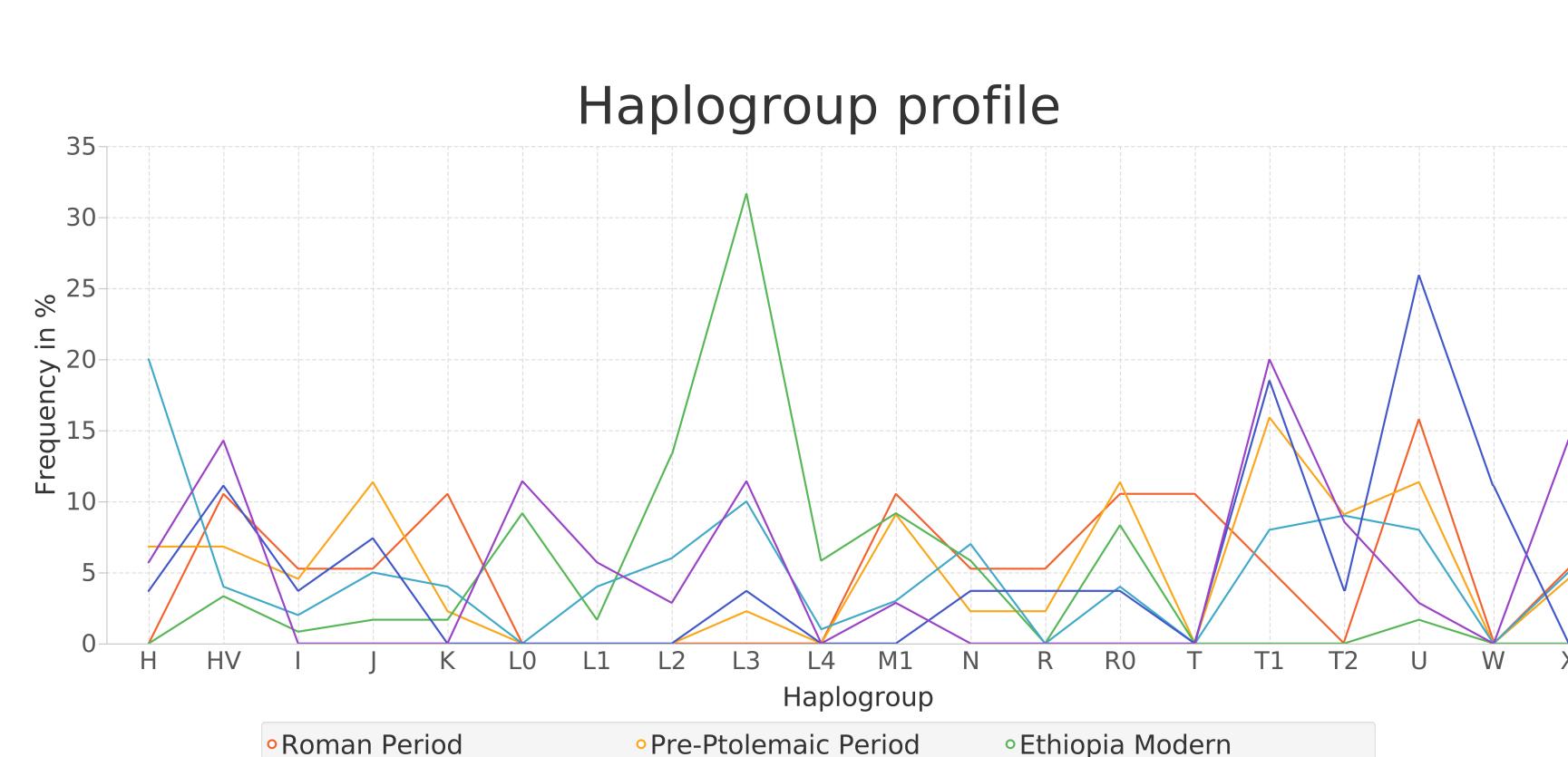
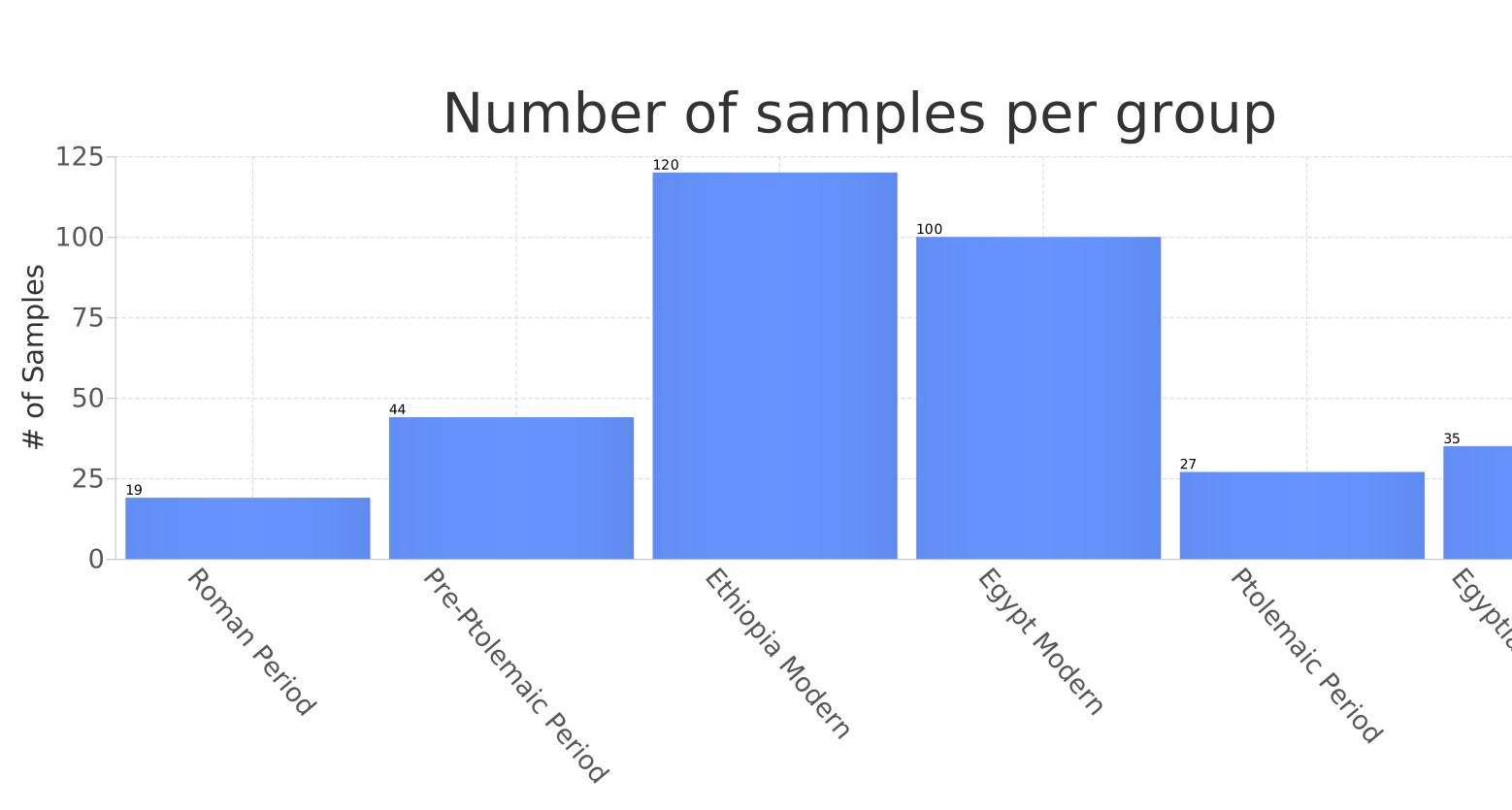
- Haplogroup filtering / frequencies
- Haplotype filtering / frequencies

### Data grouping

- group data based on shared features (e.g. time period, location)
- user defined grouping
  - ⇒ allows analysis of Haplogroup distribution between different groups

### Exemplary study: Analysis of 90 ancient Egyptian mummy mtDNA genomes<sup>4</sup>

→ Did Haplogroup frequencies change over time? If so, how?



## Outlook

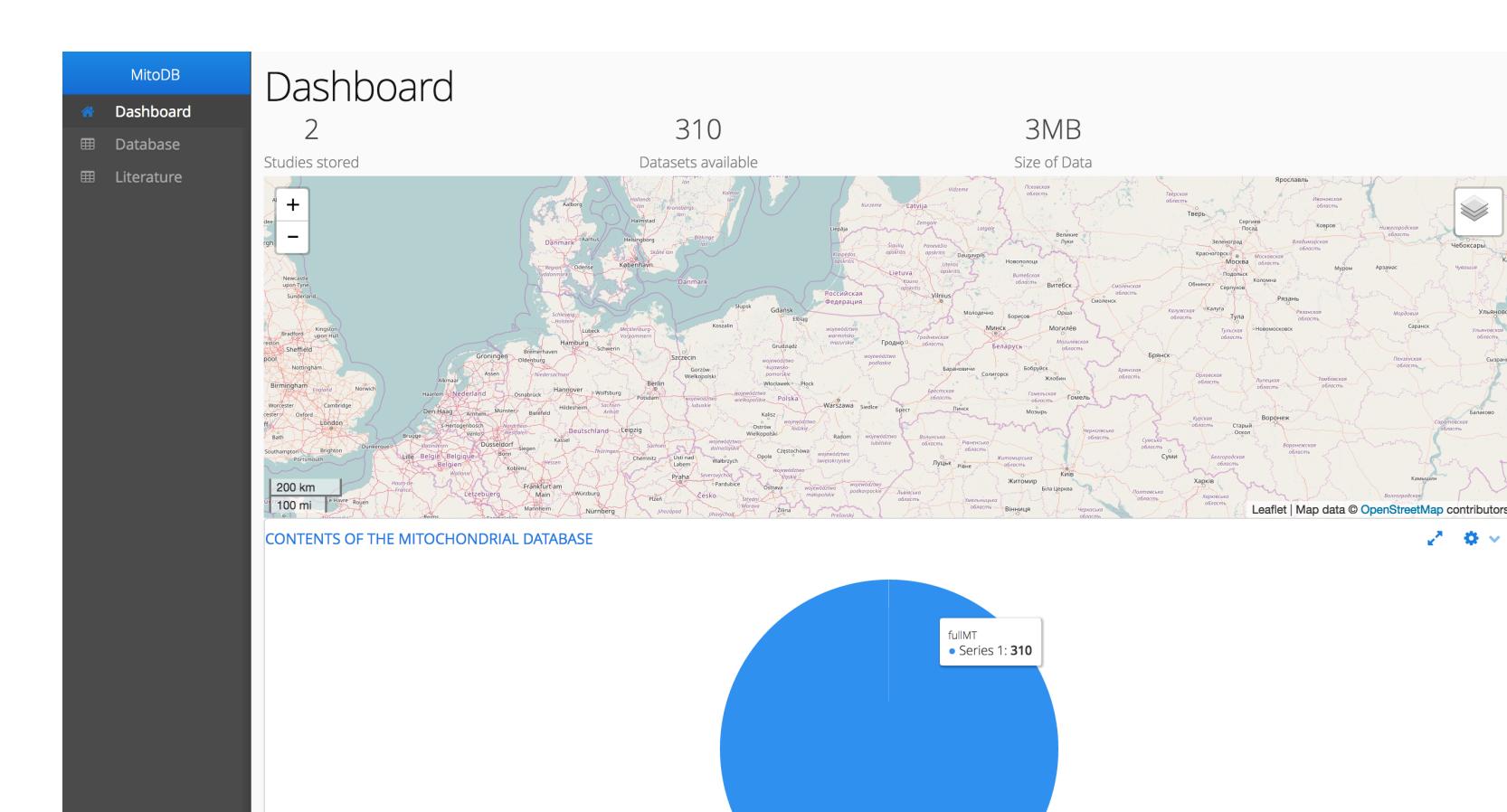
### mitoBench

- Provide methods for downstream analysis (e.g.  $F_{ST}$  calculations, Founder Analysis)
- Offer more visualizations (e.g. geographical maps for origin of samples)



### mitoDB & web interface

- Export/import functionality
- Web-based dashboard to explore database
- More public datasets: 1000G, GenBank, ...



## References

1. Drummond, A. J.; Suchard, M. A.; Xie, D.; & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST. *Molecular biology and evolution*, 29(8), 1969-1973.
2. Excoffier, L. & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses. *Molecular ecology resources*, 10(3), 564-567.
3. Van Oven, M. & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human mutation*, 30(2), E386-E394.
4. Schuenemann, V.J. & Peltzer, A. et al. (2017). Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. *Nature Communications*, 15694.