Haploid DNA Markers in Forensics, Berlin, April 22-24, 2010

EMPOP 2

Walther Parson
Institute of Legal Medicine
Innsbruck Medical University
Austria

Snapshot of EMPOP´s history

1999 - project plan at the ISFH Meeting in San Francisco

Only small, local mtDNA databases, some erroneous

Lack of harmonization (typing strategies, nomenclature, ...)

2000-2006 - Milestones

1) Inventory of existing data(bases)
2) Identification of problematic areas in forensic mtDNA analysis
3) Research for improved laboratory protocols and database management
4) Programming of database and website

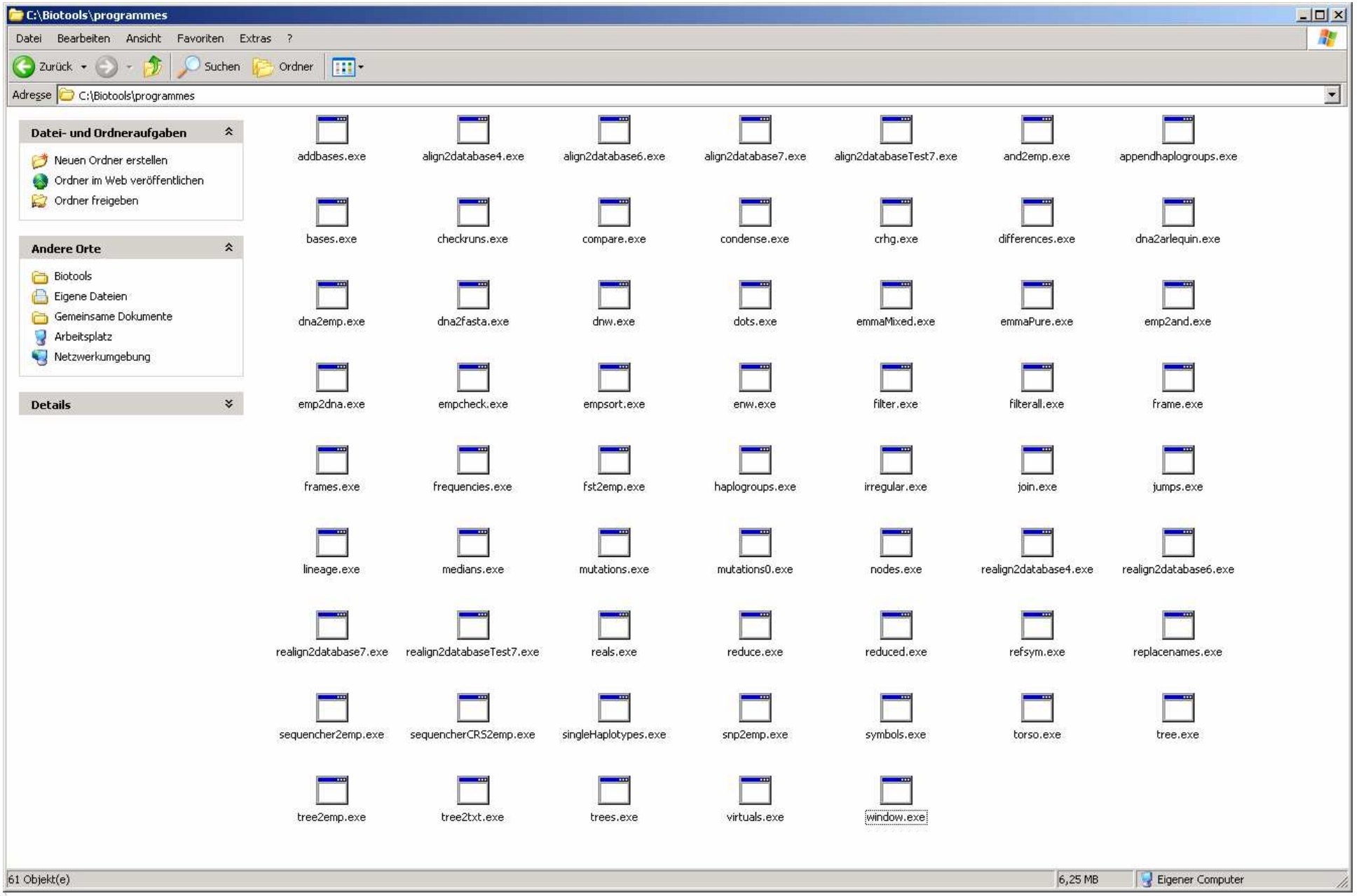
2006 - EMPOP 1 launch (N=5.173)

2006-2010 - Milestones

1) Additional tools for quality control - complete V1 review
2) String alignment
3) Output - Statistics

2010 - EMPOP 2 launch (N=10.970)

Software aiding QC of mtDNA data



Database updates

Re-analysis of EMPOP1 (N=5.173)

Summary of changes EMPOP 1 -> 2:

- 3 haplotypes deleted (close relatedness)
- Change of one mutation in 2 haplotypes (0.039% error rate)
- 10 changes in alignment (phyloTree)

Changes will be specified in release history

Continuous data control and transparency - quality criterion

Alignment of mtDNA

Cut the Gordan knot about mtDNA nomenclature discussion

Reporting and database searching are dissociated

Procedure:
Query sequence -> string of sequences

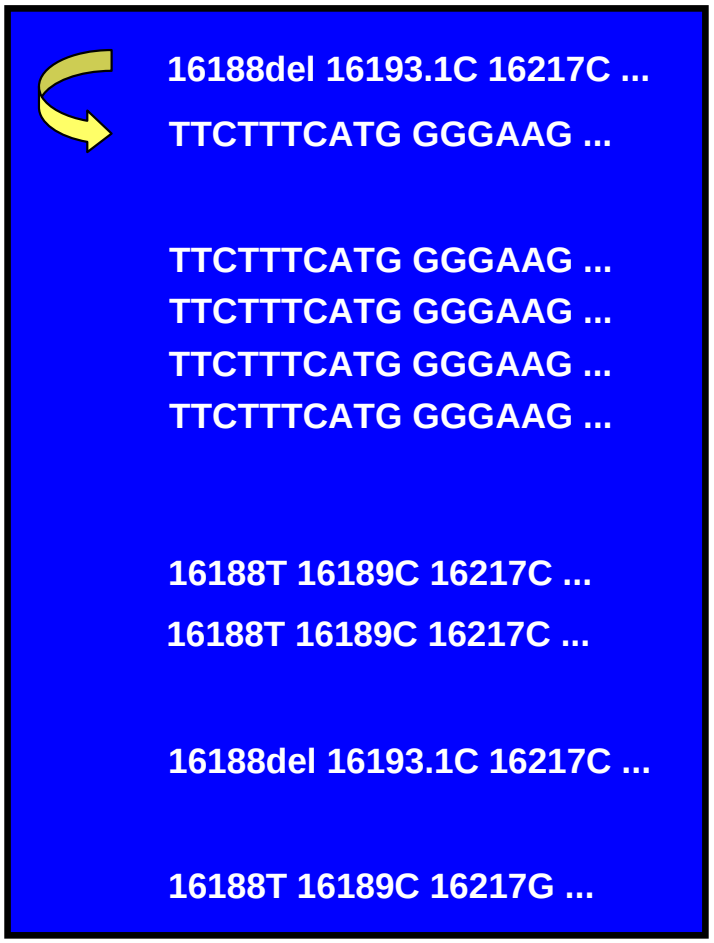
Database sequences -> strings of seq.

Search procedure - result

Transcripts of (near) matching strings
- Maximum weighted parsimony
- Maximum likelihood

Query is echoed

Database annotation in details



STRING Alignment in EMPOP 2

Calculation of (near) matching transcripts is quite time consuming

Hotspots (around 16193, 309, ...) can be ignored in a query, which increases calculation time, as the additional distance needs to be taken into consideration for the query (selected hotspots are then ignored in the output)

Even though the performance of the query software has been optimized a search takes longer than a difference-coded haplotype comparison (20`` - 50``)

Compromise of time and search quality. The latter is certainly better as the search is independent of alignment and finds matching sequences irrespective of the notation

The EMPOP haplotypes follow phylogenetic nomenclature (phyloTree)

Output - Statistics

1) Counting Method

2) Frequency

uncorrected frequency (95% CI)

“n+1 rule” (95% CI)
(e.g. SWGDAM)

“n+2 rule” (95% CI)
(Balding and Nichols, 1994)

C.L. f. Zero proportion
(Holland and Parsons, 1999)

3) Likelihood

Matching likelihood ratio
(Brenner 2010)

Acknowledgments

EMPOP Team Innsbruck

Martin Bodner
Liane Fendt
Theresa Harm
Gabriela Huber
Daniela Niederwieser
Alexander Röck
Bettina Zimmermann

former:
Sylvain Amory
Anita Brandstätter
Tanja Göbel
Jodi Irwin
Konrad Schwarz
Janine Wagner
Verena Lubei

EMPOP-IT Innsbruck

Stefan Troger, Martin Pircher
former: Andreas Erhard

external testers

Jodi Irwin
Lourdes Prieto
Sabine Lutz-Bonengel
Tomasz Grzybowski

collaborating laboratories

see website

Thomas Parsons
Hans-Jürgen Bandelt
Arne Dür

Richard Scheithauer