22nd Conference of the International Society of Forensic Genetics
Pre-congress seminar: mtDNA typing - EMPOP
Copenhagen, August 20th, 2007

EMPOP Release 2
Towards Standardization on the Use of Haplogroups in Forensic Science

Walther Parson[1], Anita Brandstätter[1], Alexander Röck[2]
[1] Institute of Legal Medicine, Innsbruck Medical University
[2] Institute of Mathematics, University of Innsbruck
Austria

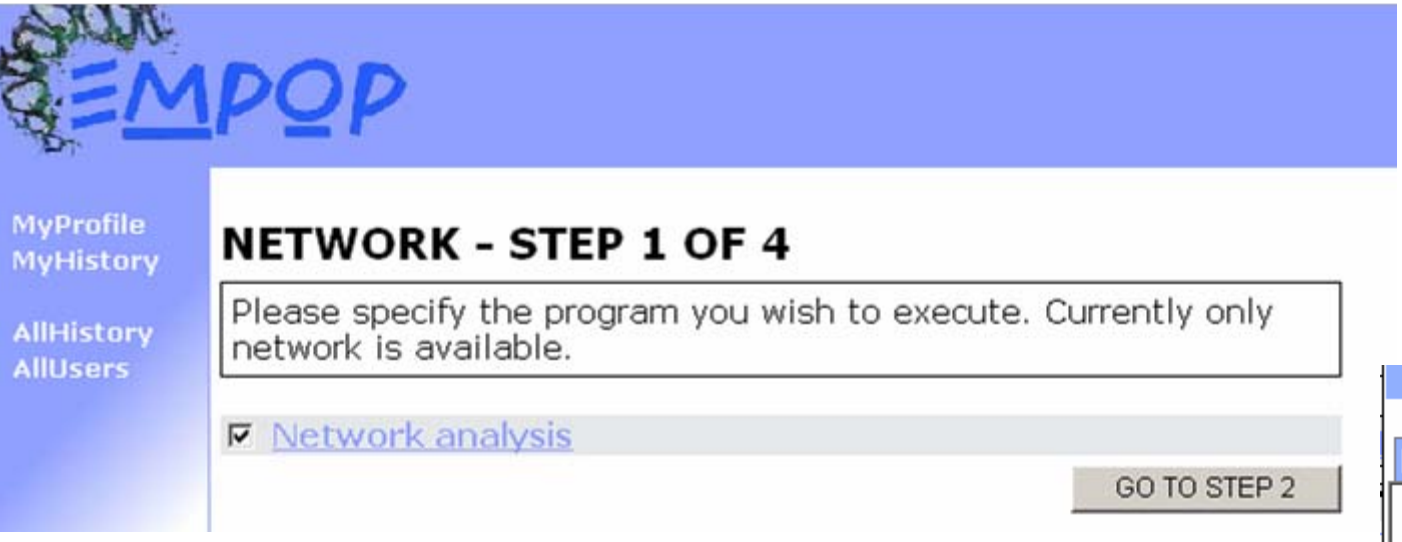
EMPOP Network Analysis

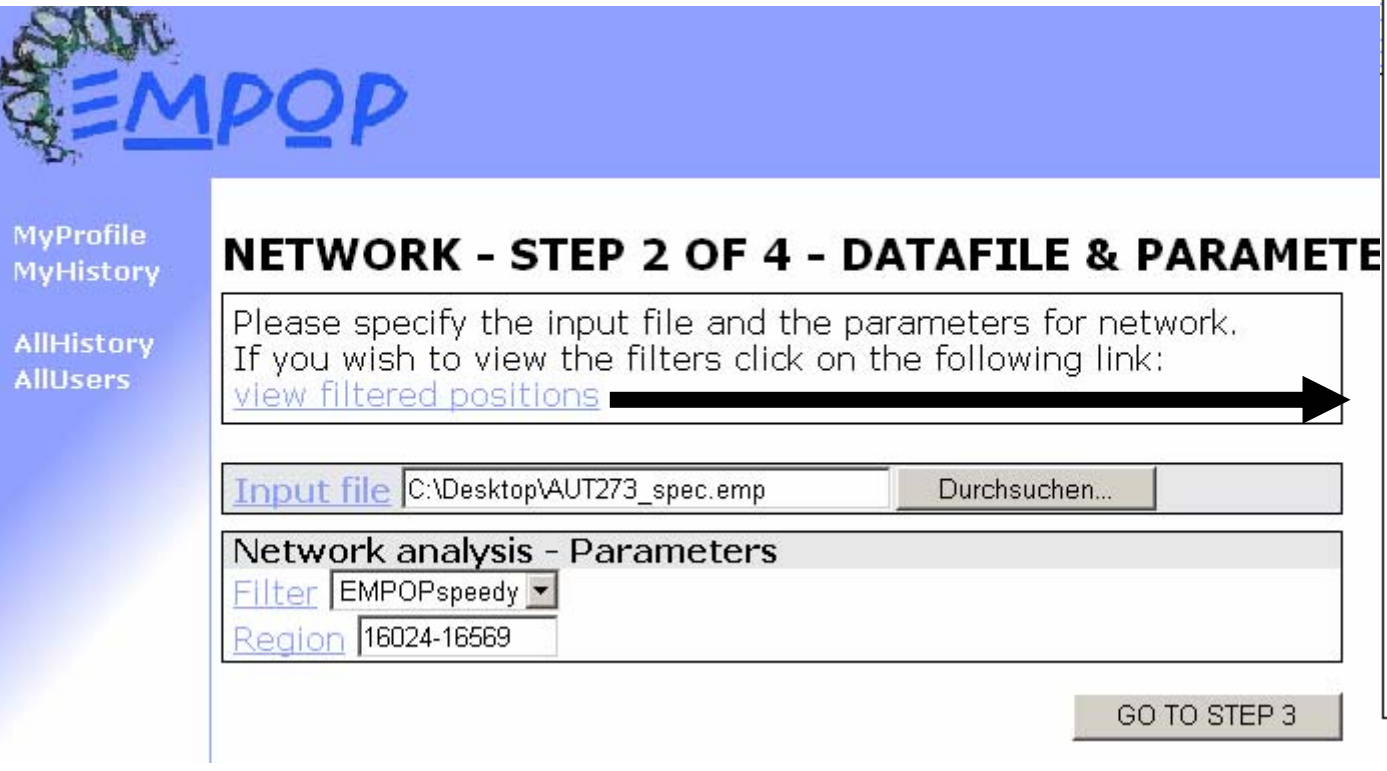
C:/Network/

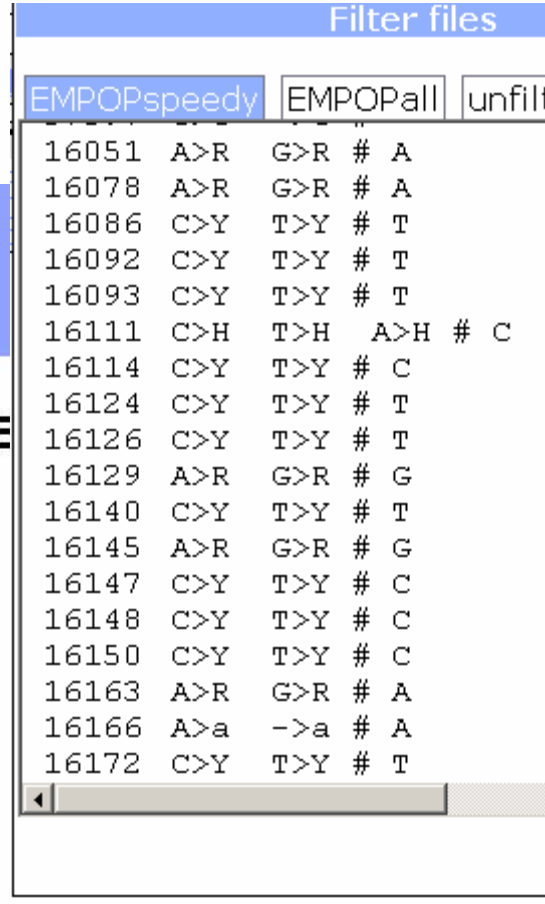
Microsoft Excel - AUT273_spec.emp

	A	B	C	D	E	F	G
1	# This is an example for the emp format. Find details in Brandstätter et al. (2006) FSI doi:10.1016/j.for						
2	# Ambiguous positions (sequence heteroplasmy) are resolved by reporting the dominant sequence v						
3	# Sample name	# Haplogroup	# Number	# Polymorphisms			
4	AUT0500000	K1a	1	73G	263G	309.1C	315.1C
5	AUT0500001	?	1	235G	263G	309.1C	309.2C
6	AUT0500002	H*	1	263G	309.1C	309.2C	315.1C
7	AUT0500003	J1c	1	73G	185A	188G	228A
8	AUT0500004	J1c	1	73G	185A	189G	228A
9	AUT0500005	H*	1	263G	315.1C	340T	523DEL
10	AUT0500006	K2b	1	73G	151T	152C	199C
11	AUT0500007	U5b	1	73G	150T	263G	315.1C
12	AUT0500008	H*	1	73G	187T	263G	315.1C
13	AUT0500009	H*	1	189G	263G	315.1C	16519C
14	AUT0500010	H1c	1	263G	315.1C	477C	16093C
15	AUT0500011	H5	1	146C	195C	263G	309.1C
16	AUT0500012	U4	1	73G	150T	195C	263G
17	AUT0500013	H1c	1	263G	309.1C	315.1C	477C

Executing EMPOP Network







Executing EMPOP Network

MyProfile
MyHistory

AllHistory
AllUsers

NETWORK - STEP 3 OF 4 - CHECK INPUT

Please check your input from the previous step.
Press "SUBMIT INPUT" if you wish to execute network.
Press "EDIT INPUT" if you wish to change parameter(s) and/or input file.

Input file AUT273_spec.emp

Parameters
Filter EMPOPspeedy
Region 16024-16569

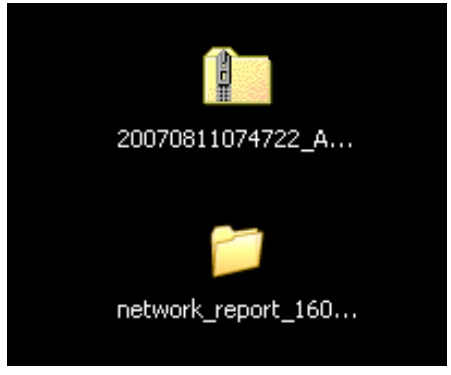
MyProfile
MyHistory

AllHistory
AllUsers

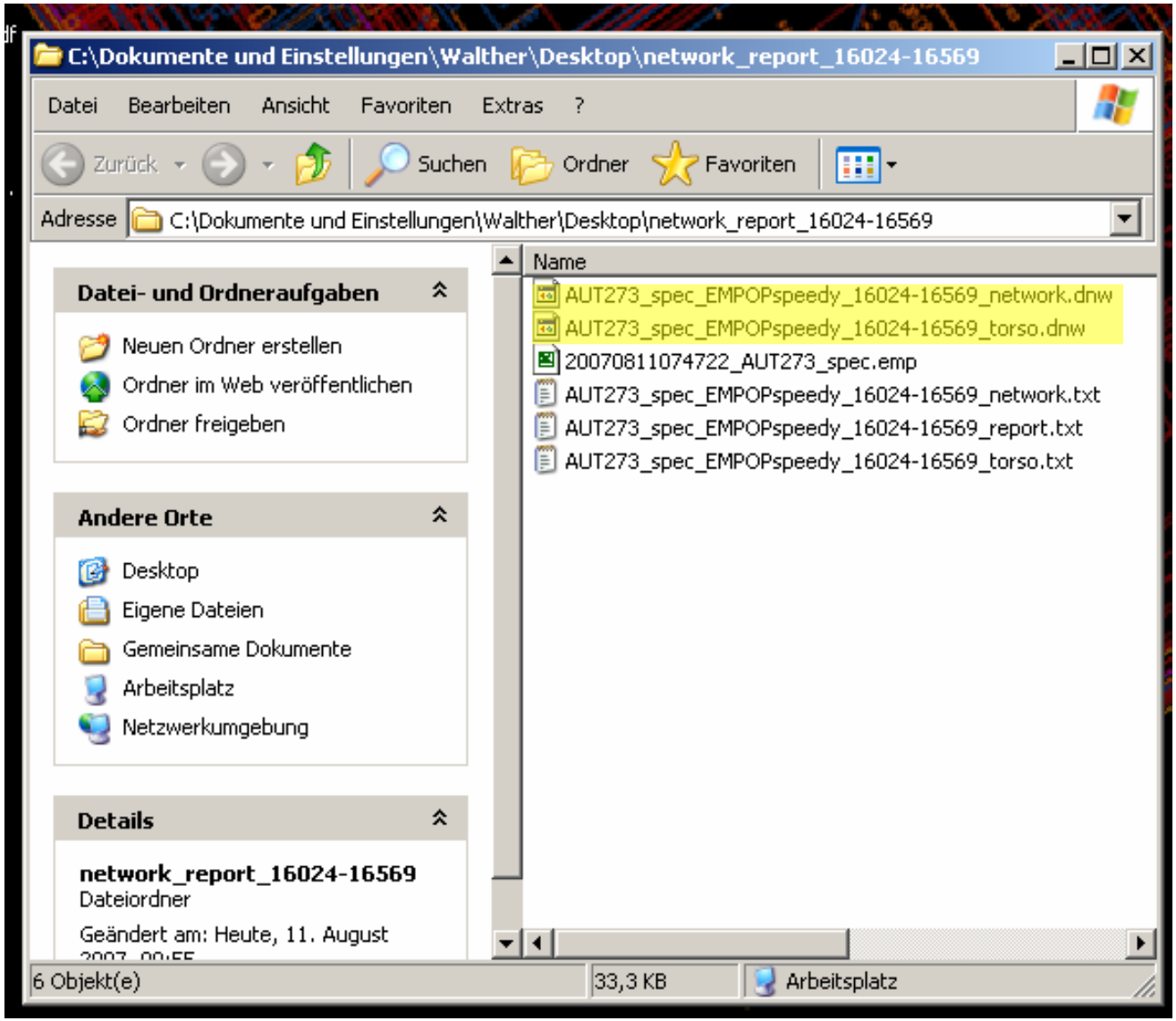
NETWORK - STEP 4 OF 4 - RESULT

Thank you for using network.
It took 2 updates ≡ 16 seconds.

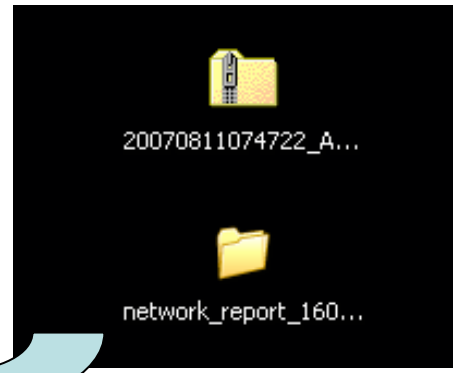
Use the link below to download the result file.
20070811074722_AUT273_spec_EMPOPspeedy_16024-16569.zip (zip, 8.581 kb)



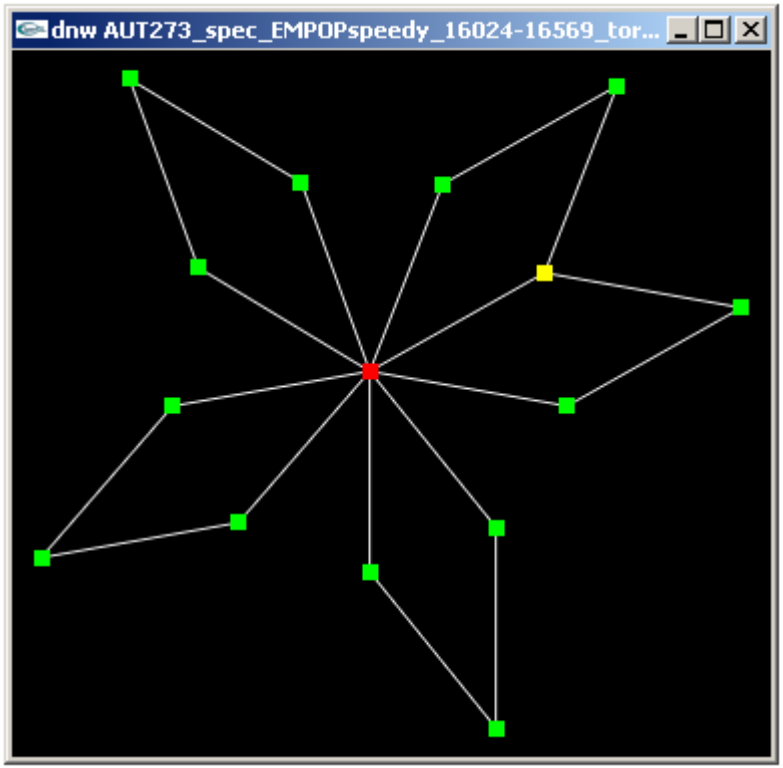
Drawing the Network

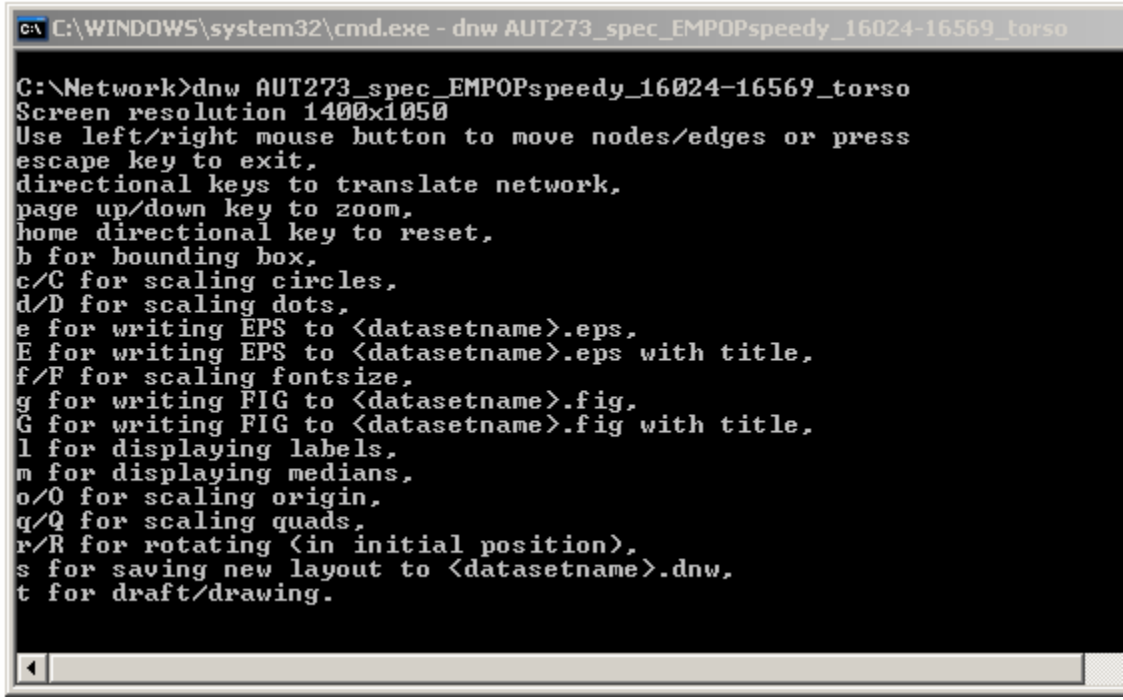


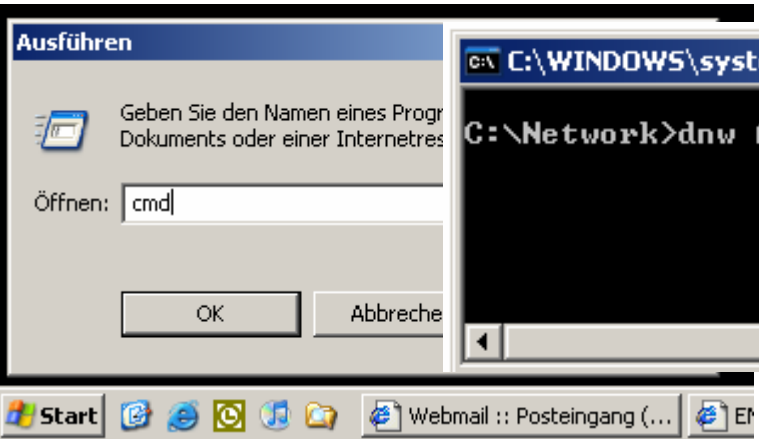
C:/Network/



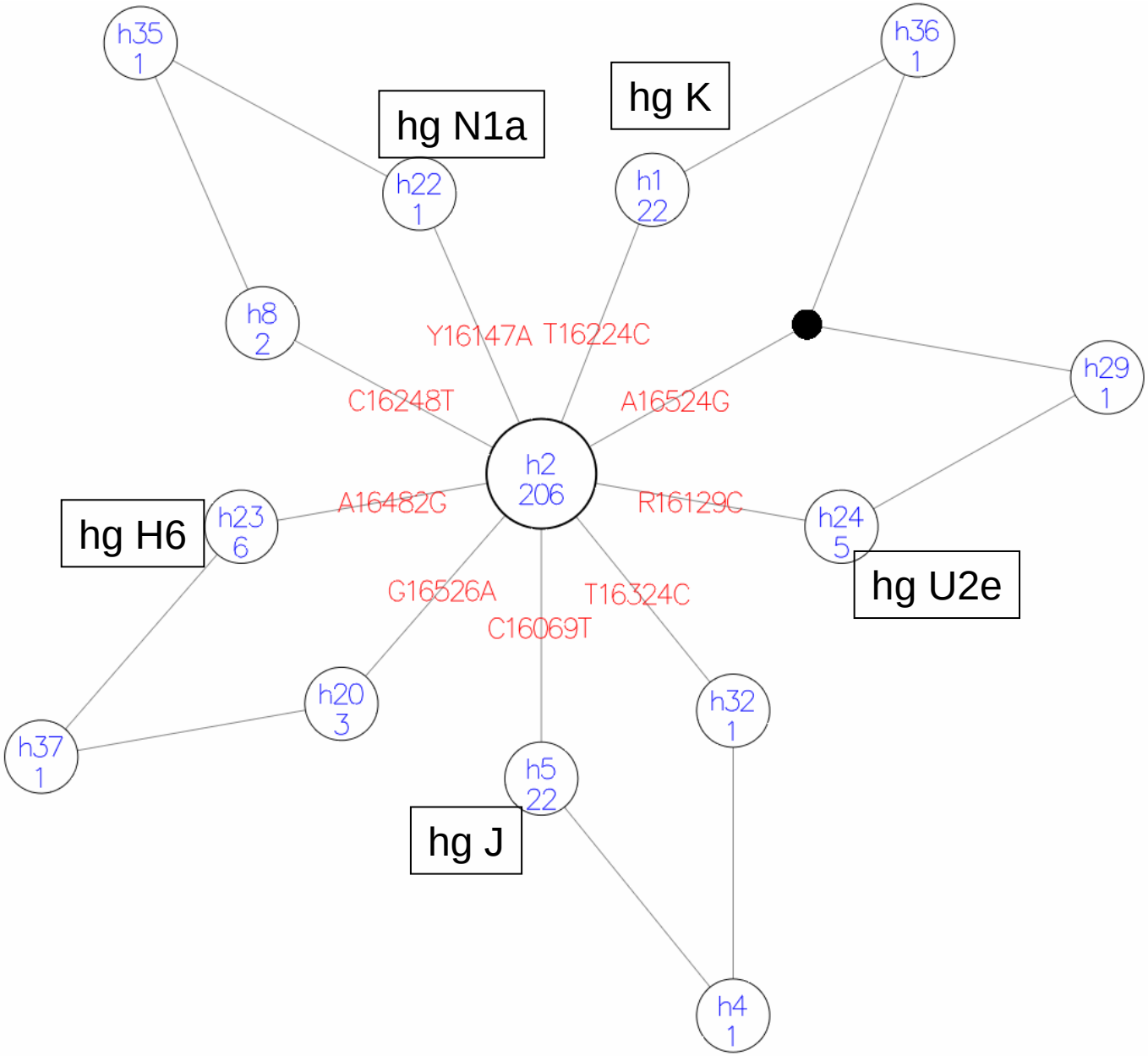
Drawing the Network – Torso HVS-I



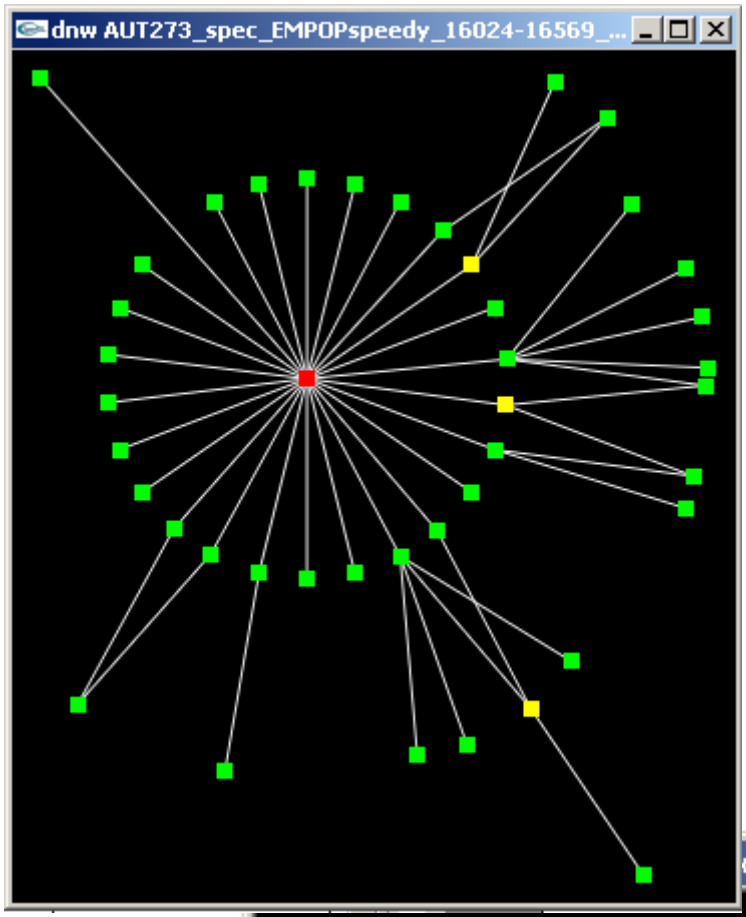


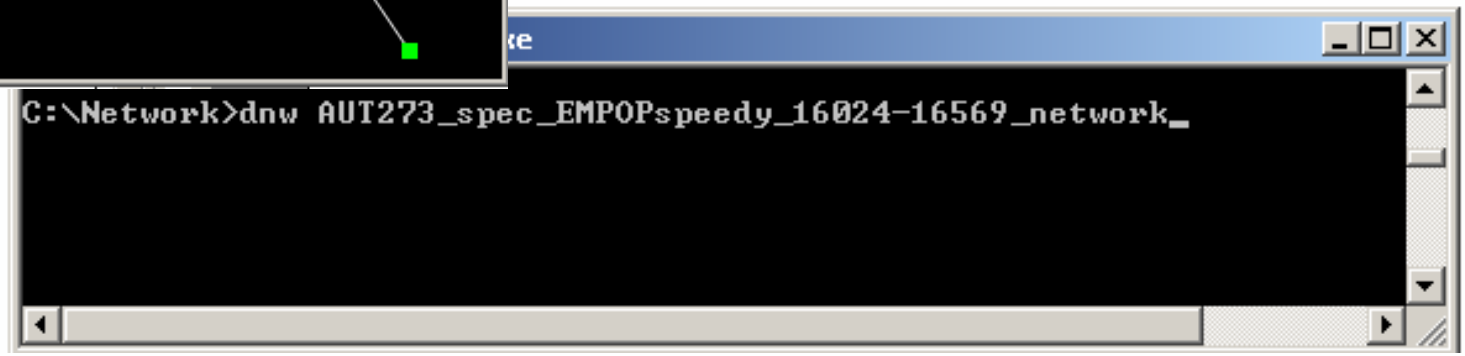


HVS-I Torso Interpretation

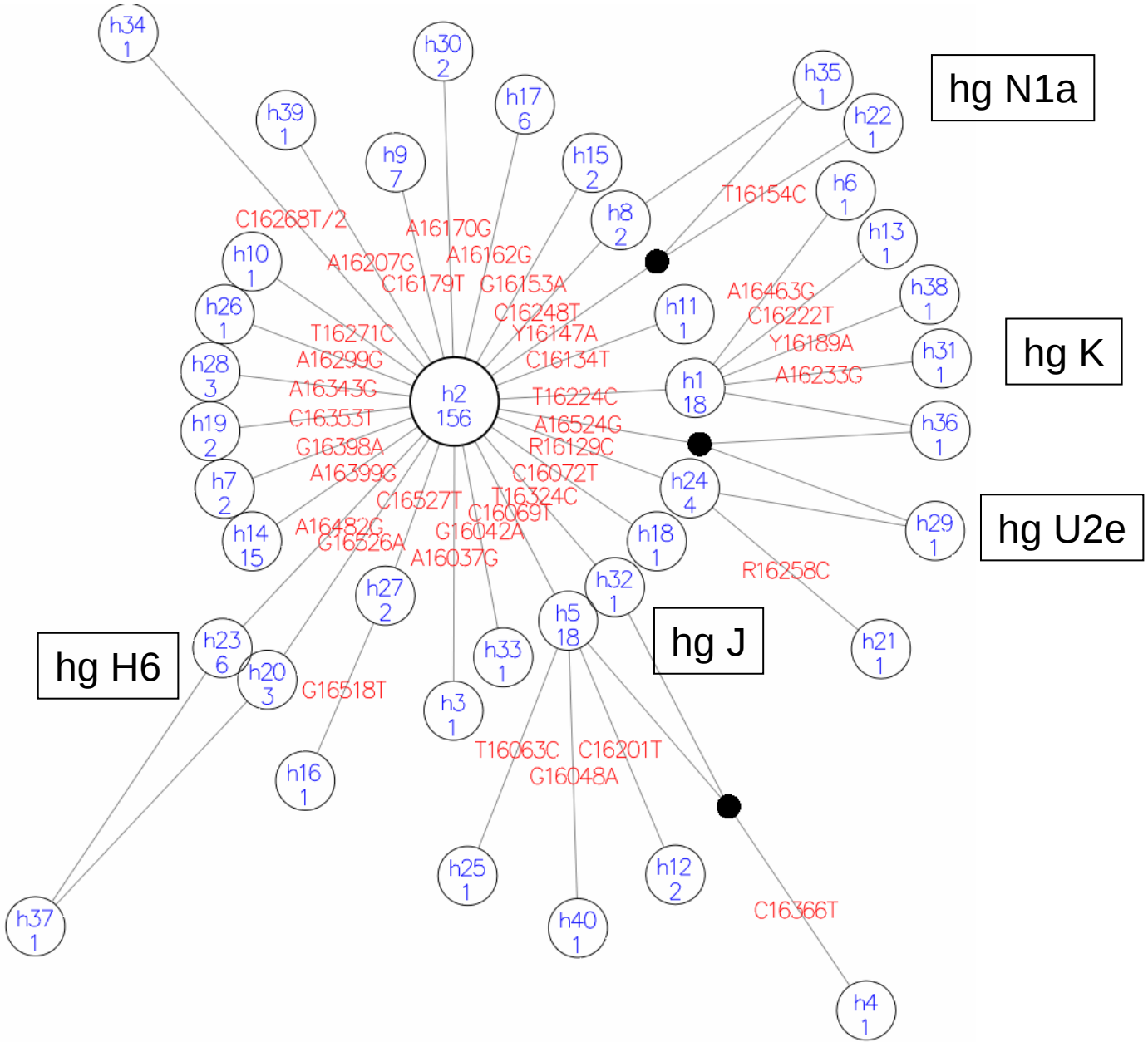


Drawing the Network - entire HVS-I Network





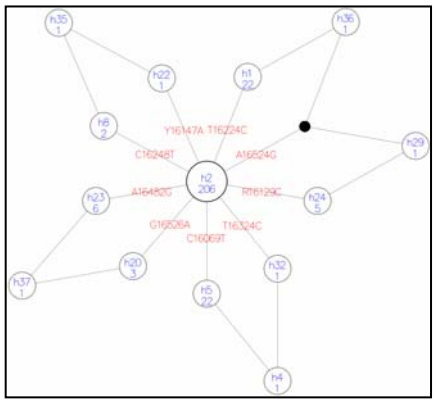
HVS-I Network Interpretation

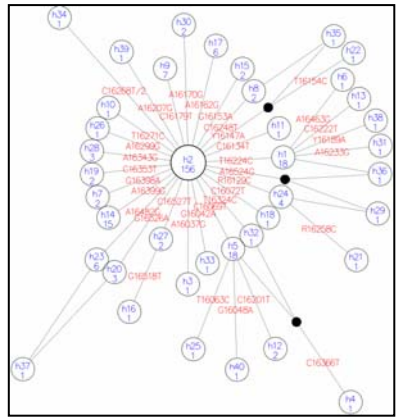


Torso - Network

torso – network without the individual haplotypes; rough overview, less complex

network – contains all unfiltered mutations; harbours all most parsimonious reconstructions for any tree
Highlights clerical errors, phantom mutations, unusual alignment, reference bias, wrong reference sequence





ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Molecular Phylogenetics and Evolution 42 (2007) 256–271

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Translating DNA data tables into quasi-median networks for parsimony analysis and error detection

Hans-Jürgen Bandelt [a,*], Arne Dür [b]

Let's test an example – 50 sequences from South Asia

Microsoft Excel - ExampleSAsia_50.emp

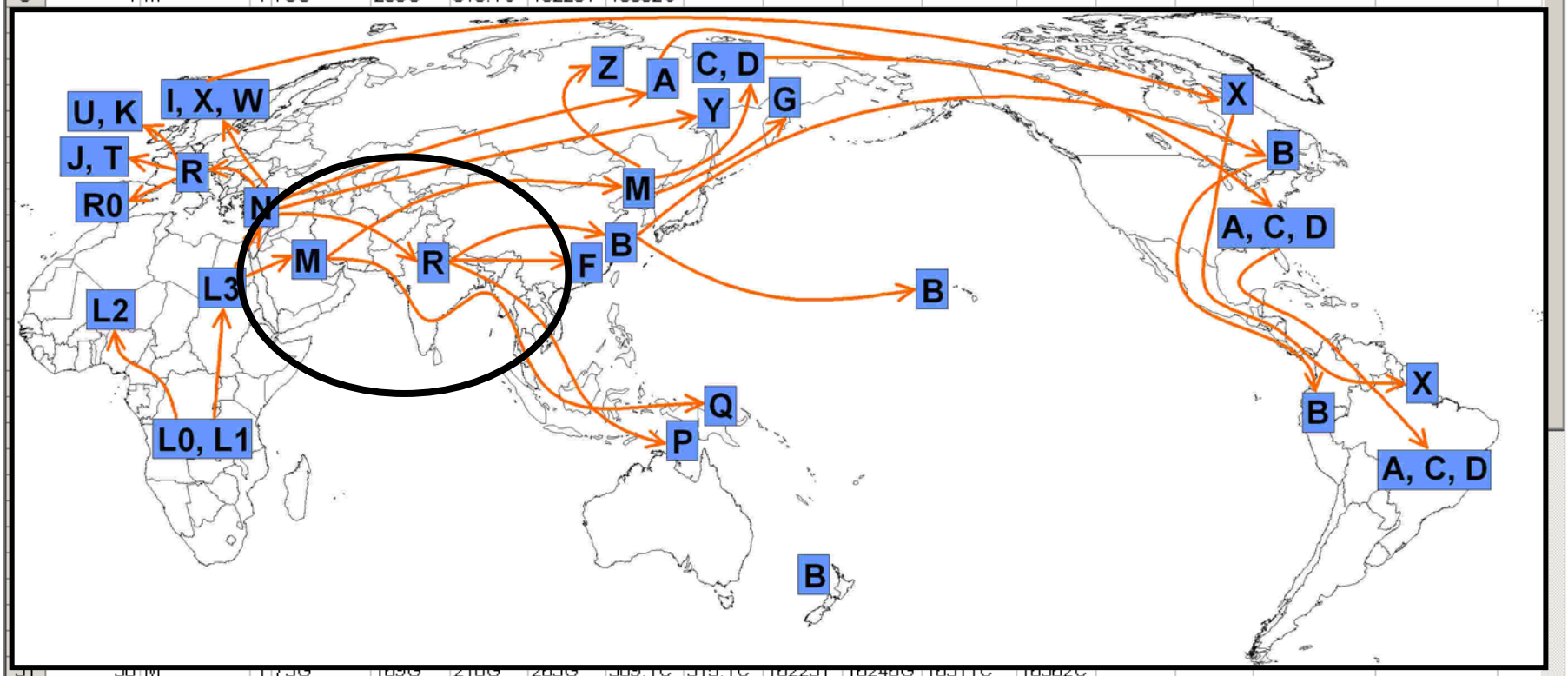
Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster ? Adobe PDF

Frage hier eingeben

100% Aus Arial 10

AC33

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	# Sample	Haplogroup	f	Haplotype													
2	1	M7	1	73G	150T	195C	263G	315.1C	16192T	16223T	16297C						
3	2	M	1	73G	106A	263G	298T	309.1C	315.1C	16223T	16274A	16262.1N					
4	3	M7	1	73G	185A	263G	315.1C	16223T	16272G	16297C	16362C						
5	4	M	1	73G	263G	315.1C	16223T	16362C									
32	31	N	1	73G	263G	315.1C	16189C	16223T	16355T								
33	32	N	1	73G	146C	263G	309.1C	315.1C	16189C	16223T	16355T						



ExampleSAsia_50

HVS-II Torso

Report...

Haplotype designation in the network:

h1: 1 3 4 6 8 9 11 12 15 16 21 24 29 31 32 33 34 35 36 37 41 45 46 47 48 49 50

h2: 2

h3: 5 7

h4: 10 14 27 28

h5: 13 19 20 23 39 42 44

h6: 17

h7: 18

h8: 22

h9: 25

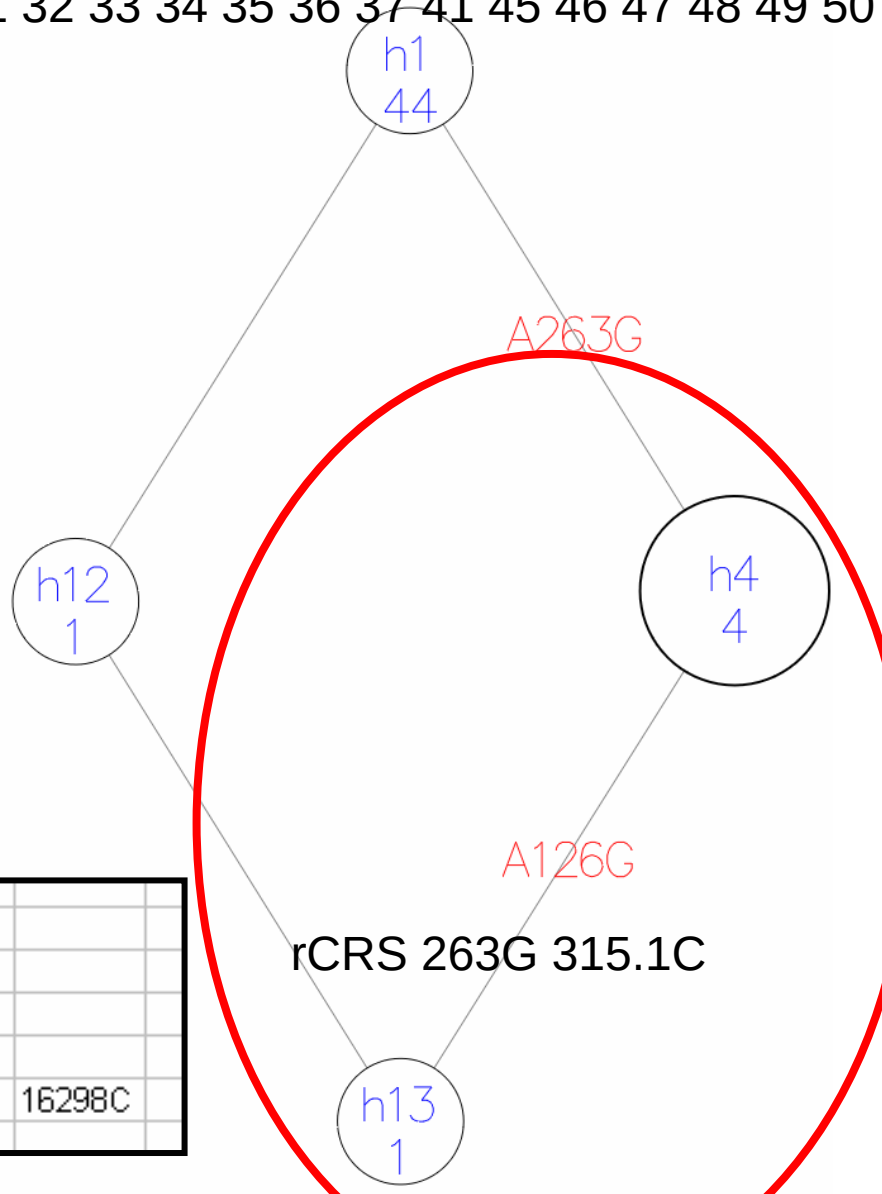
h10: 26

h11: 30

h12: 38

h13: 40

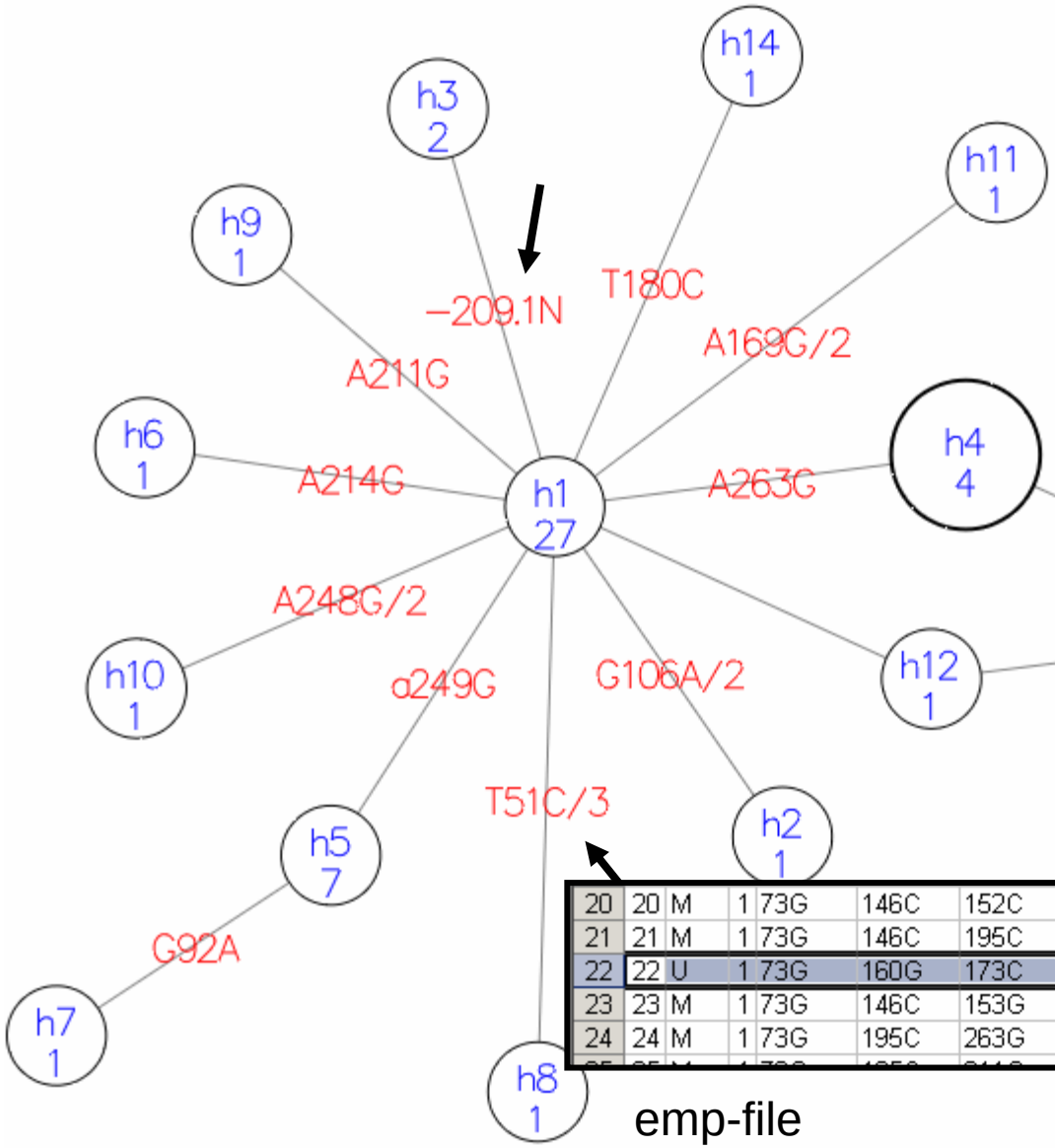
h14: 43



emp-file

13	10	R	1	16213A	16246G	16311C	16362C				
14	14	M	1	16179T							
15	27	M	1	16223T	16258G	16261T					
16	28	M	1	16223T	16266T						
17	40	M8	1	126G	152C	235G	315.1C	16223T	16260T	16298C	

HVS-II entire Network



Network legends:

Character designation in the network: Positions with identical partitions are merged into weighted characters. Each weighted character is represented by its first position in the following lists. The symbol preceding the position specifies the Anderson nucleotide, the symbol following the position shows all observed nucleotides using the IUB code where small letters indicate additional deletions.

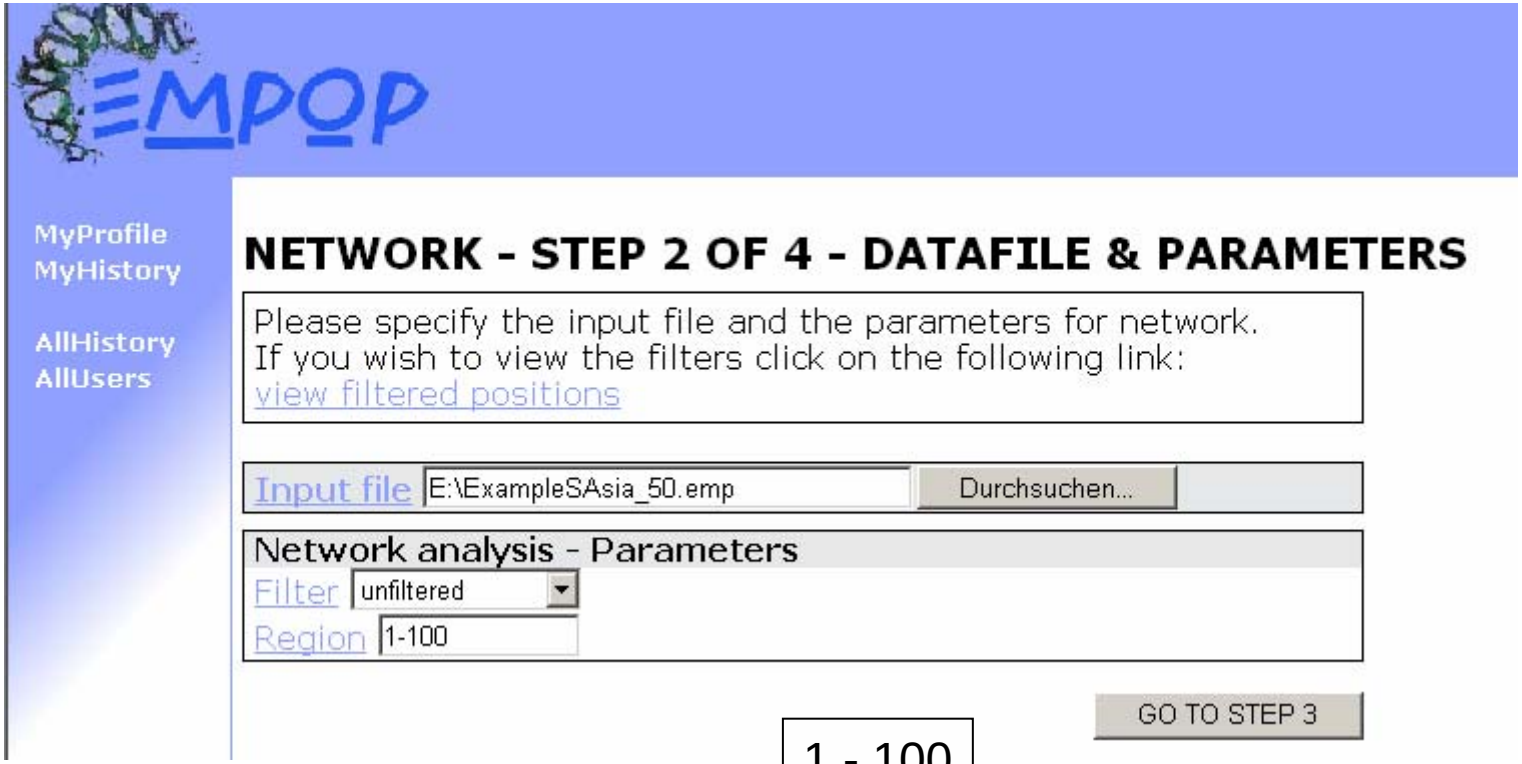
T51Y A160R T173Y
G106R C298Y
A169R A210R
A248R A287R
(5 duplicates)

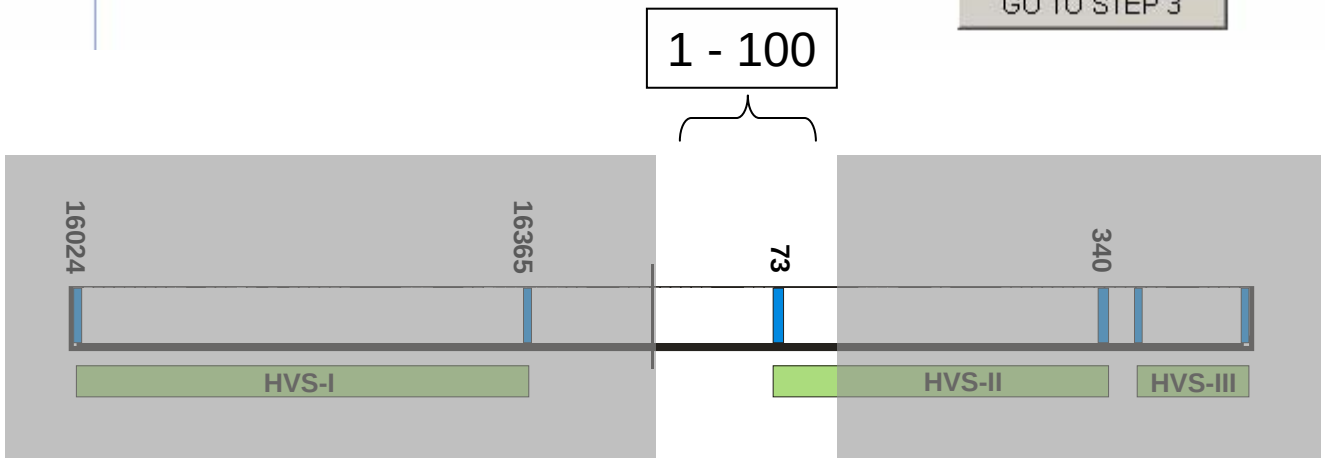
20	20	M	1	73G	146C	152C	249G	263G	315.1C				
21	21	M	1	73G	146C	195C	263G	309.1C	315.1C	16136C	16217C	16223T	16319A
22	22	U	1	73G	160G	173C	204C	263G	315.1C	51C	16168T	16224C	16326G	16357C
23	23	M	1	73G	146C	153G	204C	249G	263G	16223T	16398A			
24	24	M	1	73G	195C	263G	315.1C	16223T	16246G					

emp-file

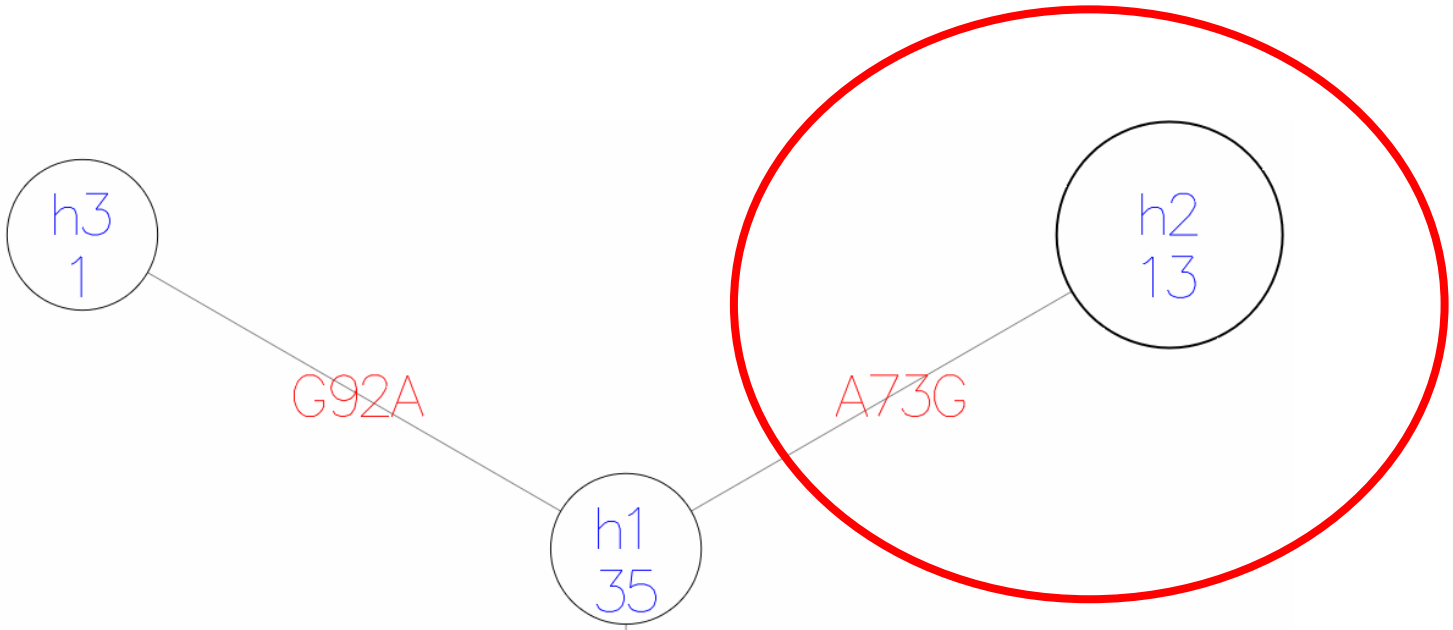
16051?

Detailed analysis – unfiltered – 1-100





Detailed analysis – unfiltered – 1-100

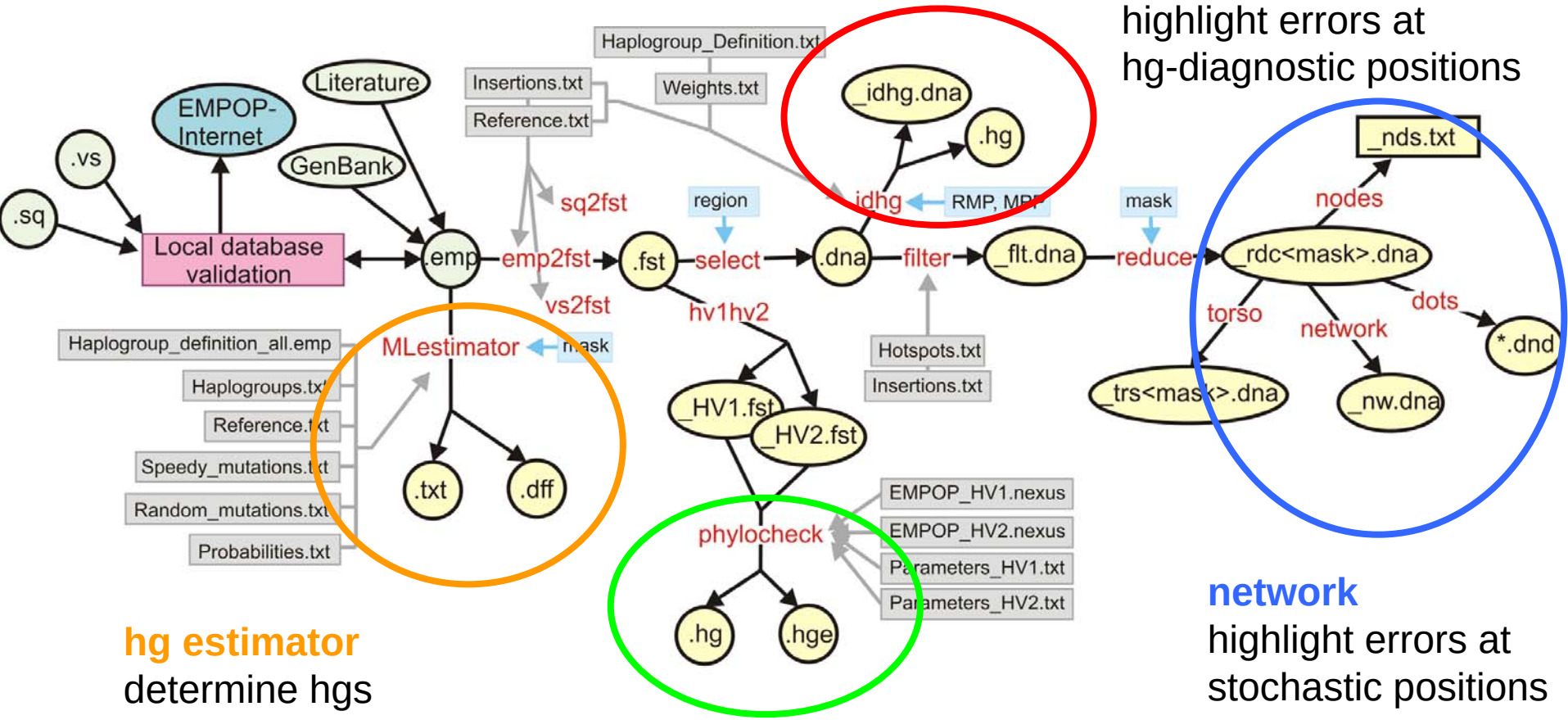


Report
Haplotype designation in the network:

h1: 1 2 3 4 5 6 7 8 9 11 12 13 16 17 19 20 21 23 24 25 26 29 30 31 32 41 42 43 44 45 46 47 48 49 50
h2: 10 14 15 27 28 33 34 35 36 37 38 39 40

51	33	A	1	152C	235G	263G	309.1C	315.1C	16223T	16274A	16290T	16304C	16319A	16356C	16362C
52	34	M	1	263G	309.1C	315.1C	16189C	16223T	16301T	16362C					
53	35	M8	1	146C	153G	263G	315.1C	16223T	16298C	16327T					
54	36	N	1	152C	235G	263G	315.1C	16129A	16223T	16252G	16284G	16290T	16319A	16362C	
55	37	N	1	152C	235G	263G	309.1C	315.1C	16129A	16223T	16319A	16362C			
56	38	M8	1	126G	150T	152C	263G	309.1C	315.1C	16185T	16189C	16193T	16223T	16260T	16298C
57	39	M9	1	152C	249G	263G	309.1C	315.1C	16145A	16223T	16316G				
58	40	M8	1	126G	152C	235G	315.1C	16223T	16260T	16298C					
59															

EMPOP bioinformatics chart



haplogroup-ID
highlight errors at
hg-diagnostic positions

network
highlight errors at
stochastic positions

hg estimator
determine hgs
find omitted mutations

phylocheck
find artificial HV1/HV2
recombinants