# Drawing and interpreting quasi-median networks with EMPOP DrawNetWork v1.24

# A short introduction

1. Quasi-median networks basics

2. How to perform a QM network analysis

3. How to interpret QM networks

4. Networking via EMPOP means that...

5. Short keys

Bettina Zimmermann, Alexander W. Röck, Walther Parson
Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria
http://empop.org
info@empop.org
September, 2012

# 1. Quasi-median (QM) networks basics

Quasi-median networks are used here to effectively check the quality of mtDNA data.

The application of quasi-median networks enables the examination of mtDNA haplotypes by graphically representing the genetic structure of the lineages in the data set. Hence, data idiosyncrasies in regard to homoplasy and potential artifacts which are caused by sequencing or data interpretation problems can be detected. Filtering of highly recurrent mutations prior to network analysis is required to reduce the complexity of the resulting network. The quasi-median network displays the filtered, condensed and reduced haplotypes of the analyzed data set. For the torso of the network all pendant subtrees are collapsed into their base nodes. Use the **torso** for a **quick data check**, the **network** is useful if a **closer look** on the data is preferred.

Here the quasi-median network torso of 202 mtDNA haplotypes from the example data set WE_Etalon with the analysis range set to 16024-16569 and filtered with the WE filter is displayed.
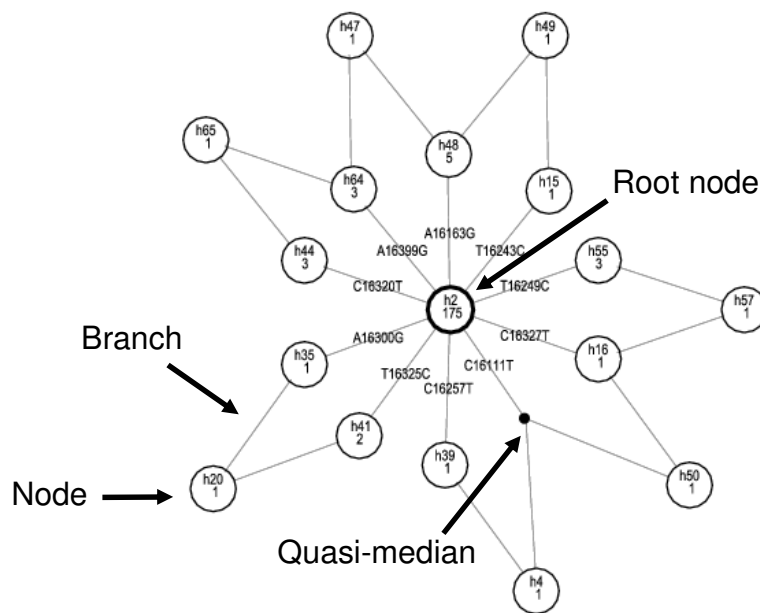


→ The **nodes** correspond to filtered, reduced and condensed haplotypes.

→ The **root node** is displayed in bold; with West Eurasian data it usually corresponds to the rCRS haplotype;

→ The **branches** represent mutational events and are read starting at the root node; mutational events given on the branches are colored either in black for transitions or in red for all other mutations (insertions, deletions and transversions); parallel branches carry the same mutation and are labeled only once.

→ **Quasi-medians** are represented by small black dots and represent virtual haplotypes; quasi-medians are required to link haplotypes within a QM network and are calculated by the network software.
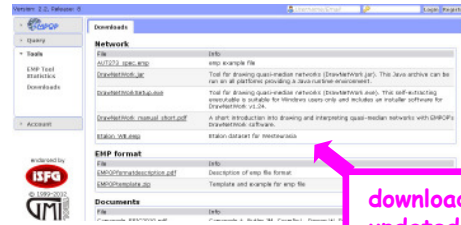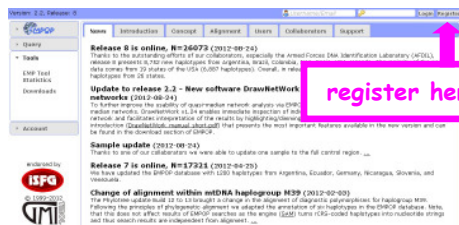
# 2. How to perform a QM network analysis?

**Preparations**
Register and login at http://empop.org
Download, unzip and install the drawing program DrawNetWorksetup.zip from the section „Tools → Downloads"
Prepare your data in emp file format (see „EMPOPtemplate.zip in the „Downloads"-section)



register here



download info and updated drawing software here

**Perform the network analysis**
Open the Network link in the „Tools"-section
Input your emp file and the required analysis settings (i.e. filter, analysis range) and proceed
Download and unzip the resulting zip file that includes the analysis results

**Draw the network**
By double-clicking the „*_network.dnw" file (or drag and drop) the image of the network will automatically be drawn by the installed drawing software DrawNetWork. The same for „*_torso.dnw". The quasi-median network (A) displays the filtered, condensed and reduced haplotypes of the analysed data set, for the torso of the network (B) all pendant subtrees have been collapsed into their base nodes which facilitates identification of complex structures.



A

B
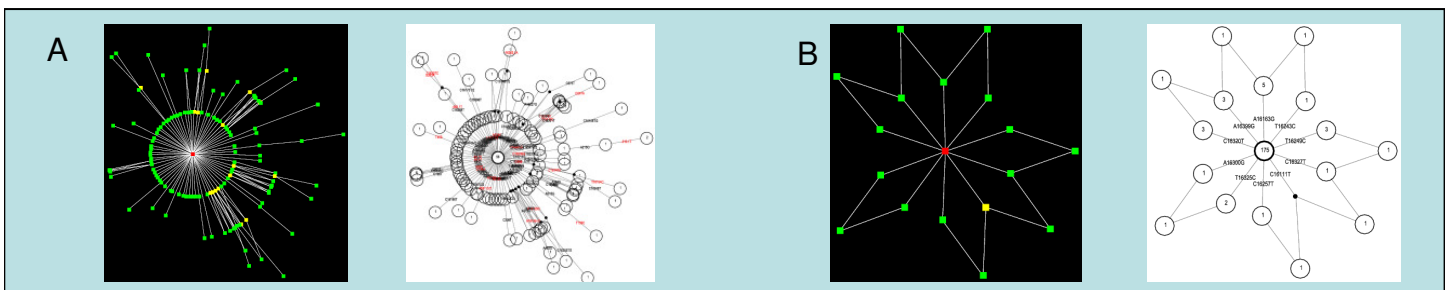
**Configure the network -** Change the appearance of the network/torso
Pressing „t" changes the draft version to the drawing version, „l" changes the labeling of the nodes and coloring of the mutations, „h" changes the labels by adding haplogroup information



t    l    h
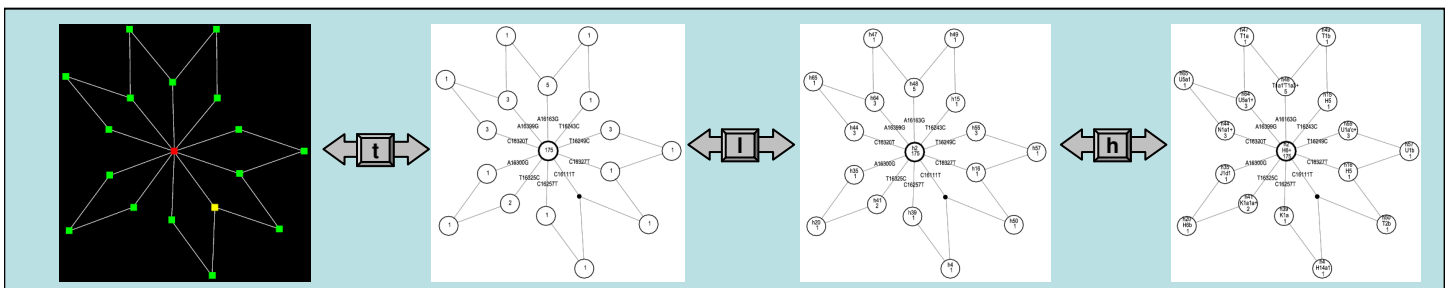
Move nodes (left mouse button) and branches (right mouse button) by clicking and holding the center of nodes and labels, respectively. Further modifications see 5. Short keys

# 3. How to interpret the QM network?

## Example 1 – Confirming mutations

Two samples of the West Eurasian Etalon data set contain the transition **C16111T**. In the network torso the two samples are named **h4 and h50**. The two samples are separated by the mutations **C16257T** (h4) and **C16327T** (h50). To connect these two samples in this network the quasi-median is required.

> **Quasi-medians or complex reticulations in the network may indicate the presence of ambiguities in the data. Inspection of the raw data is indicated!**

## Where do I find information about the haplotypes shown in the network?

The new GUI allows immediate check of included haplotypes h4 and h50
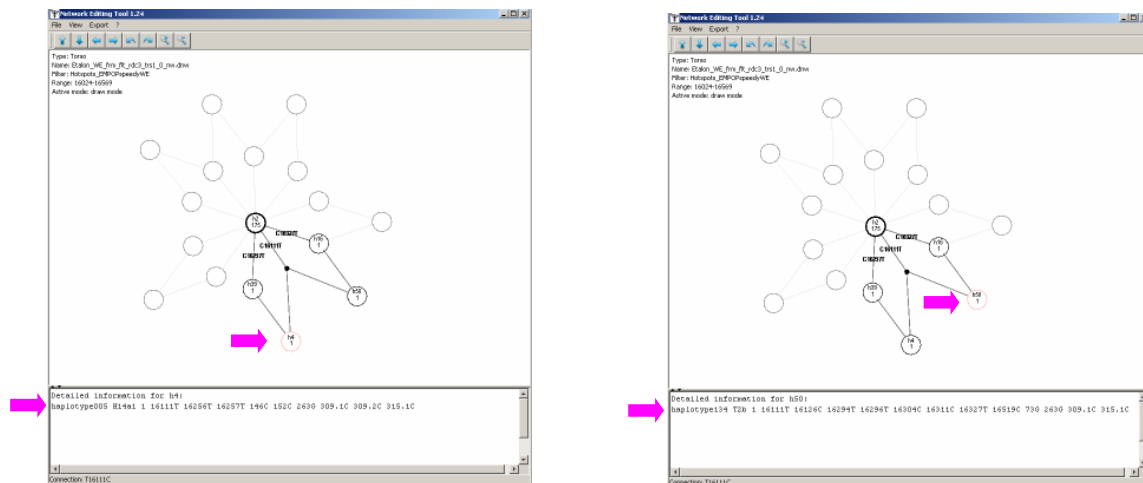


h4 and h50 both show mutation 16111T that causes a reticulation in the torso of the QM network

Check the raw lane data of samples „haplotype005" and „haplotype134" to confirm the transition C16111T in both samples



| **h4** = haplotype005 (Hg H14a1) | **h50** = haplotype134 (Hg T2b) |

→ the virtual haplotype (quasi-median) and the corresponding mutation is confirmed

# 3. How to interpret the QM network?

**Example 2 – finding errors**

This example shows a complex network of 320 Westeurasian haplotypes that were submitted to EMPOP for quality control (QC). The EMPOPall filter was applied, that results in a simple QM network or a single node. The high number of quasi-medians and reticulations indicate ambiguities.

> **Quasi-medians or complex reticulations in the network may indicate the presence of ambiguities in the data. Inspection of the raw data is indicated!**
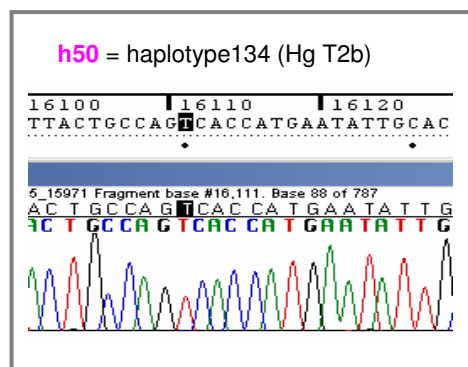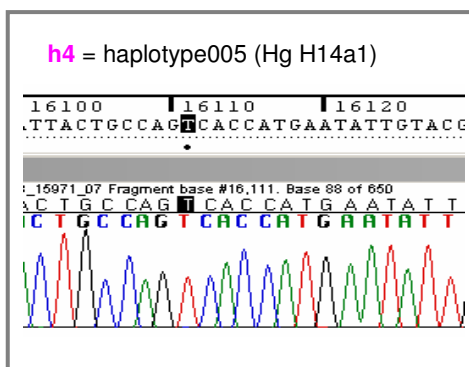


Nodes that are linked with quasi-medians are of primary interest. The node including 15 haplotypes (arrow) reveals mutation 366A that is causing this complex structure.



These 15 haplotypes and corresponding mutation patterns are displayed within the new GUI. The raw data indicate that 366A is indeed a phantom mutation. The G at 366 is masked by interfering signals originating from length heteroplasmy around position 310 with 7C and 8C length variants.

→ **366A represents a phantom mutation!**

# 4. Networking via EMPOP means that...

- Registration and software download/installation is required only once

- The calculation and drawing of a network can be repeated any time directly via EMPOP

- Network analyses are stored in your account history and can be reviewed at a later date (or deleted)

- Large data sets (> 300-500; depending on the population structure) may cause complex networks, especially when different major phylogenetic lineages are present (e.g. hgs L and M and N)

- You can use the high quality Etalon data set available in the download section to embed your data (every difference to the Etalon network should be evaluated)

- You should use different analysis ranges up to the entire control region to better understand the quality of your data (depending on sample size)

- Complex reticulations in the network can be caused by
  **errors** → check mutations that connect quasi-medians with nodes
  **sample size** → reduce sample size and try again
  **phylogeny** → remove phylogenetically distant haplotypes and/or use the appropriate filter and try again

- You should filter your data according to their phylogenetic background (e.g. Westeurasian data needs the Westeurasian filter); try the "EMPOPall" filter to pinpoint unobserved mutations; additional filters are added regularly

- Editing your network with the drawing software trying to find those samples within the network that are responsible for reticulations, quasi-medians and unknown mutations

- There is also a more detailed manual available via EMPOP in case you are curious now that you have read this short introduction! And last but not least, contact us via info@empop.org if you have detailed questions concerning your network analysis

# 5. Short keys

Move the camera: arrow keys
Zoom in: +
Zoom out: -
Center the camera: c
Bounding box: b
Rotate the graph: r/R

Export current layout as EPS: e
Export current layout as SVG: s
Export current layout as GIF: g

Change font size: f/F
Change font size of connections: x/X
Change the size of quasi medians: q/Q
Change the size of nodes in draft mode: n/N
Change the size of nodes in draw mode: d/D

Toggle display of quasi medians: m
Toggle display of haplogroups: h
Toggle display of labels: l
Toggle display of mode: t
Toggle display of logo, date, and time: i
Toggle display of description: p
Toggle display of legend: o
Toggle display of multiple connections: j

Reset draw settings: u
Exit application: ESC