

## Visualizing a Mitochondrial Haplotype Network in Gephi

Elizabeth J. Sbrocco

[National Evolutionary Synthesis Center](#)

[elizabeth.sbrocco@nescent.org](mailto:elizabeth.sbrocco@nescent.org)

November 1, 2012

### About Gephi

From the website... “Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs. Runs on Windows, Linux and Mac OS X. Gephi is open-source and free.”

In other words - Gephi is a program that allows you to visualize networks -- all kinds of networks, including haplotype networks. Note that Gephi is ONLY a visualization tool -- you still need to use a program like Arlequin to tell you how things are related first.

### Required Programs

[Arlequin](#) - version 3.5.1.3 is currently the most up-to-date version

[Gephi](#) - 0.8.1-beta is currently the most up-to-date version

Microsoft Excel

Adobe Illustrator (helpful)

### Create a Minimum Spanning Tree in Arlequin

This tutorial assumes you already have an Arlequin Project File ready to go...

- 1 - Open Arlequin - click “Open Project” and navigate to your \*.arp file to open it
- 2 - Click the Settings tab and select Calculation Settings → Genetic Structure → AMOVA
- 3 - Check the box for “Standard AMOVA computations (haplotypic format)”
- 4 - Check the box for “Computer Minimum Spanning Network (MSN) among haplotypes”
- 5 - In the first drop down menu, select “Compute distance matrix”
- 6 - In the second drop down menu, select “Pairwise difference” unless you have a reason to use a different model of molecular evolution
- 7 - Click the Start Button

Look for the MINIMUM SPANNING TREE in your out put file. Depending on the other settings you have selected, Arlequin may have computed MSTs for haplotypes within each population, but the one you want is for all of your haplotypes, so make sure that the number of OTUs equals the number of haplotypes you think you have. It is usually located just after your AMOVA results. If you chose the “pairwise difference” option above, the “connection length” is equal to the number of

mutational differences between two haplotypes. Note that this “minimum spanning tree” is only one of several possible trees that result in the same tree distance. There are other ways to connect haplotypes that result in the same size tree, and Arlequin shows you these “alternative connections” that you can use to extend your “minimum spanning tree” into a “minimum spanning network.” Some people show MSTs in the literature (Paul Barber, for example), some people prefer to show MSNs (Rob Toonen - and he uses a different program to make them). I leave it up to you whether you will include these alternative connections in your final network or not.

## Create your Gephi Input Files

- 1 - Copy and paste your minimum spanning tree (and alternative connections if desired) into Microsoft Excel. Reformat the data so that you have OTU1 in the first column, OTU2 in the second column, and Connection Length in the third column.
- 2 - Rename the column headings “Source”, “Target”, and “Mutations” without the quotation marks. Create two new columns named “Type” and “Weight”
- 3 - In the “Type” column, fill the column with “undirected” - again, without the quotes.
- 4 - In the “Weight” column, fill the column with the formula “=1/CR” (as always, without quotes), where C equals the column and R equals the row number of the mutations between the pair of haplotypes (example: =1/C2). Copy and paste this down the column. Essentially, this gives you a column that equals the inverse of the connection length between each pair of haplotypes.
- 5 - Save as a comma separated value file named something like “YOUR SPECIES edges.csv.” Your final “edges” file should look something like this:

Source	Target	Mutations	Type	Weight
AP_Hap01	AP_Hap05	2	undirected	0.5
AP_Hap05	AP_Hap18	1	undirected	1
AP_Hap05	RajaAmpat.1939.14	1	undirected	1
AP_Hap05	TC.3218.19	1	undirected	1
AP_Hap05	TC.3218.22	1	undirected	1
AP_Hap05	AP_Hap04	2	undirected	0.5
AP_Hap04	AP_Hap10	1	undirected	1
AP_Hap04	AP_Hap19	1	undirected	1
AP_Hap04	AP_Hap24	1	undirected	1
AP_Hap19	AP_Hap25	1	undirected	1
AP_Hap04	TC.1365.02	1	undirected	1
AP_Hap04	TC.3288.08	1	undirected	1
AP_Hap05	TC.1336.02	2	undirected	0.5
AP_Hap05	TC.1336.05	2	undirected	0.5
TC.1336.05	TC.3187.01	1	undirected	1
TC.1336.05	TC.3218.09	1	undirected	1
TC.1336.05	TC.3264.07	1	undirected	1
TC.1336.05	TC.3264.17	1	undirected	1
TC.3187.01	TC.AM747157	1	undirected	1
AP_Hap05	TC.3218.06	2	undirected	0.5
AP_Hap18	TC.3218.20	2	undirected	0.5
AP_Hap05	TC.3265.01	2	undirected	0.5
TC.3265.01	TC.3218.10	1	undirected	1
AP_Hap04	TC.4296.09	2	undirected	0.5
TC.3218.10	TC.AM747149	2	undirected	0.5
AP_Hap05	AP_Hap12	3	undirected	0.333333333
AP_Hap12	AP_Hap23	2	undirected	0.5
AP_Hap05	AP_Hap06	3	undirected	0.333333333

6 - Create a new workbook with the haplotype names in one column and the frequency of each haplotype in a second column. This can be accomplished in a number of ways -- if you already have a list elsewhere you can just copy and paste (just make sure that your haplotype names are spelled EXACTLY THE SAME as how they are in your Arlequin input file!!). Otherwise, you can go back into your Arlequin Output File and scroll down to the section (near the top) called "Haplotype frequencies in population." If you have multiple populations, then each haplotype will be listed along with its frequency in each population -- make sure you use "frequency" and not the "relative frequency." You want a count, not a proportion. Copy and paste this into an excel workbook and create a new column that sums the frequencies over all your populations to get the overall frequency for each haplotype. Copy the "total" column and paste the values (Paste Special → Values) then delete the columns for the individual populations. Rename the "Haplotypes" column as "ID" and the "Total Frequency" column as "N." Save this file as a comma separated values file named something like "YOUR SPECIES nodes.csv." Your final "Nodes" file will look something like this:

	A	B	C
1	ID	N	
2	AP_Hap01	5	
3	AP_Hap02	4	
4	AP_Hap03	4	
5	AP_Hap04	4	
6	AP_Hap05	4	
7	AP_Hap06	3	
8	AP_Hap07	2	
9	AP_Hap08	3	
10	AP_Hap09	3	
11	AP_Hap10	3	
12	AP_Hap11	3	
13	AP_Hap12	3	
14	AP_Hap13	2	
15	AP_Hap14	2	
16	AP_Hap15	2	
17	AP_Hap16	2	
18	AP_Hap17	2	
19	AP_Hap18	2	
20	AP_Hap19	2	
21	AP_Hap20	2	
22	AP_Hap21	2	
23	AP_Hap22	2	
24	AP_Hap23	2	
25	AP_Hap24	2	
26	AP_Hap25	2	
27	Halmahera.21	1	
28	Halmahera.22	1	
29	Halmahera.22	1	
30	Kimbe.AM745	1	
31	Kimbe.AM745	1	
32	Kimbe.AM745	1	
33	RajaAmpat.19	1	
34	RajaAmpat.19	1	
35	RajaAmpat.96	1	
36	RajaAmpat.96	1	
37	Solomons.289	1	
38	Solomons.289	1	
39	Solomons.299	1	

## Draw your Minimum Spanning Tree (or Network) in Gephi

I got too lazy to type up all of these instructions - so I made a [video tutorial!](#)