Using the NETWORK Package to display your Family Connections

by Dr John S Plant and Prof Richard E Plant (members 4890 and 6100)

NETWORK is a freely available software package which can be used to create network diagrams of the relationships among surname families based on Y chromosome data. It is very widely used by both professional and amateur researchers. In an article in the Members' Section of the Guild website http://one-name.org/members/DNA/Plant_article_network_package.pdf we provide help with using NETWORK: we give step by step instructions to create diagrams based on Y-STR (short tandem repeat) data, such as can be downloaded by the FTDNA (Family Tree DNA) Project Administrators. All that is required is a basic knowledge of Microsoft Excel and a text editor such as Notepad. In this article we provide a brief description of the software and how it works.

The NETWORK software package

The NETWORK package can be downloaded from the Fluxus Engineering site located at http://www.fluxus-engineering.com/sharenet.htm. To distinguish between the NETWORK package and the networks it constructs, we will capitalize the former. An example of a relationship network created by using the package is shown in Figure 1.
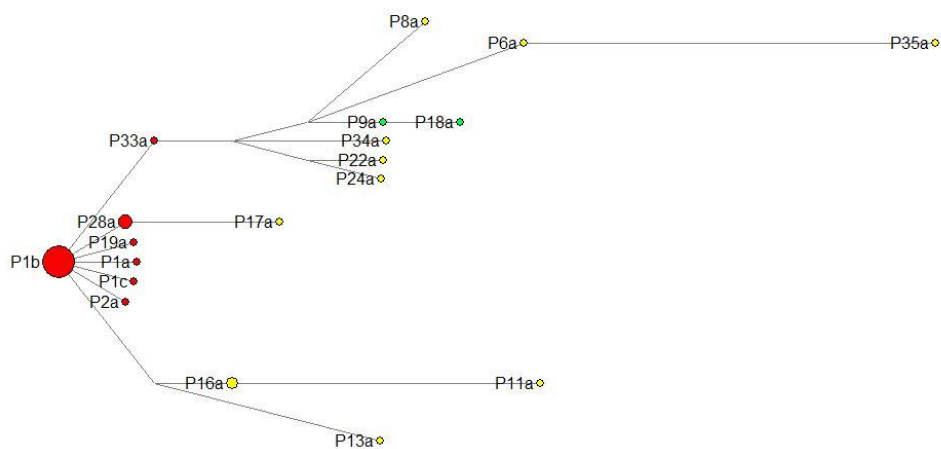

Figure 1: Diagram of relationships among members of the Plant Family History Group based on 12 marker Y-STR data.

Such networks can be used to show various pieces of information about a family or surname, such as shared ancestry and the number of individuals with matching DNA results. In this particular network, the circles represent individual men or groups of men with matching Y-STR values at the level of data determination (e.g. for 12 measured Y-STR markers for each man in this case). The size of each circle represents the number of men in the matching group and the label for each circle represents a code which identifies him or, in the case of groups with more than one member, a representative individual from the matching group. In Figure 1, the colours are adjusted so that red circles represent probable members of the main family, yellow circles represent individuals not identified as members of a family, and the green circles represent individuals who may form a second, smaller family. Some of this information, such as the size of the circles and the identifying codes, are generated by the software, and some, such as the colours, are adjusted manually. It is important to note that these network diagrams are not family trees: there is no identified initial ancestor or line of descent. They simply show who is related to whom and how closely.

How NETWORK constructs its diagrams

NETWORK uses an algorithm called *median joining* developed by H-J Bandelt et al. [1]. We can illustrate the basics of median joining with the very simple example shown in Table 1.

	I	II	III	IV
A	15	10	16	24
B	15	11	17	25
C	15	11	16	24
D	16	11	16	25

Table 1: A simple example consisting of four subjects, denoted A, B, C, and D. Each subject is tested at four markers, denoted I, II, III, and IV. The STR values are shown in the table.

In the simple example shown in Table 1, the genetic distance from A to B is 3, from A to C is 1, from A to D is 3, from B to C is 2, from B to D is 2, and from C to D is 2. The key to the algorithm is to introduce new fictitious individuals, called *median vectors*, whose STR values provide the most parsimonious means of creating a network. In the example of Table 1, the median vectors can be formed by taking the actual medians of the subjects' STR values. There are four possible combinations, ABC, ABD, ACD, and BCD. Table 2 shows the median value for each of the markers for each of these potentially useful median vectors.

	I	II	III	IV
ABC	15	11	16	24
ABD	15	11	16	25
ACD	15	11	16	24
BCD	15	11	16	25

Table 2: Median marker values of each of the three combinations of three subjects.

Two of the above hypothetical medians, ABC and ACD, are identical to C and, hence, they add nothing useful. Two others, ABD and BCD, are identical to each other, so we will ignore BCD. We choose ABD as a median vector and add it to the set of subjects. If we compute the genetic distances among the augmented set of subjects we find that the distances from B, C, and D to ABD are all 1. Since the distance from A to C is 1, we can construct a network, by including ABD, where the

distance from any one subject to another subject is 1. This network, shown in Figure 2, is called a *minimum spanning network*.
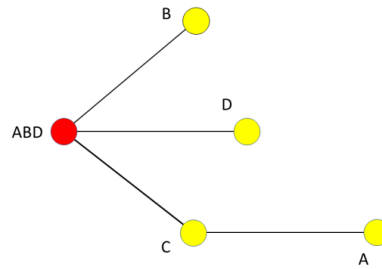
Figure 2: A minimum spanning network including the four original subjects and the median vector ABD.

This is a very highly oversimplified view of the process of median joining, but it does convey the general idea. It may appear that we have created a sort of "virtual person" in ABD. Indeed, from the way Figure 2 is drawn, the combination of marker values that are present in ABD might seem to represent the original ancestor, and then that B, C, and D each have one mutation, and thereafter A has a second mutation. There is, however, no directionality implied in the network, and without other evidence an equally plausible interpretation is that A has the marker values of the original ancestor, that C has one mutation, and that B and D represent two separate lines of descent, with ABD possibly never having existed. This simple example excludes a number of important points, but it does convey the general idea.

Using the NETWORK package

Our description of the use of NETWORK is intended for Windows users. However, NETWORK also runs on the Apple Macintosh and the ideas we describe can be easily extended to that platform. It is possible to use NETWORK as a stand-alone program, but we have found that when working with Family Tree DNA data it is easier to use NETWORK in conjunction with Microsoft Excel and a text editor such as Notepad. If you have never used Notepad, in Windows 7 it is in the Accessories folder of the Windows Start menu.[1] Only a basic familiarity with Microsoft Excel is required. If you are a Family Tree DNA administrator and want to avail yourself of this powerful tool, we encourage you to read our full paper in the Members' Area.

References

[1] Bandelt H-J, Forster P, and Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

---

[1] For Version 8.1 of Windows for example, it is under Windows Accessories in the Apps section of the Windows Start screens. For Version 10, click the Start button and type "Notepad" in the Search box.