

## Setting up your data file: the Nexus format

(slightly modified from <http://paup.phylosolutions.com/tutorials/quick-start/> )

PAUP reads your data in a NEXUS file format. All NEXUS files must begin with:

### - NEXUS file declaration

#NEXUS

If you open up the file “*Veronica\_example.nex*” (use Notepad++ for it), you will notice that it is divided into blocks of text, delimited by the words “begin” and “end.” The word following “begin” defines the block-type. Notice that the block-type is always followed by a semicolon “;”.

The following types of blocks can be used: data, assumptions, and paup. In the example, only data is used. In fact, one of the advantages of the NEXUS format is that applications will simply skip over blocks that they do not recognize.

### -DATA block:

Between the “begin” and “end” in a block you will find the AFLP genotypes (DATA block) that PAUP will read and execute. At a bare minimum PAUP requires your Nexus file to have “taxa” and “characters.” These are easily defined with the dimensions command in the data block, as shown below.

```
begin data;  
dimensions ntax=141 nchar=1127;  
format datatype=standard interleave=no gap=-;  
matrix
```

Here the number of taxa and characters are defined with the dimensions command and the options “ntax” and “nchar”, respectively. In this example file, the data block also contains several other commands including the “matrix” command that contains genotypes for AFLP loci (code: 1 presence, 0 absence). If you scroll down the file to the end of the genotypes you will see where the data block is terminated with an “end;”. Notice that both the “matrix” and “end” are followed by a semicolon.