Concatenating Data Sets and Running Analyses

Concatenating Molecular Data Sets

Mesquite can concatenate molecular data sets. It is also possible to concatenate molecular and morphological data matrices (aka, "total evidence" matrices) by hand. Open up the two molecular matrices that you want to concatenate.

Get the taxa in the two matrices to be identical. This means that you have to change the names so that they are the same. If the data sets do not have exactly the same taxa add taxa without any characters to the data sets to make them identical. The taxa do not have to be in the same order.

1. Save both matrices and close them both.
2. Open up one of them again.
3. While in the data editor window, go to **File>Link File…**
4. Select the other matrix. It will now open.
5. In the first matrix data editor window go to **Matrix>Utilities>Concatenate Other Matrix.**
6. Select the only available choice, it should be called unnamed matrix. It will then say something about the taxa list is different, select to rearrange the list.
7. The two matrices should now be concatenated. Go to the end of your matrix and make sure that it is. Take note of the character numbers that refer to each of the separate data sets.
8. Save the file with a new name.

Concatenating Morphological and Molecular Data Sets

The new version of mesquite may be able to do this as easily as the two molecular data sets. If you want you should check it out.

Once again get your two matrices so that they have identical taxa. This time they have to be in the same order too. You can do this in Mesquite.

Look at the two files in a text editor. Copy the matrix from your morphology file and paste it underneath the matrix from your Molecular file. Leave a line between them. This may sound easy, but you have to get all the semicolons and other notation right for this to work. It should say matrix once at the beginning of the combined matrix, and there should be one semicolon at the end of the combined matrix.

At the beginning of the **Characters block** change **nchar** in the **dimensions** commands to reflect the characters you have added. Take note of which characters are standard and which are molecular.

Also add the word **interleave** to the **format** commands. This allows you to have one block of characters after the other, rather than having them all in a row.

After you have done this, whenever you open this file in a program make sure to look at the characters it has read to see if it read them all.

Mixed Analysis in PAUP*

You can not analyze morphological data using likelihood. Instead you have to do the whole analysis using parasimony. Thus you have to make PAUP* think that all the DNA characters are just standard data. This is easy to do.

Change your **Format** commands at the beginning of the **Characters** block to read: **Format Datatype = standard missing = "?-" symbols="01ACGT" interleave;**

You may need to add more symbols to the list, depending on your matrices. You may need more numbers for your morphology, and your molecular matrix may have N R Y and any number of other letters. You can put the "-" in either the missing list or the symbols list, depending on whether you want to view gaps as homologous or unknown.

If you make the matrix in Mesquite it will have a bunch of other blocks after the Characters block. If they're giving you trouble, when you run them in PAUP*, then just erase them.