Tue 15.xi.

IPS-164 Introduction to phylogenetics
Introduction to some functions of the **GenBank** and database **TreeBASE** (treebase.org). Demonstration of the programs **winclada** (diversityoflife.org/winclada/) and **nona** (lillo.org.ar/phylogeny/Nona-PeeWee/). Check readme.htm for contents and instructions when you visit the web-page to download the programs into your own computer.

1. With the browser of your choice go to **GenBank** @ ncbi.nlm.nih.gov, on the right hand "column" click **BLAST**. A new window will open and there click the link **Nucleotide BLAST**. Again a new window will open and it includes a box where you can paste a sequence, or you can choose a file with your sequence that you wish to compare with those in the database. Try the functionality and scroll down and press the blue **BLAST** button. Depending on the ongoing activity of the server how fast your results will be shown varies. After a couple of (tens of) seconds a page will be updated with the results. By default, it will show a list of the taxa that have a sequence resembling the one that you uploaded. Check also the other options to show the results by clicking and checking the info under the other tabs available. Finally, contemplate on why this operation is important part of the work-flow of phylogenetic analyses?

2. Now return to the opening page of the GenBank, and in the left-hand column click link **Taxonomy**. A new window appears and there scroll down a bit and click the link **Taxonomy Browser**. Again a new window appears with an empty box in the upper left hand corner where you can type in a scientific name of your interest. In the case you do not have any personal preference type **Pogonatum**. Check what kind of information you will be able to get. Explore the different options available to save and display the info as well as different levels of taxonomical groups available.

3. Very good practical guide for winclada has been written by prof. Diana Lipscomb and you can download it @ www2.gwu.edu/~clade/faculty/lipscomb/Winclad.doc
Download the manual, open it and keep open during this session. Please, note that the program has been updated since this guide was written and thus citation of the program also has changed accordingly.

4. Open the web-page of **TreeBASE** (treebase.org). In the column on the left (top) click **Search data**. You will be presented a search view and you can type in a search-term, e.g. author, taxon etc. If you do not have any particular study or taxon you wish to use you can type as a search term my family name, choose Author from the pull-down menu, and then click **Search** radio button. Among the results you will be presented a study by Koskinen & Hyvönen 2004. Use the second icon on the right-hand side of the line (**Download Reconstructed NEXUS File**).

5. Open the downloaded file with a basic text-editor (e.g. **Notepad++)** and take simply a look on its contents, how they are organized, what kind of information you can find there etc.

6. Open the program **winclada.** Open now the matrix you downloaded from **TreeBASE** (**File > Open file**).

7. In **winclada** you have two alternate main windows (**Interface > WinClados/WinDada**). These enable you to switch between matrix and tree(s).

8. If you cannot see the names of your terminals completely you can enlarge the width of this column simply by pressing **Shift** and tapping arrow to the right, and naturally also make it narrower with the arrow to the left.

9. Under **Matrix > Matrix Title edit** you can insert some brief description of the matrix.

10. Analyze the matrix by using a heuristic search using the default values (**Analyze > Heuristic**). You will be presented a new small window. At this point you can leave other options with their default values, but increase the number of **Maximum trees to keep (hold)** to 1 000, **Number of replications (mult*N)** to 10, and **Starting trees per rep (hold/)** to 2.

11. The actual search will be performed by the program **nona** and its activity can be seen in a separate DOS-window that will automatically appear in the center of the winclada window. After the search is completed result will appear in WinClados window of the **winclada**. You can see the number of trees in the cache memory plus the essential statistics (length, Ci & Ri) of the tree(s) at the bottom bar of the window. Please, note that **winclada** keeps also the trees that have formerly in the cache (in this case already in the matrix that you downloaded) and thus you probably see differences in these values when you scroll through them (**Trees > Next tree**). On the lower left-hand corner of the window you can see which one of the trees you are viewing. By default trees will be rooted by the first terminal of the matrix. Check the difference of length between the trees now in cache. Any differences?

12. Choose now the view for the matrix (**Interface > WinDada**). You can now learn some useful tools for editing matrices. With **winclada** you can easily, for example, remove part of the matrix. This is needed in the cases

when you are working with concatenated matrices and you wish to study any part of the matrix independently. Please, open now the file again using **Notepad++**. Under BEGIN SETS; you will see how it is divided between different genes and morphology.

13. Using the information about limits between different genes edit now the file using **winclada**. Cut away other parts of the matrix leaving only the gene *rps*4, and another with only ITS2 (or some other genes if you have downloaded a different matrix of your choice). Please, note that numbering differs between winclada- and the nexus-file and thus you should adjust this in **winclada** using *Chars > Number chars from 1 (Illogical)* in order to make it easier to follow the partition of the nexus-file. You can select characters by double-clicking anywhere in the matrix with the left button of your mouse. The whole column will be highlighted. Now by keeping down the left button you can highlight other characters by moving mouse left or right. Under *Chars > Unselect all chars* you can cancel your choices. You can also select characters to the right or left of the character that you have highlighted first (*Chars > select chars RIGHT OF CURSOR* or *Chars > select chars LEFT OF CURSOR*). Use these tools and create two smaller matrices with information from only one gene.

14. In **winclada** you can have several matrices open at the same time and switch between them using *Matrix > Go to next matrix (F12)*. Using *Chars > Mop uninformative chars* you will be able to reduce the size of the matrix, or simply check the percentage of the informative characters. By performing this separately for each partition you can compare different sources of the data included in the concatenated matrix and how much information they contain.

15. By using *Chars > Make sel chars ADDITIVE (farris)* you can also make some of them additive/ordered. In this you can modify your matrix and see how this affects results of your analyses if you have such characters that can be treated as additive (ordered).

16. You can also edit character and ch. state names using *Chars > Character dialog*. A new window will appear that enables you to edit these. If you have a large matrix you can also go to a certain character by *Chars > Goto Character*.

17. Especially with gene matrices it is also useful to visualize how well, or poorly, the nucleotides have been aligned. You can see this better using *View > Color Code states*. After this you can see them better by using *View > change state (matrix) BACKGROUND color*. Here you can assign different nucleotides colors of your choice, or simply highlight the default colors that are used also in many other programs for each nucleotide.

18. Another useful function of **winclada** is the ability to COMBINE matrices. You can do this even for matrices of unequal size, i.e. you have a certain gene or morphology coded only for part of the terminals you are studying. However, it is critically important that the names of the terminals are written EXACTLY in the same way. In order to combine matrices you first have to open at least two. Use *Matrix > New matrix merge*, a new window will appear with matrices currently open listed. Choose primary and secondary matrices to be combined. It is safest to switch under *Terminal match* to *Match terminals by name*. Under *Lights! Action!* press *Merge matrices*. A new combined matrix will appear. You can continue this until you have all the matrices combined into one. It is VERY IMPORTANT to CHECK CAREFULLY THE NUMBER OF TERMINALS and ESPECIALLY CHARACTERS in order to keep close track about the process! Best to use conventional book-keeping on paper or a text-file!

19. Please, note that only part of the published matrices has been uploaded to TreeBASE. Other data repositories are for example **datadryad.org**, **figshare.com** and **morphobank.org**. They *include data from various fields* and thus finding phylogenetic matrices is not as simple as in the TreeBASE.

20. Nowadays it is increasingly common that they are available as electronic supplements of the original articles. Using *TreeBASE* and *scholar.google.fi* search now a matrix/matrices of your interest to be analyzed on Thu.

Thu 17.xi.

IPS-164 Introduction to phylogenetics

Some additional features of **winclada** (version 1.61; diversityoflife.org/winclada/) and **nona** (lillo.org.ar/phylogeny/Nona-PeeWee/)

1. Go **TreeBASE** (treebase.org) and search for an example matrix of your interest using the **Search data** function of the left-hand column. As a search term use, for example, genus level scientific name of the taxon of your interest. Then choose **Abstract** from the pull-down menu of the next box, and then click **Search**. (As an alternative you can also check recent articles in some journal and check the supplementary material of the article of your choice).

2. In **TreeBASE** You will be presented with a list of studies. If none, use some other name of your interest. Choose one of the studies by clicking its box on the left preceding its code. The box will turn red with a tick mark. At the bottom click ***Discard Unchecked Items*** radio button. Now choose **Matrices** sheet above, and you will be presented a matrix/list of matrices related to this particular study. On the right-hand side choose the symbol in the middle (***Download Reconstructed NEXUS file***)

3. Open the program **winclada** and after that open the file you have downloaded from TreeBASE, or as a supplementary material of some journal article.

4. Using **Notepad** edit the matrix manually so that you move the outgroup taxon as the first one also in the matrix. Check the figures/text of the original publication, or the trees uploaded to TreeBASE (check **Trees** sheet for these) to determine which one (if several) of the outgroup terminals is the outermost one. You can do the same also using **winclada** by double-clicking the line of the terminal (will be high-lighted turquoise). After that ***Terms > Move selected terms > To beginning***. You can move there ALL outgroup terminals if you wish but, please, remember that by default trees will be rooted by the branch linking the **first taxon** of the matrix with all the others.

5. Analyze the matrix with default values (***Analyze > Heuristic***). You will be presented a new small window. At this point you can leave other options with their default values, but increase the number of ***Maximum trees to keep (hold)*** to 1 000, ***Number of replications (mult*N)*** to, say, 10, 50, and finally 300, and ***Starting trees per rep (hold/)*** to 2. The program creates files with extensions .out and .tre but unfortunately the default .out-file is not informative about actual operations. Thus a separate book-keeping is very important to mark down your search settings and also connect them with the appropriate .tre-file naming also the latter accordingly.

6. Save the trees you have obtained by ***Trees > Save ALL trees to file***. After this you have to clear the cache memory explicitly after each search (e.g. ***Trees > DELETE all trees***) in order to keep results of different searches separate. If, in some cases you wish to select individual trees this is performed by pressing ***S*** when that particular tree is shown in **winclada**-screen. Blue box in the left-hand upper corner with the text ***Selected*** will appear, and now under from pull-down menu ***Trees*** you can, for example, delete this particular tree.

7. Did you mark any differences between results obtained using different number of replications? If length of the trees obtained is the same for them all it is easiest to see differences between the results of separate searches by making first consensus trees of the trees obtained in different searches using ***Trees > Consensus Compromise > Consensus (strict)***. Save these consensus trees as separate tree files. After this you can compare two consensus trees of the first analyses by opening both and after this making consensus as above. If the program reports any nodes to be collapsed this informs about differences between the two trees. Add the third one and make similar comparison.

8. Another useful tool to compare results is to check whether trees differ is by ***Trees > Select all duplicate (nonunique) trees***. After this ***Trees > Delete selected trees***. If the number of trees is reduced from the original one that you had opened this indicates presence of identical trees.

9. Expand your search by using larger number of starting trees per replication, for example 20. Try this, save the obtained trees, make consensus and compare with the consensus of consensus trees (see above). Are the results identical? If not, how many nodes got collapsed when you made a new consensus?

10. **Winclada** can also be used to calculate support values for the nodes using ***Analyze > Bootstrap/Jackknife/CR with NONA***. You will be presented with a new window where you can adjust parameters of your search. You have to compromise between time and thoroughness of the search

– more thorough search will lead to more reliable values. In order to save these values it is best to save the obtained tree as graphics file **File > Create Tree Metafile**). You can then edit/place values to the final tree to be published as a figure with a graphics program you prefer.

11. Please note that you can change the view of the tree by using arrow keys. You can also adjust the size of the tree (both horizontally and vertically) using buttons of the grey bar by the upper margin of the window

First two ones on the left can be used to make the whole tree smaller or larger. With the fourth button from the left (light blue) you will be presented with a character diagnoser window. Pressing again will remove this window and original black coloration will be retained, however, note that also thicker branches will be retained in the view.

Fifth button will show branches as proportional to changes on each branch and also show characters changing on each branch as hash-marks. However, hash-marks will be shown ONLY on branches shown as originally, thin lines. In order to show them as such you have close the file and open it again. Using **HashMarks > Ellipse hashmarks (Toggle)** you can change from the default (ellipse) to square ones.

Next button towards right will supplement hash-marks with ch. numbers. If your matrix is large, please, note that you have to enlarge your tree horizontally (see below). And pressing the next one will show also the ch. state, and by pressing it once more will also show the direction of the change, and once more this info will disappear.

Next button will highlight names of the terminals. After that you have five buttons that you can use to change the appearance of the tree, and after these four buttons to stretch/shrink tree vertically and horizontally. Next two buttons will allow you to scroll between trees if you have several, and the last button on right will take you to the matrix view.

12. **Winclada** can also be used to create trees that can be used for figures to be published. You can adjust font of the taxon names using **View > Fonts > Taxon font**. Normally scientific names are written in *italics*, and this you can also choose using this tool. You can also adjust other fonts. It is advisable to save the tree as a graphics file (**File > Create Tree Metafile**) that you can open using a graphics program of your choice for further advanced editing.

13. You can also generate a group membership ch. matrix using **Analyze > Generate group membership matrix**. But you have to have saved at least one tree before this operation. Do you remember how these characters can be used?

14. Like in the matrix you can also adjust colors of different ch. states of the tree by **View > change (tree) state colors**. Please, remember to take into account also background color. Branches and text characters are NOT bordered.

BASIC different file formats (you can edit also these but **ONLY** with BASIC text-editors like Notepad, **NEVER** with MS Word or comparable!). On the left format used by nona/winclada (also TNT read these), and on the right NEXUS-format, the one required by TreeBASE, and finally fasta-format widely used by many programs for sequence level characters.

```
xread                                    #NEXUS
13 4                                      BEGIN DATA;
lungfish    00001100?0010                DIMENSIONS   NTAX=4 NCHAR=13;
frog        0110110000011                MATRIX
bird        1231112001100                lungfish        00001100?0010
mammal      1[12]41111000010             frog            0110110000011
;                                        bird            1231112001100
                                         ;
                                         END;
```

```
                    >Taxon_0
                    ACCGTTAGGCCCTGGTTAAGGCGGATTAGGCGG
                    >Taxon_1
                    ACCGTTTTGCCCTGGTTAAGGCGTAGGCG---G
                    >Taxon_2
                    CCCGTTAGGCCCTGGTTAAGG—CCGGT--TTTG
                    ……
                    >Taxon_n
                    CCTTTTAGGCTTGGCGGTATTAACGGCC–TTTC
```

Fri 18.xi.

IPS-164 Introduction to phylogenetics
Demonstration of the program **TNT**

1. A PowerPoint presentation of the different functions of TNT can be found @ lillo.org.ar/phylogeny/tnt/. Find the '**QuickTutorial**' link under '**Documentation**'. Keep this file open just in case you need to check certain functions. In order to display list of various commands available type **help** at the bottom of the window after **Enter command:**

2. Open the program **TNT** and after that open the file you have downloaded from **TreeBASE** or as a supplementary material of some journal article.

3. Most likely the file is NOT in tnt-format and you have to open it first with **winclada**, and after this save the file in nona format (**Matrix > Save As (Nona format**).

4. To open your data file **File** > **Open input file.** Find your data file saved in nona-format (extension .ss). If program complains about parsing long taxon names and alien characters: simply click **OK**.

5. The program provides option to generate log-file of all search activities. Under **File > Output > Open output file** you can give a name for this file. This is very helpful, and essential in order to keep track of the settings of different searches.

6. Under **Format > Data format > Read gaps as missing** you can define if you wish to treat gaps as missing as they are most commonly treated. However, missing entries are always problematic and it would be better to treat as a fifth ch. state. This is also a default setting of the program. You should notice that how gaps are treated WILL affect your results.

7. **File > Merge/Import data > Read a Fasta file** enables you to open also files in this format that is commonly used by sequence alignment programs. You can then save these files in TNT-formats.

8. You can also save matrix in nexus-format using **Data > Export data > Nexus.**

9. To increase the data storage capacity of the analysis: **Settings > Memory a**nd increase General RAM to 1000 Mbytes. Increase **Max. trees** to 10000. In a real analysis you want to use 99999 trees but to constrain tree search time, we will use 10000 in our exercises instead.

10. If your matrix is small (20-30 terminals) you can also try to find a tree that is sure to be optimal by using **Analyze > Implicit enumeration.** Here you will be using branch and bound -algorithm.

11. For most real matrices you have to resort to heuristic searches. For the matrices with less than 200 terminals different kind of traditional searches and their combinations would suffice. Under **Analyze> Traditional search** you will be presented with a new window and this provides you choices about the settings for your search. First you can choose whether you wish to start search with a certain number of **Wagner trees**, or, if you wish to use trees of some previous searches that you have opened before your analysis (**trees from RAM**). In this way you can divide your search into smaller parts. By using **0** as **random seed** you define program to use CPU time as seeds to randomize the order of terminals. You can also define **number of replications**. Next part of the window allows you to choose between different swapping algorithms and how many trees you wish to save per replication. In the next part you can choose how trees you find will be treated. Default setting is that trees found in a new search will replace those already in RAM. However, you can save all optimal trees by ticking this option off, or, you can even keep suboptimal trees by ticking **keep all trees found**. Collapse trees after search enables you to use different rules of collapsing for them. Consult different manuals/tutorials of details about these.

12. Perform now a couple of traditional searches for your material with different settings and open separate log-files for each search. After each search you have to close the log-file using *File > Output > Close output file*, and then open a new log-file for new search.

13. In order to save trees of each search you have to define a file for this by *File > Tree Save file > Open, parenthetical*. After you have saved the trees you have to close it by *File > Tree Save file > Close tree file*. Otherwise the file will not in a format required by the program.

14. Although new technology search will not be needed for the small matrices you can rehearse this kind of search as well using *Analyze > New Technology search*. You will be presented with a new window that enables you to define settings for different kind of searches.

15. You can also get support values calculated using *Analyze > Resampling*. You will be presented a new window where you can choose a method (Bootstrap/Jacknife/Symmetric resampling). The latter one is specifically for matrices with different weights on characters. Try to calculate **bootstrap** or **jackknife** values. Mark as a *Cutoff* value **50** that is normally used in these analyses, and *Output results as*… *absolute frequencies*, again following settings that are most commonly used.

16. Press letter "*p*" in order to print the obtained tree with support values. Then you can choose to save it as a pdf-file, or as postscript-file that you can then edit further using a graphics program of your choice. By pressing "*Esc*" you will return back to normal window, if you press "*s*", tree, with support values, will be shown also in the normal window of the program and you will be able to store it within your log-file.

17. In order to make a search using constraints you have two options. You can edit a tree manually for example using **Mesquite**. Keep on mind that you have to remove ALL resolution of the tree except the clade as monophyletic that you wish to use as a constraint. Store this tree/trees and then export the file in TNT-format. Open it using e.g. Notepad++ and edit by removing all other parts preceding the tree except the text tread and the tree in parenthetical notation that you wish to use as a constraint. If you have only one tree remember to remove also asterisk at the end. Save now the file with informative name and an extension .tre. You can now open the file if you have the matrix that accompanies it already open by *File > Open input file*, and then choosing *Tree files (*.tre)* from the pulldown menu.

18. You can also include several groups in your constraint file and choose them to be used, one by one in your analyses using *Data > Define constraints*. In a new window you can define constraints for monophyly or non-monophyly and choose the tree (in case you have several). By clicking *use selected groups* radio button you will be presented with a window showing the tree and there you can choose which clade you wish to use as a constraint. Clicking the node will highlight and inflate it in **RED**. Press "*OK*", and in the preceding window also "*OK*", and now you will return to normal log window with constraints listed at the bottom. When using constraints keep on mind that monophyletic groups are in many trees nested within each other and thus it is important to edit trees as detailed above, and use only ONE constraint at a time.

19. In order to make navigation easier in the display buffer you can also clear it using *Format > Display Buffer > Clear display buffer*. You can also adjust there for example fonts for the buffer. By reducing the size from the default 12 you can see more of it at once without need to scroll.

To read more about the functions of the program you can take a look @ https://groups.google.com/g/tnt-tree-analysis-using-new-technology

For more advanced use of the program also scripting is available with documentation available @ www.lillo.org.ar/phylogeny/tnt/scripts/General_Documentation.pdf