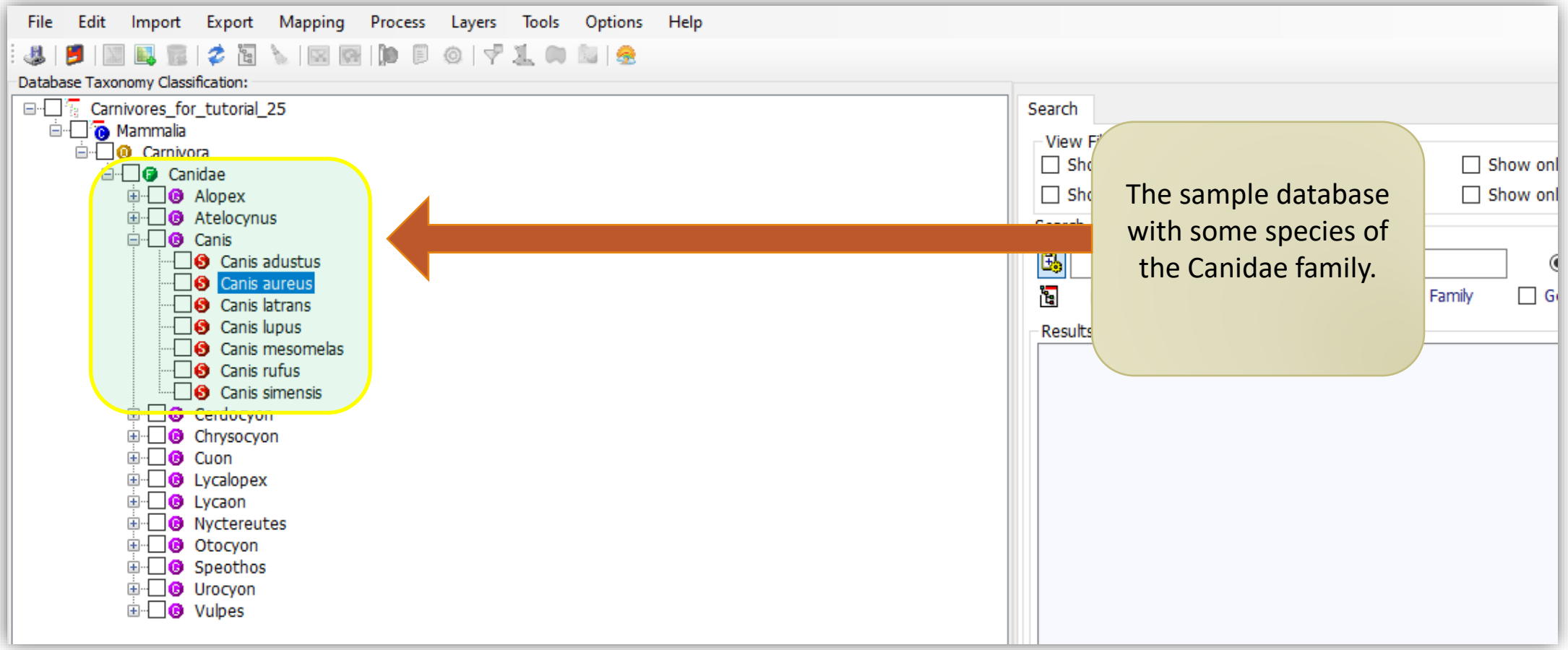
Step by step tutorial: Doing phylogenetic trees with ModestR and BOLD

What do you need for this tutorial:

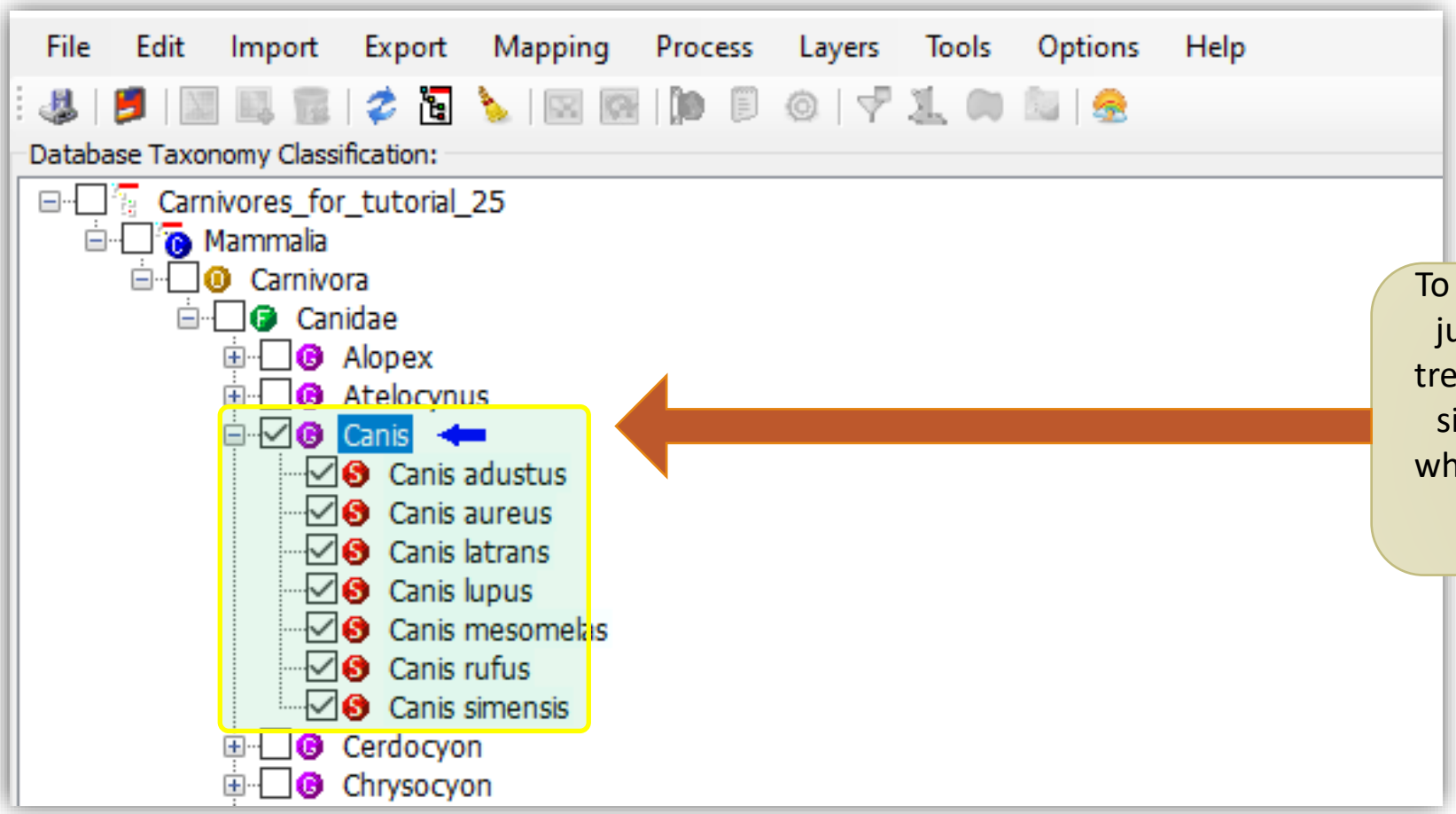
1. ModestR 6.5 or later
2. Internet connection
3. About 25 minutes

We'll describe how to create phylogenetic trees with ModestR and BOLD data.
Follow the next steps!

Run ModestR DataManager program. You must open an existing database or create a new one. To create a new database, you can see the Tutorial 1 of ModestR tutorials : How to create a ModestR database.
You may also use **this sample database** that contains a taxonomy of some species of the Canidae family.

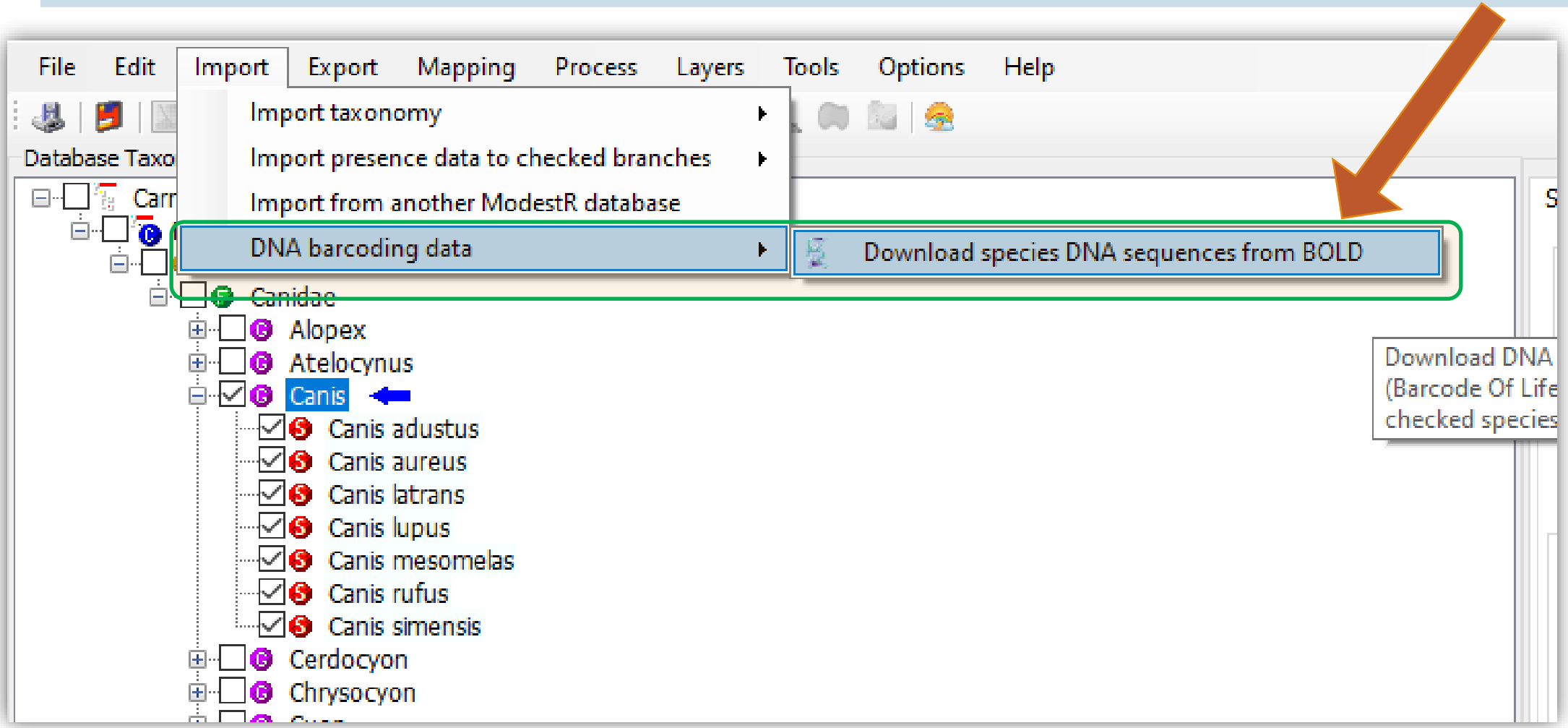


The first step is selecting the species we'll want to use to create the phylogenetic tree. For this example, we will select all species of *Canis* genus.



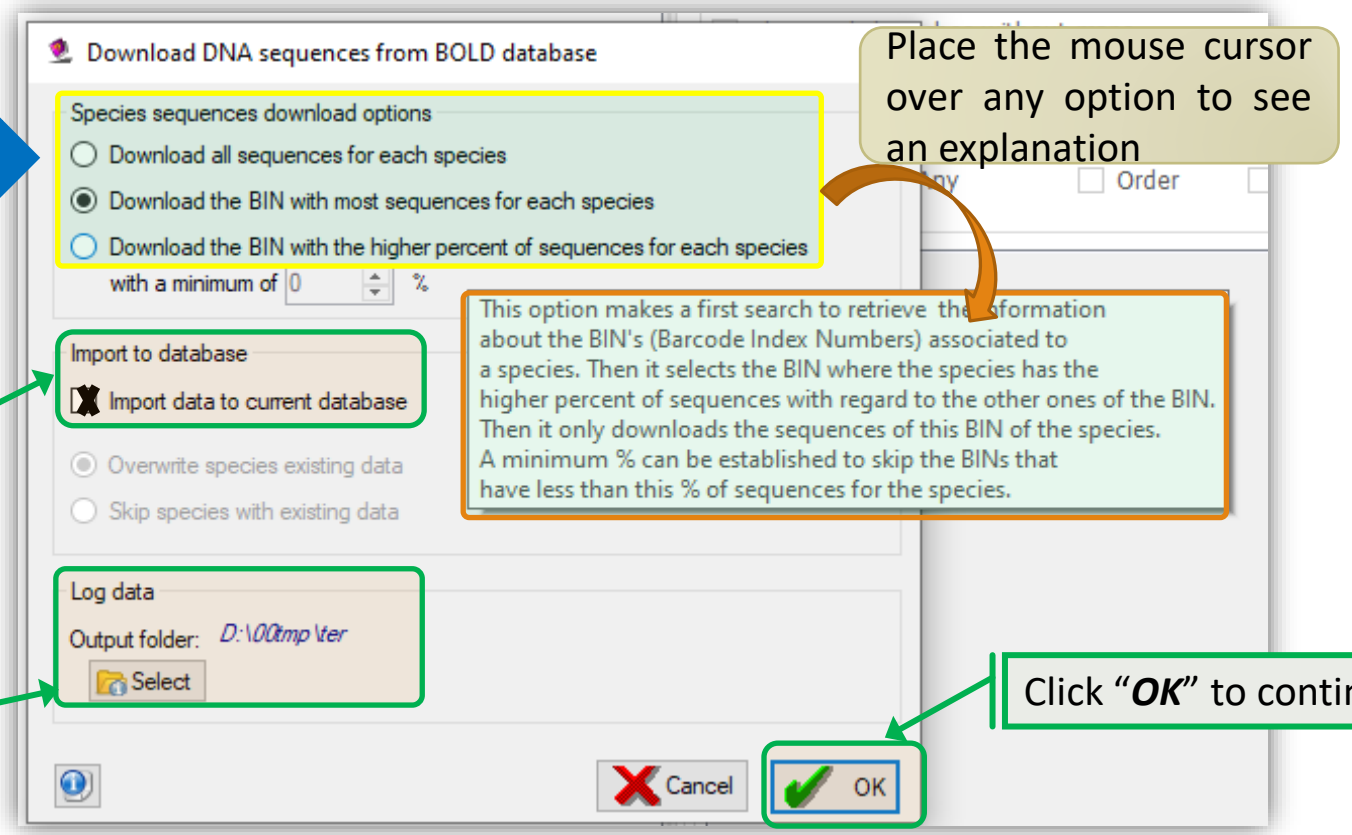
To select a taxonomy just check it in the tree. You can check a single species or a whole level such as a genus.

Then we must obtain DNA data for those species. ModestR can currently download data from the BOLD database (Barcode Of Life Data System) which provides sequence data and a BIN database (Barcode Index Numbers).
To download data from BOLD, just go to menu *Import/DNA barcoding data/Download species DNA sequences from BOLD.*



A dialog box will appear, allowing to select the data to be downloaded. You may choose to download all sequences labeled with the species name, or only the sequences from a specific BIN for each species. A BIN contains several different barcodes, that is, gene fragments, that have been algorithmically clustered together. If you are not familiar with BOLD BIN's, please refer to BOLD documentation.
You must check the option to import data to the current database. And, to select a folder where a copy of downloaded data will be saved.

You may choose to download all sequences labeled with the species name, or only the sequences from a specific BIN for each species.
To select the BIN for a species you may choose the BIN that contains most sequences labeled with the species name, or the BIN with the higher % of sequences for the species.



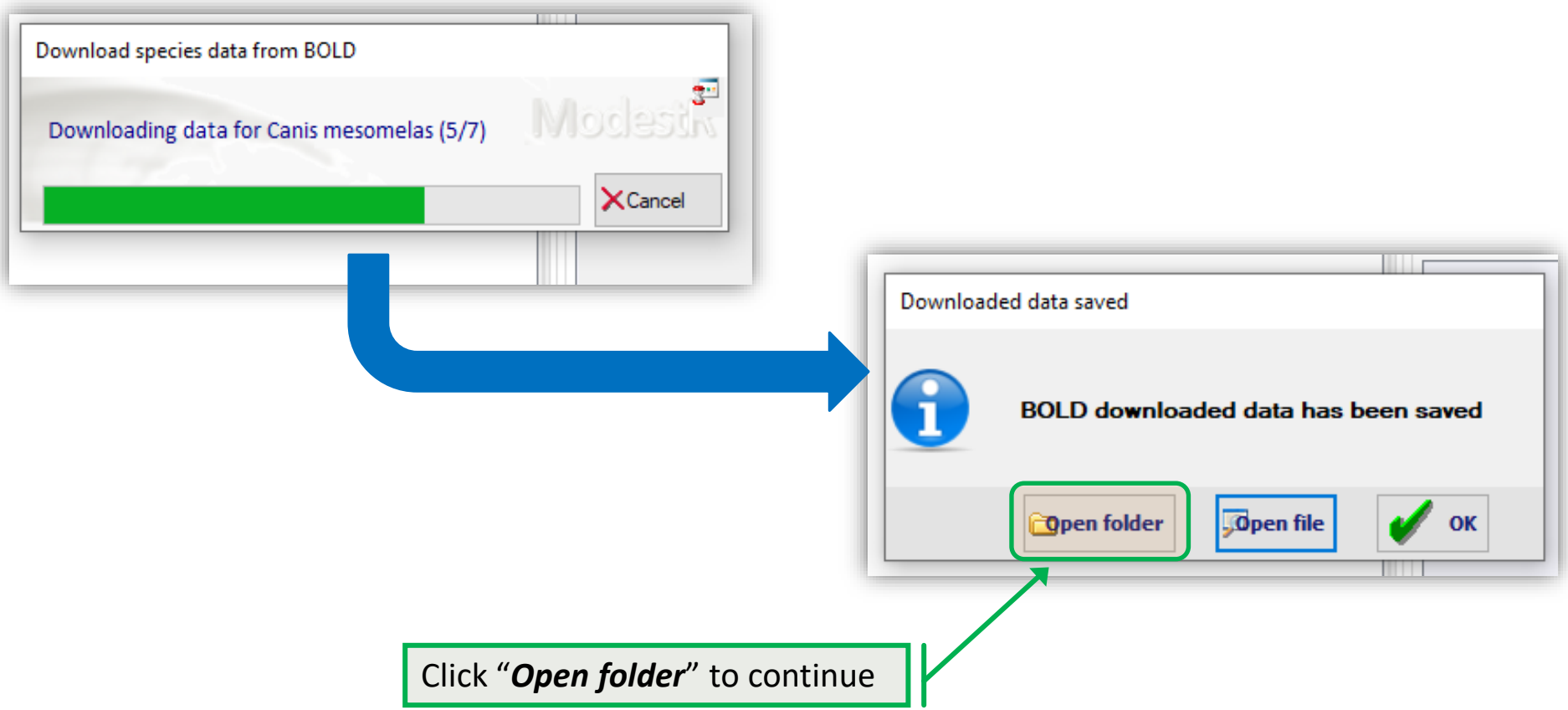
Place the mouse cursor over any option to see an explanation

Check the option to import data to the current database

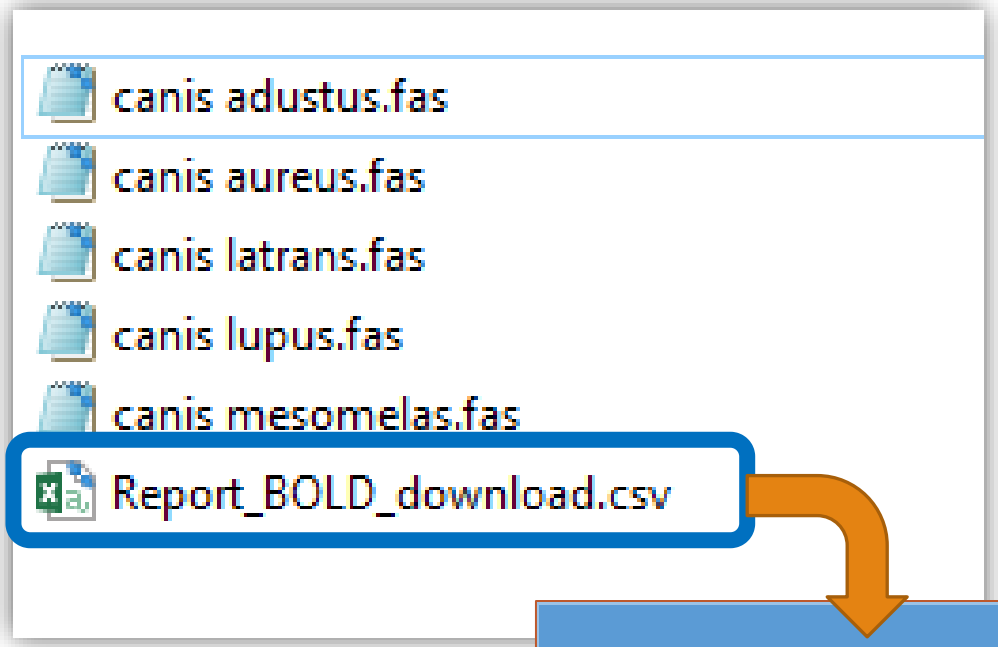
Select a folder where a copy of downloaded data will be saved

Click "***OK***" to continue

Downloading will start. Once completed, a dialog box will be displayed, where you can select opening the folder where a copy of the downloaded data, opening the report file, or just accept ("OK") and continue. Let's select "Open folder"

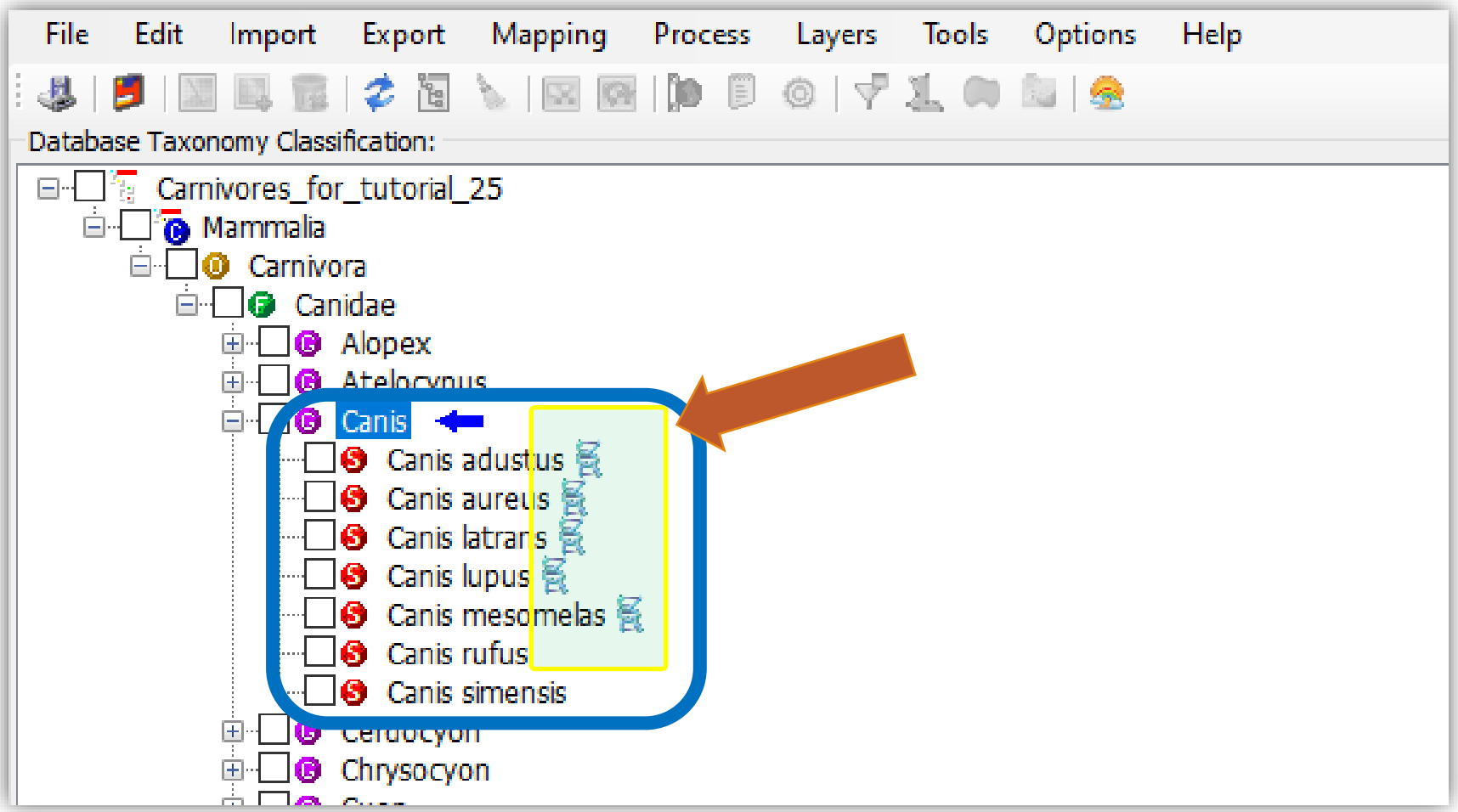


In the folder you'll find a copy of the downloaded data for each species in FASTA format and a CSV file that summarizes the data downloaded: number of samples for each species, and, if BIN downloading selected, number of BINS where it appears, the BIN id selected for the species, etc.



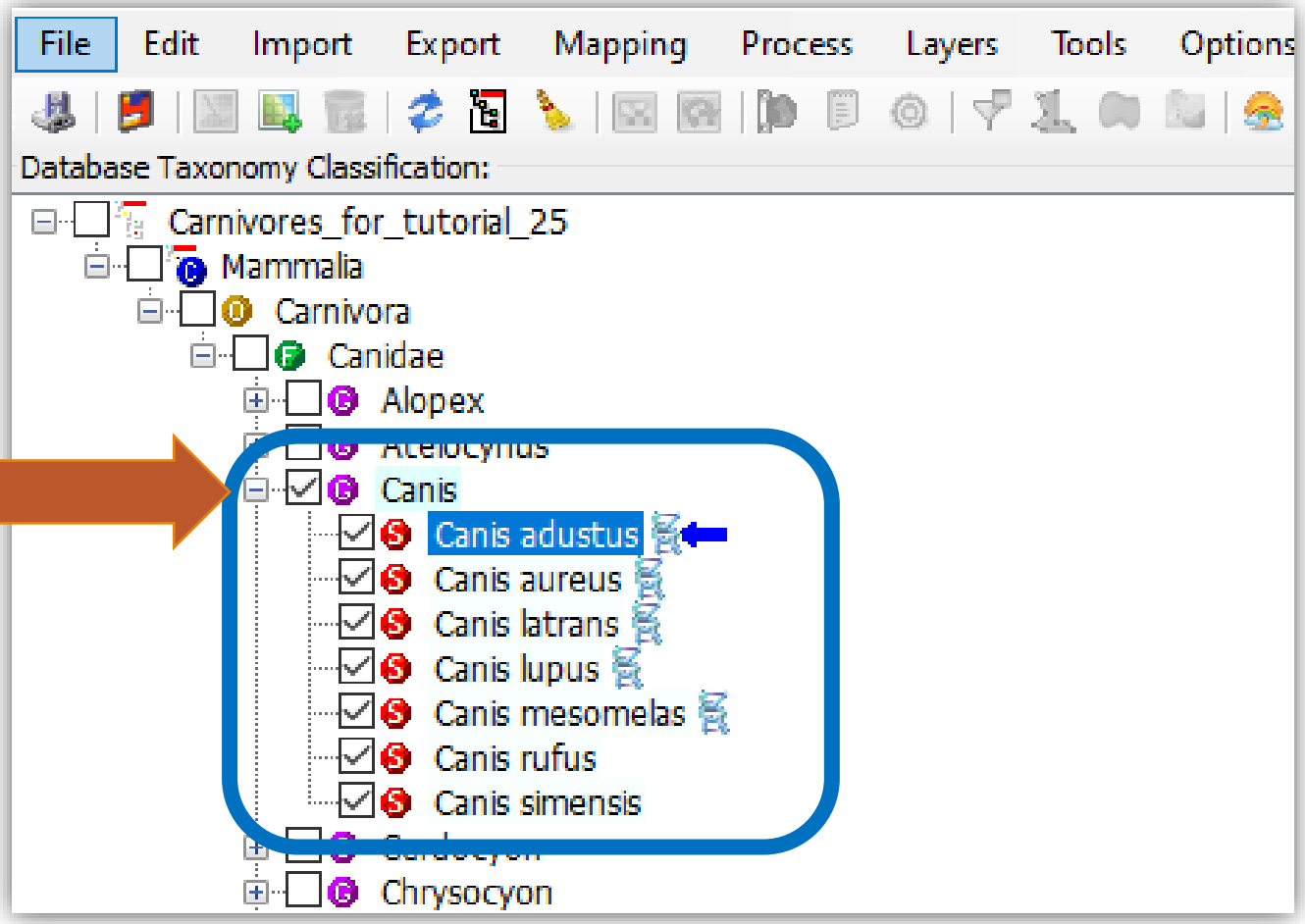
Species	Num.Samples	Num.BINs	Higher.Sample.Count.BIN(HSCB)	Num.species.in.HSCB	Num.sequences.in.HSCB	Sequences.for.species.in.HSCB	Other.BINS
Canis simens	2	0		0	0	0	
Canis adustu	8	2	BOLD:ACQ1122	1	5	5	BOLD:AEE9377
Canis aureus	11	2	BOLD:ADM0647	1	3	3	BOLD:AAA1542
Canis latrans	17	1	BOLD:AAC5017	3	33	25	
Canis lupus	1705	2	BOLD:AAA1542	12	1738	1414	BOLD:AAC5017
Canis meson	5	3	BOLD:ADK6525	1	1	1	BOLD:ADW2121,BOLD:ACR0823
Canis rufus	0	0		0	0	0	

If we come back to DataManager, we'll see that now a DNA-shaped icon appears aside some of the species, indicating that now there are sequences stored in the database for those species. Some of the species we selected don't have this icon, because no data was found in BOLD for them.



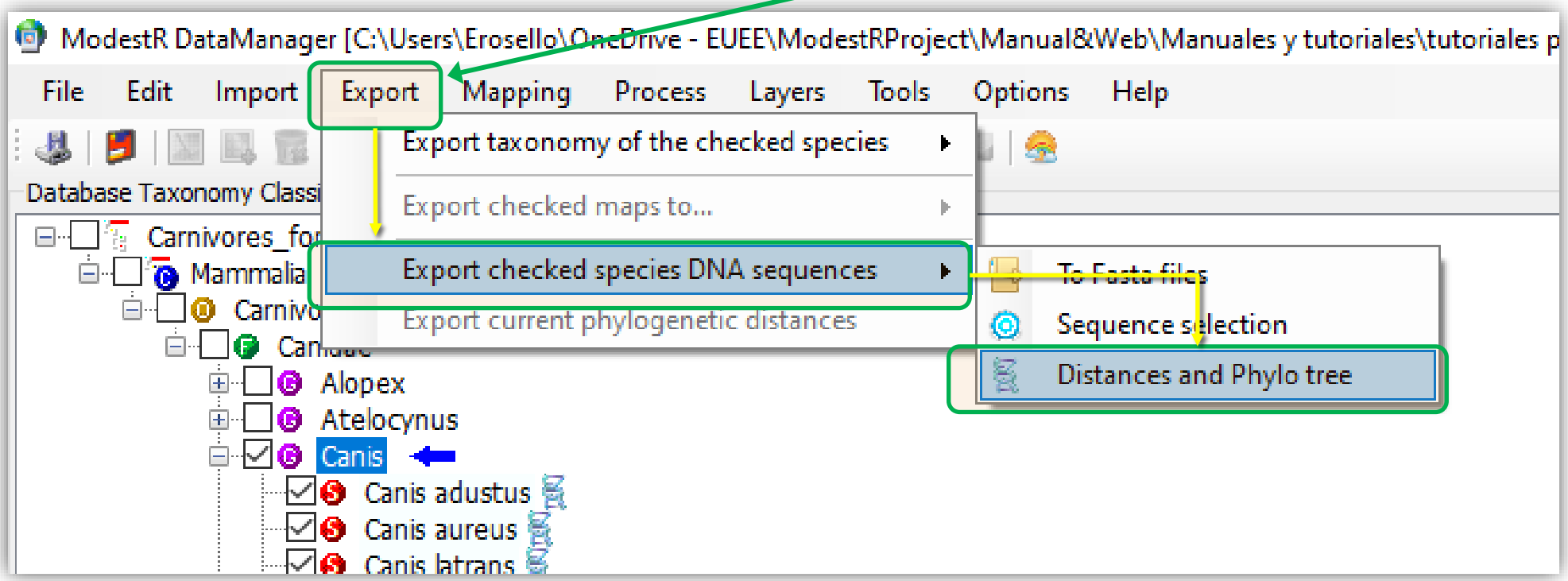
Now we must select the species to build their phylogenetic tree. In this example we'll select the whole *Canis* genus. It doesn't matter if some species of the selected species don't have DNA data. Only species with DNA data will be used to build the tree.

To select a taxonomy just check it in the tree. You can check a single species or a whole level such as a genus.

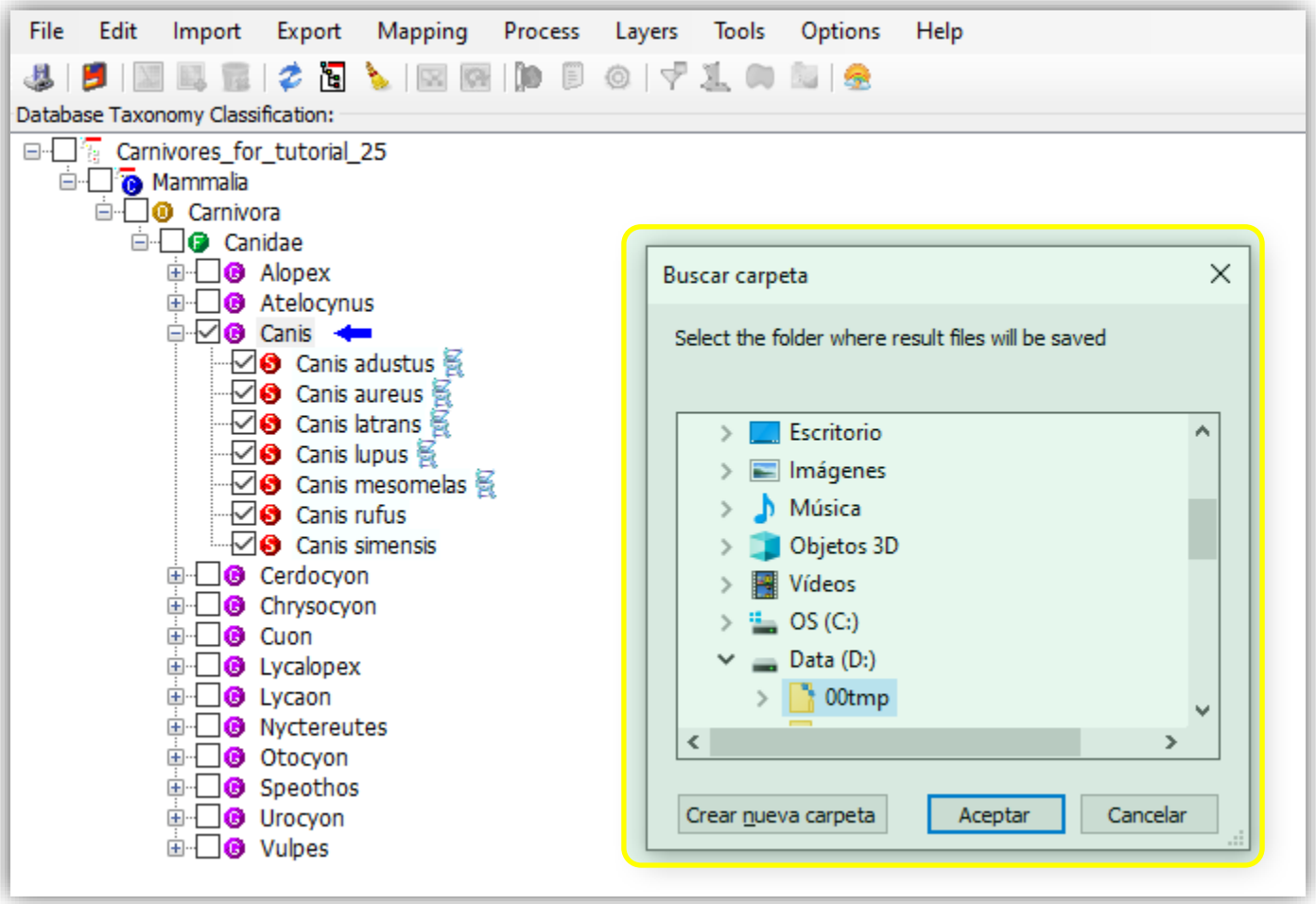


Next, we'll go to menu *Export/Export checked species DNA sequences/Distances and Phylo tree.*

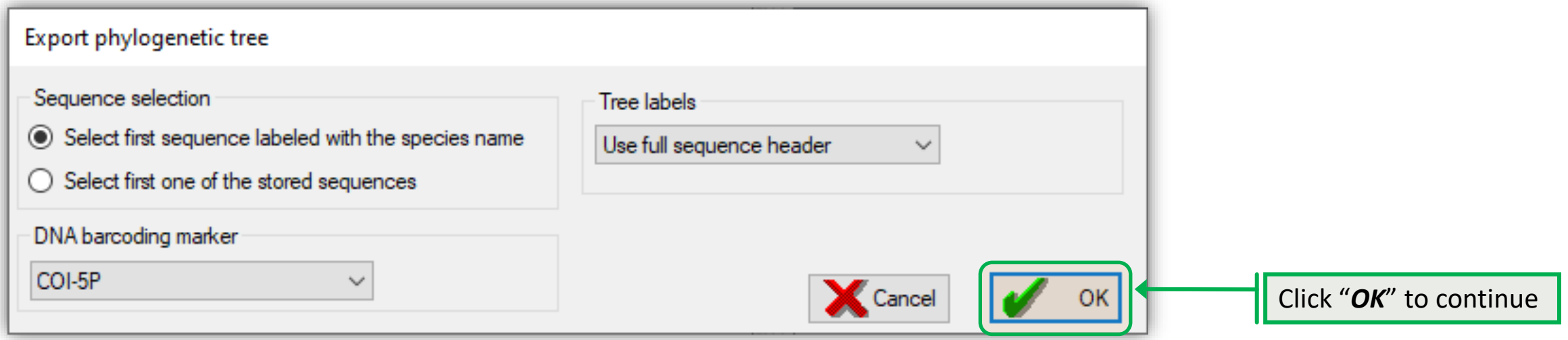
Select menu "*Export*" -
/Export checked species DNA sequences
/Distances and Phylo tree



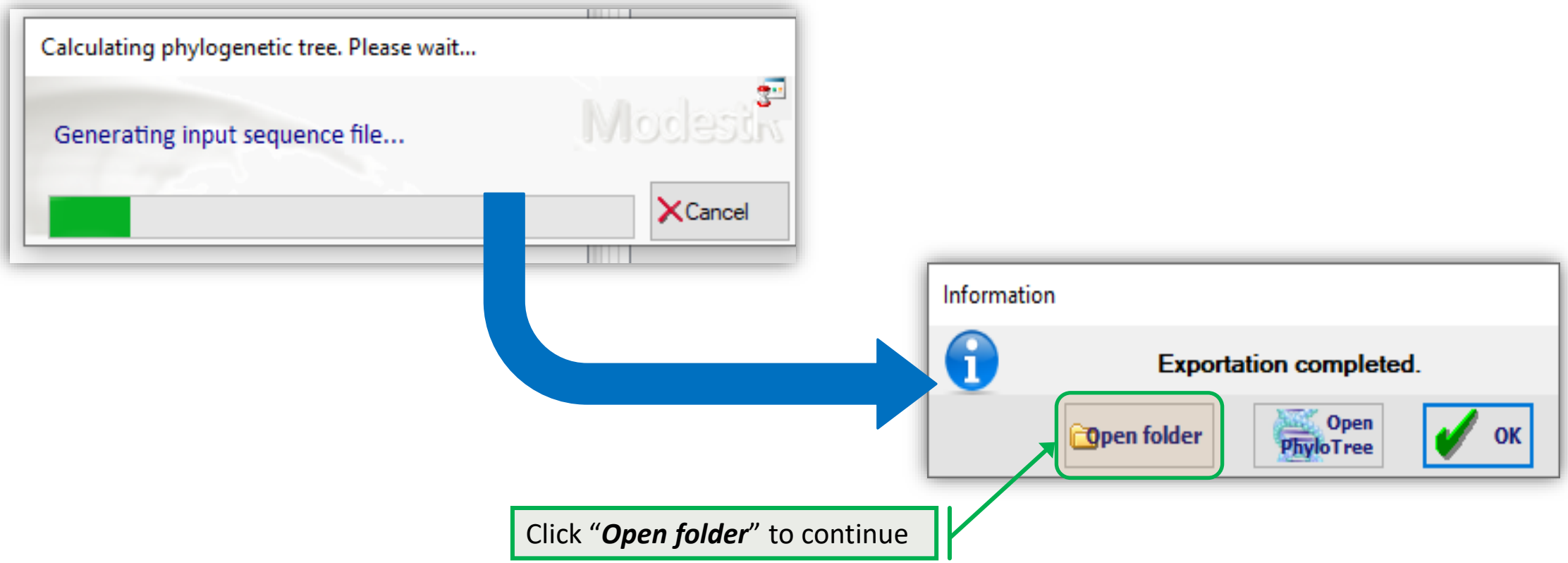
The first step is selecting a folder where all input and output data will be stored during the process. We recommend to select an empty folder to avoid accidental overwriting.



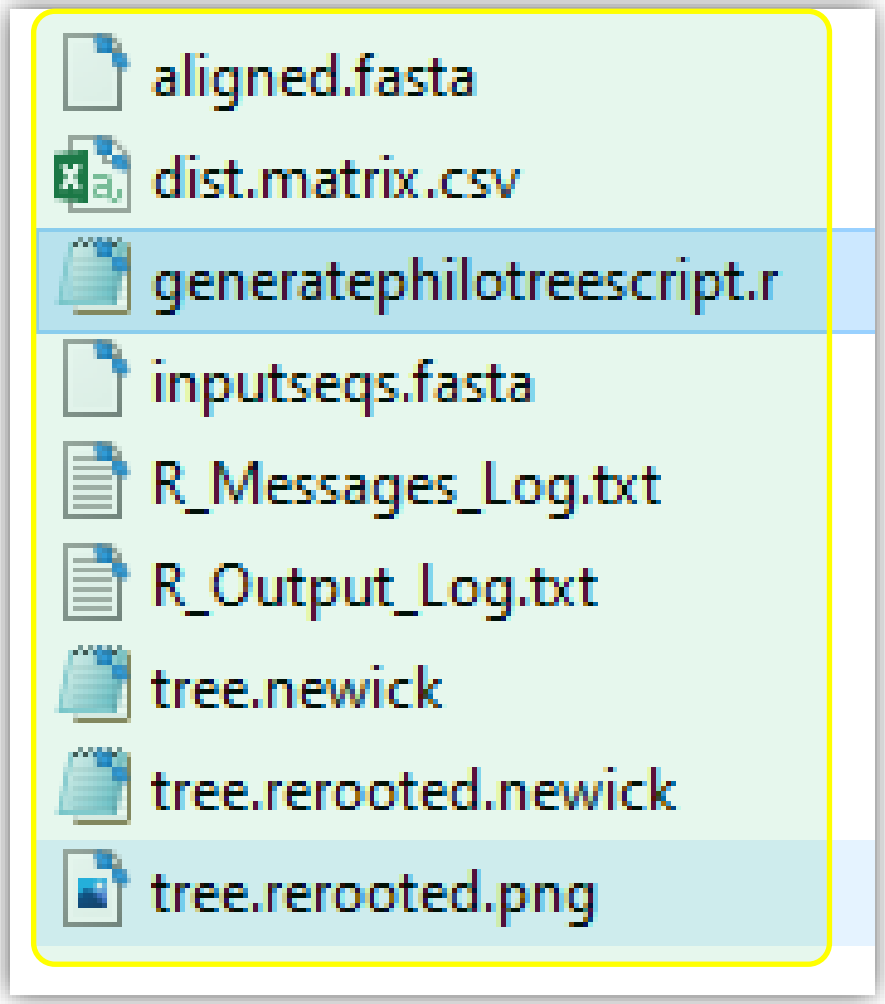
The next step is setting how the sequence to be used for each species to build the tree will be selected, the style of label for the tree, and the type of sequences marker to be used. Just place the mouse cursor over any option to see an explanation.
For this example, we'll use the default settings. Click on "OK" to continue.



Once the process completed a dialog box will be displayed, where you can select opening the folder where data has been saved, opening the resulting tree, or just accept (“OK”) and continue. Let’s select “Open folder”



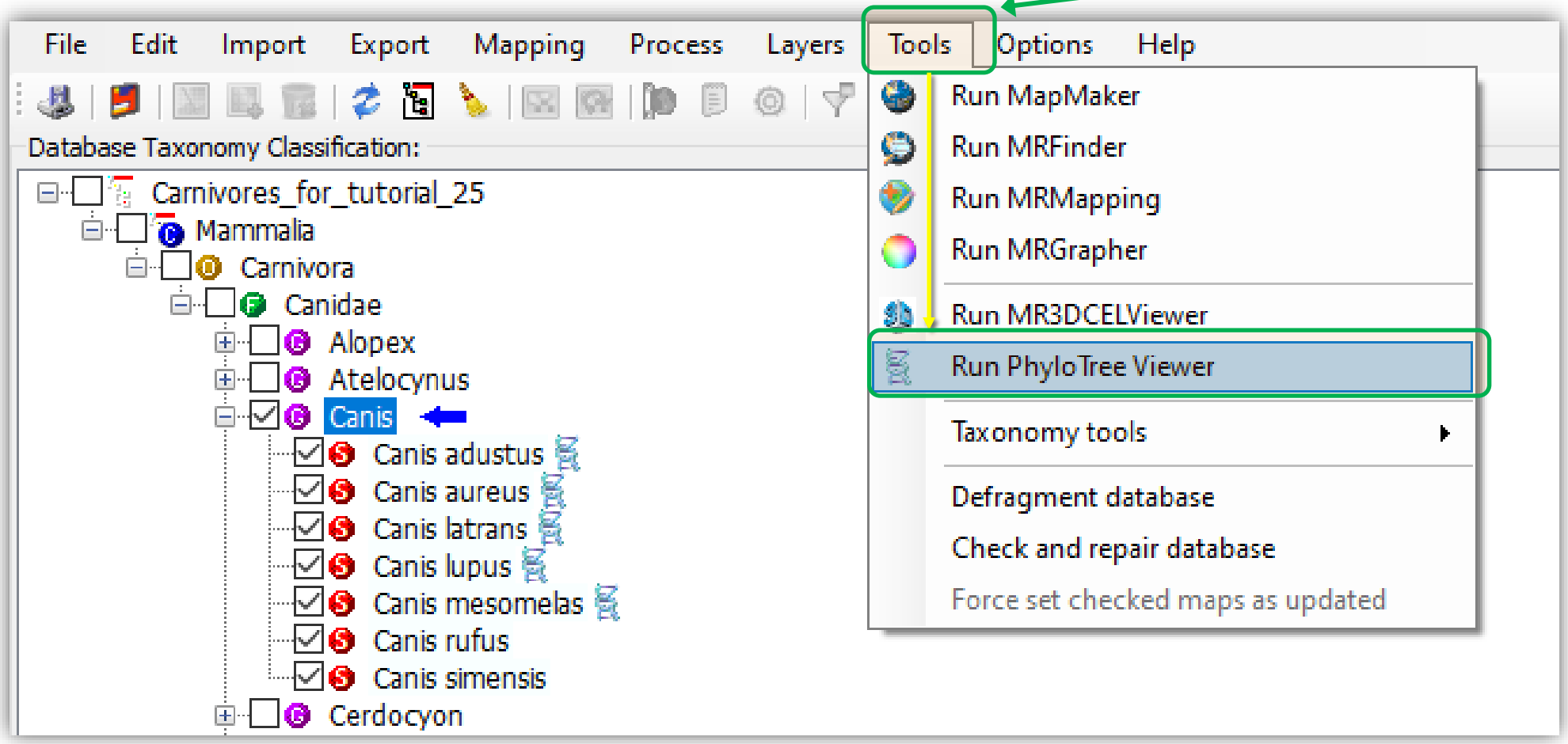
In the output folder we'll find the following files:



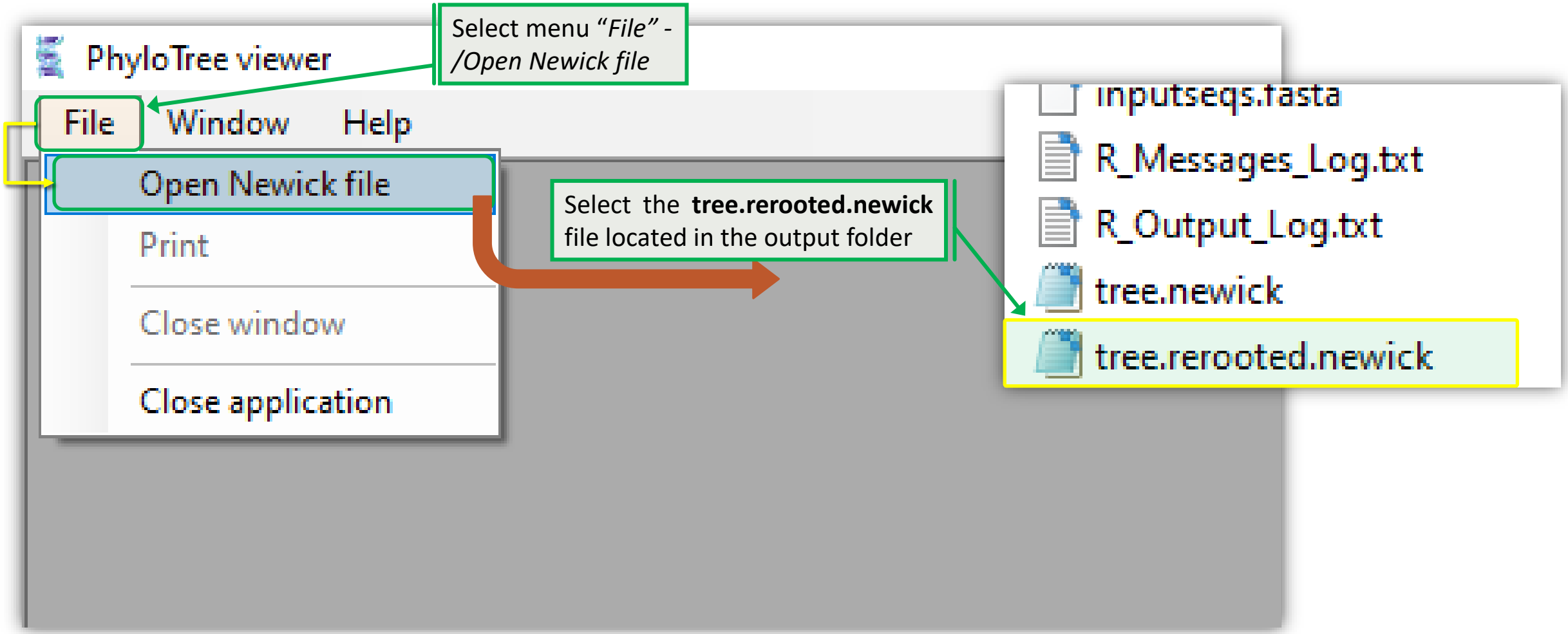
- **aligned.fasta:** The result of the sequence alignment performed before building the tree
- **dist.matrix.csv:** the obtained distance matrix between each sequence
- **generatephilotreescript.r:** ModestR internally runs an R script which uses several packages of the Bioconductor project and others to perform the sequence alignment and phylo tree. In this file you can find this full R script, which can be run in R or Rstudio, provided you install the required packages
- **inputseqs.fasta**: the collection of sequences selected (one by species) to build the tree
- **R_Messages_Log.txt** and **R_Output_Log.txt:** the output and the messages generated in R when running the script.
- **tree.newick:** the obtained tree in Newick format.
- **tree.rerooted.newick:** the obtained tree once rerooted, in Newick format
- **tree.rerooted.png:** aa graph of the obtained tree once rerooted, in png format

Let's come back to Data Manager and go to menu *Tools/Run Phylotree viewer.*

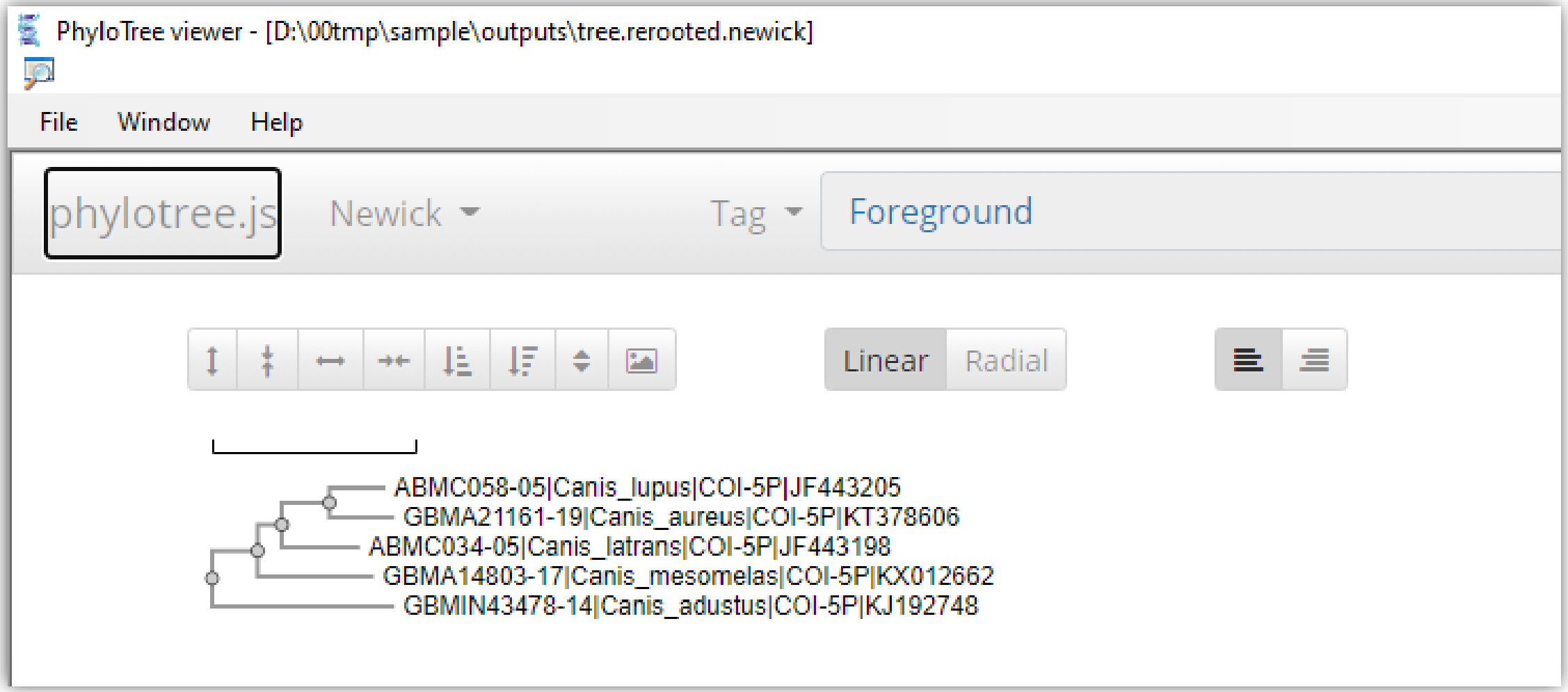
Select menu "*Tools*" - */Run PhyloTree Viewer*



A new window will be opened. It is a simple phylogenetic tree viewer that can display any tree in Newick format. Go to menu *File/Open Newick file* and select the **tree.rerooted.newick** file located in the output folder seen before.



Once loaded, the tree will be displayed. This viewer allows some options such as changing branch spacing, using radial tree, etc.



By the ModestR team

It was the Step by step tutorial:

Doing phylogenetic trees with ModestR and BOLD

You can find this one and other tutorials in http://WWW.MODESTR.ES

Thank you for your interest.