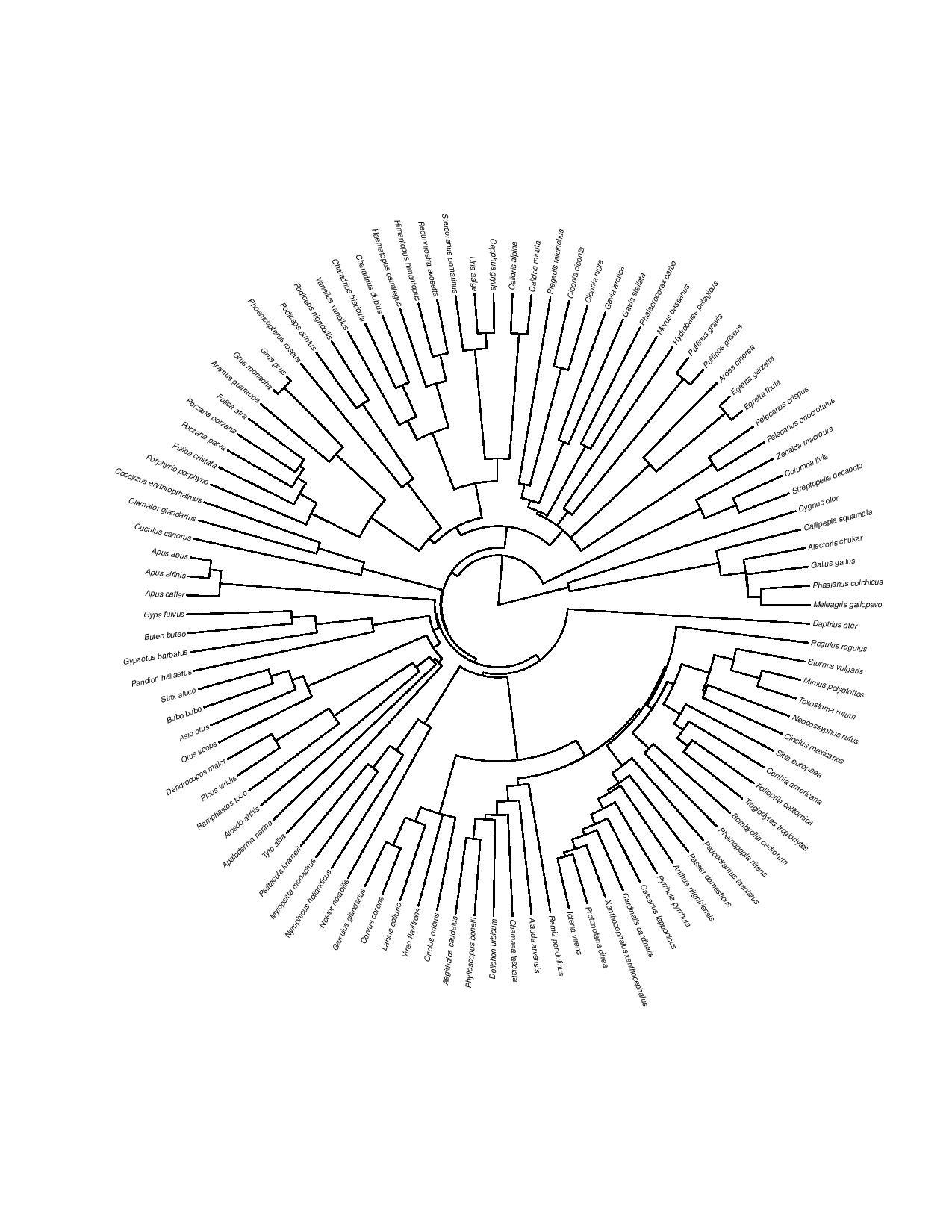
Episode#3 : The Final Concert

We started by building a phylogenetic tree of all the species we used, in order to visualize the diversity of the species we had chosen in our database. We were able to build it thanks to birdtree.org. We selected all the species we were studying, and generated 100 trees. We generated a consensus using the phytools library on R and the least-squares method. We chose a fan representation of the tree.



*Figure 1 : Phylogenetic tree of the species we used*

Last week we started writing a code for sound analysis. So we started the week by improving it and debugging it. First, we added maximal and minimal values as well as the standard deviation for each variable in the output dataframe. It also now suppresses, for a given species, all lines with NA values when calculating means for different variables, as these lines correspond to moments where there is too much noise and not enough signal. We also added lines that allow the code to save the spectrograms as pdf. Finally, we put our algorithm in a function, in order to put it in a loop and automatize the analysis for all the sound files. We thus only had to start the codes once, and it analyzed all our sound files.

We encountered a “fatal error” from R during the sound analysis, as one of the sound files (Common starling) was too heavy (36Mo, 15min of recording), so we cut out the last fourteen minutes in order to work with a much shorter file. We also modified the names of the sound files: we replaced the “-” in the vernacular names by spaces, as the code segments the file names using the “-”.

We commented on the code to help people understand how it works. We wrote a READ ME to give some tips for the use of the code, and to help other people to run it more smoothly. The READ ME also gives the meaning of each column in the final dataframe obtained.

By looking at the output dataframe “birds.csv” created by the code, we saw that for some species, some variables still had an “NA” mean value. By looking at the spectrograms we understood that it corresponded to the recordings with a lot of background noise. For some species (for example black swift Apus Apus), we also saw that the analyzed dominant frequencies did not correspond to the dominant frequencies we could observe on the spectrograms. Thus, we concluded that our method for sound analysis does not function well with recordings with too much noise.

We did not have the time to check for the quality of the analysis of each recording, by comparing the dominant frequency to the spectrogram, but it is something that we should do if we still had time on the project. It would help us know for what species it would be interesting to find a better quality recording, eventually on another database than Xeno Canto.

We also did not have the time to put the output dataframe totally in English. Indeed some file names were in French and others in English, so in the “name” column we find vernacular names in French and others in English. But it is something crucial to do for everyone to be able to handle our analyzed sound database with ease.

Finally, the next steps of the study would be eventually to gather a bigger variety of sounds for each species, to compare the analysed sounds between to species, to gather a syrinx morphology database and to finally try to make a link between syrinx morphology and the characteristics of the sounds produced.

Of course, all of these steps will take a lot of time, and we knew from the beginning we could bring only a small stone to this research. But we were honoured to be able to do so.