Tutorial 2 Work

Tutorials Explanation: Each tutorial presentation is expected to be at least 20 minutes. Tutorials will consist of 2 people. Tutorial code must be posted to Github and a link sent to the instructor to get credit. Every week all students will come up with questions about the tutorials and the next week, everyone will be quizzed on the previous week’s tutorials. Making quiz questions will be recorded as class participation, while the quizzes will be graded.

**For each tutorial presentation, you will have slides describing the algorithms (compare/contrast).. that’s “half”.**

Then have on the GitHub, the input files, the procedure AND/OR Code, the output files.

Instructions for the Datasets:

HMP Data: /ifs/groups/eces450650Grp/data/HMP

Class GitHub: <https://github.com/EESI/bio-course-materials>

Connecting to Picotte: ssh -l jpp75 picottelogin.urcf.drexel.edu

(Must use the VPN first)

Username: jpp75

Password: See index card

Pulling Files from the Cluster: (This pulls the fasta file we wanted)

scp jpp75@picottelogin.urcf.drexel.edu:/ifs/groups/eces450650Grp/data/Haemophilus\_influenzae\_16S.fasta /mnt/c/Users/jpalk/OneDrive/Desktop/Drexel\ Stuff/Senior\ Year\ College/Spring\ Term/Statistical\ Analysis\ of\ Genomics\ \(ECES\ 650\)/Tutorial\ 2\ Work

Assignment Prompt: Tree Inference

Students will align the sequences using MAFFT and build the trees using RaxML and FastTree, all using Cipres (<http://www.phylo.org/>). The students will comment on differences in the trees due to the different alignment and tree methodology and the biological differences. The trees under comparison are the RAxML v.8 on XSEDE and FastTree algorithms.

In these tutorials, the students will be expected to dedicate a large portion of the class to how the underlying algorithms work for tree inference. An analysis of the pros and cons of each of the 2 methods mentioned above are expected. Please correspond with the instructor.

Visualize Final Tree with iTOL (interactive tree of life): <https://itol.embl.de/>

In these tutorials, the students will be expected to dedicate a large portion of the class to how the underlying algorithms work for both alignment and tree inference. An analysis of the pros and cons of each of the 4 methods mentioned above are expected.

*Dataset: 16S for Haemophilus influenzae: /ifs/groups/eces450650Grp/data/Haemophilus\_influenzae\_16S.fasta*

MAFFT Info

* <https://watermark.silverchair.com/btq224.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAArwwggK4BgkqhkiG9w0BBwagggKpMIICpQIBADCCAp4GCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQM6by-HrQhLJcfjKMWAgEQgIICb_YwtD11nvl32fHIy5yGjeewuSdJAW71oMywHPsngCCY5w67qx3mOrW0R_klc4hLtYe6gr04-9DA1eUdp0aCncfmROWX9c8aqmK1fV5kkosEO9N87ICJxH3QKBNmXP-bOjeqnYGWSv3Kmirhhrq75vlqVSImUuAaujVSyiR_FDn2OBcuPBVBBssqulrgjUm_EYx1Ds9HI-4oHtClyeFuMgMWMguMugPwznhSXhlUNhjWdyemFOPTjekM0JJp0_xC7gTQFzqRE3aplCerjlRPwzlJw0XNQvKTBJFPfH8cSwG93YODtGD-wIkxiEVAlqF_IV7foTsQSHigPt-9VGNuXZP280vg_6Nd7qJ43imwBM9hvI04T3yU9sXBk25n4DXXhSySvdoPFfQlX43ZBQGcxNU26zdBf2xlMJ0LCjfocuaoDtX_LYlNnCKaHwSYOU0hBO5U3vLN4_Pt5Lbn16M23V94pskB1BZ964n-iIhf1zzI6XURHZy3HHil3SvsQQS51s5B2c6cKdHp1KHrnWW23FYO4gY7SMJBnCd39INvT-SYlqyP1_TKERP__Oj4m7e0M52uVUUsNdUx5mrziIWmBcvFeEylOSKVaaFbcb26LOMAdCfmUSSknX_wPIuV5alM3UPfWbvSULbW5-9b5vKb4hcY3d0oChOp8M516sOEPjgCTg6CSEHpE6-vaTpOyeaGHrUYPVpOs_3yX5UgBAj5royq0KHo_MGmQXvIUbajZBBNZ7deacAf_QP8KXHPKKd3hyOHnmkUqo3JtOZxZ3IgNHcIegeE2bWRa-jwIwdoguOjmIpb_NsZakH27DzdNrEE>

FastTree Info

* Visualizing the .tree file using iTOL

Question:

1. Since we are only doing Tutorial 2, then the final paragraph isn’t necessary for us to satisfy correct? Since we only are analyzing and comparing the 2 methods FastTree and RAxML using the MAFFT alignment technique?