

Assignment 4.2

Phylogenomics

Advanced Bioinformatics for NGS

Stefano Pellegrini (mlq211)

November 4, 2020

- 1 Why would one want to experiment with excluding third codon positions in protein-coding data for phylogenetic inference? (1 sentence suffices, 5 points).

Because due to the degeneracy of the genetic code, third codon positions in protein coding genes show high rates of substitution. Therefore, if these positions are included in protein-coding data for phylogenetic analysis, the accumulation of their mutations may lead to wrong phylogenies.

- 2 Paste the IQ-TREE command here (3 points).

```
./iqtree -s regier.nex -m HKY+I+G -spp part1_2.nex -bb 1000 -pre  
regier_result_bb -nt 4
```

- 3 Paste the resulting newick tree string here (1 point).

```
(FauNEOPT:0.1017518013,((((((((((((LfoCHIL0:0.0517843414,SpoCHIL0  
:0.0596634175)100:0.0138597968,Sco1CHIL0:0.0643139555)96:0.0077601372,  
Ctas2CHIL0:0.0671335186)100:0.0221751078,(((NamDIPL0:0.0614123425,  
AmaDIPL0:0.0571100179)100:0.0127688481,Pge2DIPL0:0.0720595408)  
100:0.0156106338,Pol2DIPL0:0.0933139627)99:0.0112858969)  
97:0.0082962403,((Han2SYMPH:0.0628348228,Scu3SYMPH:0.0498753515)  
100:0.0400128732,EuryPAUR0:0.1280348179)98:0.0108155998)  
100:0.0124618642,(((LpoXIPHOS:0.0099219349,Cro2XIPHOS:0.0105526066)  
100:0.0705008925,((((MgaARACH:0.0429761733,StpARACH:0.0557465085)  
100:0.0138168671,Pma2ARACH:0.0443863935)100:0.0107171267,AchARACH  
:0.0735802887)99:0.0091154728,(HariARACH:0.0200045750,HspARACH  
:0.0194701043)100:0.0568396939)100:0.0126325360,(LnigARACH  
:0.0931353427,EgigARACH:0.0598012358)91:0.0128400519)25:0.0051027403,(  
Crp2ARACH:0.0673165286,PwhARACH:0.1054433640)56:0.0102326988)  
38:0.0066531376)74:0.0087770561,((Amb2ARACH:0.0851034285,IpumARACH  
:0.1120923511)98:0.0124216541,Din2ARACH:0.1517313091)58:0.0053661289)  
100:0.0246981794,(((ThulTARD:0.1086224027,MtdTARD:0.1102345784)  
100:0.1503076777,((Pno2ONYCH:0.0199356437,ErwONYCH:0.0070678237)  
100:0.0311237026,PepONYCH:0.0355350334)100:0.0632907453)  
99:0.0194011467,((((TorPYCNO:0.0488901323,AeliPYCNO:0.0334343454)  
99:0.0104748209,AhiPYCNO:0.0354717767)100:0.0087770354,ElePYCNO  
:0.0541314874)99:0.0130161482,Col2PYCNO:0.0550416511)100:0.0779050544)  
86:0.0098531996)99:0.0102688124)100:0.0338822491,((OstOST:0.1426611059,  
DtyMYSTACO:0.1377125853)97:0.0133118752,((Arg2BIURA:0.0923967470,  
AarPENTA:0.1538624979)100:0.0429726462,(HapaOST:0.0718470109,SkleOST  
:0.0506496014)100:0.0693582038)98:0.0117211395)100:0.0130650956)  
96:0.0074028317,((((MesoCOPE:0.0228475365,A369COPE:0.0370117180)  
100:0.0652056403,EafCOPE:0.1255038820)100:0.0600057434,(NheMALA
```

```
:0.0826533687,(((LemMALA:0.0569321013,Avu3MALA:0.0943971203)
85:0.0111495466,NeoMALA:0.0682312160)100:0.0160860781)
100:0.0649642448,(((BbaTHECOS:0.0239868960,CfrTHECOS:0.0240980495)
100:0.0088457240,LeanTHECOS:0.0400796447)100:0.0256895529,LoxTHECOS
:0.1082629435)100:0.0727023586)86:0.0100468691)100:0.0104813310,(((
TloBRANCH:0.0934686781,((DmaBRANCH:0.0680778399,Lle2BRANCH
:0.0454890328)100:0.0177956170,LynBRANCH:0.0796798335)100:0.0156455368)
100:0.0171223789,(Asa3BRANCH:0.0563087566,ufsBRANCH:0.0332232919)
100:0.0925505090)100:0.0212989911)95:0.0090367340)99:0.0070391209,(
StuREMI:0.0781422293,HmaCEPHAL:0.1054077541)98:0.0125891280)
99:0.0080544819,((PaqCOLL:0.0959780661,(OimCOLL:0.0722450138,Tom2COLL
:0.0579987899)100:0.0207673979)100:0.0733782203,(JapDIPLUR
:0.0713122246,EfrDIPLUR:0.1062881698)100:0.0158215458)92:0.0073000487)
100:0.0154347415,(PsaARCHEO:0.0125488883,MbaARCHEO:0.0109489174)
100:0.0660268394)93:0.0091802186,(CliZYGEM:0.0487247748,NmeZYGEM
:0.0455974211)100:0.0213463782)98:0.0054343111,(MayEPHEM:0.0363747322,
EinEPHEM:0.0413048133)100:0.0551779822)87:0.0048647347,(IveODONAT
:0.0394205424,LlyODONAT:0.0323574881)100:0.0564801303)
98:0.0060184563,((ApaukNEOPT:0.0405081502,CpoNEOPT:0.0289825196)
100:0.0258903138,PquNEOPT:0.0368953584)100:0.0996905351)
100:0.0103205880,(PamNEOPT:0.0364590173,AdoNEOPT:0.0599289260)
100:0.0111738043);
```

- 4 Looking at the terminals ending in XxxNEOPT (short for Neoptera, a group of winged insects), does this tree support that Neoptera are a monophyletic group (= a clade)? (Answer with Yes/No. 5 points)

Yes.

- 5 Looking at the terminals ending in XxxARACH (short for Arachnida, spiders), does this tree support that spiders are a monophyletic group (= a clade)? (Answer with Yes/No. 5 points).

No.

- 6 Run ASTRAL as we did in class on the regier.gene.trees file. Paste the ASTRAL command here (3 points).

```
java -jar Astral/astral.5.15.1.jar -i regier.gene.trees -o regier.species.
tree -T 2
```

- 7 Paste the resulting newick tree string here (1 points).

```
(AarPENTA,(Arg2BIURA,((((HmaCEPHAL,((DtyMYSTACO,((JapDIPLUR,(OstOST,(
LoxTHECOS,(LeanTHECOS,(BbaTHECOS,CfrTHECOS)0.9:0.2579462326917244)
1:0.6131044728864088)1:2.015354133102803)0.7:0.13919500218232522)
0.94:0.2617485026787862,(EafCOPE,(A369COPE,MesoCOPE)
1:2.179525000236819)1:0.6309259731654561)0.5:0.0640038494982401)
0.47:0.10439583424976343,(PquNEOPT,(ApaukNEOPT,CpoNEOPT)
1:0.9270779974933916)1:2.1711067312023538)0.8:0.20492161245496154)
0.77:0.15659725710423306,(NheMALA,(Avu3MALA,(LemMALA,NeoMALA)
0.8:0.13855364459033945)0.63:0.09299136469708288)1:1.3969055987237824)
0.38:0.04913238675333127,((Asa3BRANCH,ufsBRANCH)1:2.8455295153324665,(
TloBRANCH,(LynBRANCH,(Lle2BRANCH,DmaBRANCH)1:0.4582052925409832)
0.81:0.16444962380621866)1:0.7381386032175303)1:0.3235588711977975)
0.73:0.13491000171501777,(HapaOST,SkleOST)1:1.0118872735596771)
0.62:0.08879940999650603,(((StuREMI,(((IveODONAT,LlyODONAT)
1:1.788827001369514,(FauNEOPT,(AdoNEOPT,PamNEOPT)
```

```

0.66:0.0940259144302094)0.99:0.2996854074557852)
0.84:0.14868880340129356,(EinEPHEM,MayEPHEM)1:1.8853248309082922)
0.44:0.07089754999969072,(CliZYGEM,NmeZYGEM)1:0.4179195293172652)
0.7:0.1088044146904491,(MbaARCHEO,PsaARCHEO)1:3.4231762883809314)
1:0.4959192795473894)0.6:0.06918113355472015,(EfrDIPLUR,(PaqCOLL,(
OimCOLL,Tom2COLL)0.84:0.20321688304635002)1:1.3708234817375347)
0.74:0.14051735950832323)0.69:0.09063459188189088)
0.83:0.14368948752087068,((((Din2ARACH,(PwhARACH,((HariARACH,
HspARACH)1:3.091042453358317,(AchARACH,(Pma2ARACH,(MgaARACH,StpARACH)
1:0.5200044000945292)1:0.4120015092708607)0.57:0.06710876048911354)
1:0.36070037123300874,(Cro2XIPHOS,LpoXIPHOS)1:3.3929166902645442)
0.88:0.1816951471044377)0.37:0.007842276045730205,(LnigARACH,EgigARACH)
0.45:0.03105243345556902)0.95:0.21025065363247483)
0.88:0.18680232319521617,(Crp2ARACH,IpumARACH)0.61:0.08559258044288205)
1:0.3614172184956892,((Col2PYCNO,(ElePYCNO,(TorPYCNO,(AeliPYCNO,
AhiPYCNO)0.71:0.13361626298733612)0.98:0.28776589704785277)
0.61:0.08254583789165482)1:2.3506457279215396,(PepONYCH,(ErwONYCH,
Pno2ONYCH)1:1.7047480922384253)1:2.716839140927625)
0.39:0.03430571175380222)1:0.31642311244111654,(EuryPAURO,((Pol2DIPL0,(
Pge2DIPL0,(AmaDIPL0,NamDIPL0)0.84:0.2504116510748734)
0.9:0.20841720892919038)0.59:0.07141936560271873,(Ctas2CHIL0,(ScolCHIL0
,(LfoCHIL0,Sp0CHIL0)0.95:0.25820608923625615)0.95:0.22181491556424696)
1:0.4297521010375983)0.74:0.11195926010977736)0.2:0.07269074245584187)
1:0.3294316951496192,(Han2SYMPH,Scu3SYMPH)1:0.7899852185996011)
0.52:0.04905348323418594,(Amb2ARACH,(MtdTARD,ThulTARD)
1:2.0734044383934935)0.43:0.04094979932086465)0.91:0.2339324044375095)
1:0.8876216504443147):0.0);

```

- 8 Calculate the Robinson-Foulds (RF) distance between the ASTRAL species tree you computed and the IQ-TREE concatenation tree. It does not matter which one of the two trees you use as the reference tree (parameter -r). Use TreeCmp and paste your command below (3 points).

```

java -jar TreeCmp_v2.0-b76/bin/treeCmp.jar -r regier_result_bb.treefile -i
regier.species.tree -d rf -N -o tree_comparisons/
regier_bb_iqtree_vs_species_astral.txt

```

- 9 Give the normalized RF distance (RF(0.5)_toUnifAvg) as we did in class (give the value, 2 points).

0.4033