Assignment 4.2

Mahdi

November 3, 2020

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

Since mutations in the third codon can result in the same amino acid, for a given sequence, all third codon positions will have a higher frequency of mutations which lead to phylogenetic itnerference.

Paste the IQ-TREE command here (3 points)

```
./iqtree -s regier.nex -spp part12.nex -bb 1000 -m HKY+I+G -nt 4
```

Paste the resulting newick tree string here (1 point):

```
(FauNEOPT:0.1017632793,(((((((((((((((LfoCHILO:0.0517900947,
SpoCHILO:0.0596696512)100:0.0138623296,ScolCHILO:0.0643209688)92
:0.0077607125,Ctas2CHILO:0.0671410943)100:0.0221772753,(((NamDIPLO:
0.0614186509,AmaDIPLO:0.0571159717)100:0.0127720194,Pge2DIPLO:
0.0720672424)100:0.0156120351,Pol2DIPLO:0.0933238822)100:
0.0112866217)98:0.0082966167,((Han2SYMPH:0.0628417954,Scu3SYMPH:
0.0498808881)100:0.0400156882,EuryPAURO:0.1280457762)99:0.0108188775)
100:0.0124627412,((((LpoXIPHOS:0.0099220636,Cro2XIPHOS:0.0105538516)
100:0.0705069978,((((((MgaARACH:0.0429801360,StpARACH:0.0557521371)
100:0.0138186359,Pma2ARACH:0.0443903755)100:0.0107176192,AchARACH:
0.0735871521)99:0.0091173737,(HariARACH:0.0200065872,HspARACH:
0.0194725998)100:0.0568444314)100:0.0126329089,(LnigARACH:0.0931431589,
EgigARACH:0.0598060329)86:0.0128430064)24:0.0051035126,(Crp2ARACH:
0.0673214497,PwhARACH:0.1054517942)52:0.0102356618)42:0.0066550510)76:
0.0087779327,((Amb2ARACH:0.0851098833,IpumARACH:0.1121020037)97:
0.0124247404,Din2ARACH:0.1517437953)57:0.0053674835)100:0.0247012198,
(((ThulTARD:0.1086335328,MtdTARD:0.1102446998)100:0.1503201028,
((Pno2ONYCH:0.0199373605,ErwONYCH:0.0070683567)100:0.0311263641,
PepONYCH:0.0355392950)100:0.0632936457)99:0.0194084967,((((TorPYCNO
:0.0488956154,AeliPYCNO:0.0334375263)99:0.0104753534,AhiPYCNO:
0.0354756204)100:0.0087792370,ElePYCNO:0.0541370416)99:0.0130171390,
Col2PYCNO:0.0550477195)100:0.0779125346)87:0.0098540820)99:
0.0102691410)100:0.0338865973,((OstOST:0.1426718843,DtyMYSTACO:
0.1377234439)96:0.0133174945,((Arg2BIURA:0.0924049946,AarPENTA:
0.1538746813)100:0.0429816588,(HapaOST:0.0718568405,SkleOST:
0.0506489763)100:0.0693764858)98:0.0117189861)99:0.0130682271)94:
0.0074031375,((((MesoCOPE:0.0228513467,A369COPE:0.0370151881)100:
0.0652130667,EafCOPE:0.1255172451)100:0.0600130504,((NheMALA:
0.0826626943,((LemMALA:0.0569383639,Avu3MALA:0.0944074514)91:
0.0111517920,NeoMALA:0.0682388504)100:0.0160880251)100:0.0649714011,
```

```
(((BbaTHECOS:0.0239891689,CfrTHECOS:0.0241001970)100:0.0088464847,
LeanTHECOS:0.0400847585)100:0.0256921046,LoxTHECOS:0.1082752325)100:
0.0727106910)88:0.0100494010)100:0.0104822341,((TloBRANCH:0.0934786854,
((DmaBRANCH:0.0680857763,Lle2BRANCH:0.0454941927)100:0.0177979471,
LynBRANCH:0.0796887701)100:0.0156469865)100:0.0171245849,(Asa3BRANCH:
0.0563148333,ufsBRANCH:0.0332261127)100:0.0925606848)100:0.0213016253)
92:0.0090380108)99:0.0070401621,(StuREMI:0.0781495002,HmaCEPHAL:
0.1054191311)98:0.0125925753)99:0.0080550921,((PaqCOLL:0.0959893110,
(OimCOLL:0.0722535122,Tom2COLL:0.0580055369)100:0.0207694641)100:
0.0733868717,(JapDIPLUR:0.0713198271,EfrDIPLUR:0.1062999147)100:
0.0158242648)94:0.0073012445)100:0.0154370626,(PsaARCHEO:0.0125509637,
MbaARCHEO:0.0109494277)100:0.0660343280)93:0.0091810400,(CliZYGEN:
0.0487300847,NmeZYGEN:0.0456022167)100:0.0213496477)99:0.0054349021,
(MayEPHEM:0.0363786057,EinEPHEM:0.0413094183)100:0.0551843045)88:
0.0048652120,(IveODONAT:0.0394258749,LlyODONAT:0.0323607093)100:
0.0564866731)99:0.0060189987,((ApaukNEOPT:0.0405135834,CpoNEOPT:
0.0289857526)100:0.0258927184,PquNEOPT:0.0368994209)100:0.0997021377)
100:0.0103219864,(PamNEOPT:0.0364628431,AdoNEOPT:0.0599355859)100:0.0111761621);
```

Looking at the clade formed by terminals ending in XxxDIPLUR (short for Diplura or bristle-tails), which clade is their sister group? (List the terminal names of the sister group. 5 points).

The terminal names are PaqCOLL, OimCOLL and Tom2COLL.

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, they include only terminals with NEOPT.

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, they include terminals with XIPHOS.

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
```

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,(CliZYGEN,NmeZYGEN)
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,((Pol2DIPLO,
(Pge2DIPLO,(AmaDIPLO,NamDIPLO)0.84:0.2504116510748734)0.9:0.20841720892919038)
0.59:0.07141936560271873,(Ctas2CHILO,(ScolCHILO,(LfoCHILO,SpoCHILO)
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);
```

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 part12.nex.treefile -i
regier.species.tree -d rf -N -o tree_comparisons/iqtree_vs_astral.txt
```

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

0.4033