

ECOGEO 'Omics Training: 4.4 Phylogenetics Version 2

Nina Dombrowski and Kiley Seitz

Abstract

Provides a short introduction to phylogenetics and how to build a phylogenetic tree from concatenated ribosomal marker proteins.

Open this protocol inside the virtual machine (details in 'Start Instructions') for easy copy, paste of commands into the command line terminal window.

Citation: Nina Dombrowski and Kiley Seitz ECOGEO 'Omics Training: 4.4 Phylogenetics. **protocols.io**

dx.doi.org/10.17504/protocols.io.fjabkie

Published: 18 Aug 2016

Before start

Before starting, please visit the ECOGEO website for more information on this "Introduction to Environmental 'Omics" training series. The site contains a pre-packaged virtual machine that can be downloaded and used to run all of the protocols in this protocols.io collection. In addition to the VM, the website contains video and presentations from our initial "Intro to Env 'Omics" workshop held at the Univ. of Hawai'i at Manoa on 25-26 Jul 2016.

Please email 'ecogeo-join@earthcube.org' to join the ECOGEO listserv for future updates.

Protocol

Identify Marker Genes

Step 1.

Demo for identifying marker genes using [RiboDB](#).

Marker genes from reference genomes Hug et al (2016). 16 Ribosomal proteins: L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17, S19

Flavobacteria and Prochlorococcaceae → RpL14 (uL14), RpL22 (uL22), RpS8 (bS8), RpS18 (uS18)

 **LINK:**

<https://ribodb.univ-lyon1.fr/ribodb/ribodb-in.cgi>

cmd **COMMAND**

```
$ cd /home/c-debi/ecogeo/secrets/secretphylodir
```

Secret directory with pre-computed files. To access, open a new terminal window and cd.

 **NOTES**

Elisha Wood-Charlson 10 Aug 2016

Paste in Flavobacteriia (yes, with 2 i's)

Identify Marker Genes

Step 2.

Move compressed file to /home/c-debi/ecogeo/phylogenetics

cmd **COMMAND**

```
$ mv R-PROTS.tar.gz /home/c-debi/ecogeo/phylogenetics
```

Identify Marker Genes

Step 3.

Decompress file:

cmd **COMMAND**

```
$ tar zxvf R-PROTS.tar.gz
```

 **NOTES**

Elisha Wood-Charlson 10 Aug 2016

Contains 4 folders – bS18, uL14, uL22, uS8

Each contains a number of files – 3 nucleotide and 3 protein FASTA files

Identify Marker Genes

Step 4.

Navigate to each folder and copy the *_prot.fst files up two directories to /home/ecogeo/phylogenetics

cmd **COMMAND**

```
$ cp *_prot.fst /home/c-debi/ecogeo/phylogenetics
```

 **NOTES**

Elisha Wood-Charlson 10 Aug 2016

Navigate to each folder and use:

```
$ cp *_prot.fst /home/c-debi/ecogeo/phylogenetics
```

Identify Marker Genes

Step 5.

Clean up header names & repeat:

```
cmd COMMAND  
$ cut -f1 -d "~" bS18_prot.fst | sed 's/\\.//' | sed 's/|///' > temp1_bS18  
$ rm bS18_prot.fst  
$ mv temp1_bS18 bS18_prot.fst
```

🔗 NOTES

Elisha Wood-Charlson 10 Aug 2016

When renaming:

Check that you get unique names!

Perform & Trim Alignments

Step 6.

MUSCLE performs an alignment of the sequences and generates an aligned FASTA file output

```
cmd COMMAND  
$ muscle -maxiters 8 -in bS18_prot.fst -out bS18_prot.aln
```

Perform & Trim Alignments

Step 7.

Trim alignments in [Jalview](http://www.jalview.org/).

🔗 LINK:

<http://www.jalview.org/>

```
cmd COMMAND  
$ Jalview
```

Perform & Trim Alignments

Step 8.

TrimAl - performs automated trimming for which various parameters can be set for columns to include or exclude

```
cmd COMMAND  
$ trimal -in bS18_prot.aln -out bS18_prot.trim.aln -automated1
```

Concatenate and Tree Visualization

Step 9.

For a reasonable number of sequences possible to use [Galaxy](#). Follow steps in presentation for manipulating data in Galaxy. After you download the results, change the name to "ribomarkers.concat.fasta"

Combine aligned files (specifically for Galaxy format):



LINK:

<https://galaxyproject.org/>

cmd **COMMAND**

```
$ cat *mfa | cut -f1 -d "/" > Galaxy_input_aligned.fasta
```

Concatenate and Tree Visualization

Step 10.

Construct a maximum likelihood tree with FastTree.

Remove 5 sequences from ribomarkers.concat.fasta in Text Editor

-candidatus_sulcia_muelleri

-Flavobacterium_indicumGPTS100-9+DSM17447

-Flavobacterium_psychrophilum

-Riemerella_anatipestiferATCC11845+DSM15868

-Prochlorococcus_spMIT0801

Concatenate and Tree Visualization

Step 11.

Execute FastTree:

cmd **COMMAND**

```
$ FastTree -gamma ribomarkers.edit.concat.fasta > ribomarkers.concat.newick
```

NOTES

Elisha Wood-Charlson 10 Aug 2016

If this command can not read ribomarkers.edit.concat.fasta, editing may have altered the file in the previous step. Copy the file from /secretphyloDir, run command again, and confirm your newick file has data.

Concatenate and Tree Visualization

Step 12.

In Dendroscopy, "Enter Tree in Newick Format" requires copying and pasting the text version of ribomarkers.concat.newick in to the window

Visualize tree in Dendroscope:

cmd **COMMAND**

```
$ Dendroscope
```

NOTES

((Aequorivita_sublithincolaDSM14238:0.08326,(Dokdonia_sp4H-3-7-5:0.00305,Dokdonia_spMED134:0.01017)1.000:0.13088)0.921:0.03356,((Croceibacter_atlanticusHTCC2559:0.09962,(Zunongwangia_profundaSM-A87:0.04152,Gramella_forsetiiKT0803:0.05448)0.790:0.01993)0.747:0.01917,((Psychroflexus_torquusATCC700755:0.19204,Nonlabens_dokdonensisDSW-6:0.13910)0.915:0.03830,((Polaribacter_spMED152:0.11994,((Owenweeksia_hongkongensisDSM17368:0.20782,((((Prochlorococcus_marinusNATL1A:0.00237,Prochlorococcus_marinusNATL2A:0.00015)0.999:0.07150,(Prochlorococcus_marinus_subsp_marinusCCMP1375:0.05352,(Prochlorococcus_marinusMIT9303:0.00497,Prochlorococcus_marinusMIT9313:0.00211)1.000:0.07224)0.868:0.03076)1.000:0.11900,(Prochlorococcus_marinus_subsp_pastorisCCMP1986:0.00017,Prochlorococcus_marinusMIT9515:0.02187)0.961:0.03246)0.800:0.01717,(Prochlorococcus_marinusMIT9215:0.00779,(Prochlorococcus_marinusMIT9312:0.00778,(Prochlorococcus_marinusMIT9301:0.01215,(Prochlorococcus_marinusAS9601:0.00241,Prochlorococcus_spMIT0604:0.00015)0.000:0.00015)0.739:0.00187)0.657:0.00499)0.650:0.00683)1.000:1.17436,Fluviicola_taffensisDSM16823:0.18731)0.918:0.11519)0.903:0.05702,(((candidatus_uzinura_diaspidicolaASNER:0.35795,Endosymbiont_of_llaveia_axin_axin:0.25635)0.925:0.06383,(((candidatus_sulcia_muelleriSULCIA-ALF:0.03902,(candidatus_sulcia_muelleriGWSS:0.01425,candidatus_sulcia_muelleriDMIN:0.00529)0.784:0.01706)1.000:0.10330,((candidatus_sulcia_muelleriCARL:0.03192,candidatus_sulcia_muelleriPSPU:0.02297)1.000:0.14189,candidatus_sulcia_muelleriSMDSEM:0.23520)0.556:0.03894)1.000:0.38125,((Blattabacterium_spBLATTELLAGERMANICABGE:0.08500,(Blattabacterium_spBLABERUSGIGANTEUS:0.09600,(Blattabacterium_spNAUPHOETACINEREA:0.07443,Blattabacterium_spPANESTHIAANGUSTIPENNISSPADICABPAA:0.10253)0.495:0.02189)0.996:0.07934)0.915:0.05540,((Blattabacterium_spPERIPLANETAAMERICANABPLAN:0.01704,Blattabacterium_spBLATTAORIENTALISTARAZONA:0.01387)0.997:0.09216,(Blattabacterium_punctulatus:0.13409,Blattabacterium_spMASTOTERMESDARWINIENSISMADAR:0.11546)0.977:0.05685)0.543:0.03409)1.000:0.26701)0.996:0.16276)1.000:0.18036,(Weeksellia_virosaDSM16922:0.14099,((Ornithobacterium_rhinotrachealeORT-UMN88:0.0,Ornithobacterium_rhinotrachealeDSM15997:0.0):0.13612,(((Elizabethkingia_spBM10:0.00222,(Elizabethkingia_meningosepticaFMS-007:0.0,Elizabethkingia_anophelisNUHP1:0.0):0.00015)0.978:0.02005,(Chryseobacterium_spSTRB126:0.02239,Flavobacteriaceae_bacterium_3519_10:0.08219)0.967:0.02133)0.935:0.02107,((Riemerella_anatipestiferCH3:0.00015,(Riemerella_anatipestiferRA-CH-2:0.00232,Riemerella_anatipestiferRA-CH-1:0.00017)0.444:0.00232)0.792:0.00218,Riemerella_anatipestiferRA-GD:0.00222)0.901:0.01427)1.000:0.13750)0.293:0.03498)0.734:0.04156)0.414:0.03365)0.996:0.09624)0.984:0.06089,(((Maribacter_spHTCC2170:0.03440,Zobellia_galactanivorans:0.02677)0.976:0.03030,(Robiginitalea_biformataHTCC2501:0.06675,(Muricauda_lutaonensis:0.06165,Muricauda_ruestringensisDSM13258:0.06440)0.984:0.03474)0.918:0.01935)0.403:0.01022,((Cellulophaga_lytica:0.00015,Cellulophaga_lyticaDSM7489:0.00221)0.997:0.05540,((Cellulophaga_baltica18:0.0,Cellulophaga_balticaNN016038:0.0):0.00015,Cellulophaga_algicolaDSM14237:0.00886)0.998:0.03681)0.945:0.02148)0.765:0.02283)0.909:0.02506)0.218:0.00705)0.628:0.01463,(((Capnocytophaga_cani_morsusCC5:0.02621,Capnocytophaga_ochraceaDSM7271:0.05283)1.000:0.08911,((Myroides_profundi:0.0,Myroides_spA21:0.0):0.03621,((Flavobacterium_psychrophilumFPG3:0.00214,(Flavobacterium_psychrophilumFPG101:0.0,Flavobacterium_psychrophilumJIP02:0.0):0.00015)1.000:0.06831,(Flavobacterium_columnareATCC49512:0.05563,(Flavobacterium_branchiophilumFL-15:0.04473,Flavobacterium_johnsoniaeUW101:0.00169)0.331:0.00942)0.916:0.01402)0.870:0.02492)0.990:0.04233)0.941:0.03338,(Siansivirga_zeaxanthinifaciensCC-SAMT-1:0.04040,(Winogradskyella_spPG-2:0.08403,Lacinutrix_sp5H-3-7-4:0.04083)0.598:0.01215)0.873:0.02404)0.811:0.01568);