

Phylogenetics Tutorial 2

Using Phylogenies: Genomic Epidemiology

Finlay Maguire

Faculty of Computer Science

Table of contents

1. Overview
2. Installation
3. Phylogenomics
4. HGT

Overview

- Perform a core-genome phylogeny (phylogenomics)
- Use mixture models
- Perform topology comparison tests
- Use phylogenetics to investigate an HGT

Requirements

- Internet connection
- mafft
- trimal
- aliview
- iqtree
- FigTree
- snp-sites

Installation

If you don't have miniconda

<https://docs.conda.io/en/latest/miniconda.html>

```
conda create -n phylo -c bioconda mafft trimal fasttree iqtree roary  
snp-sites
```

```
conda activate phylo
```

Unfortunately, not everything is in bioconda:

- AliView
<https://github.com/AliView/AliView/releases>
- FigTree
<https://github.com/rambaut/figtree/releases>

Phylogenomics

- Get your genomes.
- Annotate those genomes (e.g. prokka)
- Find the annotations shared between the genomes and align core genome (e.g. roary)
- Extract the variant sites (snp-sites)
- Infer your phylogeny

```
scp transfer@veles.research.cs.dal.ca:/home/transfer/  
core_gene_alignment.aln .
```

password is 'phylogenetics'

```
snp-sites -o listeria_snps.fna core_gene_alignment.aln
```

Infer ML phylogeny with mixture model

```
iqtree -m "MIX{JC,GTR}+G4" -alrt 1000 -bb 1000 -nt AUTO -s  
listeria_snps.fna -pre listeria_mix
```

Infer ML phylogeny with GTR

```
iqtree -m "GTR" -nt AUTO -alrt 1000 -bb 1000 -s listeria_snps.fna -pre  
listeria_gtr
```

Tree Topology Tests

```
cat listeria_mix*.treefile listeria_mix*.bionj \  
  
listeria_gtr*.treefile listeria_gtr*.bionj > trees  
  
iqtree -nt AUTO -s listeria_snps.fna \  
  
-m "MIX{JC,GTR}+G4" -z trees -pre tree_comp \  
  
-n 0 -zb 1000
```

Visualise "Best" Tree

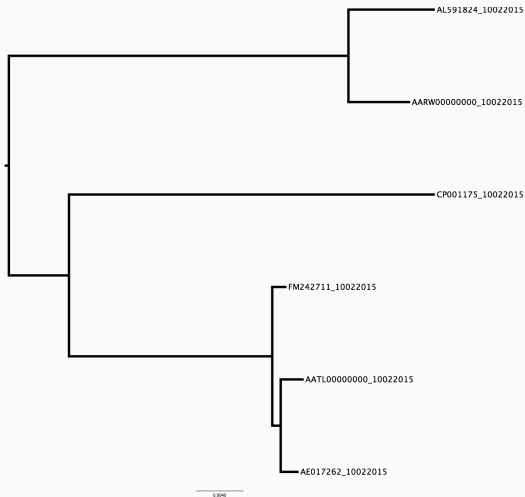


Figure 1: Roary Tutorial

HGT

Starting Sequence

hypothetical protein A5816_002916 [Enterococcus sp. 3G1_DIV0629]

GenBank: OT022244.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS	OT022244	1279 aa	linear	BCT 16-MAY-2017
DEFINITION	hypothetical protein A5816_002916 [Enterococcus sp. 3G1_DIV0629].			
ACCESSION	OT022244			
VERSION	OT022244.1			
DBLINK	BioProject: PRJNA313452			
	BioSample: SAMN04634015			
DBSOURCE	accession NGI_T01000004_1			
KEYWORDS	.			
SOURCE	Enterococcus sp. 3G1_DIV0629			
ORGANISM	Enterococcus sp. 3G1_DIV0629			

Customize view

Analyze this sequence

- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence

Related information

- BioProject

<https://www.ncbi.nlm.nih.gov/protein/OT022244.1>

BLAST® » blastp suite

Standard Protein BLAST

[blastn](#)

[blastp](#)

[blastx](#)

[tblastn](#)

[tblastx](#)

BLASTP programs search protein databases using a protein query.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

[Clear](#)

Query subrange

OT022244.1

From

To

Or, upload file

[Browse...](#)

No file selected.

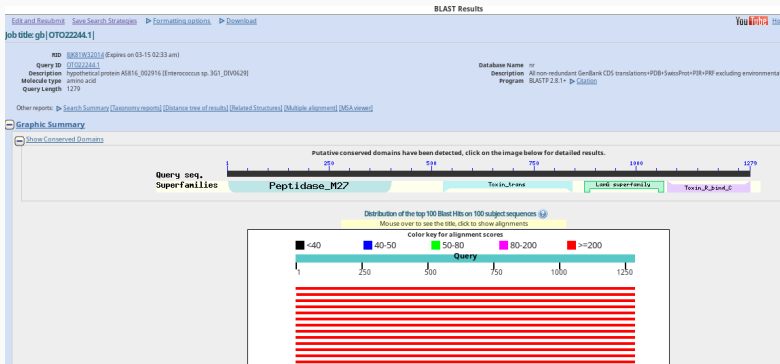
[?](#)

Job Title

Enter a descriptive title for your BLAST search

[?](#)

Get Sequences



- Align sequences: mafft
- Mask alignment: trimal
- Infer phylogeny (and fit model): iqtree
- Visualise phylogeny: FigTree

Questions?