



Research Topic

Microbial Taxonomy, Phylogeny and Biodiversity

Submission closed.

18 156

GET_PHYLOMARKERS, a Software Package to Select Optimal Orthologous Clusters for Phylogenomics and Inferring Pan-Genome Phylogenies, Used for a Critical Geno-Taxonomic Revision of the Genus *Stenotrophomonas*

Pablo Vinuesa , Luz E. Ochoa-Sánchez and
Bruno Contreras-Moreira

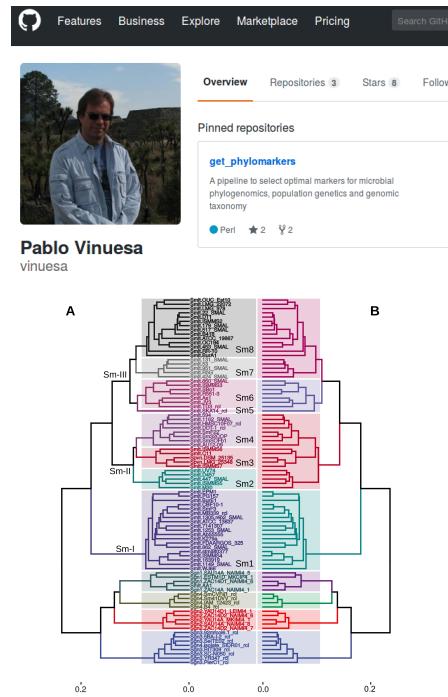
Original Research The massive accumulation of genome-sequences in public databases promoted the proliferation of genome-level phylogenetic analyses in many areas of biological research. However, due to diverse evolutionary and genetic processes, many loci have ...

Published on 01 May 2018
Front. Microbiol. doi: 10.3389/fmicb.2018.00771

1,223 total views Altmetric 27

Slides for qBio18: <http://www.qbio18.mx/>

https://github.com/vinuesa/get_phylomarkers



Slide 2 07/24/2018 19:05

The search for “well-behaved” molecular markers to infer genome phylogenies

... due to diverse evolutionary forces and processes, many loci in genomes have undesirable properties for phylogenetic reconstruction. If undetected, these can lead to erroneous or biased estimates (Shen et al., 2017; Parks et al., 2018), although, ironically, with strong branch support (Kumar et al., 2012). Their impact is particularly strong in concatenated datasets (Kubatko and Degnan, 2007; Degnan and Rosenberg, 2009), which are standard in microbial phylogenomics (Wu and Eisen, 2008). Hence, robust phylogenomic inference requires the selection of well-suited markers for the task (Vinuesa, 2010).

From Vinuesa et al. 2018, Front. Microbiol.

The search for “well-behaved” molecular markers to infer genome phylogenies

GET_PHYLOMARKERS is a software pipeline for the automatic and robust estimation of genome phylogenies using state-of-the art methods.

The emphasis of the pipeline is on selecting top-ranking markers for the task, based on the following quantitative/statistical criteria:

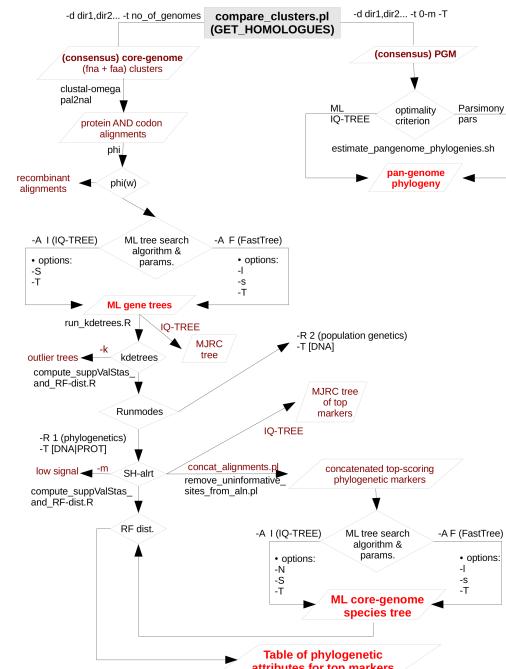
- (i) they should not present signs of recombination,
 - (ii) the resulting gene trees should not be anomalous or deviating from the distribution of tree topologies and branch lengths expected under the multispecies coalescent model
 - (iii) they should have a strong phylogenetic signal.

The top-scoring markers are concatenated to estimate the species phylogeny under the ML optimality criterion using either FastTree (Price et al., 2010) or IQ-TREE (IQT) (Nguyen et al., 2015).

From Vinuesa et al. 2018, Front. Microbiol.

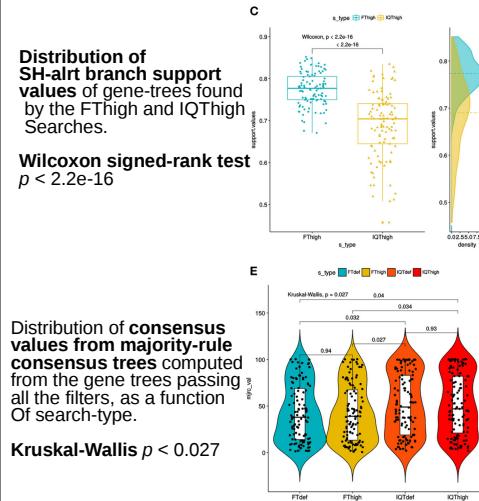
Flowchart showing how the GET_PHYLOMARKERS package integrates with

GET_HOMOLOGUES

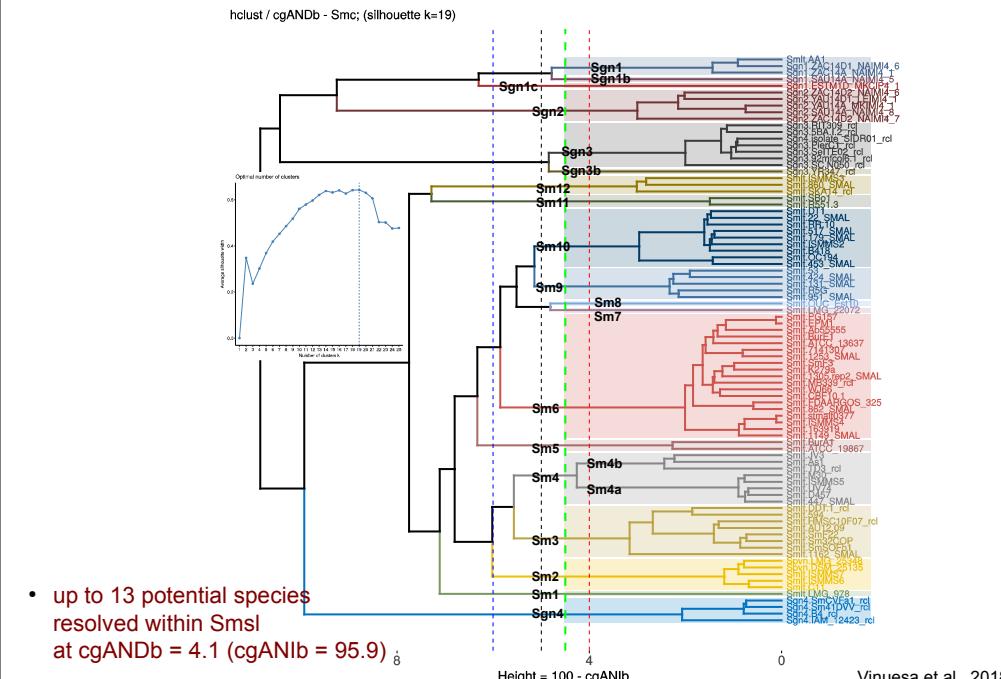


Benchmark analyses of the phylogenetic performance of FT vs IQ-TREE

Consensus core-genome (n=239) computed with GET_HOMOLOGUES



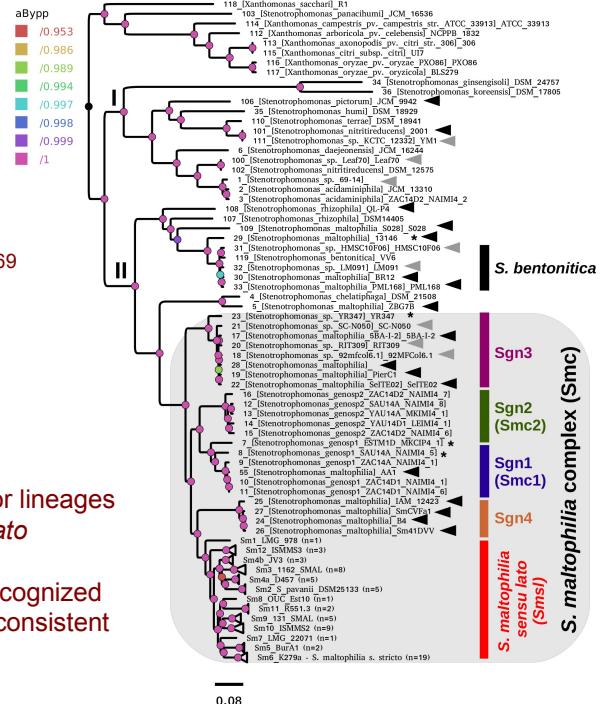
Unsupervised learning methods to find groups within the Smc and *S. maltophilia* s. lato lineages based on cgANDb (= 100 - cgANib)



Slide 5 07/24/2018 19:05

ML species tree

- 118 genomes
- Top 52(231) markers
- GTR+ASC+F+R7

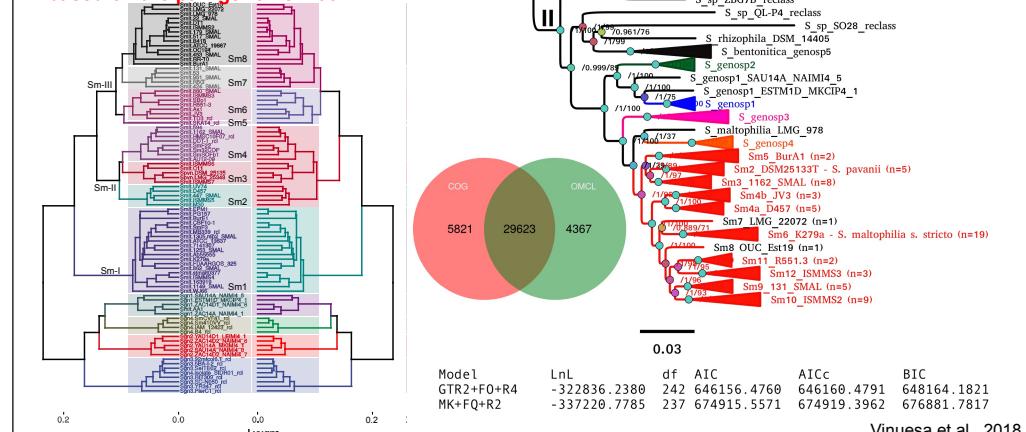


Slide 6 07/24/2018 19:05

ML pan-genome tree

- 118 genomes
- 29,623 markers
- GTR2+F0+R4

Unsupervised learning methods to find species-like clusters in the Smc based on the pan-genome matrix



Hands-on tutorial on GET_PHYLOMARKERS

source code and tutorial can be found at

https://github.com/vinuesa/get_phylomarkers