

GET_PHYLOMARKERS: open-source tools for phylogenomics

Pull requests Issues Marketplace Explore

Pablo Vinuesa
vinueza

I'm a microbiologist interested in genomics, ecology and evolution. I develop open-source code for these topics in collaboration with @ead-csic-combio

@CCG-UNAM
Cuernavaca, Morelos, Mexico
<http://www.ccg.unam.mx/~vinueza/>

Contribution activity

February 2019

Created 21 commits in 3 repositories

- vinueza/OMICAS_UAEM 16 commits
- vinueza/get_phylomarkers 3 commits
- vinueza/get_Darwinian_genes 2 commits

Edit

<https://github.com/vinueza>

Slide 3 03/07/2019 20:41

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

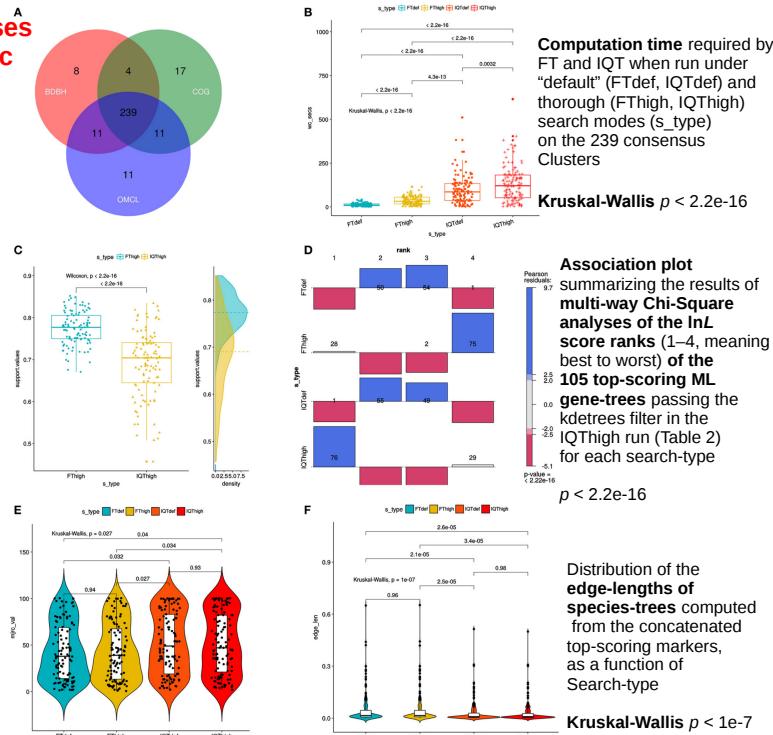
Consensus core-genome (n=239) computed with GET_HOMOLOGUES

Distribution of SH-alrt branch support values of gene-trees found by the FThigh and IQThigh Searches.

Wilcoxon signed-rank test $p < 2.2e-16$

Distribution of consensus values from majority-rule consensus trees computed from the gene trees passing all the filters, as a function Of search-type.

Kruskal-Wallis $p < 0.027$



Frontiers In Microbiology Research Topics



Research Topic

Microbial Taxonomy, Phylogeny and Biodiversity

Submission closed.

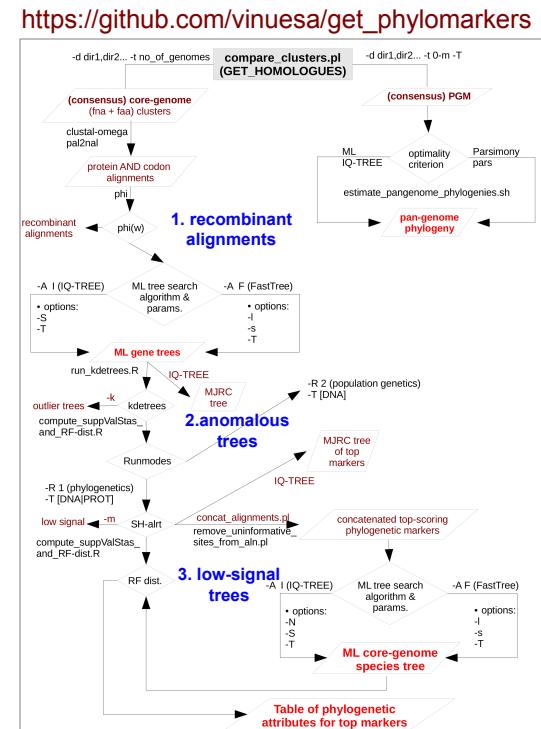
Overview Articles Authors Impact Comments

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



ML species tree

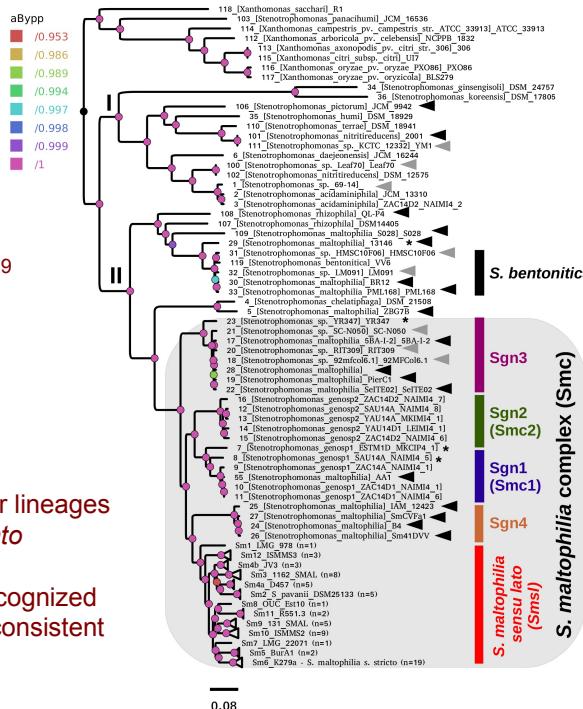
- 118 genomes
- Top 52(231) markers
- GTR+ASC+F+R7

- ~8% misclassified genomes 13/169
- ~8% unclassified genomes 14/169

- reclassified 27 RefSeq genome sequences!

- The Smc is split into 5 major lineages including *S. maltophilia* s. lato

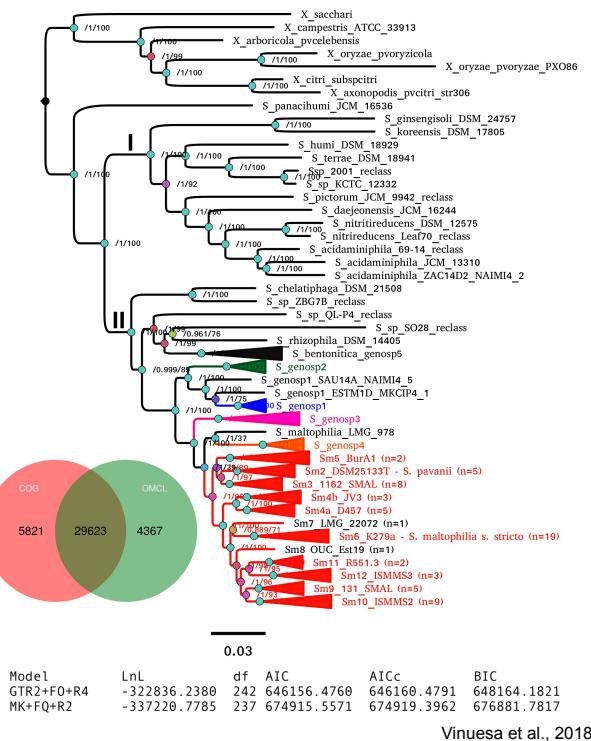
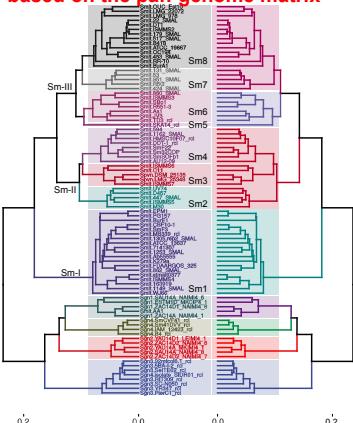
- 13 clades (species?) are recognized within the latter, which are consistent with a 95.9% cgANIB cutoff



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



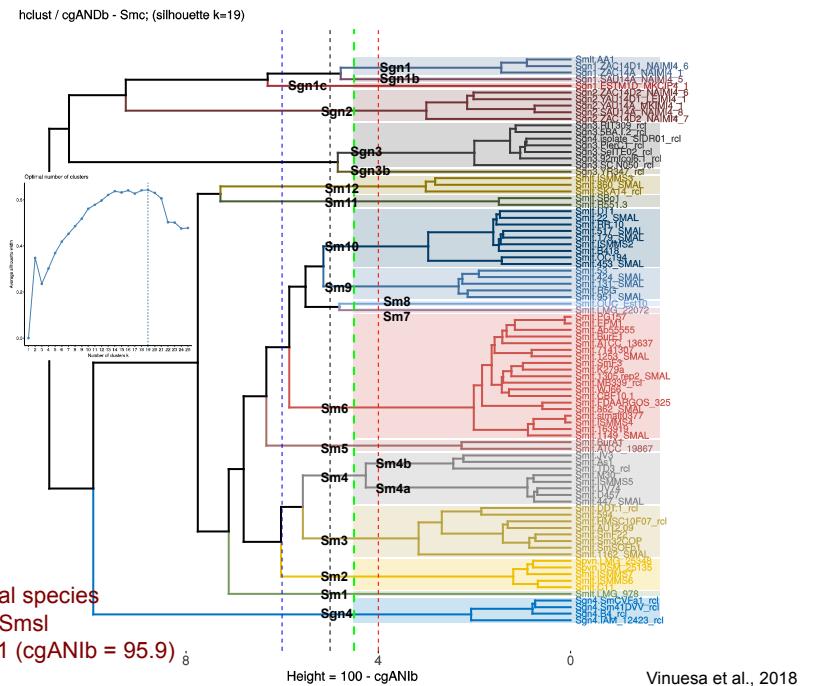
Model 1
GTR2+F0+R4
MK+FQ+R4

LnL	df	AIC	AICc	BIC
-322836.2380	242	646156.4760	646160.4791	648164.1821
-337220.7785	237	674915.5571	674919.3962	676881.7817

Vinuesa et al., 2018

Slide 6 03/07/2019 20:41

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



- up to 13 potential species resolved within SmcI at cgANDb = 4.1 (cgANlb = 95.9)