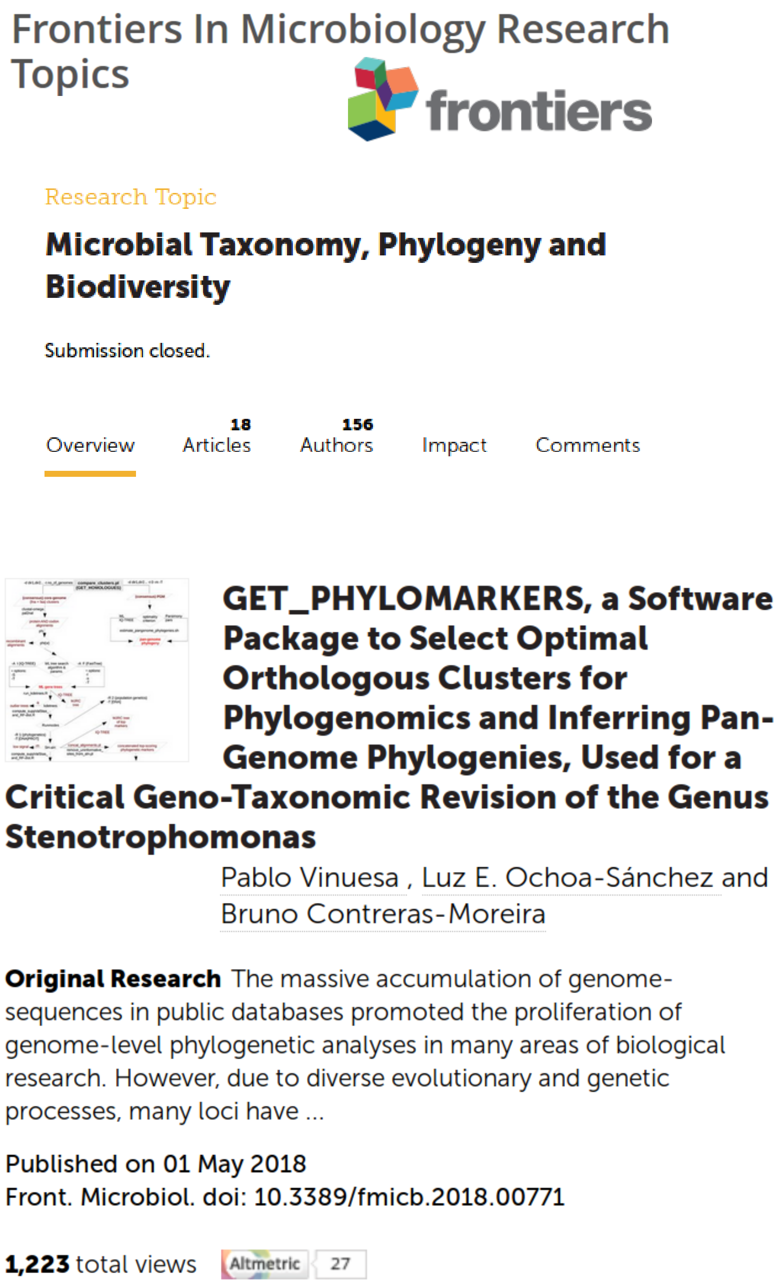


Frontiers In Microbiology Research Topics

frontiers

Research Topic

Microbial Taxonomy, Phylogeny and Biodiversity

Submission closed.

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

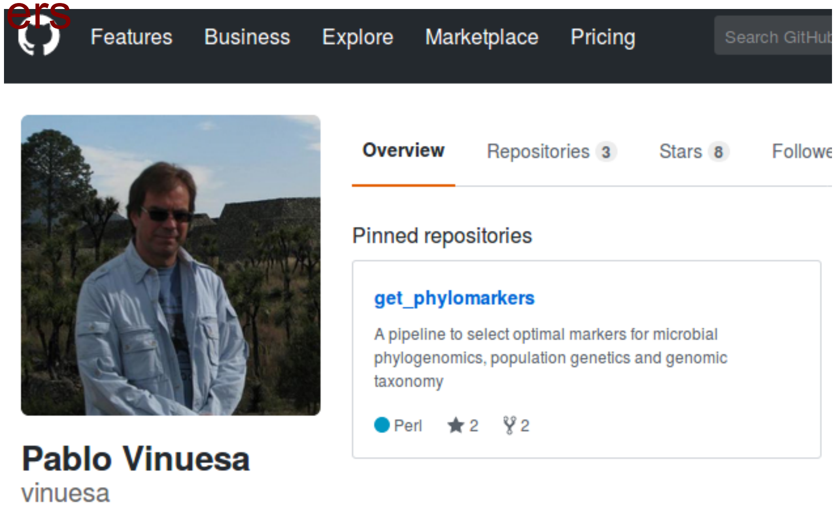
Pablo Vinuesa , Luz E. Ochoa-Sánchez and Bruno Contreras-Moreira

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

https://github.com/vinuesa/get_phylomarkers

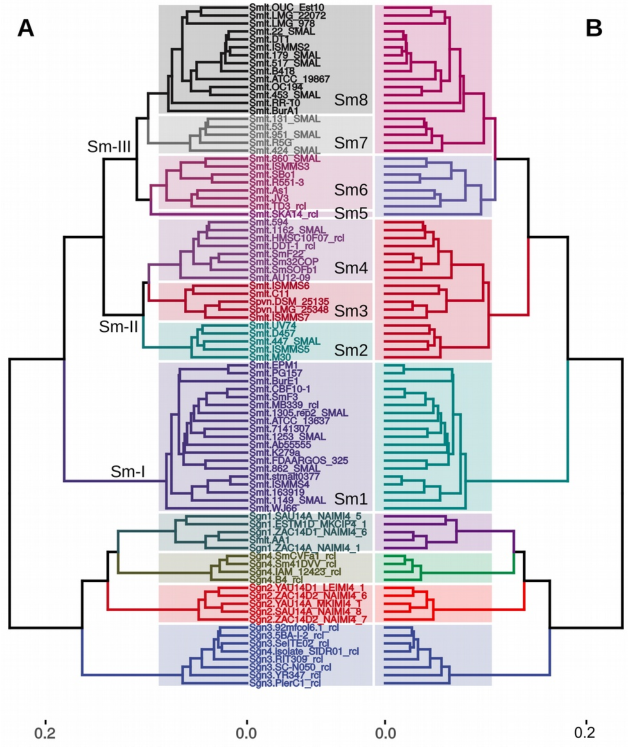


Pinned repositories

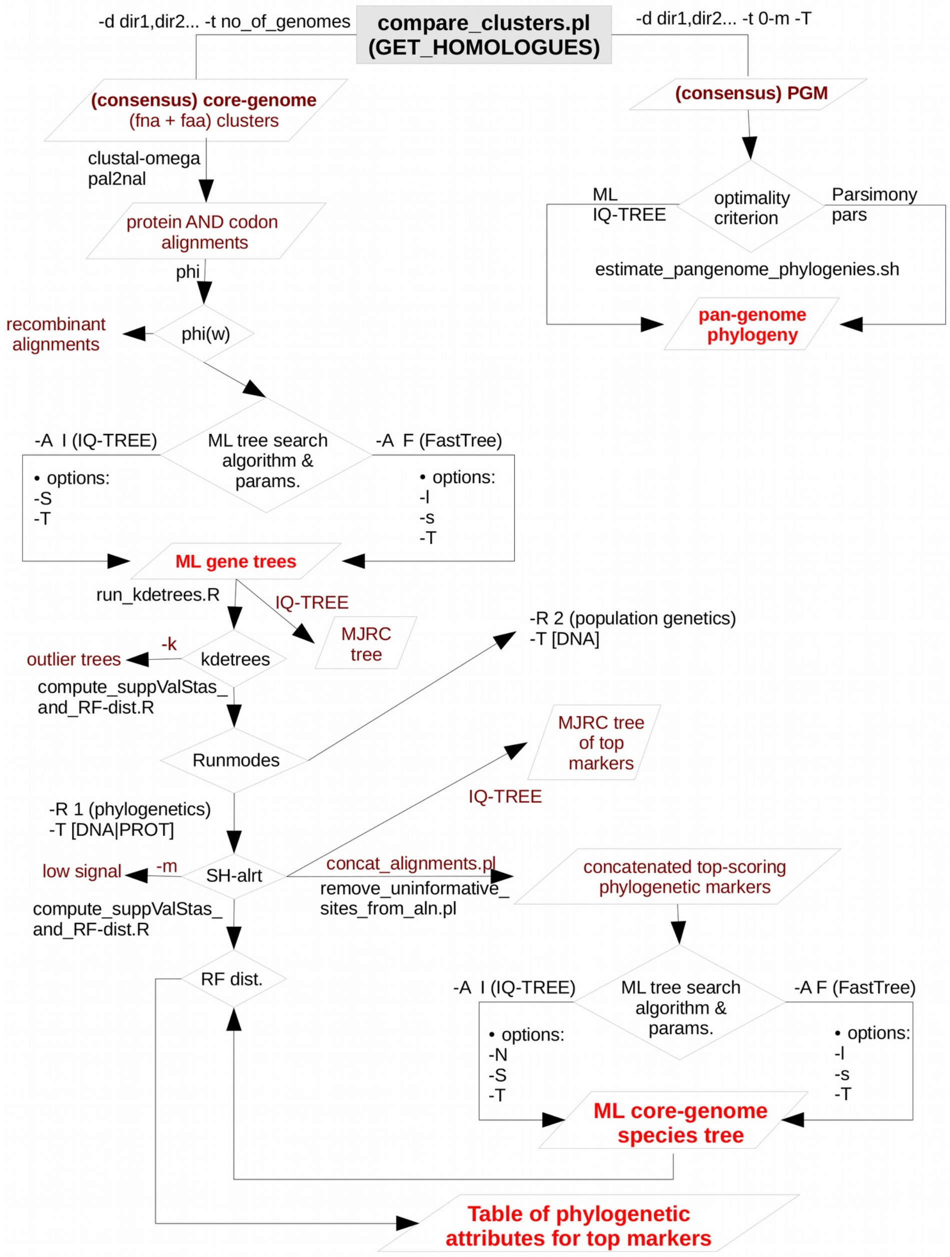
get_phylomarkers

A pipeline to select optimal markers for microbial phylogenomics, population genetics and genomic taxonomy

Pablo Vinuesa
vinuesa



Flowchart showing how the GET_PHYLOMARKERS package integrates with GET_HOMOLOGUES



-d dir1,dir2... -t no_of_genomes

compare_clusters.pl (GET_HOMOLOGUES)

-d dir1,dir2... -t 0-m -T

(consensus) core-genome
(fna + faa) clusters

(consensus) PGM

clustal-omega
pal2nal

protein AND codon alignments

phi

ML
IQ-TREE

optimality criterion

Parsimony
pars

estimate_pangenome_phylogenies.sh

pan-genome phylogeny

recombinant alignments

phi(w)

-A I (IQ-TREE)

ML tree search algorithm & params.

-A F (FastTree)

- options:
-S
-T

- options:
-l
-s
-T

ML gene trees

run_kdetrees.R

IQ-TREE

MJRC tree

-k

outlier trees

kdetrees

compute_suppValStas_and_RF-dist.R

-R 2 (population genetics)
-T [DNA]

Runmodes

MJRC tree of top markers

IQ-TREE

-R 1 (phylogenetics)
-T [DNA|PROT]

low signal

-m

SH-alrt

concat_alignments.pl

remove_uninformative_sites_from_aln.pl

concatenated top-scoring phylogenetic markers

compute_suppValStas_and_RF-dist.R

RF dist.

-A I (IQ-TREE)

ML tree search algorithm & params.

-A F (FastTree)

- options:
-N
-S
-T

- options:
-l
-s
-T

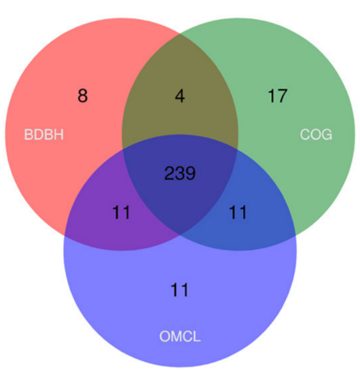
ML core-genome species tree

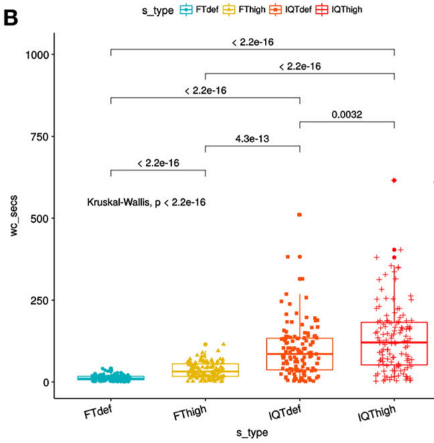
Table of phylogenetic attributes for top markers

Vinuesa et al., 2018

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

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



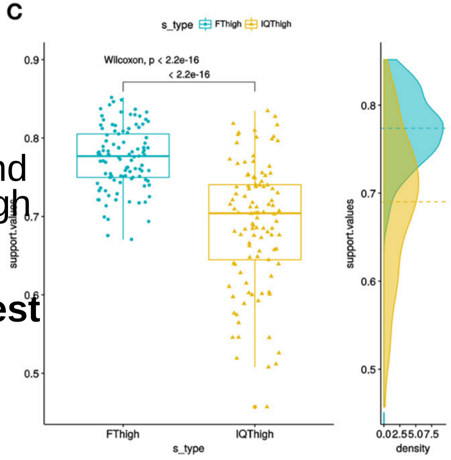


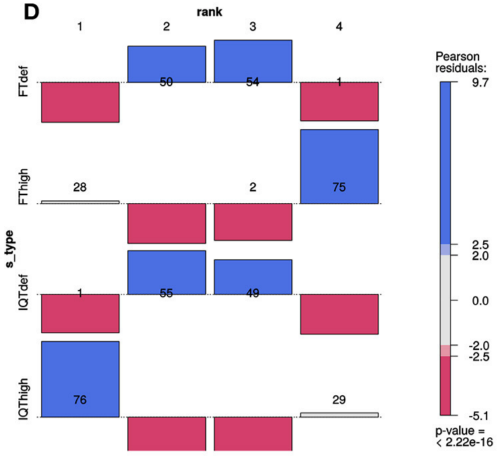
Computation time required
FT and IQT when run under
“default” (FTdef, IQTdef) and
thorough (FThigh, IQThigh)
search modes (s_type)
on the 239 consensus
Clusters

Kruskal-Wallis *p* < 2.2e-16

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



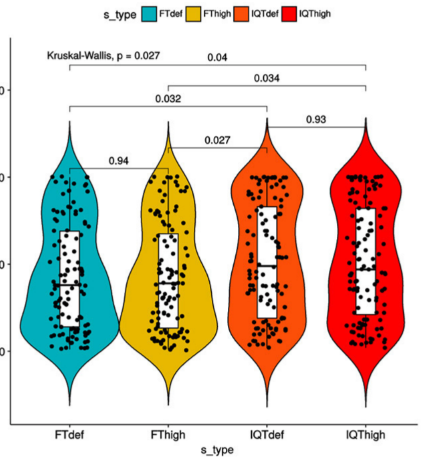


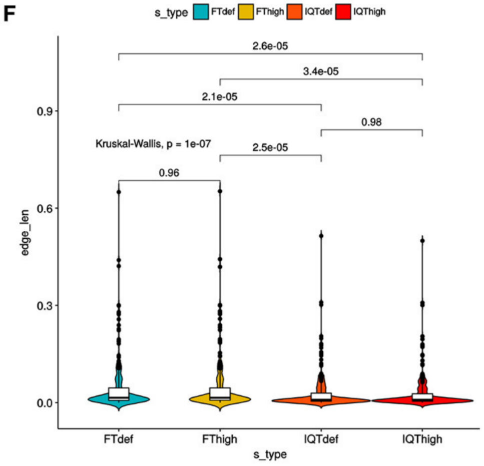
Association plot
summarizing the results o
multi-way Chi-Square
analyses of the ln*L*
score ranks (1–4, meanin
best to worst) **of the**
105 top-scoring ML
gene-trees passing the
kdetrees filter in the
IQThigh run (Table 2)
for each search-type

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





Distribution of the
edge-lengths of
species-trees computed
from the concatenated
top-scoring markers,
as a function of
Search-type

Kruskal-Wallis *p* < 1e-7

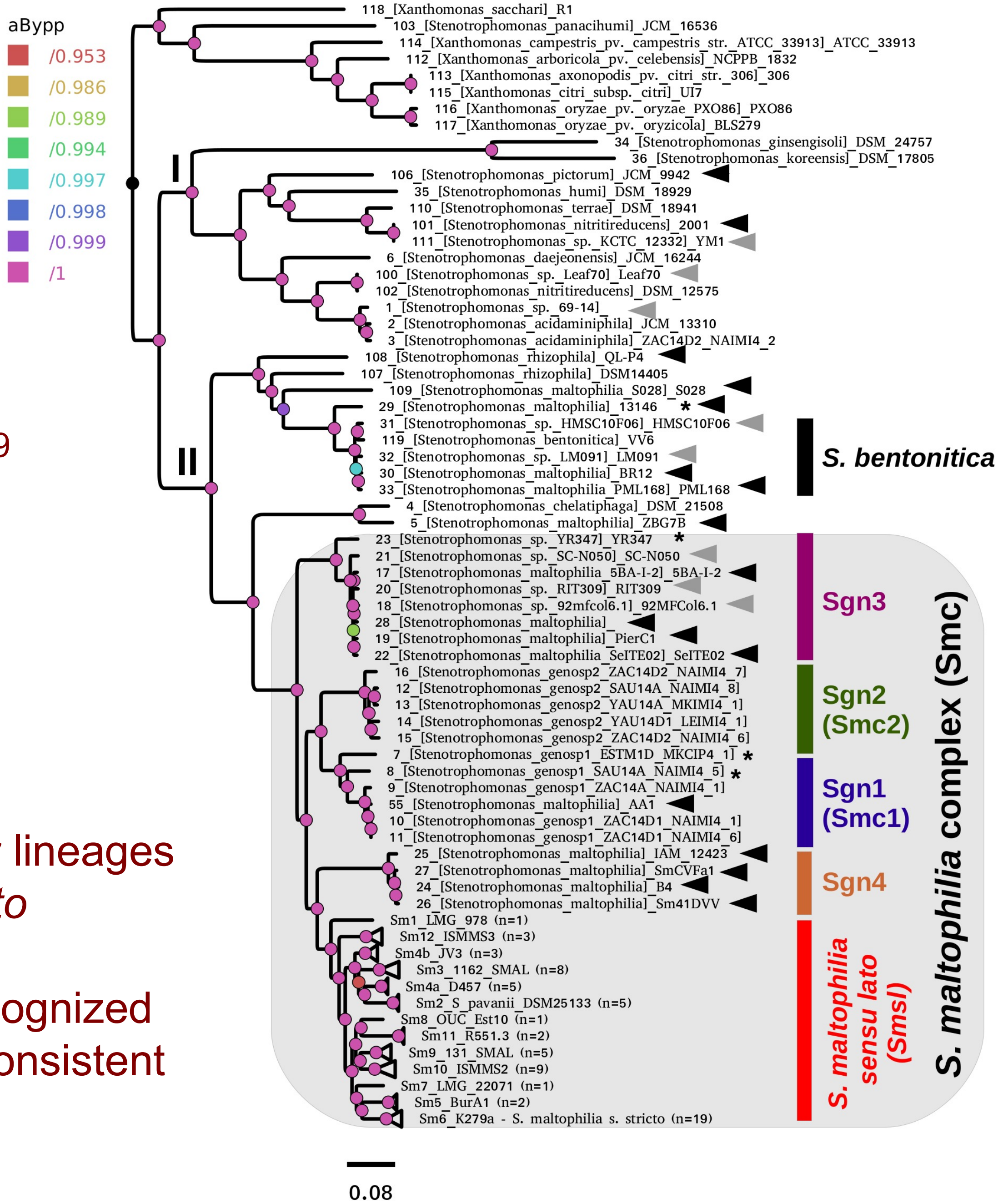
ML species tree

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

► ~8% misclassified genomes 13/169

► ~8% unclassified genomes 14/169

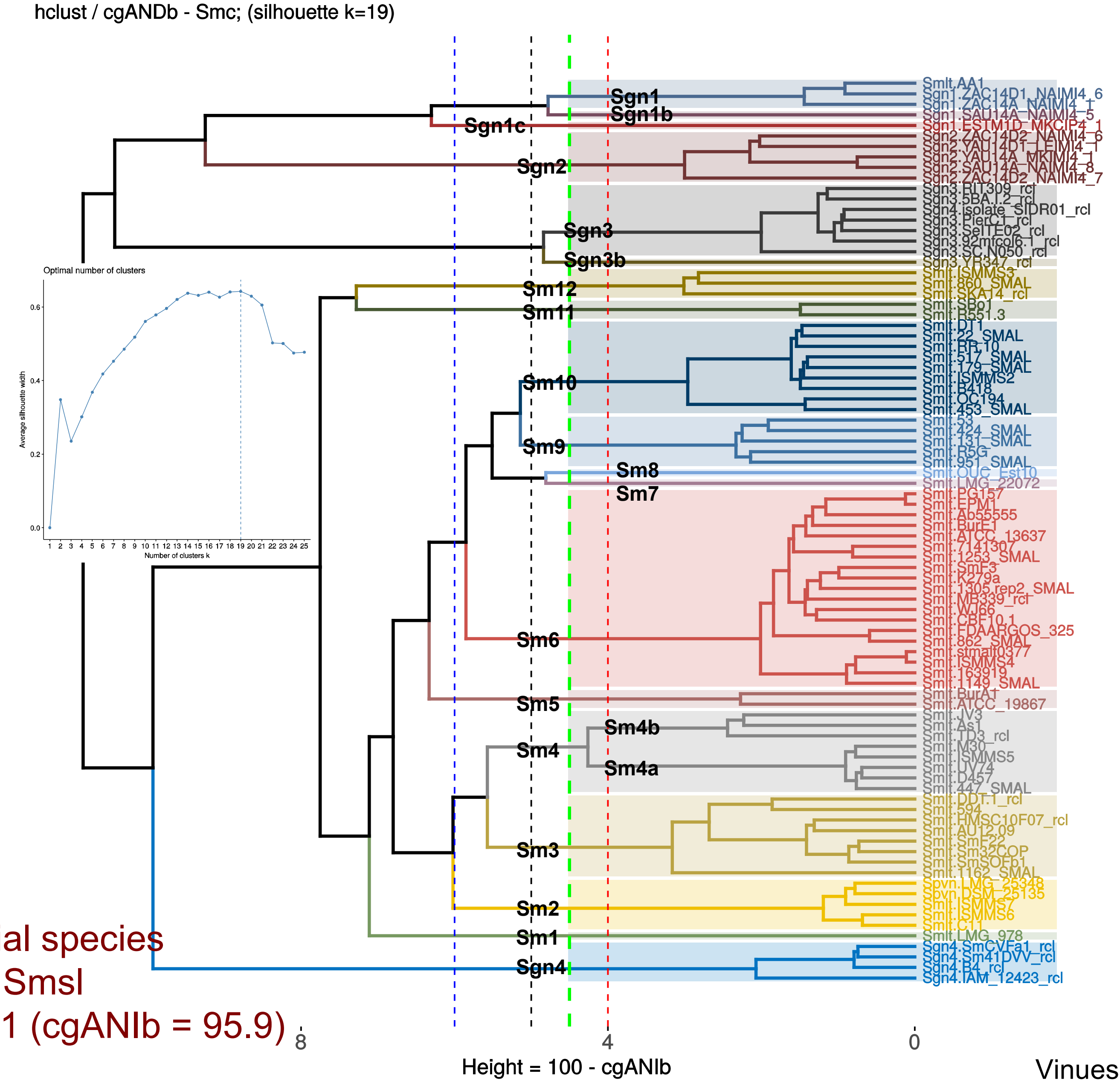
- reclassified 27 RefSeq genome sequences!



- 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

Vinuesa et al., 2018

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



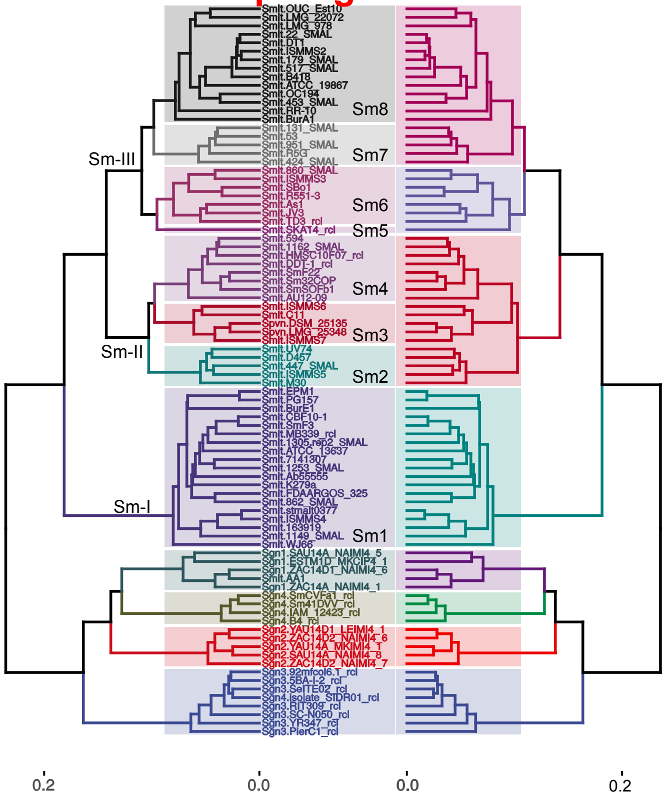
- up to 13 potential species resolved within Smsl at cgANDb = 4.1 (cgANIb = 95.9)

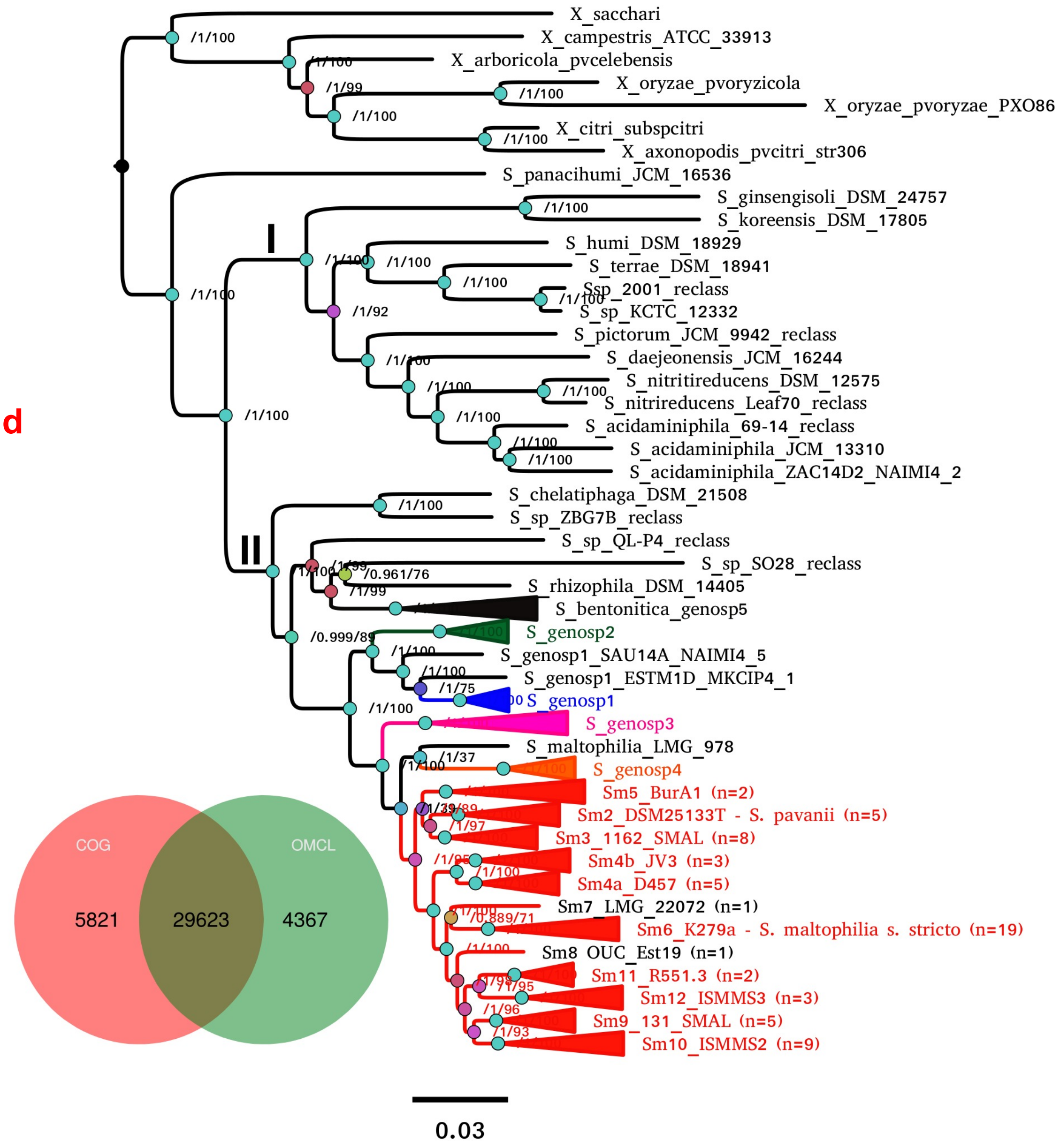
Vinuesa et al., 2018

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         LnL           df  AIC          AICc         BIC
GTR2+FO+R4    -322836.2380  242 646156.4760  646160.4791  648164.1821
MK+FQ+R2      -337220.7785  237 674915.5571  674919.3962  676881.7817
```

Vinuesa et al., 2018