

# SAGeR-P a FAIR resource enabling large scale functional comparative genomics



Jasper J Koehorst, Jesse CJ van Dam, Edoardo Saccenti,  
Vitor AP Martins dos Santos, Maria Suarez-Diez, Peter J Schaap  
Laboratory of Systems and Synthetic Biology  
Wageningen University & Research

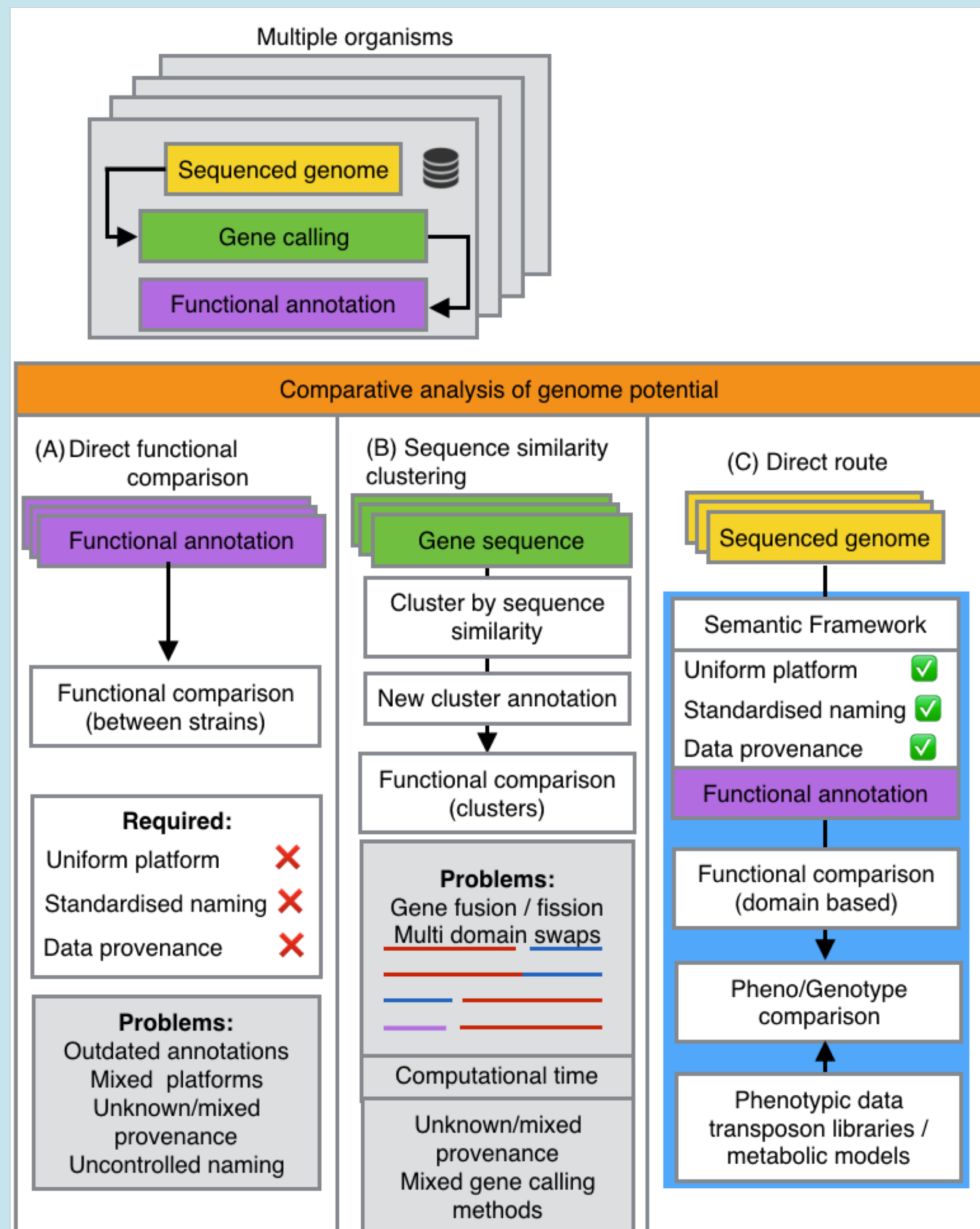
2.127.549.787

Triples available

6968

genomes available

## Approaches for comparative genomics



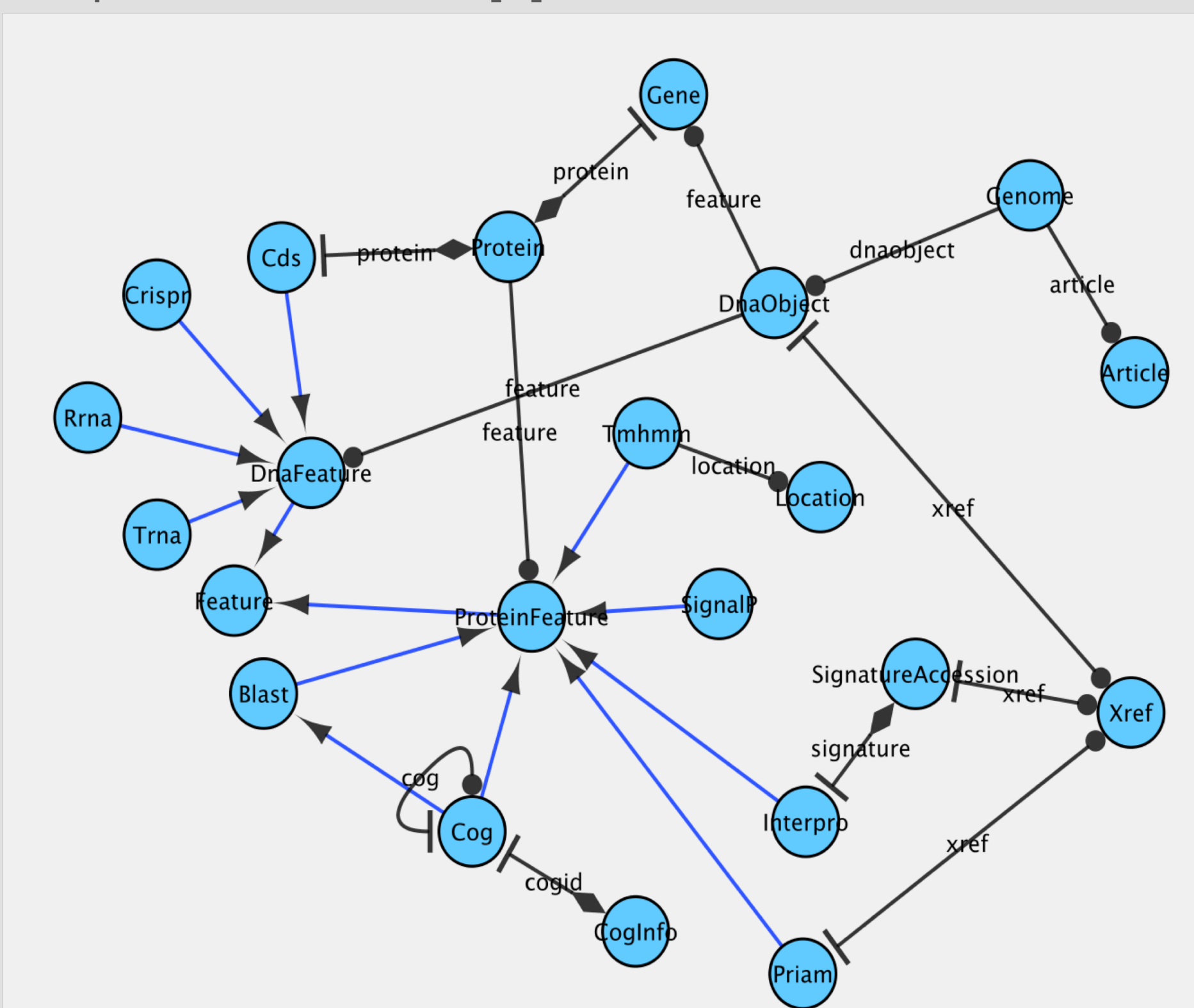
### Direct route: A,C

### Indirect route: B

A direct route is only feasible through a unified and standardized annotation framework. The Semantic Annotation Platform for Prokaryotes (SAPP) has been developed for this purpose.

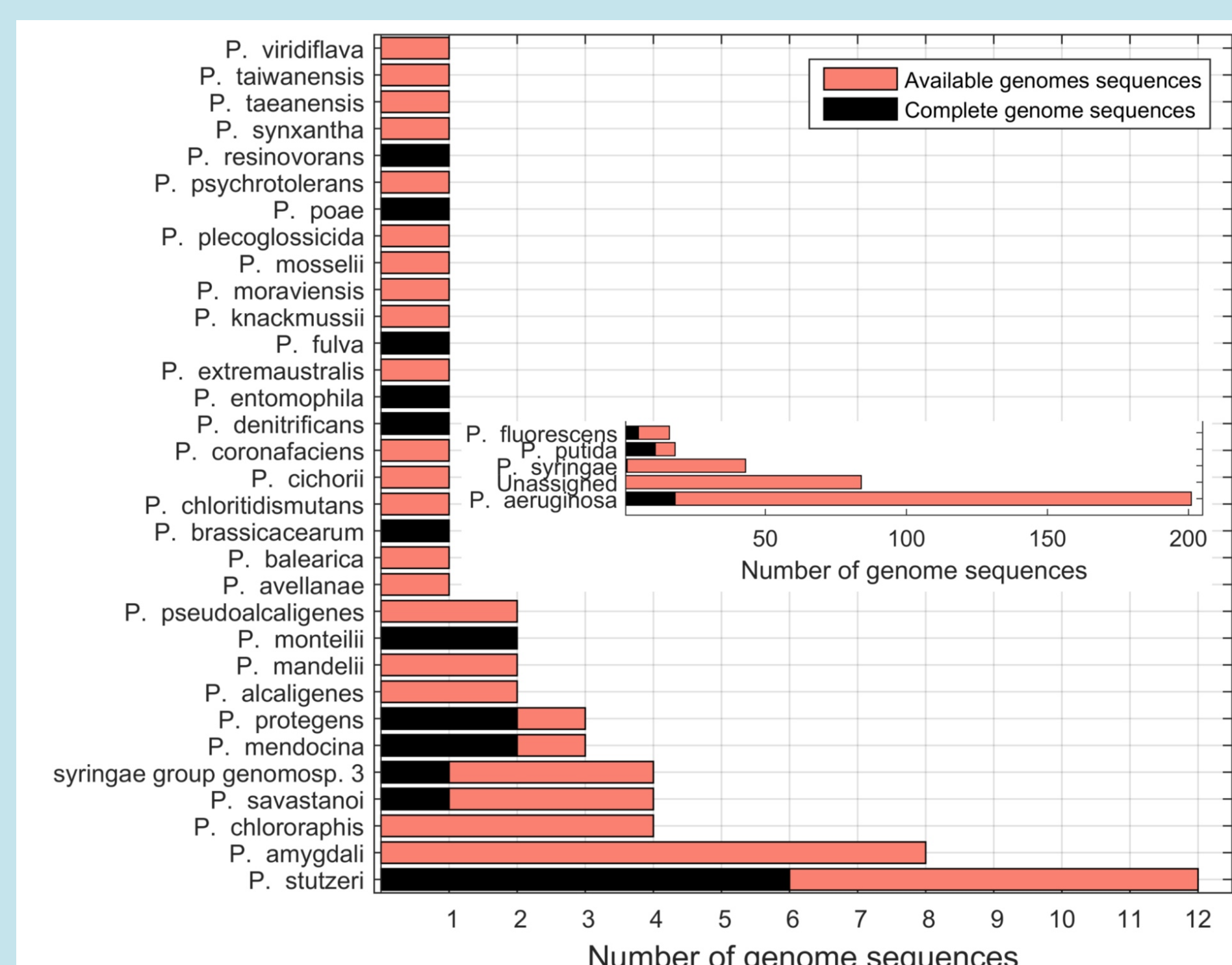
## Semantic representation

A schematic representation of the RDF resource generated with SAPP. The structure has been automatically retrieved using RDF2Graph and is compliant with GBOL [1].



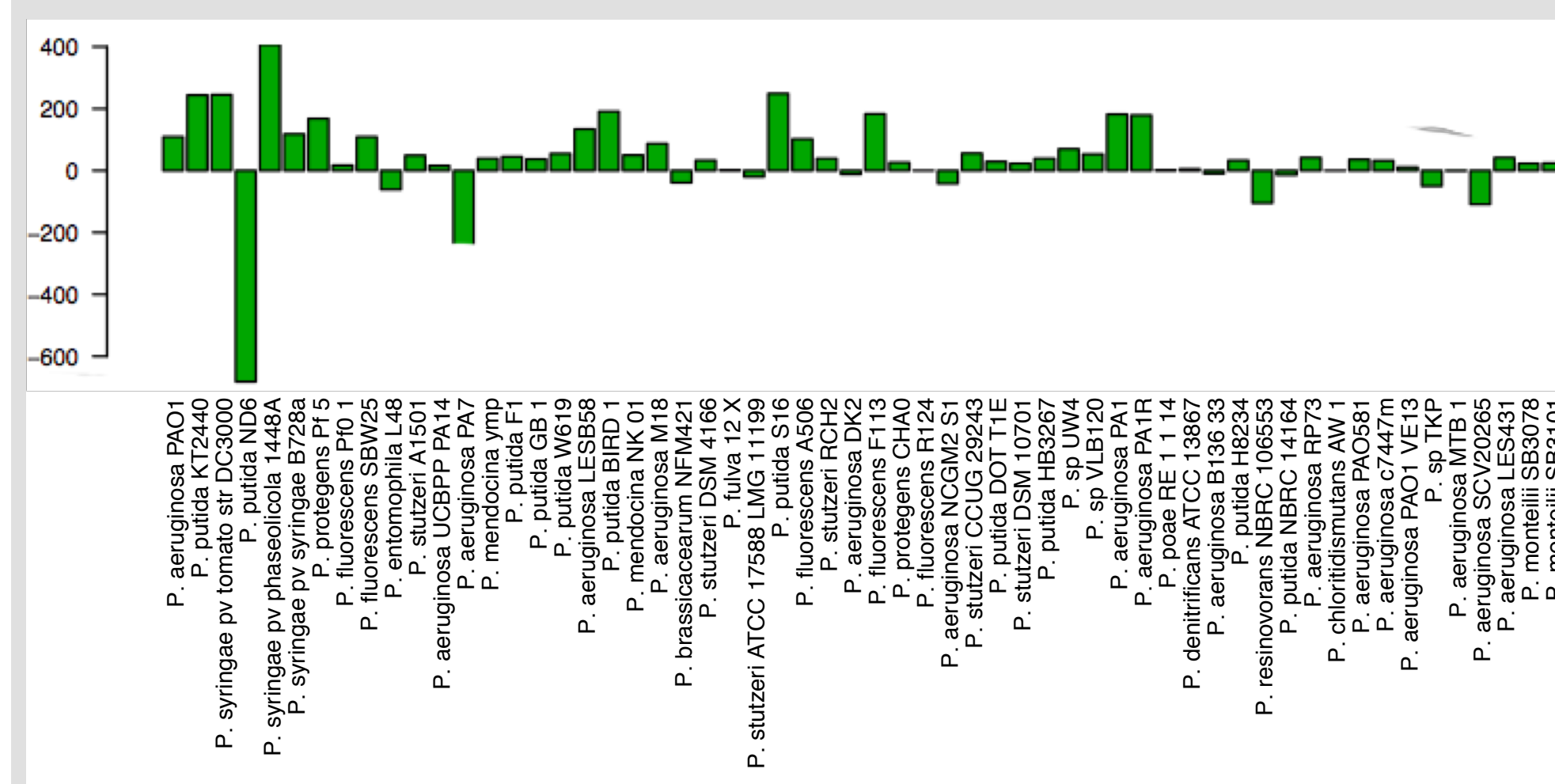
## Pseudomonas functional comparison

Species distribution of 432 *Pseudomonas* genomes



## Functional annotation

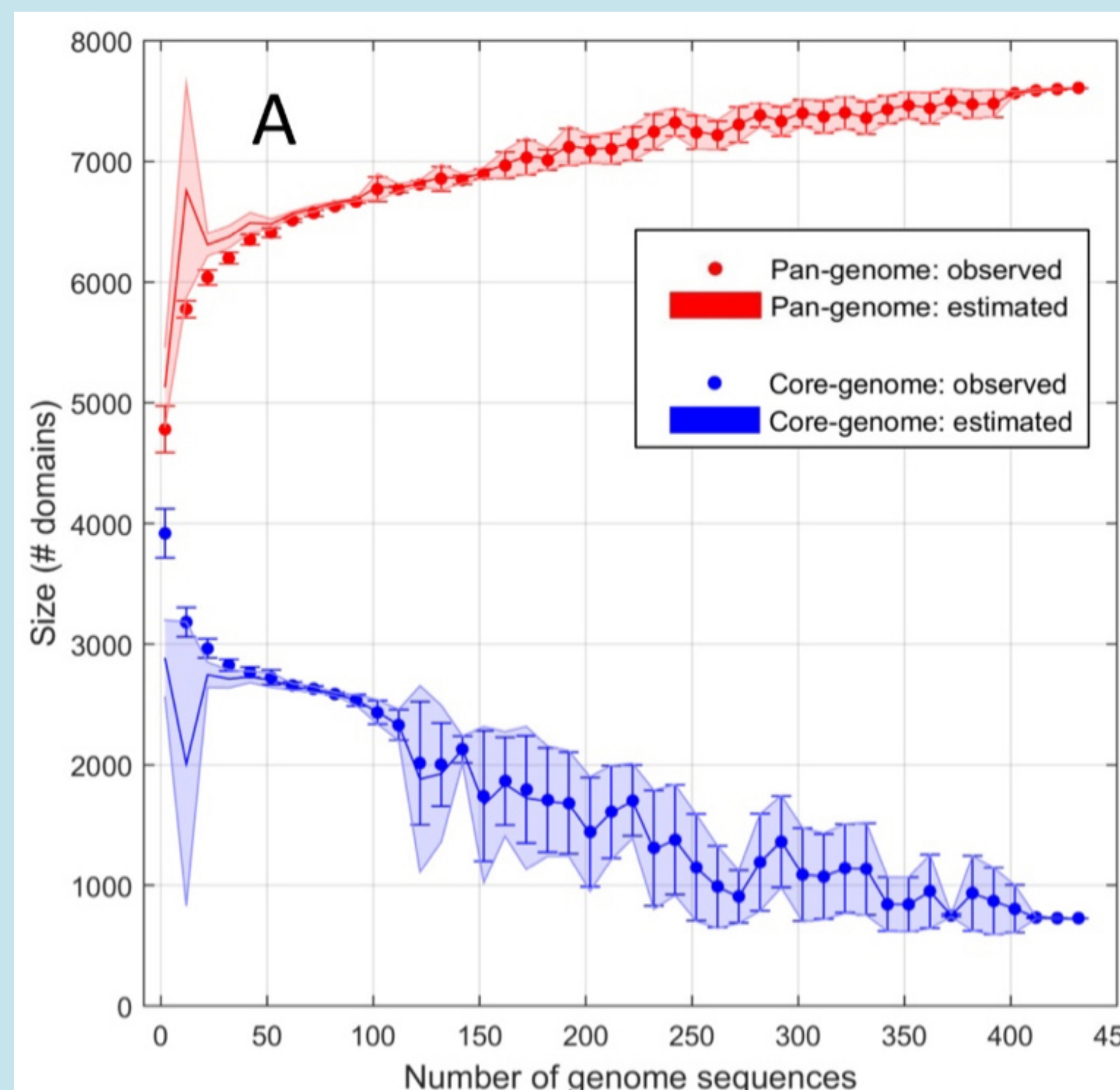
*Pseudomonas* genomes were *de novo* annotated using SAPP, revealing variation in quality of the deposited annotations.



Consistent genome re-annotation: total of newly predicted genes versus total number of genes extracted from the published annotation file

## Pan- Core- genome

For a large set of genomes obtaining pair-wise sequence similarity scores is unfeasible and orthologue detection can be hampered by domain shuffling and domain duplication events. Sizes of the pan- and core-genome of *Pseudomonas* was estimated through the presence of protein domains.



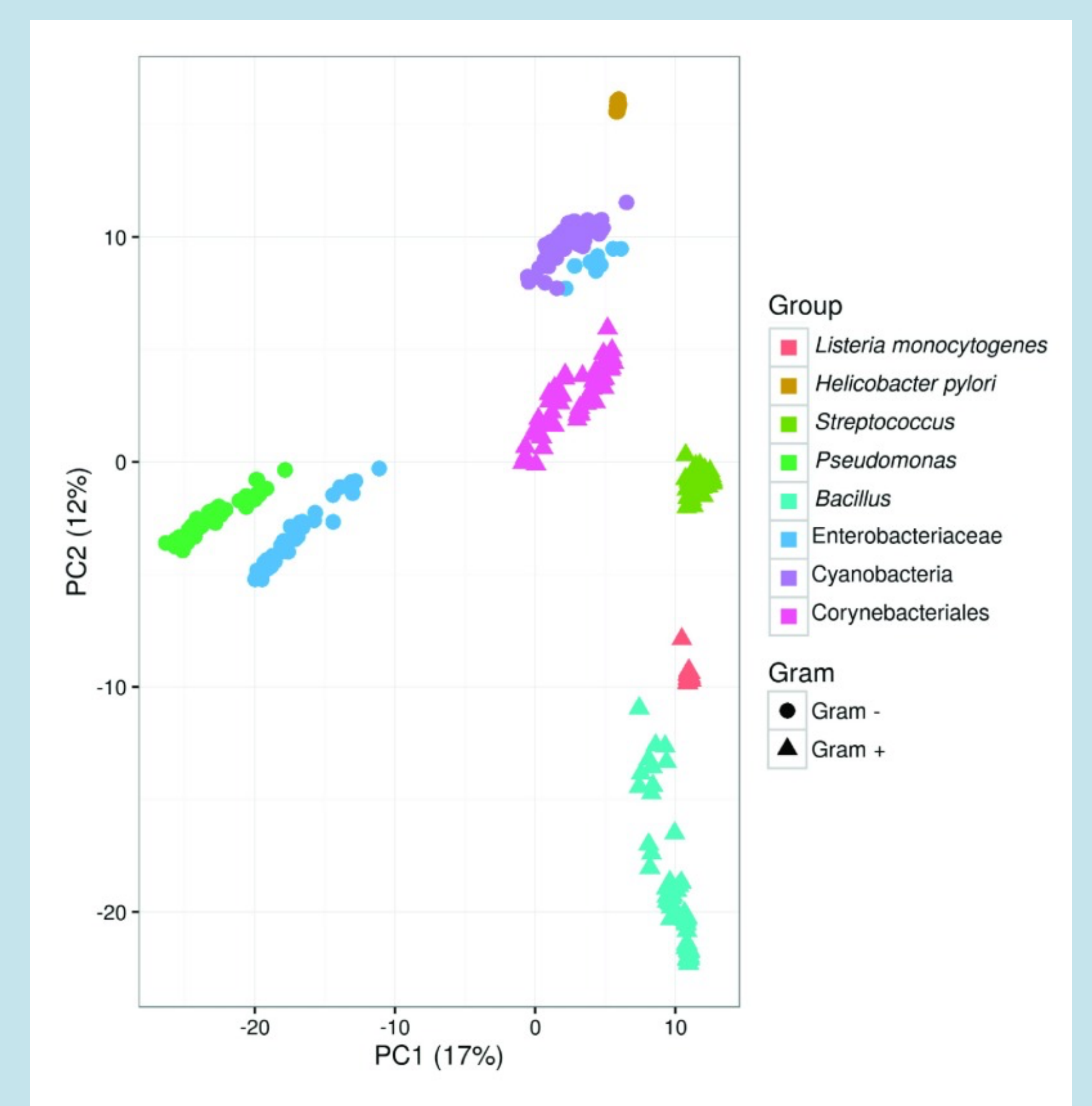
## Web Framework for functional comparative genomics

SAGeR-P enables browsing through selected genomes and live functional comparisons of selected strains through a intuitive shiny interface.

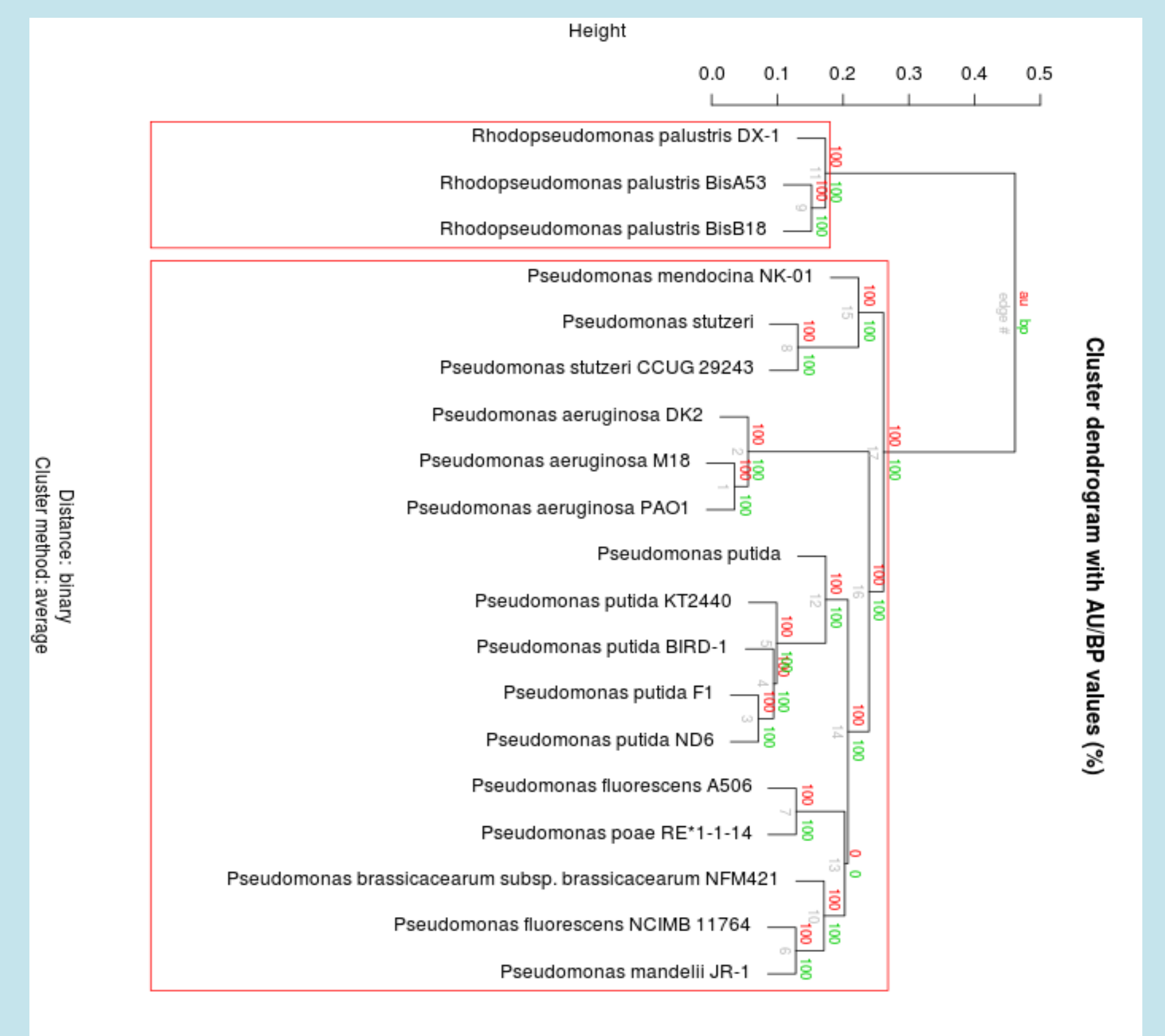
sample	name	lineage
1	GCA_000003645	Bacillus cereus m1293
2	GCA_000003925	Bacillus mycoides DSM 2048
3	GCA_000003955	Bacillus cereus AH1273
4	GCA_000005825	Bacillus pseudofirmus OF4
5	GCA_000005845	Escherichia coli str. K-12 substr. MG1655
6	GCA_000006155	Bacillus anthracis str. A2012
7	GCA_000006605	Corynebacterium jeikeium K411
8	GCA_000006625	Ureaplasma parvum serovar 3 str. ATCC 700970
9	GCA_000006645	Yersinia pestis KIM10+

## Functional comparison

Functional similarities of 446 genomes based on the presence/absence of domain architectures.



On the fly visualization of functional similarities in SAGeR-P



## Concluding remarks

- Development of a Semantic framework has been established for functional comparative genomics
- Pseudomonas* functional core is closed
- Domain persistence is higher than sequence persistence
- SAGeR-P enables integrational analyses through a web framework

## References

- JCJ van Dam et al. - Journal of Biomedical Semantics – 2015
- Koehorst et al. – F1000 PMC5031134 - 2016
- Koehorst et al. – Scientific reports - 2016

## Acknowledgements

This work was supported by the European Commission-funded FP7 project INFECT and was carried out on the Dutch national e-infrastructure with the support of SURF Foundation.

## Contact Information

jasper.koehorst@wur.nl  
http://semantics.systemsbiology.nl  
http://sagerp.systemsbiology.nl

