

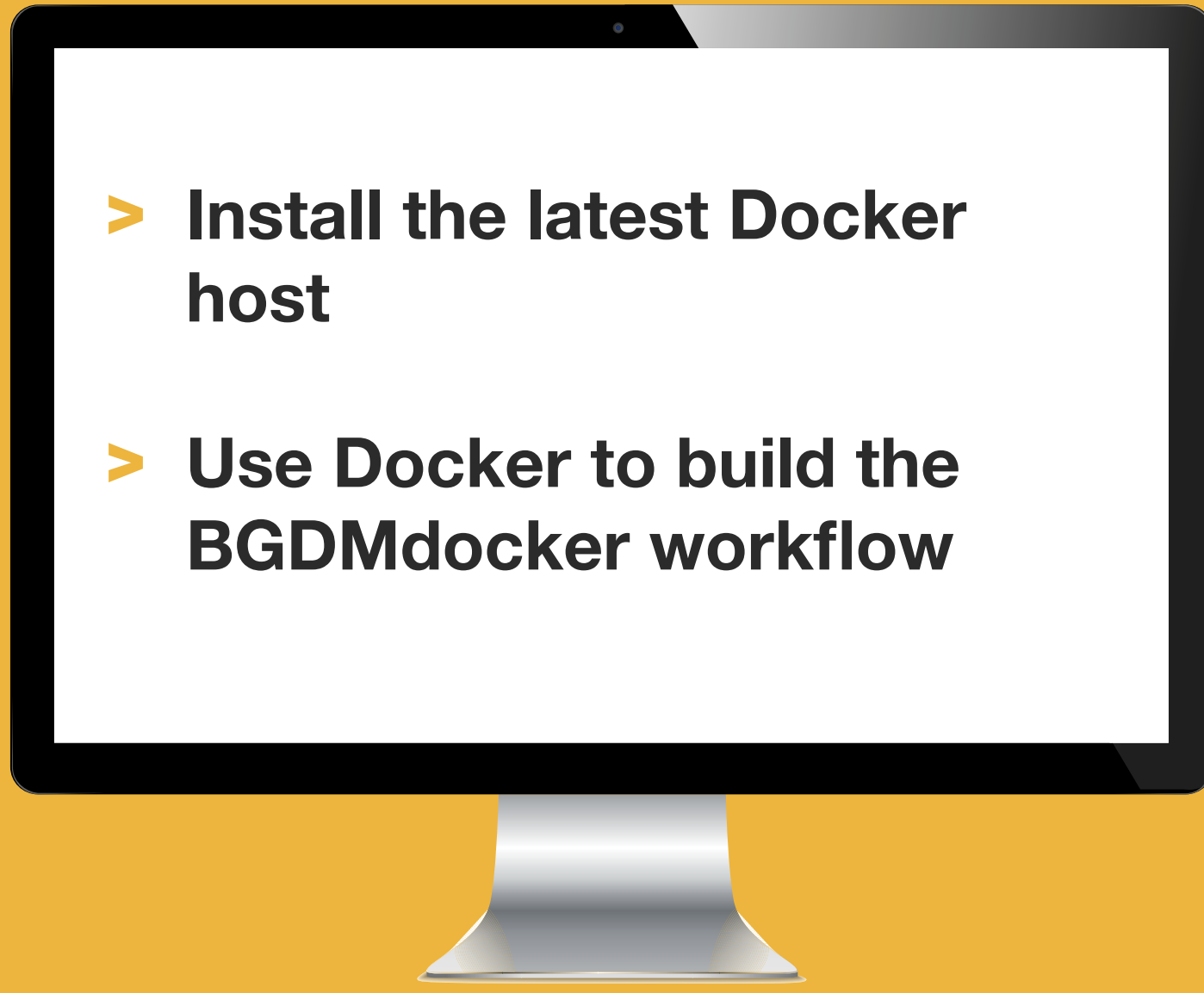
# BGDMdocker: a Docker workflow for data mining and visualization of bacterial pan-genomes and biosynthetic gene clusters

## INTRODUCTION

Recently, **Docker technology has received increasing attention throughout the bioinformatics community**. However, its implementation has not yet been mastered by most biologists; accordingly, its application in biological research has been limited.

In order to popularize this technology in the field of bioinformatics and to promote the use of publicly available bioinformatics tools, we introduce here a **complete and accurate bioinformatics workflow based on Docker**.

## METHODS

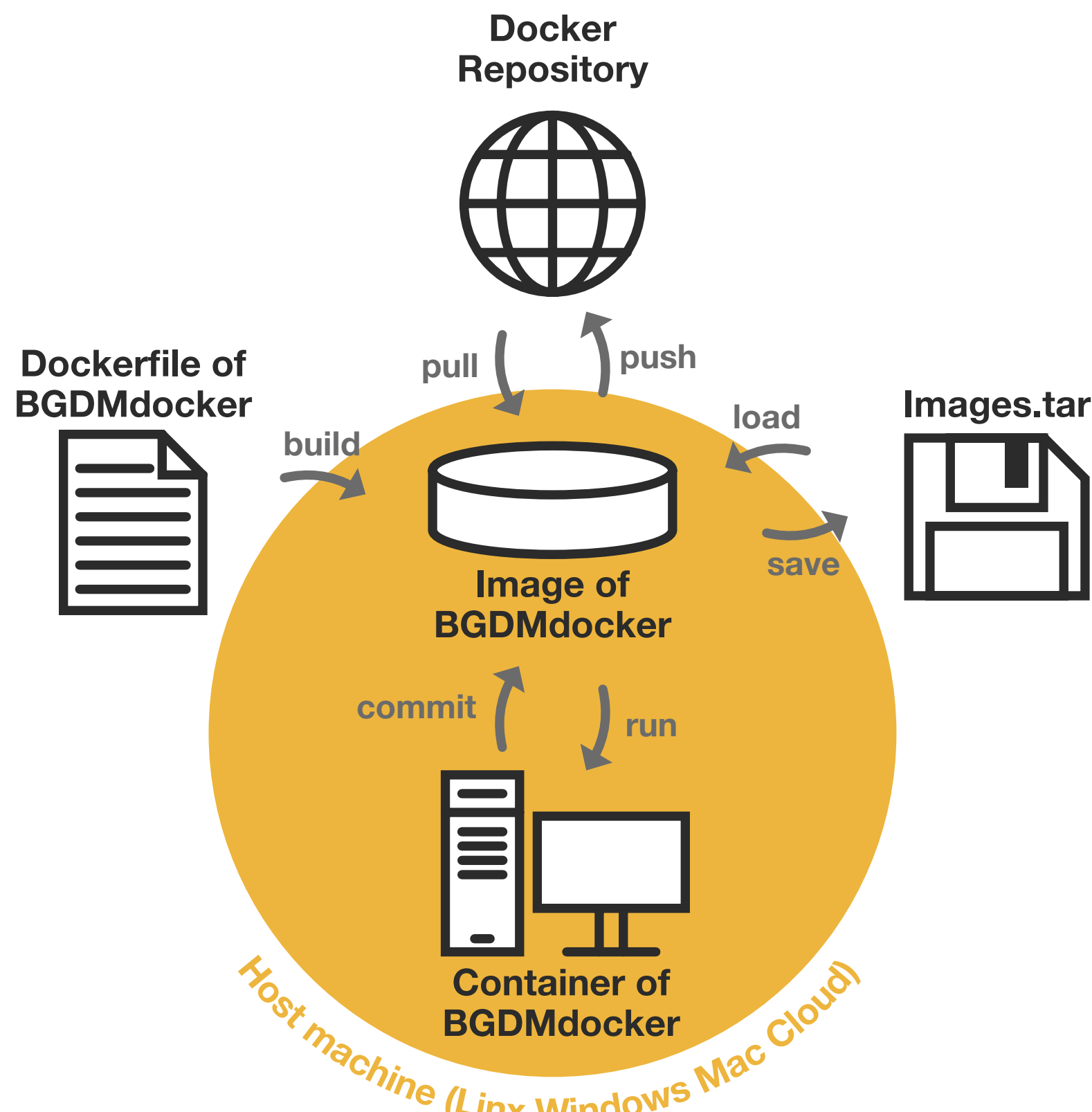


The present workflow enables analysis and visualization of pan-genomes and biosynthetic gene clusters of bacteria.

For detailed commands: <https://peerj.com/articles/3948/#supp-1>

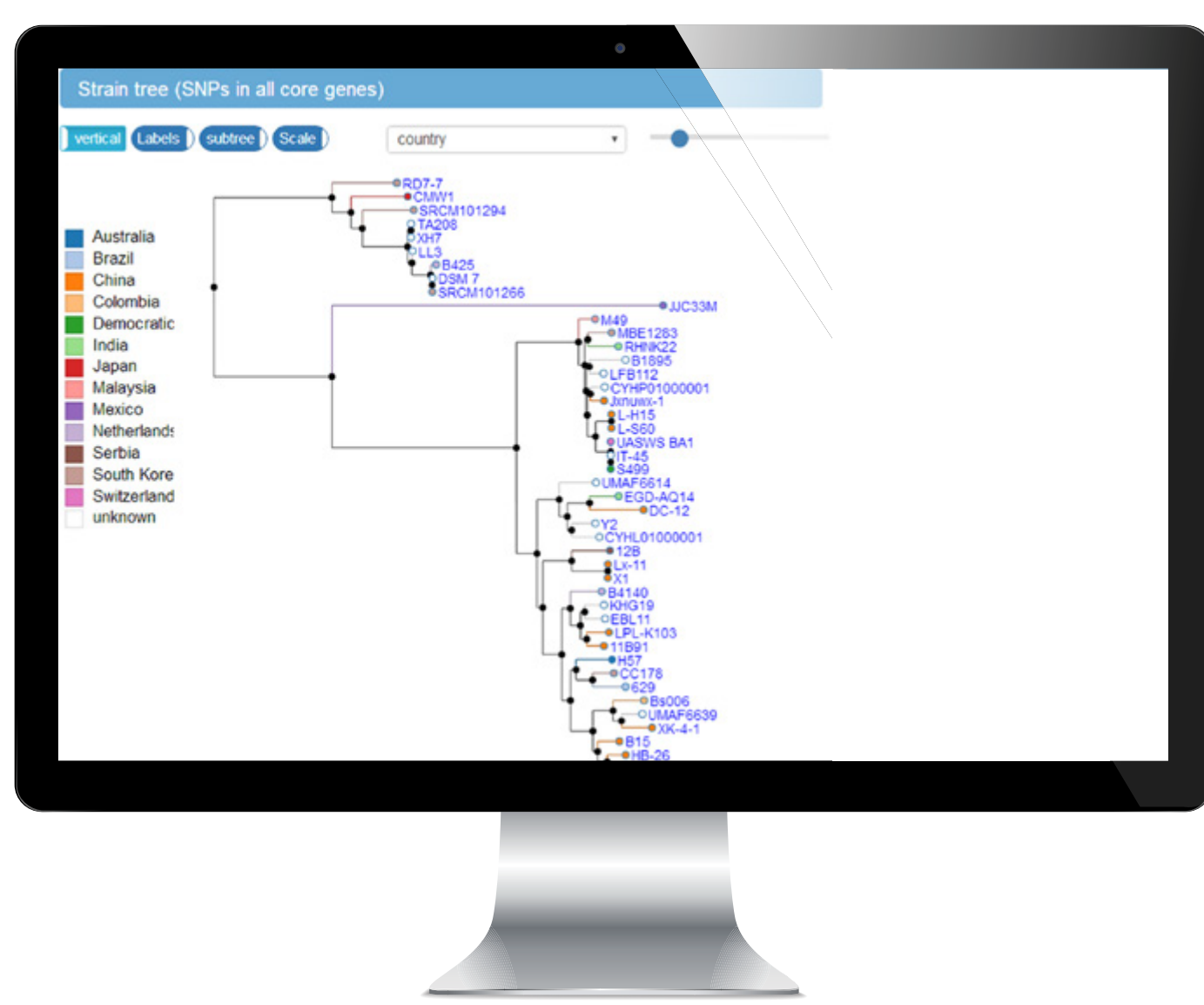
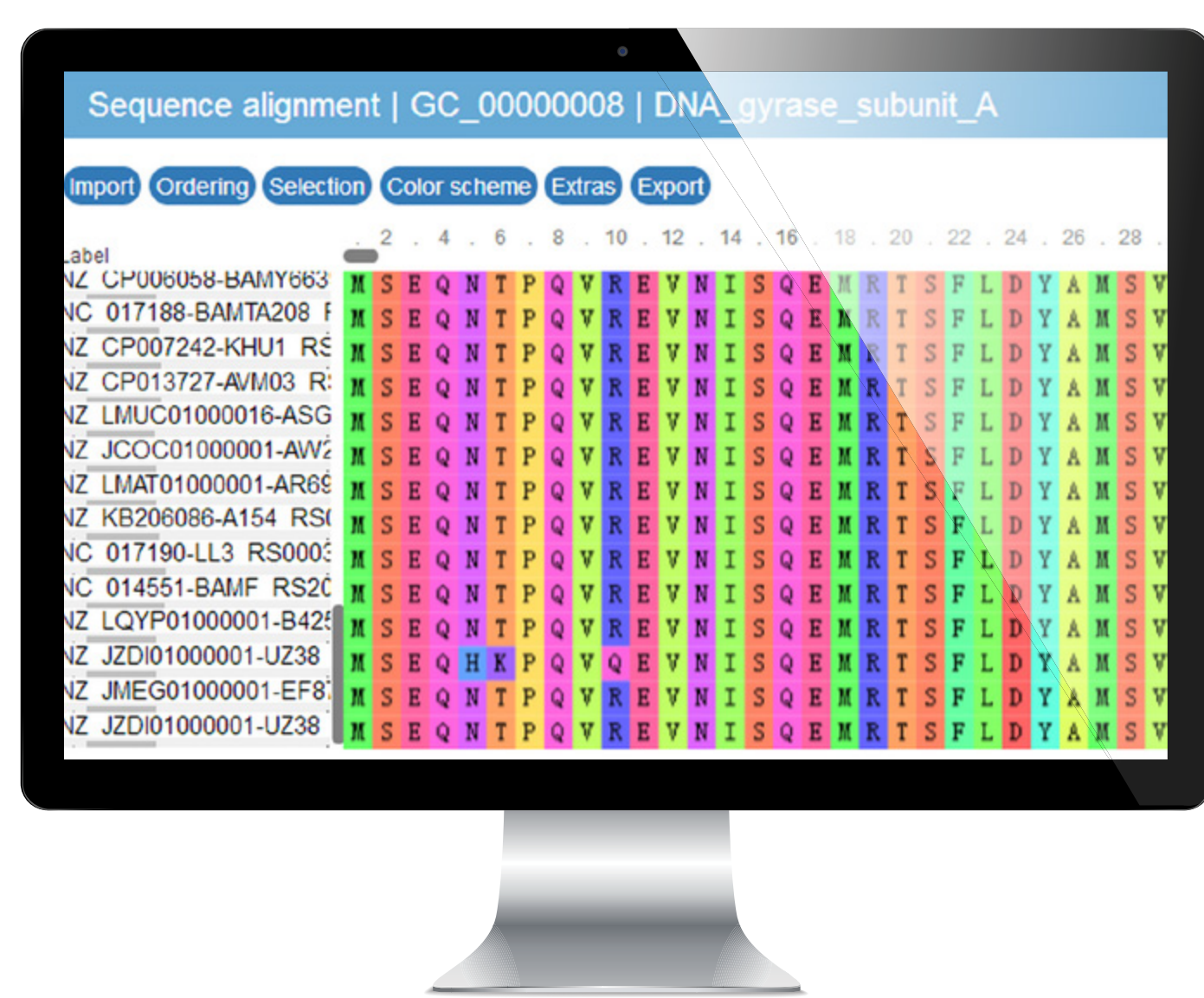
## RESULTS

**Fast and reproducible building of the BGDMdocker workflow across computing platforms using Docker**

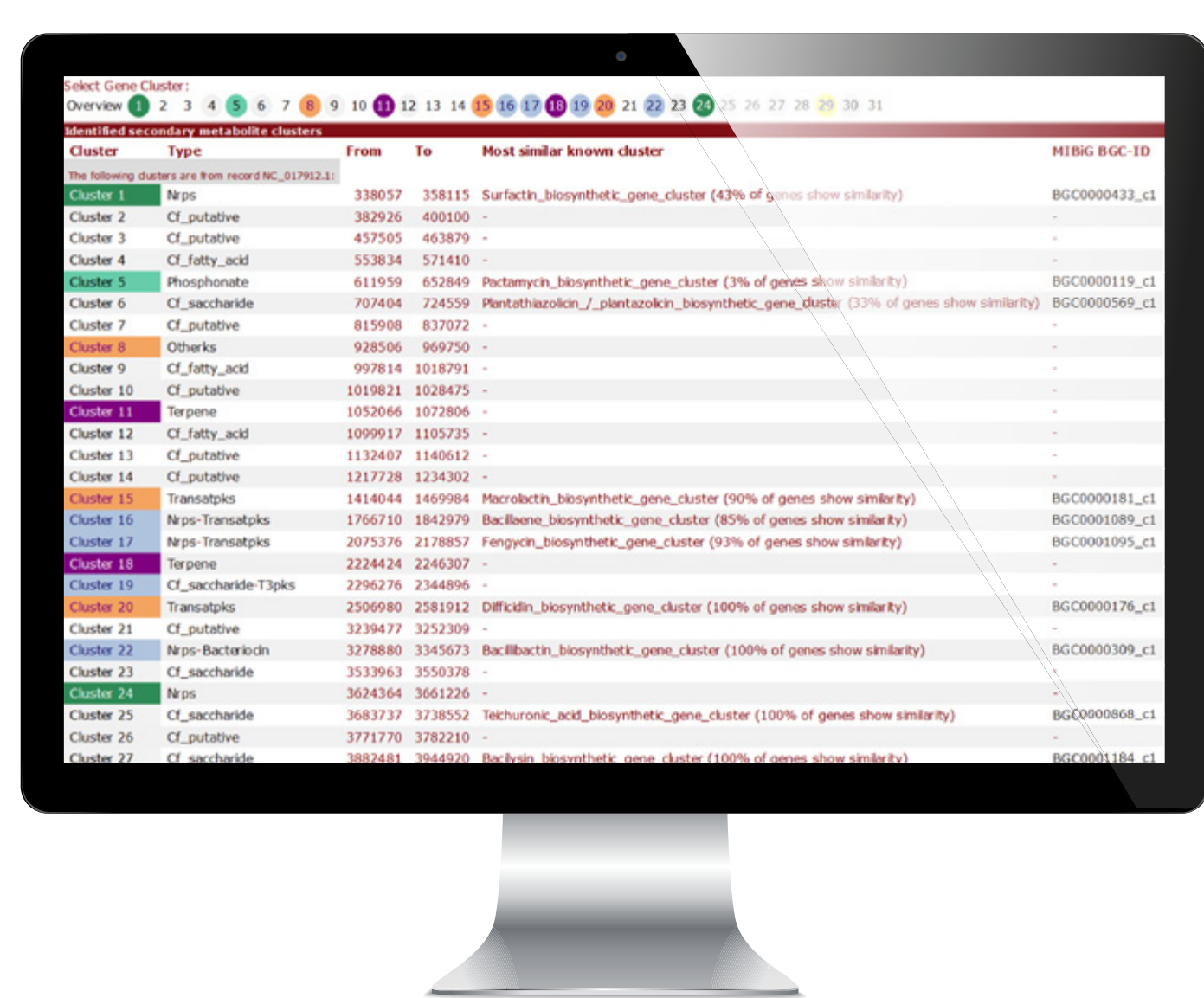


**Datamining and visualizing the pan-genomes of *B. amyloliquefaciens***

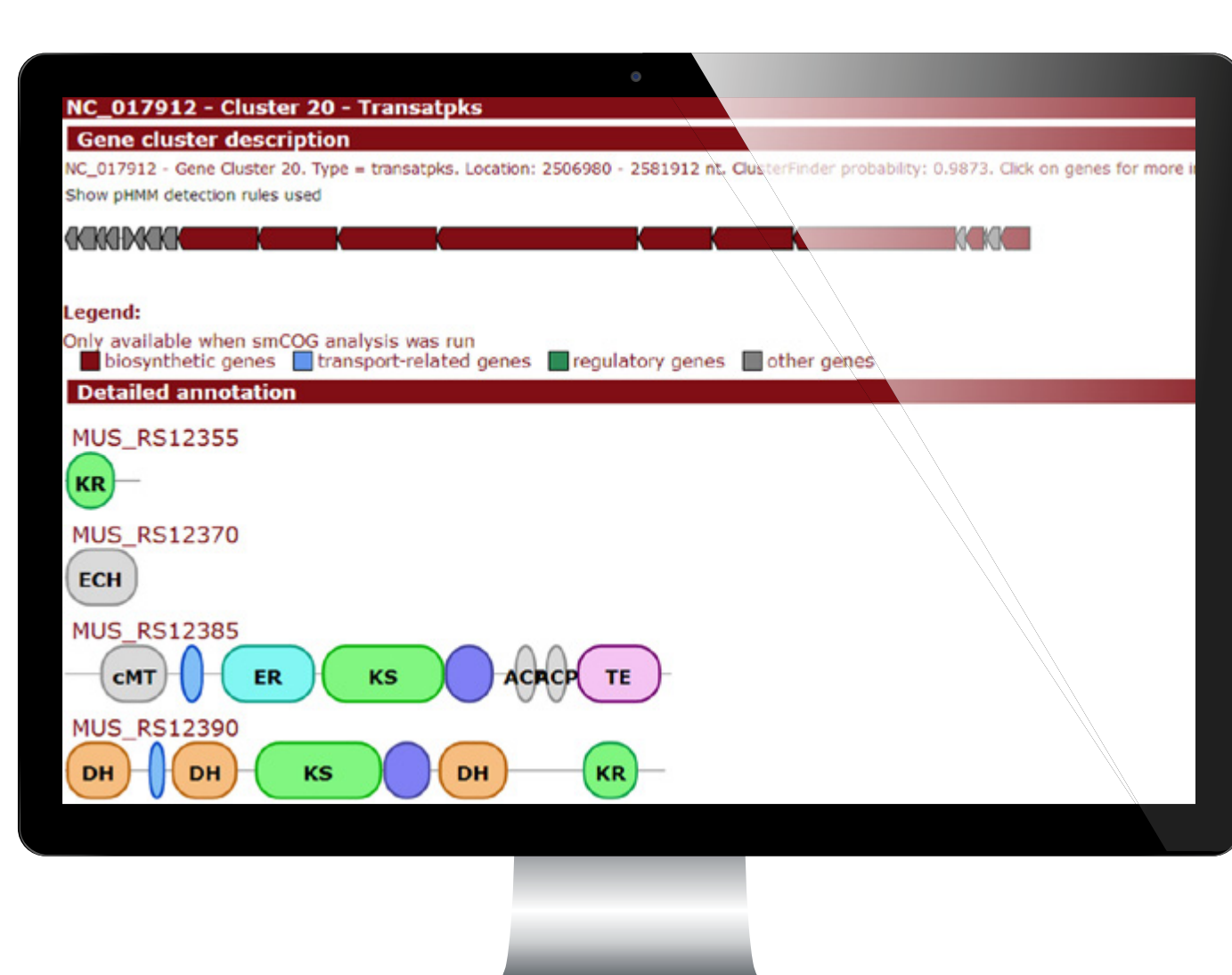
In order to explore the result data, a website (<http://bapgd.hygenomics.com/pangenome/home>) was built for the interactive exploration of the *B. amyloliquefaciens* pan-genome and biosynthetic gene clusters using the BGDMdocker workflow. Visualization allowed for the rapid filtering and searching of genes.



**Datamining and visualizing of biosynthetic gene clusters of *B. amyloliquefaciens***



Cluster	Type	From	To	Host similar known cluster	HGNC ID
Cluster 1	nc	138637	158113	Surfactin_biosynthetic_gene_cluster (43% of genes show similarity)	BGC0000423_c1
Cluster 2	CI_sulfate	382930	400280	-	-
Cluster 3	CI_sulfate	437505	462879	-	-
Cluster 4	CI_sulfate	533834	574493	-	-
Cluster 5	Phosphonate	611959	652849	Phosphonate_biosynthetic_gene_cluster (3% of genes show similarity)	BGC0000119_c1
Cluster 6	CI_saccharide	704246	724559	Phenylacetate_biosynthetic_gene_cluster (27% of genes show similarity)	BGC0000595_c1
Cluster 7	CI_sulfate	819068	837072	-	-
Cluster 8	Others	928296	969720	-	-
Cluster 9	CI_sulfate	1019821	1028475	-	-
Cluster 10	nc	1053066	1072966	-	-
Cluster 11	nc	1099917	1107259	-	-
Cluster 12	CI_sulfate	1132487	1140612	-	-
Cluster 13	CI_sulfate	1217758	1225252	-	-
Cluster 14	Transatpks	1414644	1469584	Transatpks_biosynthetic_gene_cluster (90% of genes show similarity)	BGC0000181_c1
Cluster 15	nc	1760710	1842979	Bioamine_biosynthetic_gene_cluster (85% of genes show similarity)	BGC0001089_c1
Cluster 16	nc	2052129	2178517	nc	BGC0000105_c1
Cluster 17	nc	2248244	2246307	-	-
Cluster 18	nc	2296276	2344896	-	-
Cluster 19	CI_saccharide	2509590	2581912	Diffridin_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000176_c1
Cluster 20	nc	3239477	3252309	-	-
Cluster 21	nc	3278660	3350579	Baibactin_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000309_c1
Cluster 22	nc	3539863	3550378	-	-
Cluster 23	nc	3624364	3618258	-	-
Cluster 24	CI_saccharide	3687117	3738032	3-hydroxy-acyl_biosynthetic_gene_cluster (100% of genes show similarity)	BGC0000868_c1
Cluster 25	CI_sulfate	3771703	3782310	-	-
Cluster 26	CI_saccharide	3834416	3865910	nc	BGC0001168_c1



## CONCLUSION

This provides a **new solution for bioinformatics mining of big data from various publicly available biological databases**. The present step-by-step guide creates an integrative workflow through a Dockerfile to allow researchers to build their own Image and run Container easily.

The visual display of data provided in this study may be completely duplicated. All resulting data and relevant tools and files may be downloaded from our website (<http://bapgd.hygenomics.com/pangenome/home>) with no registration required.