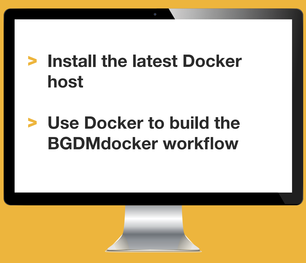
BGDMdocker: a Docker workflow for data mining and visualization of bacterial pan-genomes and biosynthetic gene clusters

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

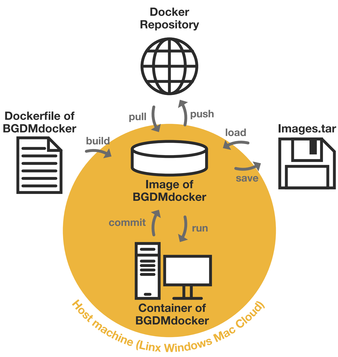


The present workflow enables analysis and visualization of pan-genomes and biosynthetic gene clusters of bacteria.

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

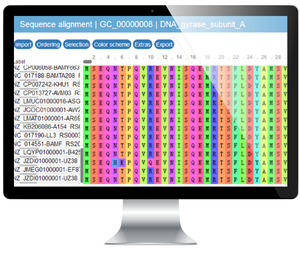
RESULTS

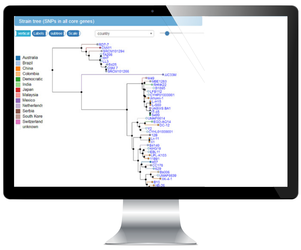
> **Fast and reproducible building of the BGDMdocker workflow across computing platforms using Docker**



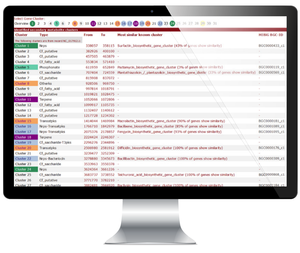
> **Datamining and visualizing the pan-genomes of *B. amyloliquefaciens***

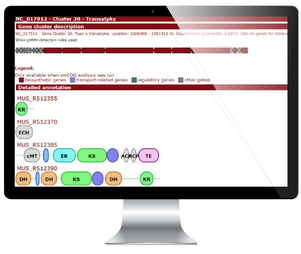
In order to explore the result data, a website (http://bapgd.hygenomics.com/pangenome/home) was built for the interactive exploration of the *B. amyloliquefaciens* pan-genome and biosynthetic gene clusters using the BGDMdocker workflow. Visualization allowed for the rapid filtering and searching of genes.





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





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.

PeerJ

BGDMdocker: a Docker workflow for data mining and visualization of bacterial pan-genomes and biosynthetic gene clusters. PeerJ 4:e3948 DOI: 10.7717/peerj.3948 https://peerj.com/articles/3948/