Title

Subtitle

Author Name

More text

more text

Department Name

University Name

Date

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur porta enim a ipsum hendrerit aliquam. Phasellus at enim gravida, lobortis ante id, vestibulum velit. Aliquam iaculis maximus ligula quis ornare. Nunc ut orci id ante malesuada posuere id sit amet dolor. Nulla facilisi. In vitae est eu lacus finibus tincidunt vitae in neque. Aliquam tincidunt ipsum sed ligula rhoncus vestibulum gravida at erat. Quisque ligula elit, laoreet ut aliquet vel, finibus ac orci.

Introduction

Methods

A total of 31 HQ genomes (29 SGBs and 2 reference genome) were used in this study.

The species' pangenome was obtained using Roary (citation) based on Prokka annotation files (.gff).

Results and Discussion

Genome Annotation

Pangenome Analysis

Pangenome analysis found 4670 total genes (figure 2a), among which 1017 were attributed to the *core* (above 90% prevalence in MAGs), 2074 to the *shell* (from 15% to 89% prevalence), and 1579 were classified as *cloud* (from 0% to 15% prevalence). The results were robust with respect to many rounds of computation.

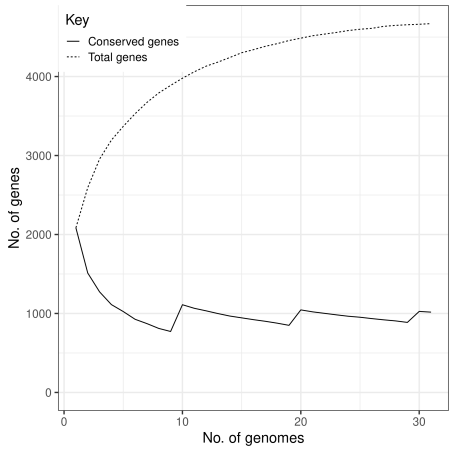
The number of conserved genes appears to reach a plateau (figure 1a) when the number of MAGs increases, suggesting that this species has a closed pangenome. This is further confermed by the trend of unique genes plotted against the number of genomes (figure 1b).

The plot of pangenome frequencies ?? (figure 2b) shows the typical shape observed in microbiome samples. It is U shaped since there are conserved functions (core genes) that are present in every strain, and other genes that are very specific and are present in only one or few strains (unique genes), with very few genes in between.

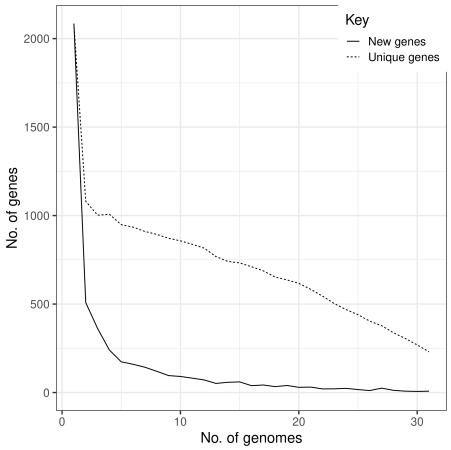
Phylogenetic Structure

Taxonomic Annotation

Conclusion

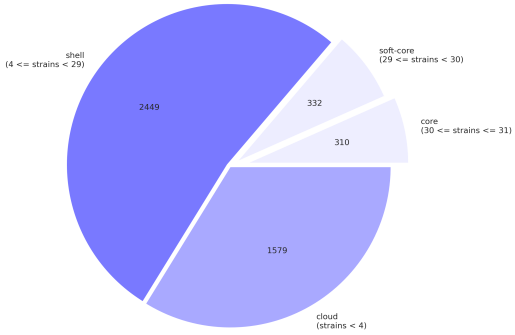


(a) The trend of conserved genes suggest that the species analyzed has a closed pangenome

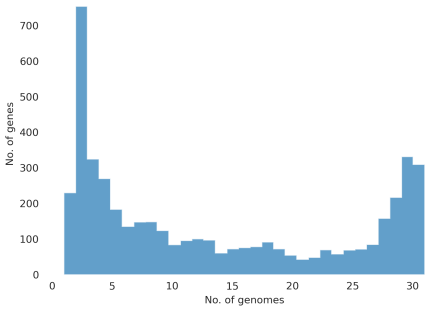


(b)

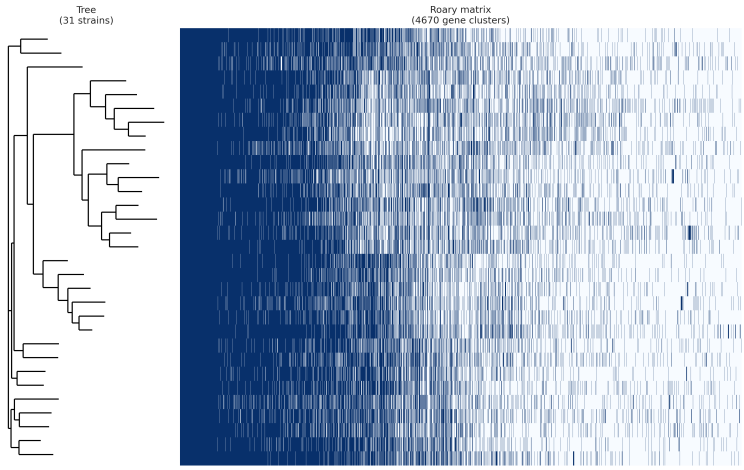
Figure 1



(a)



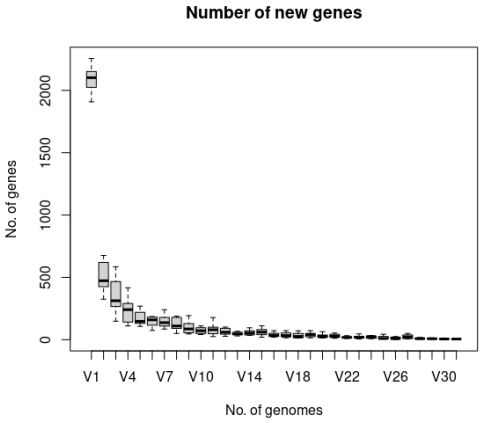
(b)



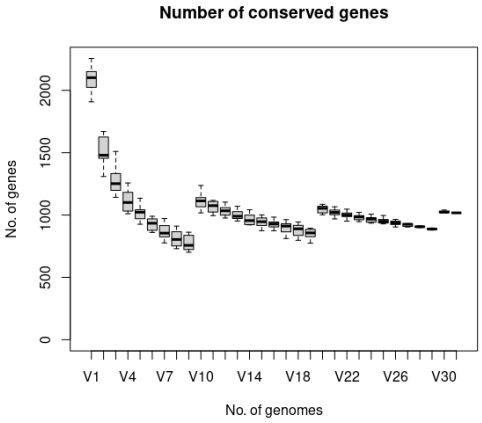
(c)

Figure 2: Three simple graphs

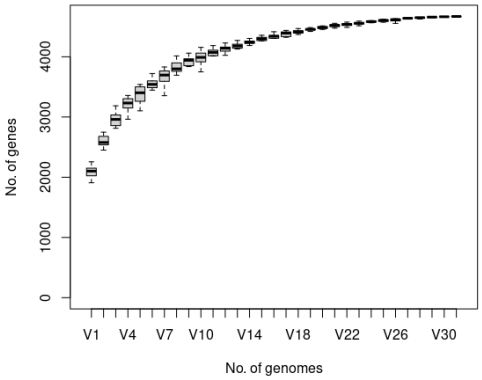
Supplementary data



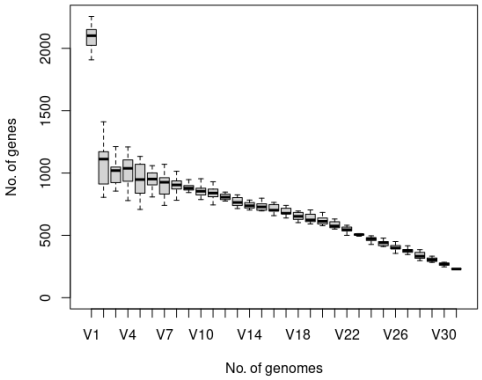
(a) Number of new genes



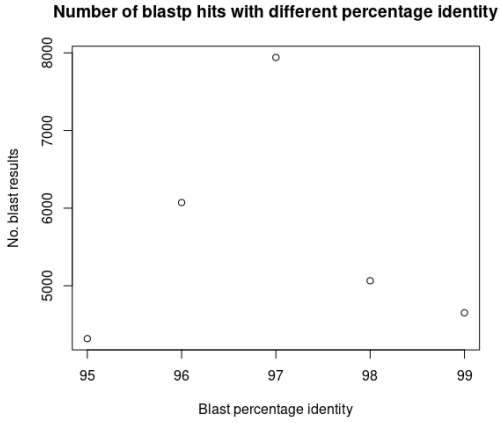
(b) Number of conserved genes



(c) Number of genes in the pan-genome



(d) Number of unique genes



(e) Number of blastp hits with different percentage identity

Figure 3: Three simple graphs

References

[1] O. Tange. “GNU Parallel - The Command-Line Power Tool”. In: *The USENIX Magazine* (2011). DOI: 10.5281/zenodo.16303. URL: http://www.gnu.org/s/parallel.