BITS :: Call for Abstracts 2023 - Oral communication

Type	Oral communication
Session	Metagenomics and metatranscriptomics
Title	Rediscovering pangenomic content of metagenomic assembled genomes with PanDelos-frags
All Authors	Vincenzo Bonnici(1), Claudia Mengoni(2), Manuel Mangoni(3,4), Giuditta Franco(2), Rosalba Giugno(2)

Affiliation

(1) Department of Mathematical, Physical and Computer Sciences, University of Parma, Parma
(2) Department of Computer Science, University of Verona, Verona
(3) Fondazione IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (FG)
(4) Department of Experimental Medicine, Sapienza University of Rome, Roma

Motivation

A microbiome is a community of microorganisms inhabiting an environment, including living beings. This field experienced a great boost thanks to the advent of metagenome shotgun sequencing, which made it possible to sequence the overall genetic material of a sample containing different organisms without need of isolation. Computational techniques have been developed to reconstruct nearly-complete genomes from metagenomic sequences, called metagenomic assembled genomes (MAGs)1. The increasing amount of MAGs present in online resources represent a great opportunity for studying gene families' evolution via means of analysis such as pangenomics, consisting of grouping genes into gene families within a set of bacteria belonging to the same species2. Established tools for pangenomic analysis that are currently used focus on the reconstruction of gene families from complete genomes. However, it is becoming increasingly evident that even MAGs reflecting the highest standards of quality (completeness >95%, contamination <5%) cannot be treated as complete genomes in pangenomic analysis, because this can lead to several biases, including pangenome size overestimation and core genes loss3,4. It is crucial to use tools for pangenomic analysis that can deal with fragmented genomes. In this work, we exploit available MAGs whose pangenome has been so far only studied via established pangenomic tools to determine how their pangenomic composition varies when we use PanDelos-frags, a pangenomic tool specifically implemented to deal with fragmented genomes.

Methods

In our study we consider MAGs from three species belonging to different phyla that were made available online by Pasolli et al. (2019)5 after being reconstructed and annotated. The datasets of the metagenomic experiments are composed of 17 Abiotrophia defectiva MAGs, 29 Bacteroides nordii MAGs, and 39 Pseudomonas aeruginosa MAGs. For this analysis we compute and compare the pangenomes of the three species using a pangenomic tool for complete genomes, Roary, which was also the tool of choice in the MAGs study, and a tool which can handle genome fragmentation, namely PanDelos-frags. We also test two additional tools which can work with draft genomes, GenAPI and Panaroo.
In order to compare the gene families obtained by using the different tools we exploit the concept of diffusivity of a gene family F, as the number of genomes G to which the genes in F belong to. This notion allowed us to determine the overall pangenomic distribution and see how gene families are distributed across multiple genomes. We further investigate the relationship of the obtained gene families by comparing the size of the exact same gene family across the pangenomes of Roary and PanDelos-frags.
As no ground truth on diffusivity distribution is obtainable in metagenomic samples, we performed a functional annotation analysis of the genes that were found in one tool and not in the other to belong to a specific family. Precisely, we investigate if the functional annotation of these genes reflect those of the gene families they were assigned to. This was achieved by mapping genes to the set of annotated coding sequences of the respective reference genome.

Results

Although the number of small-size gene families is comparable across tools, we see that in all the three experiments PanDelos-frags is able to retrieve a greater number of gene families, with a larger diffusivity than all the other tools.
When we compare the size of the same gene family in Roary and PanDelos-frags, we detect an increase in gene family size with PanDelos-frags, indicating an improvement in the ability of detecting homology relationships across multiple genomes. Respectively in A. defectiva, B. nordii, and P. aeruginosa, 50.8%, 56.3%, and 79.5% of the gene families have greater diffusivity in PanDelos-frags, only 0.5%, 1%, and 0.9% a smaller diffusivity, while the remaining gene families have the same size.
While investigating the function of the genes that are assigned to a family only by PanDelos-frags, we are able to see that in more than 90% of cases the genes detected are functionally coherent with the rest of the family, suggesting that the genes are correctly assigned to the families.
These tests suggest that analysing MAGs using a tool for pangenomic analysis which accounts for genome fragmentation leads to improvements with respect to traditional pangenomic tools. Namely, genes reconstructed from fragmented genomes lead gene families to span across a bigger set of genomes, thus giving a complete description of the pangenomes of the analysed species, while not compromising functional coherence in the enlarged gene families.

Info

[1] Y. Zhou, M. Liu, J. Yang, Recovering metagenome-assembled genomes from shotgun metagenomic sequencing data: Methods, applications, challenges, and opportunities, Microbiological Research (2022) 127023.
[2] H. Tettelin, V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, et al., Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial “pan-genome”, Proceedings of the National Academy of Sciences 102 (2005) 13950–13955.
[3] A. Meziti, L. M. Rodriguez-R, J. K. Hatt, A. Pe˜na-Gonzalez, K. Levy, K. T. Konstantinidis, The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived from the Same Fecal Sample, Applied and Environmental Microbiology 87 (2021).
[4] T. Li, Y. Yin, Critical assessment of pan-genomic analysis of metagenome-assembled genomes, Briefings in Bioinformatics 23 (2022) bbac413.
[5] E. Pasolli, F. Asnicar, S. Manara, M. Zolfo, N. Karcher, F. Armanini, F. Beghini, P. Manghi, A. Tett, P. Ghensi, et al., Extensive unexplored human microbiome diversity resource extensive unexplored human microbiome diversity revealed by over 150 000 genomes from metagenomes spanning age, geography, and lifestyle, Cell 176 (2019) 649–662.

filename	-
Figure	
-	
Availability	https://github.com/InfOmics/PanDelos-frags-SUPP

Corresponding Author

Name, Surname	Claudia, Mengoni
Email	claudia.mengoni@univr.it
Submitted on	26.04.2023