From kmers to genetic events: compacted De Bruijn graphs shed light on bacterial GWAS

Magali Jaillard Dancette*†[1,2], Leandro Lima[2,3], Maud Tournoud[1], Vincent Lacroix[2,3], and Laurent Jacob[2]

[1]bioMérieux – BIOMERIEUX – Chemin de l'Orme 69280 Marcy l'Etoile, France
[2]Laboratoire de Biométrie et Biologie Evolutive UMR5558 (LBBE) – Université de Lyon, Université Claude Bernard - Lyon I, Université de Lyon – UMR CNRS 5558 - LBBE "Biométrie et Biologie évolutive" UCB Lyon 1 - Bât. Grégor Mendel 43 bd du 11 novembre 1918 69622 VILLEURBANNE cedex, France
[3]ERABLE – INRIA Greoble-Rhône Alpes, Univ. Grenoble-Alpes, Laboratoire Jean Kuntzman – Inria Grenoble, Rhône-Alpes, France

Abstract

Motivation: Genome wide association study (GWAS) methods applied to bacterial genomes and phenotypes have shown promising results for new genetic marker discovery or fine assessment of marker effect. Recently, alignment-free methods based on kmer composition have proved their ability to explore accessory genome. However they lead to redundant descriptions and results which are hard to interpret.
Results: We propose to extend kmer-based approaches by taking advantage of compacted De Bruijn graphs (cDBG) and offer a graphical framework to interpret GWAS results. This approach, called DBGWAS, allows to identify genetic events at various scales, from SNPs to plasmid insertions, hence can deal with both clonal and highly dynamics bacterial genomes. Our method gathers kmers (which are the cDBG nodes) identified by the association model into components defined from their neighborhood in the initial cDBG. Their representation as colored cDBG subgraphs allows then an easy interpretation.

We illustrate the benefit of our approach on antimicrobial resistance phenotypes in several bacterial species of various degrees
of genome plasticity: *M. tuberculosis*, *S. aureus* and *P. aeruginosa*.
Availability: Tool is available at https://gitlab.com/leoisl/dbgwas

*Speaker
†Corresponding author: