Journées GDR/SFB 2017

Representing Genetic Determinants in Bacterial GWAS with Compacted De Bruijn Graphs

Laurent Jacob

UMR CNRS 5558 – LBBE, « Biométrie et Biologie Évolutive », Univ. Lyon

Antimicrobial resistance has become a major worldwide public health concern, calling for a better characterization of existing and novel resistance mechanisms. GWAS methods applied to bacterial genomes have shown encouraging results for new genetic marker discovery. Most existing approaches either look at SNPs obtained by sequence alignment or consider sets of kmers, whose presence in the genome is associated with the phenotype of interest. While the former approach can only be performed when genomes are similar enough for an alignment to make sense, the latter can lead to redundant descriptions and to results which are hard to interpret. We propose an alignment-free GWAS method detecting haplotypes of variable length associated to resistance, using compacted De Bruijn graphs. Our representation is flexible enough to deal with very plastic genomes subject to gene transfers while drastically reducing the number of features to explore compared to kmers, without loss of information. It accomodates polymorphisms in core genes, accessory genes and non coding regions. Using our representation in a GWAS leads to the selection of a small number of entities which are easier to visualize and interpret than fixed length kmers. We illustrate the benefit of our approach by describing known as well as potential novel determinants of antimicrobial resistance in *Pseudomonas aeruginosa*, a pathogenic bacteria with a highly plastic genome. A pre-print is available at http://biorxiv.org/content/early/2017/03/03/113563.