amd MIDWEST

December 2019

Upcoming Events

- **AMD Academy** - January 27-30, 2020
- **Inform 2020** - March 9-12, 2020
- **AMD Academy** - April 20 - 23, 2020
- **APHL Annual Meeting** - June 8 - 11, 2020
- **APHL ID Con** - August 11 - 13, 2020

Resources

- **The AMD-Midwest Regional Website:** https://staph-b.github.io/midwest-region/
- **The StaPH-B Website:** https://staph-b.github.io/

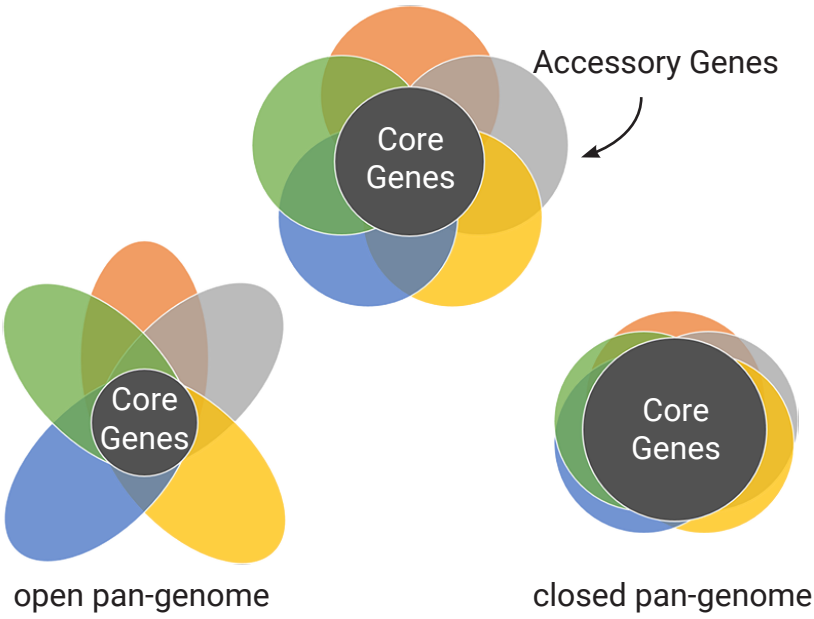
Bioinformatics Demystified

Last month's issue described how Single Nucleotide Polymorphisms (SNPs) and Multi-locus Sequence Typing (MLST) can be used to compare isolates. This month we are expanding on the concept of MLST to include cgMLST and wgMLST. If you missed last month's issue you can find it here: https://staph-b.github.io/midwest-region/newsletters/2019-11_newsletter.pdf.

All the genes!!!

Traditional MLST uses only 7 or 8 genes and as you might suspect, bacteria tend to have more than just 7 or 8 similar genes. Gene content can vary within and between bacterial species, because bacterial genomes are flexible and often change.

The pan-genome refers to the collection of all genes present within a bacterial species. The set of genes shared by all members of a bacterial species (~99% of strains) is known as the core-genome, while genes shared by only some members of a species is known as the accessory genome.



Bacteria with a small number of accessory genes are said to have closed pan-genomes, while bacteria with a large number of accessory genes are said to have open pan-genomes. For example, 21% of M. tuberculosis' pan-genome is accessory gene content. Conversely, 84% of K. pneumoniae's pan-genome is accessory gene content. So, M. tuberculosis has a closed pan-genome and K. pneumoniae has an open pan-genome.

Expanding the loci!

How would our MLST approach change if expanded the number of genes from 7 - 8 to the core-genome? This is known as the core-genome MLST or cgMLST. Taking it a step further and including both core and accessory genes is known as whole genome MLST or wgMLST. Including the accessory genes allows for a comparison that includes genes often found on plasmids or carried by transposons. In general, the more genomic content the more discriminatory the results will be. However, there are some potential caveats to wgMLST. Should an isolate be considered part of an outbreak if it loses a plasmid during transmission? Changes in the genes outside of the core-genome could impact the wgMLST results. Additionally, databases for wgMLST are very large (*E. coli* includes more than 30,000 genes) and require a lot of curation to make sure analyses are functional and up to date.

What's the best approach?

Identifying the best approach is difficult as each of these methods have various pros and cons! Many of the newer methods utilizing next-generation sequencing are capable of discriminating outbreak isolates from non-outbreak isolates to varying degrees. The process used to select SNPs or how a cgMLST / wgMLST database is curated all have an effect that could introduce bias. In practice, we have found it works best to apply multiple methods and interpret the results together.