BioNumerics®

RELEASE NOTE

ESCHERICHIA COLI - SHIGELLA SCHEMA

for whole genome sequence typing

We are proud to present a schema for true whole genome multi-locus sequence typing (wgMLST) of E. coli in BioNumerics. When used in combination with our cloud-based Calculation Engine, typing E. coli isolates up to strain level using whole genome sequencing is now easily accessible to everyone.

What is the schema exactly?

Based on the Enterobase schema recently developed by Marc Achtman and team[1], the Applied Maths scientists extended the existing core genome MLST schema to a pan-genomic schema, reflecting the known diversity of *E. coli* based on a set of 289 publicly available reference sequences. By also capturing the accessory loci, we increased the discriminatory power of the schema, permitting the detection of subtype- or outbreak-specific markers, for powerful classification and outbreak definition tools.

Which loci are present?

Starting from the 289 annotated reference genomes, our in-house developed schema uses a sampling-based multi-reciprocal BLAST procedure to determine the allele sets that make up the stable loci in the accessory genome. A per-locus allele assessment procedure then determines the central prototype allele, simultaneously defining the locus. The accessory schema, including 14837 loci, is then complemented with the 2513 core loci to obtain maximal consistency with classical and novel multi-locus sequence typing initiatives for *E. coli*. The final schema contains four typing schemes next to the complete pan genome, i.e. the core genome, synchronized to Enterobase's core definition, and three MLST schemes containing 7, 8 and 15 MLST loci respectively, as defined by Achtman[2], Pasteur[3] and Whittam[4].

How will it help you?

By using BioNumerics and the integrated powerful calculation infrastructure, analyzing whole genome sequencing data for *E. coli* has become a lot more straightforward. Our cloud-based Calculation Engine offers a high-throughput environment for all your sample processing needs. Its quality-controlled de novo assembly algorithms allow you to easily assemble whole-genome sequencing data without the need for local computing power. The two allele detection procedures (assembly-based and assembly-free) allow you to perform fast and reliable allele calling for e.g. cluster detection which can be combined with whole genome SNP analysis to obtain the utmost resolution within your sample comparisons!

The BioNumerics wgMLST schema for *E. coli* has been tested, validated and approved by our microbiologists.

Great care has been taken to create an analysis procedure that minimizes sample artifacts, while maintaining an enormous discriminatory power that supersedes the core genome schema.

With turnaround times of less than 30 minutes per sample and the ability to process multiple samples simultaneously, the power of high-performance computing will be brought to your desktop with a few clicks.

Interested?

Click on this link to request a calculation engine project or scan the QR code:

References:

(1) http://enterobase.warwick.ac.uk

(2) Wirth, T. et al. Sex and virulence in Escherichia coli: an evolutionary perspective. Mol. Microbiol. 60, 1136–1151 (2006).

(3) Jaureguy, F. et al. Phylogenetic and genomic diversity of human bacteremic Escherichia coli strains. BMC Genomics 9, 560 (2008).

(4) Qi, W. et al. EcMLST: an online database for multi locus sequence typing of pathogenic Escherichia coli. in 2004 IEEE Computational Systems Bioinformatics Conference, 2004. CSB 2004. Proceedings 520–521 (2004). doi:10.1109/CSB.2004.1332482

APPLIED MATHS

A BIOMÉRIEUX COMPANY