P0299
Paper Poster Session
Bacterial strain typing

Genomic determination of minimum multi-locus sequence typing schemas for Mycoplasma hominis to represent genomic phylogeny

Aleksey Jironkin*[1], Rebecca Brown[2], Anthony Underwood[2], Victoria Chalker[3]

[1]*Public Health England, Bioinformatics, London, United Kingdom*

[2]*Public Health England, London, United Kingdom*

[3]*Public Health England, Respiratory and Vaccine Preventable Bacterial Reference Unit, London, United Kingdom*

Background: *Mycoplasma hominis* is an opportunistic human bacterial pathogen found in the urogenital tract that has on occasion been associated with serious infection such as meningitis, endocarditis or post-operative infection. Understanding pathogen epidemiology and transmission may have importance in preventing future infections and comprehension of transmission chains. There is currently no standardised method of molecular typing of *M. hominis* and due to the fastidious growth requirements of Mollicutes genomic typing is unlikely to be available for routine practice for the foreseeable future. Previous Mollicute MLST schemas have been derived based upon the conservation of genes, which are considered to be evolutionarily stable housekeeping genes. The aim of this study was to develop a multi-locus sequence based typing schema based on bioinformatics analysis to derive the minimum number of genes required to accurately reflect genomic phylogeny.

Material/methods: Genomic sequence was obtained from a total of 18 *M. hominis* isolates including two isolates pairs from individual patients using the Illumina 2500 2x100bp platform. Phylogenetic analysis was performed by calling SNPs against ATCC27545 reference genome using bwa and GATK. High quality SNPs (min depth > 5, MQ0 > 0.05, QUAL > 40, AD ratio > 0.9) were used to as input for RAxML to construct phylogenetic tree under GTR-GAMMA model. Samples were also assembled using SPADES 3.6.0 and put through a core-accessory analysis toolkit. Multiple tree comparisons were carried out using FastTree and with the tree based on the whole genome SNPs.

Results: Core-accessory analysis of genomic assemblies of the isolates resulted in 777 genes forming a pan-genome set, of which 408 genes were found to be conserved across all isolates in the study. Exploration of the complete search space of all possible gene sets was not possible (3.5×10^{14} possible states). Therefore the set of 408 genes was reduced to 57 genes that represent the minimum required core genome to represent phylogeny determined on genomic sequence (cgMLST) by sequentially removing a single gene from all possible remaining genes. Genes were assessed for stability after 10 passages and further manual removal of genes with alleles that did not correspond to the observed epidemiological data resulted in smaller gene set allowing complete search space exploration. From the 57 genes 2 potential 7 loci-based schemas were identified as possible candidates for MLST based typing for *M. hominis* that can reflect genomic phylogeny.

Conclusions: Bioinformatics l analysis of genomic sequence data has enabled the identification of 2 candidate 7 loci MLST schemas for typing *M. hominis* that best replicate the topological phylogeny derived from genomic data. Discriminatory power of the 2 proposed MLST schemas was calculated to be 1.0 and Simpson's per locus diversity index ranged from 0.924-0.871 (Hunter-Gaston diversity index 0.99-0.933).