O032
Oral Session
Epidemiology based on whole genome sequencing

COMPARISON OF BIOINFORMATICS APPROACHES FOR WHOLE GENOME SEQUENCING ANALYSIS OF LEGIONNAIRES' DISEASE CLUSTERS

D. Harmsen[1], K. Prior[1], E. Yakunin[2], L. Valinsky[2], V. Agmon[2], I. Grotto[2], **J. Moran-Gilad**[3]

[1]Poliklinik fur Parodontologie, Universitatsklinikum Munster, Munster, Germany ; [2]Public Health Services, Ministry of Health, Jerusalem, Israel ; [3]ESCMID Study Group for Molecular Diagnostics (ESGMD) & Public Health Services, Ministry of Health, Jerusalem, Israel

Objectives: The introduction of sequence-based typing (SBT) for Legionella pneumophila (Lp) has dramatically improved Legionnaires'disease (LD) cluster investigation capabilities. Microbial whole genome sequencing (WGS) is a promising modality for LD cluster investigation but sequence data analysis methods are not standardised nor agreed. We sought to compare the analysis of Lp sequence data by calling differences in core genome alleles (core genome MLST or MLST+ approach) to an already published approach that utilises SNP mapping.

Methods: We analysed 30 Lp genomes, including 8 published reference genomes retrieved from NCBI, 7 published genomes analysed by SNP calling (Reuter et al., BMJ Open 2013), of which 5 were epidemiologically clustered and 2 outgroup strains, and 15 unpublished genomes of characterised sporadic and clustered LD cases generated by Illumina sequencing (Miseq platform). Sequence data were *de novo* assembled using CLCbio and then analysed by the SeqSphere+ software (Ridom GmbH, Münster, Germany). SBT allele sequences were extracted from the genome and compared to the database. Strain relatedness was assessed by comparing the number of differing alleles in the Lp core genome (cgMLST or MLST+ approach).

Results: The ST of each of the analysed strains was successfully extracted from the genome. ST allocation matched strain data for all analysed strains. Analysis of genomes revealed 1,623 core genome targets that could be used to generate strain comparison using MLST+. Analysis of published outbreak strains generated a nearly identical clustering of clinical and related environmental strains. Notably, SNP calling found less than 15 SNP differences within the cluster, and MLST+ has found less than 10 allele differences in that cluster, thus generating a nearly identical phylogeny. Of 15 novel strains analysed,12 were ST1 strains. Phylogenetic analysis of ST1 strains revealed a number of differing alleles in a range between 10 to 150 alleles when epidemiologically unrelated strains were compared, whilst highly related strains exhibited 0 to 1allele differences.
Conclusion: Lp genome analysis by MLST+ was readily achievable using an appropriate bioinformatics software and could easily determine classic strain STs. MLST+ successfully identified epidemiologically clustered strains and differentiated those from unrelated strains even within the same clonal complex. MLST+ seems to have a comparable discriminatory power to SNP mapping in analysis of LD outbreaks but may be advantageous as it is easier for standardisation and dissemination. Thus MLST+ has the potential of becoming a gold standard tool for Lp cluster analysis