

P2420

Abstract (poster session)

Plasmidfinder and pMLST: in silico detection and typing of plasmids

A. Carattoli*, E. Zankari, A. García-Fernandez, M. Larsen, O. Lund, L. Villa, F. Aarestrup, H. Hasman (Rome, IT; Lyngby, DK)

Objectives: With the recent rapid increase in whole-genome and plasmid sequence there is the need to identify resistance genes and plasmids using raw sequence data or contigs generated by high-throughput sequencing of entire genomes. Here we describe the design of easy-to-use web-tool useful for the rapid identification of plasmids in bacterial genomes. **Methods:** 580 plasmid sequences were downloaded from the <http://www.ncbi.nlm.nih.gov/nucleotide/>, corresponding to currently available, no-redundant, complete sequences of plasmids from Enterobacteriaceae of clinical interest (Oct 2012). Among them, 243 and 337 were small and large plasmids, respectively. Plasmid regions controlling replications (replicons) were analysed for generating a replicon sequence database, useful for the classification of plasmids in homogeneous DNA homology groups (plasmid families). Plasmid sequences were firstly aligned by BLASTn against 38 sequences of previously characterized replicons: 18 replicons from the original PCR-Based Replicon Typing scheme (Carattoli et al., 2005), and 20 replicons identified in recently fully sequenced plasmids. 282 plasmids were detected at >95% nucleotide identity, >96% coverage. The replicons of the 298 remaining plasmids (showing <95% nt identity), were classified in 62 DNA homology groups, obtaining a database of 113 specific replicon sequences. A second database was generated with the alleles of plasmid loci currently used for assigning Sequence Types (STs) by the plasmid MultiLocus Sequence Typing (pMLST), (<http://pubmlst.org/plasmid/>). **Results:** The databases were used to generate the Plasmidfinder and pMLST web tools that will be hosted at the Center for Biological Sequence Analysis of the Danish Technical University, DK (<http://www.cbs.dtu.dk/services/>) for public consultation. Web-tools were successfully tested using raw sequence data or contigs generated by high-throughput sequencing of 48 E. coli and 50 Salmonella genomes. These tools quickly detected reads and contigs containing plasmid replicons, and for plasmids belonging to the HI2, HI1, I1, N and F families, STs were also assigned. **Conclusion:** Together with other specific characteristics of the bacterial strain (resistance genes, MLST, phylogroup, serotype), replicon content is currently used as a marker for bacterial typing during epidemiological investigations. Plasmidfinder is a new web tool for quick identification of plasmid content in bacterial genome raw data.