Sequencing, assembly and comparative genomics of six Enterococcus faecium ST117, an emergent multiresistant clone responsible for an increase of bacteremia and fecal carriage in Spain

R. Tobes, M. Manrique, P. Pareja-Tobes, E. Pareja-Tobes, E. Pareja

1. Oh no sequences! Research group. Era7 Bioinformatics. Madrid, Spain

A.P. Tedim, P. Ruiz-Garbajosa, R. Cantón, F. Baquero, T. Coque

An abrupt emergence of an AmpR ST117 Enterococcus faecium (Efm) clone associated with a dramatic increase in the rates of bacteremia and fecal carriage by Efm at different Spanish hospitals since 2009. This clone belongs to the human adapted ST78 Efm lineage.

We analyze variation at genome level of phenotypically diverse Efm ST117 from patients attending in Madrid area (2009-2012) and describe the first completed ST117 genome.

After assembly of PacBio sequences with RS_HGAP_Assembly.2 [Chin-2013] we get a finished E1 genome with one chromosome and 5 plasmids (1 MegaPL, 1 medium size PL and 3 small size PL).

PacBio assembly allowed to perfectly define the plasmids, even the small size ones. The existance and the size of these 5 plasmids was experimentally tested.

Independently we sequenced E1 genome with illumina and did the assembly with velvet. In the figure we show the pair-wise alignment of the PacBio assembly and the illumina assembly. The PacBio assembly has a significant larger size. The additional sequence fragments in PacBio assembly (in the figure the white regions in the MAUVE blocks corresponding to the PacBio E1 genome) are mainly trasposases and other mobile elements and RNA operon copies that are probably colapsed in the illumina velvet assembly. PacBio sequencing is especially useful for defining all the copies of each gene.

Whole genome sequencing (WGS) with PacBio allowed to get the firsts E. faecium ST117 finished genome.

Using PacBio we have been able to solve the elements with many different copies as MGE (transposons, Insertion Sequences). These repeated MGE elements frequently bears virulence and antibiotic resistance genes that is important to solve and analyze.

Plasmids are difficult to assembly because they bear a high number of MGE. Thus PacBio is a specially useful technology for working with bacterial plasmids.

Our new comparative genomics pipelines allow the detection of differences that are not detected by classical methods.

The 6 E. faecium genomes with their BG7 annotations will be publicly available within the next months.

ACKNOWLEDGEMENTS: Work at HRyC-IRYCIS is supported by research grants funded by the European Commission (EvoTAR-282004), the Ministry of Economy and Competitiveness of Spain (NEXT-MICRO-ID-20120242, FIS-PI12-01581) and the Spanish research networks (CIBER-ESP), (REIPI RD12/0015), (PROMPT) (S2010/BMD2414), and (REDEX) (BFU 2008-0079-C03-01).

Work at Era7 Bioinformatics was supported in part by Ministry of Economy and Competitiveness of Spain (NEXT-MICRO-ID-20120242 and by Era7 that funds ohnosequences.com)