Session 1
6th May
17:00 h

Understanding the genome architecture and evolution of Shiga toxin encoding bacteriophages of *Escherichia coli*

Graça Pinto[1,2], Marta Sampaio[1], Óscar Dias[1], Carina Almeida[2], Joana Azeredo[1], Hugo Oliveira[1]

[1] CEB - Centre of Biological Engineering, University of Minho, 4710-057, Braga, Portugal
[2] INIAV, IP-National Institute for Agrarian and Veterinary Research, Rua dos Lagidos, Lugar da Madalena, Vairão, Vila do Conde, Portugal

Shiga toxin-producing *Escherichia coli* (STEC) is an important foodborne pathogen, and its major virulence factor is their ability to produce Shiga toxins. This toxin is coded by the *stx* gene, acquired through the insertion of a prophage into their genome. In our study, 179 STEC genomes were analysed for their serotype, distribution, and *stx* gene variants. Stx phages were also analysed and grouped based on shared gene content. We show that most STEC were isolated from different sources and geographical regions and belong to the non-O157 serotypes (73%). While the majority of STEC encode a single *stx* gene (61%), strains coding for two (35%), three (3%) and four (1%) *stx* genes were also found, being *stx2a* the most prevalent gene variant. PHASTER analysis found *stx* genes in intact prophage regions, indicating they are phage-borne. Stx phages from our dataset were grouped into four clusters (A, B, C and D), three subclusters (A1, A2 and A3) and one singleton, in agreement with the predicted virion morphologies. Stx phage genomes are highly diverse with a vast number of 1,838 gene phamilies (phams) of related sequences (of which 677 are orphams i.e. unique genes) and although having high mosaicism, they are generally organized into three major transcripts (structural, metabolism, lysis and virulence). There is a strong selective pressure to maintain the *stx* genes location in close proximity to the lytic cassette composed of predicted SAR-endolysin and pin-holin lytic proteins. Taken together, we demonstrate that Stx phages' genomes are highly diverse, with several lysis-lysogeny regulatory systems identified but with a conserved lytic system always adjacent to *stx* genes.