ABSTRACT

TITLE: Endolysin binding domain is highly conserved in staphylococcal phage genomes

AUTHOR(S): Hugo Oliveira, Marta Sampaio, Luís D. R. Melo, Óscar Dias, Joana Azeredo

AFFLICATION: CEB – Centre of Biological Engineering, LIBRO – Laboratório de Investigação em Biofilmes Rosário Oliveira, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

---

Bacteriophages are the most predominant and diverse entities in any ecosystem, mostly driven by constant predator–prey evolution dynamics and horizontal gene transfer. Phage population remains an untapped and uncharacterized source of genetic diversity with highly mosaic genomes and no universal genes. To better understand the genetic diversity and the evolutionary relationships of staphylococci infecting phages, we performed an *in silico* comparative study of all phage genomes infecting *Staphylococcus* genus. A total of 205 genomes, encoding 20579 proteins which could be sorted into 2137 phamilies (phams) of related sequences, 742 of which possessing only a single (orphams) unrelated sequence, where analysed. Based on shared gene content, phages could be grouped into 4 clusters (A, B, C and D), 14 subclusters (A1-A2, B1-B5, C1-C6 and D) and few singletons. Strikingly, we found that the endolysin molecular organization is shared by most members, showing an intriguing selective pressure that resists the endless cycle of coevolution between phages and the diversified staphylococcal hosts. The endolysin structure is highly conserved regardless the 3 distinct phage morphologies, the 5 endolysin synthesis modes (single gene with and without inter-lytic secondary translation site, two genes spliced by group I intron, two genes adjacent and non- adjacent) and 11 different host genera found in the analysed dataset. We found that staphylococcal phages' endolysins have several catalytic domains (AMI2, AMI3, GLUCO, PET-M23, CHAP, being the latter present in ∼95% of the cases) but only one cell wall binding domain identified (SH3, Src homology 3). Thus, SH3 domain is an intriguing biological marker of *Staphylococcus*-infecting viruses.

Overall, we give a high-resolution and update view of the staphylococcal viral genetic diversity, providing novel insights into their evolution. This analysis included a significant number of phages infecting coagulase-negative species that were limited so far and that have helped to revise previously classifications. The highly conserved endolysin binding domain represents a peculiar evolution pathway to bind the host peptidoglycan at the end of the phage lytic cycle.

We acknowledge the help of Dr. Graham Hatfull group for the development of this work.