

---

**ESTUDIO COMPARATIVO DE VARIACIONES EN LAS SECUENCIAS DE  
REGIONES CRISPRs Y HUELLAS DE FAGOS EN CEPAS DE *Vibrio*  
*parahaemolyticus*.**

**NATALIA VALDÉS PARRA  
INGENIERÍA EN BIOINFORMÁTICA**

**RESUMEN**

Estudios recientes de los genomas bacterianos han revelado diferentes componentes asociados a genomas de virus bacterianos o bacteriófagos, en la forma de fragmentos aislados, secuencias de profagos no funcionales o fagos crípticos, y también secuencias llamadas CRISPRs que le confieren inmunidad a la bacteria frente a virus y plásmidos. Todos estos elementos constituyen un registro de diferentes huellas de fagos y proporcionan una oportunidad de determinar la historia de infecciones, y sus posibles efectos en la evolución de bacterias a través de su caracterización y estudio comparativo. Acorde a los antecedentes mencionados, en el presente estudio se realizó la identificación de elementos relacionados con bacteriófagos, específicamente, secuencias CRISPR y huellas de fagos, en 8 cepas pandémicas serotipo O3:K6 de *Vibrio parahaemolyticus*, aisladas en Chile, cercanamente relacionadas. Una vez identificados los CRISPR para cada una de las cepas, se compararon de manera de determinar diferencias entre cepas. Este análisis no mostro diferencias entre las cepas chilenas. Después de esta observación se decidió extender el estudio a cepas de *V. parahaemolyticus* pandémico que pudieran tener un mayor tiempo de evolución independiente, y así mayores oportunidades de infección por fagos, como las cepas AN5030, K5034 y Peru466 encontradas en otros sitios del planeta. Además se incluyó una cepa no pandémica, filogenéticamente mucho más distantes, AQ3810. Los resultados mostraron solo la presencia de un SNV (Variante de un simple nucleótido) en esta última cepa. Lo cual sugiere que el sistema CRISPR-Cas no está actuando como un mecanismo que le confiera inmunidad a estas bacterias frente a virus y plásmidos. La identificación y posterior comparación de las huellas de fagos, si mostro diferencias entre las cepas chilenas, lo cual permitió clasificarlas de acuerdo a las regiones encontradas, es así como las cepas ATC220, RIMD2210633, y Vpkx pertenecerían al grupo de

cepas que carece de la Integrasa relacionada con fagos, ubicada en una región denominada 5k'; la cepa PMC48 se encuentra separada del resto ya que no tiene el bacteriófago f237 inserto en el genoma. Y todas las cepas restantes (ATC210, PMA109.5, PMC14.7, PMC58.5 y PMC58.7), forman un tercer grupo que posee en su genoma un total de 12 regiones pertenecientes a huellas de fagos. Además se realizó la identificación y estudio filogenético de las mutaciones puntuales encontradas en los 8 genomas de las cepas chilenas, con respecto al genoma de la cepa de referencia. El árbol filogenético obtenido por comparación de los SNV es diferente a la filogenia definida según la aparición y desaparición de huellas de fagos. Lo cual sugiere que existen relaciones evolutivas diferentes dependiendo del tipo de análisis.

**ABSTRACT**

Recent studies of bacterial genomes have revealed the existence of different viral genomes or bacteriophage components associated in the form of isolated fragments, nonfunctional sequences, or cryptic prophages, also called CRISPRs these sequences might confer immunity to bacteria against viruses and plasmids. All these elements constitute a record of different “phages footprints” and provide an opportunity to determine the history of infections and their possible effects on bacterial evolution through characterization and comparative studies. According to the above background, in this study the identification of items relating to bacteriophages, specifically the CRISPR sequences and phages footprints was performed in 8 pandemic strains called O3:K6 which are serotypes of *Vibrio parahaemolyticus* closely related, isolated in Chile. Having identified the CRISPR sequences for each one of the strains, were compared in order to determine differences between strains. This analysis showed no differences between Chilean strains. After this observation it was decided to extend the study to strains of *V. parahaemolyticus* pandemic that could have a greater time of independent evolution, and thus more opportunities for get infected by phages, the strains AN5030, K5034 and Peru466 have been found elsewhere on the planet. In addition, a non-pandemic strain phylogenetically distant called AQ3810 was included. The results showed the presence of a SNV (single nucleotide variant) only in the AQ3810 strain. This suggests that the CRISPR-Cas system is not acting as a mechanism to confer immunity to these bacteria against viruses and plasmids. The identification and subsequent comparison of the phages footprints, has showed differences between Chilean strains, which allowed to classify them according to the regions found, so as strains ATC220, RIMD2210633 and Vpkx belong to the group of strains that lacks the related phage integrase, located in a region called 5k'; the PMC48 strain was separated from the rest because it does not showed insertions of the bacteriophage f237 into his genome. All the other strains (ATC210, PMA109.5, PMC14.7, PMC58.5 and PMC58.7) form a third group, which has in its genome a total of 12 regions belonging to phages footprints. Also the identification and phylogenetic analysis of the point mutations that have found in the 8 genomes of the Chilean strains, with respect to the genome of the

reference strain was performed. The phylogenetic tree obtained by comparison of the SNV differs from the defined phylogeny by the appearance and disappearance of the phages footprints. This suggests that there exist different evolutionary relationships depending on the type of analysis.