



MECANISMO EVOLUTIVO DEL ORIGEN DE GENES SUPERPUESTOS CONVERGENTES

CAROLINA MABEL GONZÁLEZ ROSALES
INGENIERO EN BIOINFORMÁTICA

RESUMEN

La secuenciación de diversos genomas muestra que los genes se encuentran sometidos a una selección y pueden evolucionar a través del tiempo. Un ejemplo de ello es la superposición de genes, característica usualmente asociada a los virus, y que se presenta cuando una misma secuencia de ADN codifica para dos proteínas en marcos de lectura distintos. La finalidad de esta Memoria es estudiar genes superpuestos de orientación convergente mediante el análisis de su origen, de modo de explicar por qué determinados genes se encuentran separados en ciertos genomas y en otros están superpuestos. Con este conocimiento se podrá entender el modo en que surge una región codificante a partir de una que no lo es y generar un modelo de cómo nuevas secuencias de aminoácidos emergen de regiones no codificantes para determinadas duplas de genes superpuestos. Las principales herramientas de trabajo fueron los árboles filogenéticos, creados con el programa MrBayes 3.1.1, los scripts en lenguaje de programación Perl y BioPerl, la Base de Datos “*Clusters of Orthologous Groups*” (COGs) y ClustalW2 para alineamientos múltiples. Los resultados indican que cuando uno de los genes sufre una mutación, inserción, deleción o alteración del marco de lectura que altera el codón de término del gen, se produce una extensión de la secuencia hasta el siguiente codón de término, originando la superposición con el gen adyacente y la expresión de una nueva secuencia genómica. Este cambio en el gen puede permanecer por una selección positiva en el genoma, otorgando cierta estabilidad al cambio, si representa una característica positiva para el genoma. Es incluso posible que la secuencia extendida pueda adquirir nuevas propiedades con el paso del tiempo.

ABSTRACT

The sequencing of several genomes shows that genes are subject of selection and can evolve over time. An example of this are overlapping genes, a feature usually associated with viruses, which occurs when the same sequence of DNA encodes two proteins in different open reading frames. The purpose of this report is to study convergent overlapping genes through its origin analysis, which could explain why certain genes are separated in some genomes and are overlapped in others. With this knowledge it will be possible to understand how a coding region arise from a non-coding one, and generate a model of how new aminoacid sequences emerge from non-coding regions for certain pairs of overlapping genes. The main tools used were phylogenetic trees, created with MrBayes 3.1.1 program, scripts in Perl programming language and BioPerl, the “*Cluster of Orthologous Groups*” (COGs) database, and ClustalW2 for multiple sequences alignment. The results show that when a gene suffers a mutation, insertion, deletion or frame shift that modify the gene stop codon, it produces an extension of the sequence to the next stop codon, causing the overlap with the adjacent gene and the expression of a new genomic sequence. This change in the gene may remain by a positive selection in the genome, giving some stability to the modification, if its represent a positive feature for the genome. It is even possible that the extended sequence could acquire new properties with the passing of time.