



## ***Oligo-skew*: una representación gráfica que permite la detección de transferencia horizontal en procariotas**

Mira A<sup>1</sup>, Garaizar J<sup>2</sup>, Rementeria A<sup>2</sup>, San Millán R<sup>2</sup>, Bikandi J<sup>2</sup>

<sup>1</sup>Evolutionary Genomics Group, División de Microbiología, Campus de San Juan, Universidad Miguel Hernández. Alicante

<sup>2</sup>Departamento de Inmunología, Microbiología y Parasitología, Universidad del País Vasco

Los microorganismos filogenéticamente relacionados presentan en sus genomas frecuencias de oligonucleótidos similares. En este trabajo hemos comparado distintos métodos de cálculo de distancias genómicas para las frecuencias de oligonucleótidos. De estos métodos, hemos aplicado los más adecuados para estudiar la variación de la frecuencia de oligonucleótidos a lo largo del cromosoma bacteriano de todos los procariotas completamente secuenciados hasta la fecha. Como consecuencia, hemos actualizado un modelo de representación gráfica denominado *oligo-skew* que habíamos creado anteriormente con otros estadísticos. La nueva versión es capaz de mostrar con gran nitidez las zonas del genoma procariota con una frecuencia de oligonucleótidos claramente diferenciada, y cuyo origen probable es la transferencia horizontal. Ambas versiones del *oligo-skew* son accesibles en <http://insilico.ehu.es/oligoweb>.

Con respecto al cálculo de distancias, hemos observado que algunos de los estadísticos que habitualmente se usan para esta tarea son inadecuados y pueden inducir a error. Los métodos rechazados en este estudio fueron la distancia Global (Almeida, 2001), la distancia de Pearson y la de Hamming (Wang, 2005). Los métodos de cálculo que resultaron útiles fueron el coeficiente de correlación de Pearson para valores z-scores de tetranucleótidos (Teeling, 2004) y la distancia euclídea (Dufraigne, 2005). Otros estadísticos posibles están relacionados con alguno de los citados anteriormente.

La versión 1 de los *oligo-skews* es fácilmente aplicable a cualquier secuencia génica: se calcula la frecuencia de oligonucleótidos correspondiente a toda la secuencia, y se compara con la frecuencia en ventanas concretas a lo largo del genoma. El método de cálculo de distancias utilizado ha sido la distancia euclídea. Las gráficas resultantes presentan diversos picos. Algunos de los picos corresponden a fragmentos del genoma que otros autores han considerado provenientes de transferencia horizontal.

La versión 2 de los *oligo-skews* no depende únicamente de la frecuencia de oligonucleótidos en la secuencia estudiada. La idea que subyace en esta representación gráfica es la siguiente: si un fragmento del genoma de un procariota concreto ha sido adquirido mediante transferencia horizontal, es previsible que presente una mayor similitud con las frecuencias de oligonucleótidos en otros procariotas. Para crear este tipo de representación se ha comparado la frecuencia de oligonucleótidos en cada una de las ventanas con las frecuencias en todos los cromosomas procariotas secuenciados hasta la fecha, y las distancias obtenidas se han ordenado de forma ascendente. En el listado obtenido para cada ventana, se ha determinado la posición del genoma estudiado y los valores obtenidos han sido representados gráficamente. Esta representación elimina lo que podría ser considerado ruido de fondo en la versión 1 de los *oligo-skews*, y por tanto parece muy adecuada para el estudio de la transferencia horizontal.