
PREDICCIÓN DE ELEMENTOS MÓVILES DE DNA EN *Xenopus tropicalis*

**BRAULIO VALDEBENITO MATURANA
INGENIERO EN BIOINFORMÁTICA**

RESUMEN

Los elementos móviles de DNA pueden trasladar segmentos de DNA dentro de un cromosoma, o entre cromosomas. Se caracterizan por poseer una secuencia que codifica para una proteína transposasa, y, flanqueándola, poseen segmentos de DNA terminales invertidos repetidos. La transposición puede afectar a corto plazo la regulación génica, y a largo plazo, la evolución de los organismos. Hay evidencia que sugiere que los elementos móviles de DNA están relacionados con la regulación del desarrollo neural en *Xenopus laevis* y *Xenopus tropicalis*.

La tarea de identificación de elementos móviles de DNA es difícil, ya que existen 12 superfamilias distintas, las que comúnmente sufren una serie de mutaciones, perdiendo los patrones de secuencia característicos que estas poseen. Una manera de modelar una superfamilia es utilizando perfiles HMM. Estos perfiles incorporan aspectos estadísticos, dando cuenta de la divergencia de las secuencias de una superfamilia. Por esto, en este trabajo se desarrollaron perfiles HMM a partir de la curación rigurosa de secuencias de proteínas de cada una de las 12 superfamilias conocidas de elementos móviles de DNA.

Los modelos desarrollados son más sensibles y selectivos que aquellos disponibles en PFAM. Además, el análisis de las predicciones permite realizar una clasificación de los resultados en términos de presencia o ausencia de Stops, fragmento o transposasa completa, y proteína derivada de transposasa.

Las predicciones realizadas en el genoma y proteoma de *Xenopus tropicalis* muestran que la superfamilia Tc1/Mariner es la más prevalente (96% de las predicciones en el genoma y 82.4% de las predicciones en el proteoma) mientras que hay una fracción pequeña que corresponde a piggyBac, y otra fracción pequeña que corresponde a Merlin, que parece ser un elemento nuevo descrito en este organismo.

ABSTRACT

DNA mobile elements can move DNA segments inside a chromosome or between chromosomes. They are characterized by a sequence that encodes a transposase protein and flanking it, DNA segments that are inverted repeated terminals. Transposition may affect, in the short term, gene regulation, and in the long term, organisms' evolution. There is evidence suggesting that DNA mobile elements are linked to the regulation of *Xenopus laevis* and *Xenopus tropicalis* neural development.

The identification of DNA mobile elements is a hard task, since there are 12 distinct families, which often suffer mutations, losing characteristic sequence patterns. Using profile HMMs is one method to model a family. These profiles incorporate statistical characteristics, accounting for the divergence amongst a family sequences. For this reason, in this work, HMM profiles were developed from the rigorous curation of the protein sequences of each of the 12 known superfamilies of DNA mobile elements.

The models developed are more sensitive and more selective than those available in PFAM. Also, the analysis of predictions allows performing a classification of the results in terms of presence or absence of Stops, as fragments or complete transposase, and as transposase-derived protein.

The predictions performed in *Xenopus tropicalis* genome and proteome show that the Tc1/Mariner family is the most prevalent (96% of the genome prediction and 82.4% of the proteome predictions), whereas there is a small fraction belonging to piggyBac, and another small fraction belonging to Merlin, which seems to be a new described element in this organism.