

Propuesta de Trabajo Final Para la Carrera de Licenciatura en Ciencias de la Computación

Franco G. Riberi

10 de agosto de 2012

Título: Estudio de la relación de divergencia en el uso de codones
sinónimos entre virus y huésped y presencia de microRNA.

Director: Lic. Laura Tardivo.

Co-director: Daniel Gutson.

Co-director: Lic. Guillermo Biset.

Tesista: Ac. Franco Gaspar Riberi.

1. Descripción del problema

La bioinformática es una disciplina dedicada al análisis de elementos biológicos utilizando la informática como herramienta principal para generar simulaciones, probar teorías, realizar cálculos más complejos, etcétera. En particular una de las áreas estudiada es la virología, la cual estudia la acción de un virus y cómo pueden ser aplicados ciertos fármacos para su tratamiento.

Este proyecto nace como sugerencia de *FuDePAN*¹. El principal objetivo es contrastar una teoría postulada y encomendada por el Dr. Roberto

¹Fundación para el desarrollo de la programación en Ácidos Nucleicos.
<http://www.fudepan.org.ar/>

Daniel Rabinovich², que involucra principalmente la molécula de ARN³.

Para la comprensión de la hipótesis se deben tener en cuenta algunos hechos referidos al código genético y los microRNA.

- El código está organizado en tripletes o codones: conjunto de tres nucleótidos que determinan un aminoácido.
- El código genético es degenerado⁴: existen más tripletes o codones que aminoácidos, de forma que un determinado aminoácido puede estar codificado por más de un triplete. Esos tripletes son conocidos como sinónimos.
- En cada especie, se ha seleccionado una proporción de uso de esos codones que guarda relación con la proporción de RNA de transferencia correspondiente de manera de optimizar la síntesis proteica.
- Para algunos patógenos intracelulares como los virus, existe una divergencia entre el uso de codones sinónimos [12] utilizado por el virus y el huésped correspondiente. El origen de esa divergencia no está suficientemente esclarecido.
- Un microARN [10][11] (*mi*ARN o *mi*RNA por sus siglas en inglés) es un ARN monocatenario, de una longitud de entre 21 y 25 nucleótidos, y que tiene la capacidad de regular la expresión de otros genes mediante diversos procesos, utilizando para ello la ruta de ribointerferencia⁵. Se encuentran codificados en el genoma y juegan un papel importante en la regulación de la expresión proteica, en la embriogénesis⁶, procesos cancerosos e infecciones virales. La generación de los microRNA puede variar según el órgano o temporalmente.

En este trabajo se estudiará si la divergencia en el uso de codones sinónimos entre virus y huésped contribuye a disminuir la interferencia de los

²Miembro del INBIRS(Instituto Biomedico en Retrovirus y SIDA). Anteriormente CNRS, Centro Nacional de Referencia para el SIDA(Departamento Microbiología UBA). Profesor Titular de Virología (Departamento Ciencias Biológicas, CAECE). Colaborador de **FuDePAN**.

³Ácido Ribonucleico. Es un tipo de ácido nucleico compuesta por nucleótidos esencial para la vida.

⁴Implica que al menos uno de los tres nucleótidos (en general el último) puede ser distinto y sin embargo codificar para el mismo aminoácido.

⁵También conocido como RNA_i por el acrónimo del inglés RNA interference. Corresponde a un mecanismo de silenciamiento post-transcripcional de genes específicos, ejercido por moléculas de ARN que, siendo complementarias a un ARN mensajero, conducen habitualmente a la degradación de éste.

⁶Proceso que se inicia tras la fertilización de los gametos para dar lugar al embrión.

mi RNA en la expresión de los RNA mensajeros de origen viral. De esa manera se contribuirá a comprender mejor la relación virus-huésped y la evolución viral.

Estos estudios tienen también una importancia potencial en la comprensión de la patogenia viral y en el desarrollo de herramientas terapéuticas como la terapia génica.

Formalmente, dado un conjunto de small-RNA_s⁷ y una colección de RNA_m de un determinado virus, dado que hay un RNA_m por cada gen, determinar si existen small-RNA_s que se van a hibridar al RNA_m. Contabilizar la cantidad de small-RNA_s que se hibridarían a la secuencia original o a la secuencia complementaria. De manera similar, realizar el mismo estudio con el genoma humanizado⁸ contabilizando la cantidad de small-RNA_s que se hibridan.

Que se hibride o no un small-RNA_s a un determinado RNA_m involucra distintas reglas, siendo la más importante la complementariedad de bases. Pero además existen otras, como la presencia de determinadas bases, proporción de uniones GC y motivos específicos.

Para tener soporte de la ejecución distribuída de la implementación se utilizará el framework **FuD**⁹ sin tener que desarrollar específicamente esta funcionalidad y sin perder de vista el problema original. Cabe aclarar que el presente desarrollo involucrará un gran número de cálculos dada la gran cantidad de datos a manipular.

2. Recursos

Para la obtención del material bibliográfico se cuenta con la biblioteca de la Universidad Nacional de Río Cuarto en conjunto con la amplia cantidad de información que se puede encontrar en Internet y la bibliografía disponible en el Departamento de Computación mediante suscripciones a publicaciones científicas.

Además de contar con todo el material bibliográfico, se contará con un equipo interdisciplinar de trabajo, conformado por profesionales de las diferentes áreas involucradas, tanto Biología como Computación.

El sistema operativo a utilizar será GNU/Linux y el producto final deberá ser software libre (bajo licencia GPL [5], versión 3 o superior).

⁷Moléculas de RNA muy pequeñas, dentro de las cuales se encuentra la mi RNA, si RNA, entre otras.

⁸La humanización refiere al reemplazo de nucleótidos en los tripletes que forman la secuencia de RNA. Se acerca a la proporción de uso de codones utilizado en el humano.

⁹**FuDePAN** Ubiquitous Distribution[1]. <http://code.google.com/p/fud/>

3. Resultados esperados

Se espera obtener, para distintos virus, tablas comparativas de posible hibridación entre microRNA presente en el huésped humano y el RNA presente en la naturaleza y el genoma viral humanizado. Además se espera un gráfico con resúmenes estadísticos a través del cual se podrá inferir, según la tendencia de las curvas, si la divergencia en el uso de codones entre virus-huésped y la presencia de microRNA pueden estar relacionados.

Referencias

- [1] G. Biset, D. Gutson, and M. Arroyo, “A framework for small distributed projects and a protein clusterer application”, 2009.
- [2] G. Biset, D. Gutson, and M. Arroyo, “Fud: Design and implementation of a framework for small distributed applications”, 2009.
- [3] B. Meyer, “Object-Oriented Software Construction”, Second Edition, Santa Barbara: Prentice Hall Professional Technical Reference, 1997.
- [4] G. Booch, J. Rumbaugh, and I. Jacobson, “Unified Modeling Language User Guide”, Second Edition, 2005.
- [5] GNU General Public License. <http://www.gnu.org/licenses/>
- [6] H. Curtis, N. Sue Barnes, A. Schnek and G. Flores, “Biología”, Editorial Médica Panamericana S.A, 2006, ISBN: 950-06-0423-X.
- [7] B. Pierce, “Genética. Un enfoque conceptual”, Tercera Edición, Editorial médica panamericana S.A, ISBN: 978-84-9835-216-0.
- [8] A. Blanco, “Química Biológica”, Séptima Edición, Editorial El Ateneo.
- [9] Haruhiko Siomi and Mikiko C. Siomi, “On the road to reading the RNA-interference code”.
- [10] Vinay S. Mahajan, Adam Drake and Jianzhu Chen, “Virus-specific host miRNAs: antiviral defenses or promoters of persistent infection?”.
- [11] Man Lung YEUNG, Yamina BENNASSER, Shu Yun LE and Kuan

Teh JEANG, “siRNA, miRNA and HIV: promises and challenges”.

[12] Gareth M. Jenkins and Edward C. Holmes, “The extent of codon usage bias in human RNA viruses and its evolutionary origin”, 2003.

[13] Comeron JM and Aguadé M. “An evaluation of measures of synonymous codon usage bias”, 1998.