

EPIDEMIOLOGÍA MOLECULAR: USO DE ANÁLISIS FILOGENÉTICO COMO HERRAMIENTA PARA ENTENDER LA EPIDEMIOLOGÍA DE AGENTES INFECCIOSOS

Enrique Corona Barrera MVZ, PhD.¹ y Rogelio Alonso Morales MVZ, Dr.²

¹Investigador independiente. ²Departamento de Genética y Bioestadística, FMVZ-UNAM. Ciudad Universitaria, México, D.F.

enriquecoronas@gmail.com ralonsom@servidor.unam.mx

Epidemiología molecular es el estudio de las enfermedades (epidemias) usando técnicas de biología molecular. También puede entenderse como el análisis de la población de agentes infecciosos que afectan a un individuo (hospedero) o una población (hospederos).

En los últimos años, los métodos moleculares de diagnóstico de las enfermedades infecciosas en animales domésticos han incrementado su uso en medicina veterinaria, ya que han mostrado ser muy convenientes en relación a los métodos clásicos de diagnóstico que emplean el aislamiento, cultivo y diferenciación bioquímica o morfológica del patógeno.

Los métodos moleculares, como la amplificación de ADN por la reacción en cadena de la polimerasa (PCR), pueden llegar a ser menos costosos, más rápidos y sensibles. Esto, aunado al cada vez menor costo de la secuenciación de nucleótidos, ha permitido acumular una gran cantidad de información genética de diferentes patógenos de importancia veterinaria.

Con estas herramientas moleculares, adicionales al diagnóstico, se puede obtener información de las relaciones genéticas entre los aislados específicos con otras cepas del mismo patógeno y definir sus características biológicas, como su potencial patológico o variante antigénica. Esta información puede ser crucial en la epidemiología y control del patógeno.

Los estudios filogenéticos tienen como propósito reconstruir la genealogía y las relaciones genéticas presentes entre los individuos, partiendo de la idea que todos los individuos en una población están relacionados entre sí por ascendencia. Los estudios filogenéticos tienen importantes aplicaciones en taxonomía, evolución, dinámica de poblaciones, selección de vacunas y en epidemiología, entre otras más.

El pensamiento filogenético (filogenia) tiene origen en Charles Darwin quien fue el que dibujo un diagrama donde trataba de explicar la genealogía de varios taxones. En tiempos recientes, el método *cladistics* fue usado por primera vez en biología por el entomólogo alemán Willi Hennig en 1950, lo cual derivó en lo que se conoce ahora como *phylogenetics systematics* a partir de 1966. El método *Cladistics* también se ha usado en otras áreas como por ejemplo en lingüística y consiste en codificar variables en sistema binario (0, 1) para realizar un análisis. El análisis filogenético (en el caso de estudios en biología) genera un árbol filogenético (*phylogenetic tree*), el cual muestra las relaciones y formación de grupos de los individuos incluidos. La

interpretación del árbol generado tendrá lugar entonces de acuerdo al área de estudio (evolución, epidemiología), técnica molecular usada y tipo de datos (secuenciación de ADN o aminoácidos, fragmentos de ADN). Para el análisis de datos moleculares (análisis filogenético) existen varios métodos:

1. Aquellos basados en Distancia: UPGMA, WPGMA, Neighbor Joining, Minimal Evolution y Fitch-Margoliash. Los programas de computadora para realizar análisis filogenético por método de Distancia usan algoritmos distintos para ejecutar un análisis con base en modelos de evolución molecular como Jukes-Cantor, Kimura (K2P y K80), Tamura-Nei (TN93) o Felsenstein 81 (F81).
2. Aquellos basados en Caracteres: Máxima Similitud (*Maximum Likelihood*), Parsimonia (*Parsimony*) y más recientemente Análisis Bayesiano (*Bayesian Analysis*).

Cual sea el método de análisis usado para los datos moleculares, los programas de computadora generan un archivo en formato *New Hampshire/Newick* (NH), *New Hampshire Extended* (NHX) o *Nexos*.

Este tipo formato es leído del archivo generado de acuerdo al método de análisis usado por los programas como **PHYLP** (*Phylogeny Inference Program*), **PAUP** (*Phylogenetic analysis Using Parsimony*), **PROTML** (*Protein Maximum Likelihood*), **Tree-Puzzle** o **MrBayes**. Estos programas construyen la relación de los individuos y generan un árbol que puede ser visualizado de forma esquemática. Los árboles filogenéticos pueden visualizarse como cladogramas (*cladograms*) o filogramas (*phylograms*), los primeros son una representación esquemática de un estimado de la filogenia de los grupos formados en el análisis, en la que lo largo de las ramas no representan la distancia o cantidad de cambio evolutivo proporcional entre los individuos, mientras que en un filograma lo largo de las ramas sí representan la distancia o cambio evolutivo proporcional entre los individuos inferido en el análisis. La representación gráfica de los árboles filogenéticos (cladogramas o filogramas) puede ser de varias formas como rectangular, circular, radial, espacio hiperbólico, o para análisis de un gran número de individuos o taxones en plano 3D o *TreeMaps*.

El árbol filogenético puede ser con raíz (*rooted*, normalmente aparece en la parte inferior en los árboles) o sin raíz (*unrooted*). La diferencia entre estos consiste en que el árbol con raíz tiene un nodo que representa al

individuo que es el ancestro de todos los individuos representados en el árbol, mientras que un árbol sin raíz es aquel en donde no hay un nodo que represente a un ancestro, por lo que las relaciones entre los individuos se asumen sin un ancestro.

Tipos de datos (caracteres)

Carácter en filogenética se define como aquella(s) variación(es) observada(s) como características distintivas de los diferentes taxones en cuestión, o como aquellas características notorias para distinguir a un taxón de otro. Por tanto, es importante considerar el tipo de carácter a analizar, ya que aquellos caracteres cuantificables (variables continuas) no son recomendables para realizar un análisis filogenético debido a que ese tipo de datos (caracteres) requieren estadística descriptiva (media, varianza, desviación estándar) para ser descritos propiamente, por lo que los caracteres contables (variables discretas) son los adecuados para un análisis filogenético ej. Sistema binario (0 ausencia, 1 presencia), secuencias AGCT o Sin embargo, el método de Distancia permite realizar análisis usando estadística descriptiva.

El estudio filogenético en patógenos requiere de varios factores:

- A) Definir qué región del genoma del patógeno se debe secuenciar. La elección cuidadosa del segmento a estudiar puede dar información sobre el nivel de variación genética, característica de antigenicidad, grado de virulencia, resistencia a fármacos, etc.
- B) La obtención de secuencias de nucleótidos de buena calidad y confiable, así como el establecer una base de datos contra la cual se va a comparar las secuencias de los aislados.
- C) La elección del software de alineamiento de secuencias y el *software* de análisis de distancias genéticas a emplear, así como establecer los parámetros del análisis, como sería el modelo de sustitución nucleotídica, el método de análisis de confianza, y el tipo de árbol filogenético (dendrograma) elegido para representar esquemáticamente los datos del análisis y entonces interpretar las relaciones de parentesco o identidad entre las diferentes secuencias de los aislados del patógeno, así como sus niveles de diferenciación y distancias evolutivas..

Aplicaciones del Análisis filogenético.

El análisis de los árboles filogenéticos del genoma de patógenos ha permitido identificar sus orígenes y los patrones de dispersión de las poblaciones o definir especies o su posición taxonómica. También es útil para diferenciar poblaciones, diferenciar virulencia, selección de vacunas y estudio de brotes de enfermedades infecciosas (epidemiología molecular).

La estructura de los árboles puede revelar las fuerzas que han moldeado su evolución. La comparación de las secuencias de los genomas de patógenos que difieren en sus propiedades biológicas, como virulencia, tropismo, antigenicidad, transmisibilidad, nos permite identificar los genes que controlan su estilo de vida y sus características. Información de este tipo ayuda a mejorar el desarrollo de vacunas y drogas más efectivas. Mientras más sepamos sobre el origen y la estructura de la variación genética de las poblaciones del patógeno, mejor seremos capaces de controlarlos.

Referencias

Kitching IJ, Forey PL, Humphries CJ and Williams DM. *Cladistics*, The Theory and Practice of Parsimony Analysis. 2nd Edition. 2000, Oxford University Press, New York.

Holmes EC. Molecular epidemiology and evolution of emerging infectious diseases. *British Medical Bulletin* 1998, 54(3) 533-543.

McCormack GP and Clewley JP. The application of molecular phylogenetics to the analysis of viral genome diversity and evolution. *Rev. Med. Virol.* 2002, 12: 221-238.

Olvera A, Busquets N, Cortey M, de Deus N, Ganges L, Núñez JI, Peralta B, Toskano J and Dolz R. Applying phylogenetic analysis to viral livestock diseases: moving beyond molecular typing. *Vet J.* 2010, 184(2):130-7.

Pavlopoulos GA, Soldatos TG, Barbosa-Silva A and Schneider R. A reference guide for tree analysis and visualization. *BioData Mining* 2010, 3:1. <http://www.biodatamining.org/content/3/1/1>

Pybus OG and Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics.* 2009, 10:540-550.