

USO DE HERRAMIENTAS BIOINFORMÁTICAS EN LA EVALUACIÓN DE SECUENCIAS "DNA BARCODE" PARA LA IDENTIFICACIÓN A NIVEL DE ESPECIE

Fecha de recepción: 9 de agosto de 2012 • Fecha de aceptación: 16 de septiembre de 2012

USING BIOINFORMATIC TOOLS TO ASSES "DNA BARCODE" SEQUENCES FOR IDENTIFICATION AT SPECIES LEVEL

Sulma Paola Vera M¹ • Pedro Jiménez M² • Liliana Franco-Lara^{1,3}

RESUMEN

El "DNA barcoding" (código de barras de ADN) es una técnica usada para identificar la especie a la cual pertenece un espécimen biológico, usando una secuencia corta de ADN proveniente de una región estandarizada, la cual se compara con una secuencia de referencia. Los "DNA barcodes" son secuencias relativamente cortas dentro de los genomas de las especies, que se obtienen fácilmente mediante amplificación por PCR. Para que sean precisas, las secuencias "DNA barcode", deben poseer baja variabilidad intraespecífica y alta variabilidad interespecífica. Para que una secuencia sea oficialmente considerada un "DNA barcode", debe estar incluida en la base de datos BOLD (The Barcode of Life Data Systems), para lo cual debe pasar estrictos análisis bioinformáticos que valoran su idoneidad. El objetivo de este artículo es hacer una revisión de algunas herramientas bioinformáticas, propuestas por *Barcodinglife.org* para determinar si una secuencia cumple los requisitos para ser considerada un "DNA barcode". Estos métodos incluyen la construcción de árboles de genes, determinación del "barcode gap", estimación de índice de clasificación genealógica, análisis de variabilidad génica, redes de haplotipos, DNA-BAR y BLOG.

Palabras clave: Barcode gap, *Fusarium*, Bioinformática

1 Laboratorio de Biotecnología Vegetal, Universidad Militar Nueva Granada, Facultad de Ciencias Básicas y Aplicadas

2 Laboratorio de Fitopatología, Universidad Militar Nueva Granada, Facultad de Ciencias Básicas y Aplicadas

3 Autor para correspondencia: liliana.franco@unimilitar.edu.co

ABSTRACT

“DNA barcoding” is a technique used to identify the species to which a biological specimen belongs, using a short DNA sequence from a standardized region, which is compared with a reference sequence. The DNA barcodes are relatively short sequences from a species genome and are easily obtained by PCR amplification. For a “DNA barcode” to be precise, sequences should have low intraspecific variability and high interspecific variability. For a sequence to be officially considered a “DNA barcode” it has to be included in the BOLD (The Barcode of Life Data Systems) database, so it has to pass strict bioinformatics analyses to assess its suitability. The goal of this paper is to review some of the bioinformatic tools proposed by the *Barcodinglife.org* to determine if a sequence fulfills the requirements to be considered a “DNA barcode”. These methods include gene trees, the determination of the barcode gap, the estimation of the genealogical classification index, genetic variability analysis, network analysis, DNA-BAR and BLOG.

Key words: Barcode gap, *Fusarium*, Bioinformatics

INTRODUCCIÓN

La identificación y clasificación de organismos en especies, es vital para, entre otros, el mantenimiento de la biodiversidad, la protección de las especies, la prevención de pandemias, la búsqueda de nuevo germoplasma, el reconocimiento de patógenos y plagas de cultivos y de animales domésticos, identificación de cepas y variedades empleadas en la bioindustria y la industria de alimentos, ciencias forenses. La identificación taxonómica tradicionalmente ha sido dominio de especialistas, cuyo trabajo suele circunscribirse a grupos limitados de organismos, a los cuales hay que recurrir obligatoriamente para resolver muchos problemas biológicos.

Según, Schuh (2000), la sistemática, que incluye a la taxonomía, desarrolla tres actividades básicas que han cambiado poco en los últimos 250 años. La primera, es el reconocimiento de las unidades básicas de la vida, las especies. Por razones prácticas, en la mayoría de los casos el reconocimiento de las especies se hace a partir de la morfología y de otras

características observables en los especímenes. La segunda actividad es la clasificación de las especies en un sistema jerárquico y la tercera, es el estudio de la especie en un contexto más amplio que trata de definir sus relaciones evolutivas. Más recientemente, se ha incorporado el uso de marcadores moleculares, los cuales debido a los avances en las técnicas de secuenciación y bioinformática, se han convertido en una fuente adicional de información para la comprensión de las relaciones evolutivas y genéticas de las especies (Hajibabaei et al., 2007a).

La gran diversidad de especies que existen en la naturaleza, unas identificadas y otras por identificar, la poca disponibilidad de expertos para muchos grupos, la necesidad de identificar a los especímenes de forma rápida y precisa, la dificultad de reconocer juveniles de muchas especies, son algunas de las razones por las cuales se requieren estrategias alternativas de identificación que complementen y faciliten los métodos clásicos basados en características morfológicas.

En la resolución de estos retos actualmente tiene un papel importante la técnica llamada "DNA Barcoding", al permitir la asignación de individuos desconocidos a especies para las cuales se ha establecido un "DNA barcode". Algunos ejemplos exitosos incluyen aves (Hebert *et al.*, 2004), mamíferos (Hajibabaei, 2007a), peces (Ward y Holmes, 2007) y artrópodos (Smith, *et al.*, 2005). Esta herramienta ha contribuido también al descubrimiento de nuevas especies (Herber *et al.*, 2004; Smith *et al.*, 2006; Lane *et al.*, 2007).

El objetivo de este artículo es hacer una revisión de algunas herramientas bioinformáticas como construcción de árboles, determinación del "barcode gap", índice de clasificación genealógica, análisis de

2003, un grupo de investigadores en la Universidad de Guelph, Ontario, Canadá, generaron la iniciativa "DNA barcoding" (<http://www.barcoding.si.edu>), con la idea de crear un sistema universal para especies eucarióticas basado en un sistema molecular estándar, esta idea fue promovida gracias a la iniciativa internacional 'Consortium for the Barcode of Life' (CBOL) en 2004. Además, se generaron ambiciosos proyectos encaminados a identificar todas las especies dentro de grupos específicos, entre ellos peces (FISH-BOL, - www.fishbol.org), aves (ABBI, www.barcodingbirds.org) y lepidópteros (www.lepbarcoding.org).

Un marcador "DNA barcode" es una pequeña secuencia de ADN estandarizada, obtenida por PCR,

Los "DNA barcodes" son secuencias cortas dentro de los genomas, que se obtienen mediante amplificación por PCR. Para que la identificación sea precisa, deben poseer baja variabilidad intraespecífica y alta variabilidad interespecífica.

variabilidad génica, redes de haplotipos, DNA-BAR y BLOG, que permiten evaluar secuencias a través de un análisis unificado, para determinar su utilidad como "DNA barcodes", según las recomendaciones de Barcodinglife.org.

TECNOLOGÍA "DNA BARCODING"

Dadas las dificultades que se presentan para realizar la identificación taxonómica convencional de muchos grupos de organismos, durante los años 90, surgió progresivamente la idea de crear un sistema estandarizado para la identificación molecular a partir de productos de PCR (Frenzel & Leblois, 2008). En

presente en todos los taxones de interés, que muestra suficiente variación de secuencia para discriminar entre especies (Hebert *et al.*, 2003b; Ebach y Holdrege, 2005). Para identificar un espécimen, se realiza la amplificación de una región de ADN a partir de una muestra de tejido del organismo no identificado. El amplicón es secuenciado y la secuencia producida es comparada con una librería de referencia de "DNA barcodes" de especies previamente identificadas (Kress *et al.*, 2005; Savolainen *et al.*, 2005; Frézal and Leblois, 2008). La técnica se basa en la diversidad de secuencias en regiones cortas estandarizadas de uno o más loci, que contienen suficiente variación interespecífica como para reconocer especies cercanas

del mismo género, y al mismo tiempo, baja variación intraespecífica para disminuir la incertidumbre (Kress *et al.*, 2005; Ratnasingham y Hebert, 2007). Según Hajibabaei *et al.* (2007a), esta técnica es una poderosa herramienta para identificar especies, pero no reemplaza los análisis taxonómicos preliminares, por el contrario, es un excelente complemento de identificación. Esta tecnología rápida, amigable y reproducible busca un marcador universal, por lo menos en un rango taxonómico bien definido, que permita la identificación correcta a nivel de especie.

Para muchas especies, los genes mitocondriales han demostrado tener características de marcadores tipo "DNA barcode", pues gracias a su rápida evolución acumulan diferencias entre especies cercanas (Hebert, 2003). Muchos estudios han demostrado la eficiencia de éstos genes para identificar especies cercanas, como algas rojas (Saunders, 2005), crustáceos (Smith *et al.*, 2011), aves (Tavares y Baker, 2008) y reptiles (Vargas *et al.*, 2009), entre otros. Para más del 95% de las especies animales, una región de 600 pb del gen mitocondrial COX (citocromo c oxidasa) posee suficiente variabilidad y universalidad para permitir la identificación sin ambigüedad de especies de la mayoría de los clados animales (Hajibabaei *et al.*, 2007 b). Sin embargo, algunos autores han cuestionado la utilidad de genes mitocondriales, debido a factores como la herencia materna, el tamaño reducido de la población, la recombinación, la inconsistencia en la tasa de mutación, los procesos evolutivos y la existencia de pseudogenes no funcionales (Rubinoff *et al.*, 2006; Song, 2008).

A pesar de los casos exitosos de identificación mediante el gen COX, no en todos los grupos este marcador ha resultado útil. Por ejemplo, en anfibios Vences (2005) demostró que el gen COX no es útil, por la alta variabilidad en los sitios de iniciación que impedía el uso de cebadores universales para todas las especies y la superposición intraespecífica e interespecífica de las secuencias, lo cual dificulta la

definición de valores límites para identificar cada especie. Vences (2005) concluyó que era conveniente analizar otros genes mitocondriales y genes nucleares para complementar los resultados obtenidos con COX. Por otra parte, en estudios recientes en plantas, se han sugerido otras secuencias génicas como marcadores "DNA barcode" tales como el gen RCBL (ribulosa bifosfato carboxilasa) en complemento con la región espaciadora trn-HpsbA (Kress y Erickson, 2007), el gen cloroplástico trnL UAA (Taberlet *et al.*, 2007) y secuencias ribosomales (Chu *et al.*, 2006). En 2009, Gilmore *et al.*, evaluaron secuencias del gen COX como posibles "DNA barcodes" para el género *Fusarium*, pero demostraron su inutilidad debido a la presencia frecuente de intrones en diferentes especies de *Fusarium* y la presencia de múltiples copias del gen. Estos resultados fueron corroborados por Gálvez *et al.* (2011). En 2008, Balech evaluó genes como posibles marcadores "DNA barcode" para *Fusarium*, estudiando secuencias de β -tubulina, factor de elongación 1α , NADH deshidrogenasa subunidad-6 y citocromo oxidasa subunidad-1 pero ninguno de ellos cumplía con los requisitos para ser considerado un marcador tipo "DNA barcode".

Entonces, se plantea la necesidad de evaluar genes diferentes a los mitocondriales para ampliar las posibilidades de uso de esta tecnología. Por otra parte, la degradación del ADN generó un problema de procedimientos para el uso de esta técnica, lo cual se ha resuelto con el uso de varias parejas de cebadores que permitan la amplificación parcial del ADN de la muestra analizada, luego por medio de un proceso de ensamblaje construir la secuencia completa que genere la información total (Meusnier, *et al.*, 2008). La tecnología "DNA barcoding" ha sido útil en procesos de identificación taxonómica. Sin embargo, la enorme biodiversidad con la que se enfrentan los investigadores (Stoeckle, 2003), no siempre permite extrapolar los resultados más allá de los grupos para los cuales los marcadores han sido diseñados.

IMPLEMENTACIÓN DE UN NUEVO "DNA BARCODE"

La implementación de un nuevo "DNA barcode" implica formar parte de la iniciativa Código de Barras de la Vida (Barcodinglife.org) cuyo propósito es que todas las especies de la naturaleza puedan ser identificadas con una secuencia de ADN, como un método complementario de los métodos taxonómicos clásicos (Schindel y Miller, 2005, Ebach y Holdrege, 2005). Como parte del proyecto "Barcoding of Life" se ha desarrollado, desde 2004, la base de datos "The Barcode of Life Data Systems (BOLD)", una herramienta bioinformática que tiene como fin adquirir, almacenar, analizar y publicar los registros relacionados con "DNA barcodes". BOLD es una plataforma libre a disposición de cualquier investigación, que sirve como regulador para garantizar que todos los registros cumplan estrictas normas que garanticen la precisión de la identificación. La plataforma BOLD tiene tres aplicaciones (Ratnasingham *et al.*, 2007). La primera, es un depósito de secuencias correspondientes a la unidad básica de los estudios "DNA barcode". La segunda, es una mesa de trabajo que ayuda con la gestión, el control de calidad y el análisis de los datos "DNA barcode". La tercera, proporciona un vehículo de colaboración científica internacional generando un medio de comunicación y participación flexible y seguro.

Los datos que reposan en BOLD son un testimonio de la utilidad de la tecnología como estrategia complementaria de identificación. Sin embargo, las condiciones ideales para un marcador "DNA barcode" no siempre se consiguen y se han identificado limitaciones reiteradas al aplicar la técnica. Por ejemplo, a menudo no existe un gen universal que puede ser utilizado para discriminar todas las especies en una amplia variedad taxonómica, o la variabilidad intraespecífica puede ser significativa o superior a la interespecífica (Balech, 2008). Un buen "DNA barcode" debe tener las siguientes características (Stoeckle, 2003):

- Suficiente variabilidad para permitir la discriminación entre especies y una baja variabilidad entre los individuos de la misma especie,
- Ser una secuencia corta de ADN que se pueda obtener fácilmente en laboratorio,
- Ser suficientemente conservado en los extremos para anclar los mismos cebadores en todos los miembros del taxón,
- Poseer suficiente divergencia como para resolver especies estrechamente relacionadas.

Para que una secuencia sea oficialmente aceptada como un código de barras, se requiere que el procedimiento, el análisis y los resultados cumplan con las siguientes condiciones (Ratnasingham *et al.*, 2007):

1. Contar con secuencias de, al menos, tres aislados de cada especie,
2. Utilizar enzimas de alta fidelidad que garanticen la calidad de la secuencia al realizar la PCR,
3. Las secuencias analizadas deben tener la misma longitud,
4. Los organismos incluidos en el estudio deben estar identificados morfológicamente a nivel de especie y si es preciso molecularmente,
5. Contar con la información completa de cada organismo,
6. Las distancias interespecíficas de las secuencias deben ser menores que las distancias intraespecíficas, garantizando la existencia del "barcode gap"
7. Que las secuencias sean públicas y a disposición del proyecto).

Cuando un grupo de especies consiguen el "DNA barcode", nuevas secuencias del gen determinado son asignadas a una especie en particular, alimentando las bases de datos "DNA barcode". Estas secuencias pueden ser usadas como referencias de comparación, de tal forma que si se cuenta con la secuencia del gen en cuestión de un espécimen

desconocido, o no identificado plenamente por los métodos taxonómicos clásicos, la comparación con las secuencias de referencia permite la identificación por medio de la tecnología "DNA Barcoding".

ANÁLISIS BIOINFORMÁTICOS PARA EL "DNA BARCODING"

La asignación de un código de barras, para secuencias que permitan la identificación acertada de especies, depende de un soporte bioinformático que valide los resultados y demuestre que las secuencias cumplen a cabalidad los requisitos para un marcador de este tipo. Desde sus inicios, la tecnología "DNA Barcoding" ha implementado el uso de herramientas de análisis *in silico*, encaminadas a determinar relaciones entre secuencias que evidencien la cercanía entre los organismos identificados y sus secuencias almacenadas, y los especímenes desconocidos. También se han establecido estrategias generalizadas que buscan unificar los resultados mediante análisis como construcción de árboles y redes de haplotipos, determinación de "barcode gap", DNA-BAR y BLOG, cálculo del índice de clasificación genealógica y análisis de la variabilidad génica. Felsenstein (1981) propuso la construcción de árboles filogenéticos para establecer distancias entre organismos, práctica que se ha implementado y mejorado a través del tiempo y se demuestra en trabajos como los desarrollados por Saitou y Nei (1987) y Soltis y Douglas (2003). Meyer y Paulay (2005), demostraron la relevancia de establecer el "barcode Gap" dentro de un estudio "DNA barcode", lo cual ha sido ampliamente corroborado por autores como Meier *et al.*, (2008) y Seberg y Petersen (2009). Los trabajos Gómez-Zurita y Vogler (2006) plantearon el uso de análisis de reticulación como una herramienta en la evaluación de proximidad genética y, en 2012 van Velzen *et al.* presentaron una recopilación de técnicas usadas en los análisis.

A continuación, se hace una revisión de algunas herramientas bioinformáticas que permiten evaluar secuencias con el fin de determinar su utilidad como "DNA barcode". Estos análisis se realizan a partir de las secuencias derivadas de amplicones obtenidos con enzimas de alta fidelidad, incluyendo el mayor número de especies posible, y donde cada especie está representada por al menos tres muestras de especímenes diferentes.

Construcción de árboles

En primer lugar, la asignación de códigos de barras para especies se hace con base en la asociación, o pertenencia, de una secuencia a un "cluster" (grupo) dentro de un árbol. Inicialmente el análisis se realizaba con la simulación técnica Monte Carlo vía cadenas de Markov (Mau *et al.*, 1999), pero Munch *et al.*, en 2008, demostraron que el método Neighbor Joining (NJ) representaba una excelente aproximación rápida y confiable (van Velzen *et al.*, 2012). La idea es que al establecer el árbol, los miembros de la misma especie queden incluidos dentro de un mismo "cluster". Cuando esto no ocurre, se demuestra que la secuencia escogida no es conservada dentro de la especie y, por tanto no es útil como "DNA barcode". Es importante enfatizar que estos árboles de genes no reflejan necesariamente la historia evolutiva de una especie (Nichols, 2001), pues el método NJ no toma en cuenta la relación ancestro-descendiente, ni tampoco considera el principio de descendencia con modificación (Saitou *et al.*, 1987). Por lo tanto, el árbol que se obtiene usando NJ es una representación del grado de similitud de las secuencias de los terminales, lo cual corresponde al campo de la fenética (Jensen, 2009), no de la sistemática filogenética.

"Barcode Gap"

El "Barcode Gap" representa la relación entre las distancias intraespecíficas y las distancias interespecíficas de las secuencias obtenidas, para lo cual

se comparan las distancias entre secuencias de la misma especie y secuencias entre especies. Cuando una secuencia es un marcador "DNA barcode", la distancia interespecífica es mayor a la variación intraespecífica (Hebert *et al.*, 2003). Cuanto mayor sea la diferencia entre estas, la discriminación entre especies será más precisa (Meyer y Paulay, 2005). Se considera que un "barcode gap" es óptimo cuando la distancia interespecífica es 10 veces mayor (10X) que la distancia intraespecífica (Virgilio *et al.*, 2012). Sin embargo, "DNA barcode" se ha asignado a secuencias que presentan "gaps" menores, por ejemplo Meyer y Pauly (2005) propusieron umbrales de 3,2 X a 6,8 X para la identificación de gasterópodos marinos, y Lefébure *et al.* (2006), para crustáceos. La existencia de un "barcode gap" es indispensable, pues sin este el proceso de "barcoding" es inaplicable (Wiemers *et al.*, 2007).

La determinación del "DNA barcode gap" se realiza graficando el mínimo de la distancia interespecífica vs el máximo de la distancia intraespecífica, usando como referencia la recta $y = x$, conocida como la barrera "DNA barcode". Esta línea representa la igualdad entre la distancia intraespecífica (x) y la distancia interespecífica (y). Cuando la relación de distancias de las secuencias de genes tienen características "DNA barcode", se espera que los puntos de la gráfica se ubiquen sobre la barrera (Wiemers *et al.*, 2007), asegurando que la distancia interespecífica es mayor que la distancia intraespecífica (Figura 1). La determinación de las distancias se puede realizar utilizando el programa MEGA 5.0.3 en www.megasoftware.net (Tamura *et al.*, 2007).

Índice de Clasificación Genealógica

El índice de clasificación genealógica es un valor que permite cuantificar el grado de asociación exclusiva entre los grupos de secuencias que están representados en un árbol, mediante la estimación del "genealogical sorting index" (gsi). Los coeficientes

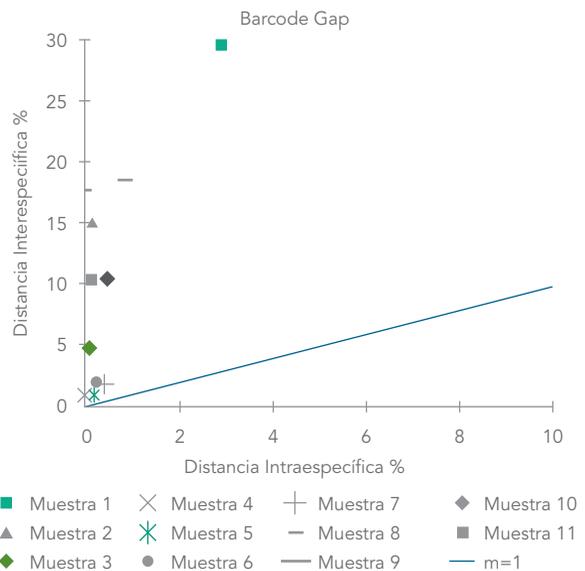


Figura 1. Ejemplo de "barcode gap". La gráfica muestra que para todas las muestras, la distancia interespecífica es mayor que la distancia intraespecífica.

de gsi están asociados a valores p, que corresponden a la probabilidad de observar valores del gsi iguales o mayores al valor determinado por los datos originales (Cummings *et al.*, 2008). El cálculo del gsi se fundamenta en la teoría de la coalescencia, en la cual se considera que los organismos, o grupo de organismos, pertenecen a la misma especie, si sus genes coalescen, es decir, que coinciden genealógicamente de forma más reciente con cualquier miembro del grupo que con otro organismo de fuera del grupo (Cummings *et al.*, 2008). La herramienta computacional utiliza simulaciones coalescentes para explorar el comportamiento del gsi, donde el mayor valor posible es 1 cuando la exclusividad dentro del grupo es monofilética. En contraste, si los valores son cercanos a 0,5 el grupo es parafilético y si es igual a 0 el grupo es considerado polifilético. Las simulaciones permiten aceptar o rechazar la hipótesis nula, la cual plantea que los grupos son parafiléticos, de manera que el gsi proporciona una forma para determinar la exclusividad incluso cuando los grupos no son monofiléticos (Cummings *et al.*, 2008, Guzmán, 2010).

La determinación del gsi se hace a través del software desarrollado en la Universidad de Maryland por Cummings et al. (2008), plataforma "Genealogical Sorting Index" (<http://www.genealogicalsorting.org>) a partir de un árbol construido con el algoritmo Neighbor Joining, en formato Phylip. Se emplea una asociación tipo cladograma, con 1000 permutaciones y sus resultados permiten evaluar si las secuencias asociadas a un grupo dentro de un árbol forman grupos monofiléticos, una condición que se requiere para un "DNA barcode" (Cummings et al., 2008). Este procedimiento ha sido ampliamente utilizado en los procesos de delimitación de especies (Sakalidis, 2011).

Variabilidad Génica

La variabilidad génica se describe en términos de cuatro índices: la diversidad haplotípica (H), la diversidad nucleotídica (π), los sitios de segregación (S) y el promedio de la diferencia entre parejas de haplotipos (K), conocidos como los Índices génicos (Librado y Rozas, 2009).

Se entiende por *Diversidad Haplotípica* (H), la probabilidad de que dos haplotipos tomados al azar de una población sean diferentes (Nei y Kumar, 2000). La diversidad haplotípica se calcula así:

$$H = \frac{n}{n-1} \sum_{i=1}^k p_i^2$$

siendo n el número de copias de genes en la muestra, k el número de haplotipos y p_i la frecuencia del i-ésimo haplotipo en la muestra.

La *Diversidad Nucleotídica* (π) es la probabilidad de que dos nucleótidos con homología de posición tomados al azar sean diferentes, y es equivalente a la diversidad génica a nivel nucleotídico (Tajima, 1989). La diversidad nucleotídica se calcula así:

$$\pi = \frac{\sum_{i=1}^k \sum_{j<1}^l P_i P_j}{L}$$

siendo π la proporción de nucleótidos diferentes entre las secuencias i y j, y L el número de nucleótidos en los haplotipos analizados.

Los *Sitios de segregación* (S) o sitios polimórficos, corresponden a sitios donde las secuencias de un haplotipo difieren, y se calculan en función del tiempo de coalescencia (Kimura, 1969) así:

$$S = 2\mu E$$

siendo μ la tasa de mutación por secuencia por generación y E el tiempo de coalescencia. El número de sitios de segregación es también el número de diferencias nucleotídicas.

Por último, el *Número promedio de las diferencias nucleotídicas por haplotipo* (K), corresponde a la medida por la cual dos secuencias homólogas al azar difieren nucleotídicamente. K se determina con la siguiente ecuación:

$$K = 4Np$$

siendo N el tamaño de la población y p la tasa de mutación (Tajima, 1993).

El análisis de la variabilidad génica tiene en cuenta todos los índices generados, entonces, es necesario evaluar el paquete completo para sugerir una idea sobre la población que se está analizando. El cálculo de la variabilidad génica se puede realizar con el programa DnaSP 5.10, (DNA Sequence Polymorphism), desarrollado por Rozas y Rozas (1999) que es adecuado para estudios a nivel genético y se encuentra disponible en www.ub.es/DNAsp, en el cual las secuencias se suben en formato phy y se agrupan por especie.

Análisis de Reticulación

Un análisis de reticulación, redes de haplotipos, o análisis "Network", consiste en evaluar la independencia entre dos linajes evolutivos. Cuando hay reticulación dos o más linajes se combinan en un cierto nivel de organización biológica que puede ser a nivel cromosómico, génico y de especie. El análisis de reticulación, genera árboles no enraizados que permiten observar relaciones entre haplotipos y sus frecuencias (Bandelt *et al.*, 1999). Este algoritmo de anidamiento agrupa haplotipos y después grupos, dependiendo del número de cambios mutacionales que existan entre las secuencias, separándolos paso a paso hasta que todos los haplotipos y clados quedan agrupados y unidos en grupos de uno, dos o más pasos mutacionales. El último nivel de anidamiento comprende la red entera (Templeton *et al.*, 1987).

Este análisis coloca las secuencias analizadas en familias operacionales, las cuales son análogas, para unirse en lugares dentro de la red proporcionando información de la relación entre ellas. La independencia de un grupo de secuencias, respecto de las demás, indica su carácter monofilético (Bittner *et al.*, 2010). Mediante esta herramienta se produce una figura que muestra sitios comunes de mutación, que representan posiciones en la secuencia en donde hay cambios de nucleótido, y sitios de vectores medios que corresponden a mutaciones relevantes para la secuencia (Fig.2). A mayor número de vectores medios entre dos haplotipos, mayor será la diferencia genética entre ellos.

La red de haplotipos se puede construir usando el software Network 4.6.0.0 con el algoritmo "median joining" disponible en: www.fluxus-engineering.com, las secuencias son subidas en formato phy (Willyard *et al.*, 2009).

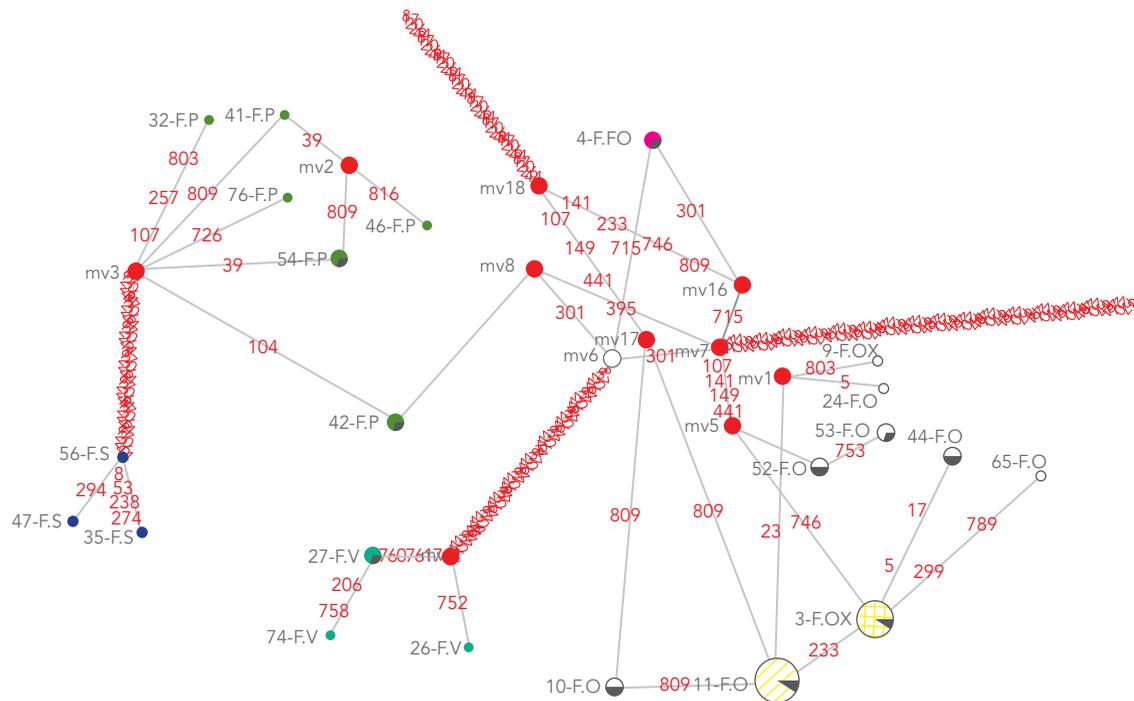


Figura 2. Gráfica de reticulación. Las muestras se representan con circunferencias de diferentes colores y su tamaño depende la cantidad de secuencias que convergen en el mismo haplotipo. La distancia entre los diferentes haplotipos se ve reflejada en la cantidad de puntos (números, diagramados en rojo) entre ellos. Los puntos rojos son conocidos como los vectores medios, que representan una conexión entre los grupos de secuencias existentes.

Prueba DNA-BAR

Es una herramienta para seleccionar grupos de diferenciadores (secuencias cortas) dentro de las secuencias "DNA barcode", utilizados para hacer análisis de identificación. El modelo puede ser visto como una cadena de ceros (0) y unos (1) asignada como un código de barras para cada organismo (DasGupta *et al.*, 2005). El análisis DNA-BAR está sujeto a un valor de redundancia que busca reducir al mínimo el número de diferenciadores reduciendo el costo computacional del ensayo. Al aumentar la redundancia aumenta la cantidad de diferenciadores (van Velzen *et al.*, 2012). En 2011, Little demostró que el análisis DNA-BAR aportaba altos niveles de precisión en la identificación de especies en secuencias que compartían estos diferenciadores.

El ensayo DNA-BAR se puede realizar en la plataforma de "Bioinformatics Lab" disponible en http://dna.engr.uconn.edu/?page_id=23, las secuencias son subidas en el formato FASTA y el análisis generalmente se realiza con una redundancia de 100.

Prueba BLOG

El propósito de BLOG (Barcoding with Logic formulas) es identificar reglas lógicas capaces de reconocer especímenes de una especie comparando su secuencia con un "DNA barcode". BLOG permite manipular los datos en los formatos típicos con los que trabaja la comunidad científica, como el formato FASTA. El archivo es convertido a un formato interno DMB con el que se realiza el análisis y el archivo de salida contiene información referente a la matriz de los datos, la etiqueta, el código y la descripción de cada espécimen e inclusive una interpretación

de los resultados mostrando las similitudes entre las secuencias y asignando un porcentaje de precisión (Bertolazzi, 2009).

El análisis BLOG se puede realizar en la plataforma de Istituto di Analisi dei Sistemi ed Informatica (Consiglio Nazionale delle Ricerche) disponible en <http://dmb.iasi.cnr.it/blog.php>.

CONCLUSIÓN

La técnica de "DNA barcoding" es de gran utilidad para la identificación a nivel de especie, pero su aplicación precisa depende de que la secuencia escogida para asociarse con una especie en particular cumpla ciertas condiciones. Entonces, es clave que los organismos a partir de los cuales se establecen las secuencias de referencia estén correctamente identificados, ya sea por métodos morfológicos, etológicos o moleculares, para lo cual se requiere de taxónomos expertos. Los métodos bioinformáticos presentados aquí, constituyen un filtro para evaluar la calidad de secuencias candidatas a "DNA barcode". A pesar de lo propuesto inicialmente, la secuencia del gen COX no es útil como marcador universal para todas las especies. Sin embargo, si dentro de un taxón determinado se encuentra una secuencia cualquiera que cumpla los requisitos exigidos para los "DNA barcodes" y esta es aceptada por la base de datos BOLD, esta se considera un marcador útil para identificación a nivel de especie.

AGRADECIMIENTOS

Los autores agradecen a la Vicerrectoría de Investigaciones de la Universidad Militar Nueva Granada por la financiación del proyecto POST CIAS 470.

BIBLIOGRAFIA

1. Balech B. 2008. An integrated molecular and morphological study to design a DNA barcode discrimination protocol for *Fusarium* species involved in dry root disease of citrus. Thesis. International centre for Advanced Mediterranean Agronomic Studies.
2. Bandelt H., P. Foster and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48.
3. Bellemain E, T. Carlsen, C. Brochmann, E. Coissac, P. Taberlet and H, Kausrud. 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol.* 10:189.
4. Bertolazzi P., G. Felici, E. Weitschek 2009. Learning to classify species with barcodes. *BMC Bioinformatics*, 10:S7.
5. Bittner L., S. Halary, C. Payri, C. Cruaud, B. de Reviers, P. Lopez and E. Baptiste. 2010. Some considerations for analyzing biodiversity using integrative metagenomics and gene networks. *Biol. Direct* 5:47.
6. Chu K, C. Li and J. Qi. 2006. Ribosomal RNA as molecular barcodes: a simple correlation analysis without sequence alignment. *Bioinformatics* 22:1690-1701.
7. DasGupta B, K. Konwar, I. Mandoiu and A. Shvartsman. 2005. DNA-BAR: distinguisher selection for DNA barcoding. *Bioinformatics Applications Note* 21: 3424–3426.
8. Feau N, A. Vialle, M. Allaire, W. Maier and R. Hamelin. 2011. DNA barcoding in the rust genus *Chrysomyxa* and its implications for the phylogeny of the genus. *Mycol.* 103:1250-1266.
9. Felsenstein J. 1981. Evolutionary trends from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17: 368-376.
10. Frézal L, R. Leblois. 2008. Four years of DNA barcoding: Current advances and prospects. *Infect Genet Evol* 8: 727-736.
11. Gálvez E, J. C, S. Restrepo, P. Jiménez, and L. Franco-Lara. 2011. Alternative oxidase gene (AOX I): a good DNA barcoding candidate for the genus *Fusarium*. *Revista Facultad de Ciencias Básicas* 2: 204-219.
12. Gilmore S., T. Gräfenhan, G. Louis-Seiz, and K. Seifert. 2009. Multiple copies of cytochrome oxidase 1 in species of the fungal genus *Fusarium*. *Mol. Ecol. Res.* 1:90-8.
13. Gómez-Zurita J. and A.P. Vogler. 2006. Testing introgressive hybridization hypotheses using statistical network analysis of nuclear and cytoplasmic haplotypes in the leaf beetle *Timarcha goettingensis* species complex. *J. Mol. Evol.* 62:421-33.
14. Hajibabaei M., G. Singer, E. Clare. and P. Hebert. 2007a. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biol.* 5:24.
15. Hajibabaei M, G.A.C. Singer, P.D.N. Hebert, D.A. Hickey. 2007b. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* 23:167-172.

16. Hebert P., A. Cywinska, S. Ball and J. DeWaard. 2003a. Biological identifications through DNA barcodes. *Proc R Soc Lond [Biol]* 270: 313–321.
17. Hebert P.D.N., E.H. Penton, J.M. Burns, D.H. Janzen, W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *PNAS* 101: 14812–14817.
18. Hebert P., S. Ratnasingham, J. DeWaard. 2003b. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proc R Soc Lond [Biol]* 270: S596–S599.
19. Huang T., Y. Yeh and D. Tzeng. 2011. Barcode-like heteroduplex DNA pattern as an aid for rapid identification of anthracnose fungi. *New biotechnol.* 28:72–78.
20. Jargeat P, Martos F, Carriconde F, Gryta H, Moreau P and Gardes M. 2010. Phylogenetic species delimitation in ectomycorrhizal fungi and implications for barcoding: the case of the *Tricholoma scalpturatum complex* (Basidiomycota). *Mol. Ecol.* 19:5216–5230.
21. Jensen R. 2009. Phenetics: revolution, reform or natural consequence?. *Taxon* 58: 50–60.
22. Kimura M. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 4:893–903.
23. Kress W. and D. Erickson. 2007. A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. *PLOS ONE* 2:508.
24. Kress, W., J., K. J. Wurdack, E.A. Zimmer, L.A. Weigt and D.H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *PNAS* 102, 8369–8374.
25. Lane C.E., C.S. Lindstrom and G.W. Saunders. 2007. A molecular assessment of northeast Pacific *Alaria* species (Laminariales, Phaeophyceae) with referenceto the utility of DNA barcoding. *Molecular Phylogenetics and Evolution* 44: 634–648.
26. Lefébure T., C.J.. Douady, M. Gouy M. and J. Gilbert. 2006. Relationship between morphological taxonomy and molecular divergence within Crustacea: Proposal of a molecular threshold to help species delimitation. *Mol. Phylogenet. Evol.* 40: 435–447.
27. Librado P. and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
28. Little D. 2011. DNA Barcode Sequence Identification Incorporating Taxonomic Hierarchy and within Taxon Variability. *PLoS ONE* 6: e20552.
29. Mau B., M.A. Newton and B. Larget. 1999. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics* 55:1–12.
30. Meier R., G. Zhang and Ali. 2008. The Use of Mean Instead of Smallest Interspecific Distances Exaggerates the Size of the “Barcoding Gap” and Leads to Misidentification. *Syst Biol.* 57: 809–813.
31. Meusnier I., G. Singer, J. Landry, D. Hickey, P.Hebert and M. Hajibabaei. 2008. A universal

- DNA mini-barcode for biodiversity analysis. *Genomics* 9:214.
32. Meyer C. and G. Paulay. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS ONE* 3:422
 33. Munch K., W. Boomsma, E. Willerslev and R. Nielsen. 2008. Fast phylogenetic DNA barcoding. *Philos. Trans. R. Soc. London [Biol]* 363:3997–4002.
 34. Nei M. and S. Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
 35. Nichols R. 2001. Gene trees and species trees are not the same. *Trends Ecol. Evolut.* 16: 358–364.
 36. Ratnasingham, S. and P. Hebert. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* 7:355–364.
 37. Rozas, J. and Rozas, R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174-175.
 38. Rubinoff D, Cameron S and Will K. 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *J. Hered.* 6:581-594
 39. Saitou N. and Nei M. 1987. The neighbor joining method - a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406–425.
 40. Sakalidis M., G. Hardy and T. Burgess. 2011. Use of the Genealogical Sorting Index (GSI) to delineate species boundaries in the *Neofusicoccum parvum*-*Neofusicoccum ribis* species complex. *Mol. Biol. Evol.* 60:333-344.
 41. Savolainen V., R.S. Cowan, A.P. Vogler, G.K. Roderick and R. Lane. 2005. Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philos. Trans. R. Soc. London [Biol]* 360: 1850–1811.
 42. Schindel D. and S. Miller. 2005. DNA barcoding a useful tool for taxonomists. *Nature* 435:17
 43. Seberg O. and G. Petersen. 2009. How many loci does it take to DNA barcode a crocus? *PLoS ONE* 4:e4598.
 44. Seifert K., R. Samso., J. Dewaard, J. Houbraken, C. Lévesque, J. Moncalvo, G. Louis-Seize and P. Hebert. 2007. Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *PNAS* 104:3901-3906.
 45. Schuh, R.T. 2000. *Biological systematics: principles and applications*. Cornell University Press. PP 3 -6.
 46. Smith M., E. Eveleigh, K. Mccann, M. Meril., P. Mc Carthy and K. Van Rooyen. 2011. Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses. *PLoS ONE* 6: 14424.
 47. Smith M., B. Fisher and P. Hebert. 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philos. Trans. R. Soc. London [Biol]* 1462: 1825-1834.
 48. Smith M. A., N.E. Woodley, D.H. Janzen, W. Hallwachs and P.D.N. Hebert, 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *PNAS* 103: 3657–3662.

49. Soltis P. and E. Douglas 2003. Applying the bootstrap in phylogeny reconstruction. *Statist. Science* 18: 256-267.
50. Song H., J. Buhay, M. Whiting and K. Crandall. 2008 Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *PNAS* 36:13486-13491.
51. Stoeckle, M. 2003. Taxonomy, DNA, and the Bar Code of Life. *Bio-Science* 53:796-797.
52. Taberlet P., E. Coissac, F. Pompanon, L. Gielly, L. Miquel C., Valentini A., Vermet T., Corthier, C. Brochmann and E., Willerslev E. 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nuc. Acids Res* 35:14.
53. Tajima F. 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.
54. Tajima F. 1993. Simple methods for testing molecular clock hypothesis. *Genetics* 135:599-607.
55. Tamura K., J. Dudley, M. Nei., and S. Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596-1599.
56. Tavares E. and A. Baker. 2008. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *BMC Evol. Biol* 8:81.
57. Templeton A., E. Boerwinkle and C. Sing. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, 117:343-351.
58. van Velzen R., E. Weitschek, G. Felici and F. Bakker. 2012. DNA barcoding of recently diverged species: relative performance of matching methods. *PLoS ONE* 7: e30490.
59. Vargas S., F. Araújo and F. Santos. 2009. DNA barcoding of Brazilian sea turtles (Testudines). *Gen. Mol. Biol.* 32:608-612.
60. Vences M., M. Thomas, R. Bonett and D. Vieites. 2005. Deciphering amphibian diversity through DNA barcoding: chances and challenges. *Philosophical Transactions of the Philos. Trans. R. Soc. London [Biol]* 360:1859-1868.
61. Virgilio M., K. Jordaens, F. Breman, T. Bäckeljau and M. Meyer. 2012. Identifying insects with incomplete DNA Barcode libraries, african fruit flies (Diptera: Tephritidae) as a test case. *PLoS ONE*, 7: e31581
62. Ward R. and B. Holmes. 2007. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (COX1) in fishes. *Molecular Ecology Notes* 7:899-907.
63. Wiemers M. and K. Fiedler. 2006. Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front. Zool.* 4:8
64. Willyard A., R. Cronn and A. Liston. 2009. Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Mol. Phyl. Evol.* 52:498-511.