

Sistemática. De la nomenclatura binomial a la genómica

*Juan Carlos Montero Castro, Deneb García Ávila,
Edgar Ismael Olvera Mendoza y Sabina I. Lara Cabrera*

Facultad de Biología, UMSNH

Resumen

Desde los inicios de la biología, clasificar a los organismos ha sido un requisito para tratar de entender la diversidad. La teoría de la evolución de Darwin como en otros campos constituye un parteaguas para entender las relaciones entre los seres vivos. Sin embargo, la aplicación de la teoría evolutiva a la sistemática tardó casi 100 años; prevaleciendo durante ese tiempo la similitud como un criterio clasificatorio. Se describen brevemente las perspectivas fenética, neodarwinista y cladista, enfatizando su importancia en sistemática, al privilegiar la ancestría común sobre criterios de similitud y divergencia adaptativa para agrupar o segregar agrupamientos taxonómicos. Adicionalmente se discuten las principales fuentes de evidencia en sistemática, como la morfología, con su reciente enfoque de fenómica y molecular desde secuenciación Sanger hasta genómica.

Palabras clave: Taxonomía, filogenia, genómica, fenómica.

Abstract

Classification has been a requirement since the beginning of biological sciences, aiming to understand diversity. Darwin's evolutionary theory was a turning point, as

in other fields, to understand living being relationships. However, it took almost 100 years for systematics to incorporate evolutionary theory; prevailing the similarity criteria in classifications. Here we describe the phenetic, neodarwinist and cladist perspectives. Emphasizing the importance in systematics of common ancestry criteria over similarity and adaptive divergence, to cluster taxa. Additionally the main sources of evidence in systematics are discussed: morphology and the phenomics, a new approach and molecular from Sanger sequencing to genomics.

Introducción

En este trabajo presentamos el desarrollo de la sistemática, que como todas las ciencias ha pasado por varias etapas conceptuales, auxiliada por el avance científico y tecnológico de otras áreas, como por ejemplo la informática. La sección *Clasificación predarwiniana*, menciona brevemente los incipientes inicios cuando el hombre realizaba clasificaciones utilitarias, hasta el ejercicio más organizado de esta disciplina en los siglos XVII al XIX, antes de la incorporación del pensamiento evolutivo darwiniano. En la sección *Clasificación postdarwiniana* además de la incorporación de las ideas evolutivas, se menciona la importancia de incorporar las ideas de herencia mendeliana, lo que produjo los primeros esfuerzos por reflejar relaciones evolutivas en las filogenias y hacer de este campo una ciencia rigurosa. De este periodo resultaron las tres principales escuelas de conocimiento taxonómico que son revisadas en las secciones respectivas: *síntesis evolutiva*, *taxonomía numérica* y *cladismo*.

En cuanto al desarrollo de las fuentes de evidencia en sistemática, en la sección *Marcadores morfológicos y el fenoma* se menciona porqué el estudio de la morfología de los organismos ha sido indispensable en la sistemática y como recientemente estudios morfométricos y anatómicos más rigurosos y la articulación de proyectos globales para reunir estos datos, le confieren a este tipo de información una importancia renovada. En la sección *Marcadores moleculares* se habla de los avances de este tipo de datos, con estudios basados en algunos cientos de pares de bases. En el análisis de este tipo de marcadores se tiene que implementar una comparación homóloga de cada nucleótido de las secuencias a confrontar, lo cual es discutido en la sección *Alineamiento y gaps*. Otro problema de homología que se ha tenido que superar es el hecho que los genes han sufrido eventos de duplicación, lo cual complica utilizarlos en esta disciplina, tema tratado en la sección *Duplicación y homología de genes*. Por otro lado, con el reciente desarrollo de la tecnología de secuenciación de última generación es viable la realización de estudios de genomas

completos en la sistemática, lo cual es introducido en la sección *Genómica*. Por último, en la sección *Algoritmos de reconstrucción filogenética*, se mencionan los programas que tradicionalmente se han utilizado en la inferencia filogenética, así como aproximaciones novedosas que realizan análisis más eficientes o que pueden manejar algunas dificultades inherentes de los datos generados por la secuenciación masiva.

Clasificación predarwiniana

Todas las culturas, han asignado nombres a los distintos organismos con los que interactúan, demostrando que desde los orígenes de la humanidad ha habido una necesidad inherente de nombrar y clasificar la diversidad biológica. Se han realizado estudios comparativos de las especies que se reconocen por la ciencia “occidental” y por integrantes de comunidades aisladas de la modernidad, encontrándose sorprendentes coincidencias; indicando que la realidad de esas entidades orgánicas va más allá de las tendencias culturales del observador (Llorente-Bouquets 1990).

Si bien el reconocimiento de la diversidad por la gente sin entrenamiento científico y la asignación de nombre vulgares es muy importante para el entendimiento de la naturaleza en cada comunidad, esos nombres, al no ser asignados con una metodología universal, tienden a ser útiles únicamente en el ámbito local. De una región a otra los nombres comunes tienden a cambiar e incluso puede ocurrir que dos especies que tengan el mismo nombre en dos regiones distantes, sean especies diferentes. Al percatarse de este problema, los primeros naturalistas buscaron un sistema universal, riguroso y predictivo para nombrar a los organismos. En ese sentido, Bauhin (1671) dio un paso importante al desarrollar la nomenclatura binomial. A partir de esta propuesta las especies se nombran por dos palabras, una corresponde al género y otra individualizada a cada especie (epíteto específico), idea popularizada posteriormente por Linneo (1753) en su obra *Species Plantarum*.

Linneo, como otros naturalistas antes que él (Aristóteles, Caesalpino o Ray), continuó con una tendencia de clasificar de acuerdo a las similitudes entre los organismos. A pesar de que la obra capital del sistema de clasificación de Linneo fue nombrada *Systema Naturae*, su sistema de clasificación fue muy artificial, enfatizando sobretudo unos pocos caracteres reproductivos. Otros autores como Adanson (1763) en el afán de encontrar un auténtico “sistema natural” se enfocaron en construir una clasificación que tomara en cuenta el mayor de número de características morfológicas y fisiológicas. Darwin (1859) se da cuenta que los

autores que siguieron esta tendencia agruparon a los seres vivos más parecidos, excluyendo en otros grupos a los más diferentes, motivados porque consideraban que esto los acercaba a conocer el plan de un creador divino.

Darwin propone que las similitudes son útiles para la clasificación solo en cuanto revelan la genealogía, y que otros caracteres con importancia adaptativa, aun cuando sean de importancia capital en la sobrevivencia del ser, carecen de valor para clasificar. Queda claro que Darwin entiende que la única forma de hacer una clasificación natural es interpretarla rigurosamente a partir de la genealogía. En ese tema, el autor concluye diciendo: “La descendencia común es el enlace oculto que el naturalista inconscientemente ha estado buscando”.

Clasificación postdarwiniana

Si bien la teoría de la evolución tuvo un rechazo por una parte de la sociedad, Mayr (2001) considera que fue aceptada con relativa facilidad entre la comunidad científica porque proporcionó, entre otros aportes, una explicación de la jerarquía de clasificación linneana. Sin embargo, la taxonomía biológica cambió poco con el advenimiento de la teoría de Darwin. Al parecer muchos seguidores de Darwin incluso tan cercanos como Lyell y Huxley no pudieron superar su visión creacionista de la clasificación (Mayr, 2001).

Uno de los primeros científicos que pretendieron utilizar las ideas de Darwin para clasificar fue Haeckel (1894). Él retoma la idea que las relaciones entre los organismos deben de ser representadas como un árbol genealógico, por lo cual una de las metas de este autor es encontrar la forma de reconstruir el árbol de la vida. Haeckel acuña la palabra “filogenia” para referirse a la historia evolutiva de un grupo de organismos. De cualquier forma, sus árboles genealógicos no fueron darwinianos, ya que se observa que sus filogenias están orientadas por una ley de progreso y perfección, dirigiéndose hacia los organismos más complejos, que tendría al hombre como meta última. Lo anterior evidencia que sus árboles en realidad fueron influenciados por las ideas de Goethe y Lamarck (Benoît, 2003).

En la época de Darwin, se consideraba que la herencia se debía a la mezcla de fluidos reproductivos, lo cual implicaba que los hijos tuvieran características intermedias a las que presentaban los padres (Sturtevant, 2001). Esta forma de entender la herencia implicaría que la variación dentro de las especies, sobre la cual actúa la selección natural, desaparecería. Con el redescubrimiento en 1900 de las leyes genéticas de Mendel se contó con la explicación que hacía falta: este autor

propuso una teoría de herencia de factores particulados, es decir, las características de los organismos están determinados por pares de unidades, hoy conocidas como alelos, que no se diluyen entre generaciones; en otras palabras, que se segregan entre los descendientes de forma discreta.

Mendel también determinó que un alelo puede expresarse con más fuerza que el alternativo, nombrados dominante y recesivo respectivamente. Entonces un individuo diploide puede presentar dos alelos dominantes, un alelo dominante y otro recesivo, o dos alelos recesivos; en los primeros dos casos el organismo tendría el aspecto determinado por el alelo dominante y en el tercer caso el individuo tendría un aspecto determinado por el alelo recesivo. Sin embargo, en los inicios del siglo XX predominaron teorías evolutivas donde el mecanismo principal de cambio evolutivo era las mutaciones genéticas. Los genetistas (p.ej. Morgan, 1916; Müller, 1927) consideraban que las grandes mutaciones a nivel de cromosomas eran la base de cambio evolutivo, a tal punto de darle un papel preponderante en el proceso evolutivo (Goldschmidt, 1940).

En cuanto a los taxónomos de esa época predominaba una metodología preevolutiva (Stevens, 1984). La mayor parte de los trabajos sistemáticos se limitaban a describir las especies y situarlas en la jerarquía linneana mediante métodos utilizados antes del surgimiento de la teoría de Darwin. Esta costumbre tiene varios inconvenientes, quizá el más grave era el "criterio de autoridad", esto es, la opinión del especialista no podía ser cuestionada, evaluada o refutada. Como una reacción en contra de esta situación, en la década de los 40s del siglo pasado, se produjo una revolución conceptual en la sistemática; los biólogos comenzaron a interesarse cada vez más, no solo por la descripción de los organismos, sino además por el esclarecimiento de la historia evolutiva de los seres vivos y así construir una clasificación objetiva. Esta nueva perspectiva produjo tres escuelas: la síntesis evolutiva, la taxonomía numérica y la sistemática filogenética (o cladismo), las cuales serán tratadas a continuación.

Escuelas de la sistemática

Síntesis evolutiva

Autores como Fisher (1930), Wright (1930), Haldane (1932), entre otros, promovieron la interrelación entre la genética, las ideas de Darwin y la taxonomía, lo cual ha sido conocido como síntesis evolutiva, neodarwinismo, gradismo o como

fue nombraba en sus inicios la nueva síntesis (Huxley, 1940). Bajo esta perspectiva los cambios genéticos son considerados como pequeñas mutaciones, que proporcionan abundante variación a las poblaciones naturales. Este enfoque considera que la sistemática debe definirse en términos de la genética y evolución dentro y entre las poblaciones, y no limitarse a comparar características morfológicas fijas. Para Mayr (1969), uno de los promotores de esta perspectiva, era importante que la sistemática no solo estuviera argumentada en la morfología, sino que incluyera datos tomados de otras disciplinas como la citología, genética, ecología, fisiología y etología, con la expectativa de centrar la atención en los procesos que originan la evolución.

La síntesis evolutiva fundamenta las clasificaciones tanto por genealogía como por similitud, además otorga una importancia capital al grado de adaptación. Los argumentos de similitud son establecidos en términos del proceso evolutivo conocido como anagénesis o modificaciones dentro de un linaje. Suponen que los integrantes de los grupos propuestos presentan una similitud sobresaliente en relación a los otros descendientes del antecesor, los cuales son excluidos. Para los practicantes de esta tendencia (p.ej. Ashlock 1979) un grupo monofilético no necesariamente tiene que incluir a todos los descendientes de un ancestro. Entonces reconocen como monofiléticos a grupos que excluyen a algunos descendientes del ancestro que presentan adaptaciones extremas como respuesta a un ambiente particular (zona adaptativa diferente), es decir, los organismos que explotan un hábitat muy particular se han especializado tanto que se consideraba que debían ser separados de otros grupos de organismos que no comparten ese hábitat.

Tanto el argumento de similitud como el de zona adaptativa están relacionados al concepto de “grado”, propuesto por Huxley (1959): el agrupamiento de especies que es justificado por presentar un nivel de complejidad evolutiva. Por lo tanto, los grupos que excluyen a algunos de los descendientes del ancestro, llamados grupos parafiléticos (Figura 1), son justificados porque ciertos descendientes del ancestro no son incluidos debido a que ellos han sufrido anagénesis excesiva (De Queiroz 1988). Los taxónomos con esta orientación, a pesar de saber que los cocodrilos y las aves comparten un antecesor común más inmediato que con respecto a las tortugas y lepidosaurios, prefieren agrupar a los cocodrilos, tortugas y lepidosaurios en un grupo parafilético (Figura 1) llamado Reptilia (De Queiroz, 1988). Por otro lado, como las aves poseen más especies que los reptiles, además de haber acumulado numerosas modificaciones, se considera que han ocupado un nicho

adaptativo diferente al de los reptiles actuales. Con ello los adeptos de esta perspectiva han justificado separarlas en un taxón excluyente.

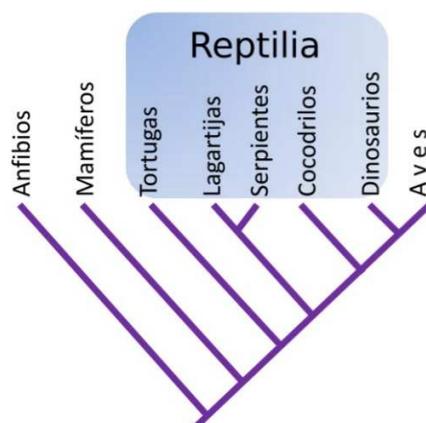


Figura 1. El área sombreada señala los integrantes del grupo Reptilia sobre la filogenia simplificada de los tetrápodos. El grupo Reptilia excluía a las aves, un descendiente del mismo ancestro del grupo, por lo tanto constituía un grupo parafilético.

Taxonomía numérica

Esta disciplina, también conocida como fenética, surgió con el fin de mejorar las clasificaciones por medio de métodos estadísticos multivariados (Sokal y Sneath, 1963). Su propósito es la construcción de agrupamientos objetivos de las unidades de estudio por medio de medidas cuantitativas de similitud. Dicho de otro modo, se construyen agrupamientos de entidades similares entre los diferentes organismos que están siendo clasificados. Esto se realiza al considerar una gran cantidad de caracteres, sin importar la naturaleza de tales características ya que se les asigna el mismo valor. Con estos datos se producen matrices de similitud que a su vez serán procesadas por algún algoritmo estadístico de agrupamiento. De los diferentes agrupamientos posibles, son preferidos los que incluyen a los organismos que presenten el mayor número de caracteres similares (Sokal y Sneath, 1963).

Los adeptos de esta disciplina consideraban que, dado que la filogenia es un artefacto meramente hipotético, entonces no debería ser el fundamento de un sistema que pretende ser objetivo. Como suponían que solo se podían conocer con certeza las similitudes morfológicas entre los organismos, estos deberían ser agrupados tomando en cuenta el mayor número de características similares entre

estos, sin importar que las similitudes sean homólogas o no, lo cual ha sido llamado la similitud global.

Una de las objeciones contra esta perspectiva es su rechazo a diferenciar entre las similitudes homólogas de aquellas que son producidas por los procesos de convergencia adaptativa. Otro inconveniente es que al considerar que la filogenia no se puede conocer, aceptan cualquier agrupación de taxa sin importar que se deriven de dos o más ancestros (Sokal y Sneath, 1963).

Aunque las similitudes y diferencias posibilitan inferir las relaciones sistemáticas, enfatizar el grado de similitud o diferencia por encima de la ancestría común niega la evolución como el rol central de la taxonomía (De Queiroz, 1988). La evolución puede haber producido las similitudes o diferencias entre los organismos, pero basar la clasificación en ellas no evidencia automáticamente sus relaciones evolutivas (De Queiroz, 1988).

Cladismo (sistemática filogenética)

Este enfoque tiene entre sus antecedentes el trabajo de Zimmermann (1931) quien escribe con una perspectiva filogenética; manejando conceptos como monofilia y homología, proponiendo métodos para determinar la evolución de los caracteres (Donoghue y Kadereit, 1992; Claßen-Bockhoff, 2001; Morrone, 2000). Tales ideas fueron retomadas por Hennig (1966), en su obra denominada *Sistemática Filogenética*, cuyo enfoque de agrupamiento es estrictamente filogenético y posteriormente fue nombrado cladista.

El principio en el que se basó Hennig fue el reconocimiento de la homología (Nixon y Ochoterena, 2001), dejando en claro que para establecer las similitudes de forma correcta deben compararse los mismos estadios de desarrollo de los organismos (Hennig, 1966). Este autor consideró que la filogenia de un grupo puede inferirse a partir del estudio comparado de un tipo especial de similitud que resulta de los eventos de especiación: las sinapomorfias, o estados derivados de un carácter, compartidos por al menos dos taxa (especies, géneros, familias o taxones superiores).

Las sinapomorfias resultan del proceso de descendencia con modificación y son heredadas del ancestro común a los linajes descendientes, por este motivo proveen la evidencia de relación filogenética cercana. De esta forma, al ubicar los caracteres compartidos entre los diferentes taxa, se pueden determinar cuáles taxones

Sistemática. De la nomenclatura binomial a la genómica

conservan las características ancestrales y cuáles presentan las características derivadas, evidenciando de este modo la relación íntima de estos últimos. La sistemática cladista, desde sus inicios, se ha utilizado para evaluar si las agrupaciones taxonómicas conocidas son grupos naturales (monofiléticos) o no. Así mismo, una hipótesis filogenética obtenida mediante el análisis de diversas fuentes de evidencia sirve de base para realizar cambios taxonómicos que reflejen las relaciones de parentesco.

TABLA 1

Resumen de las principales características de las escuelas de la sistemática surgidas en el siglo XX

| | Síntesis Evolutiva | Fenética | Cladística |
|---|---|---|---|
| Caracteres preferidos | Caracteres morfológicos, fisiológicos, genéticos, etc. | Medidas morfométricas y genéticas. | Caracteres discretos morfológicos y genéticos |
| Transformación de los datos | No transforman los datos | Datos originales se transforman a distancias | Sólo si se incluyen datos continuos se discretizan |
| Ponderación de caracteres | Los caracteres adaptativos tiene más importancia | Todos los caracteres tiene el mismo peso | Suele no ponderar, cuando pondera considera la homoplasia |
| Tratamiento de las homologías | Homologías ancestrales suelen ser más importantes | No identifican homologías | Homologías derivadas son más importantes |
| Criterio para clasificar | Por complejidad evolutiva y adaptación a nuevo ambiente | Por grandes distancias entre grupos | Por homologías derivadas |
| Delimitación de un taxón | Incluye al ancestro pero no a todos los descendientes | No considera ancestro, sólo la similitud de los integrantes | Incluye al ancestro y todos sus descendientes |
| Tipo de agrupamiento construidos | Grupos parafiléticos (grados) | Grupos polifiléticos (clusters) | Grupos monofiléticos (clados) |
| Metodología básica | Principio de autoridad del especialista en el grupo | Construcción de fenograma de distancias | Construcción Cladograma |

Hennig define estrictamente un grupo monofilético el cual incluye a un antecesor y todos sus descendientes. Bajo esta perspectiva, varios grupos monofiléticos propuestos por los adeptos de la síntesis evolutiva son en realidad grupos parafileticos. Aunque los taxa parafileticos comparten con los monofiléticos la propiedad de tener un único origen evolutivo, el hecho de no incluir algunos de los descendientes del ancestro común compromete seriamente su integridad evolutiva. Sin importar la cantidad de cambio morfológico que haya ocurrido en los linajes,

estos permanecen como parte del mismo sistema integral de ancestría común (De Queiroz, 1988). En la Tabla 1 se presenta un resumen de las características más importantes de estas aproximaciones a la sistemática.

Marcadores morfológicos y el fenoma

Los diferentes agrupamientos de especies y categorías taxonómicas se han propuestos, en gran parte de la historia de la taxonomía, con base en caracteres morfológicos. También la casi totalidad de especies descritas se han detectado con caracteres morfológicos. Aun en la época actual el reconocimiento rutinario de las diferentes especies no es realizado por datos moleculares sino por sus características morfológicas. Sin embargo, los caracteres morfológicos utilizados en estudios filogenéticos son relativamente pocos, especialmente si se comparan con los utilizados en estudios con datos moleculares. Sanderson y Donoghue (1989) reportaron que las matrices morfológicas en animales tienen un promedio de 5.5 caracteres por especie analizada, mientras que en matrices morfológicas en plantas tienen en promedio 2.5 caracteres por especie analizada. Esto hace evidente que se debe incrementar el número de caracteres en las filogenias morfológicas.

Aunque en la actualidad la mayoría de los taxónomos están de acuerdo en la búsqueda de caracteres morfológicos no utilizados en otros tiempos por los taxónomos, como los que se pueden identificar con las nuevas tecnologías en microscopia, Scotland et al. (2003) consideran que al tratar de aumentar el número de caracteres morfológicos en las estimaciones filogenéticas se incluyen caracteres que tienen altos niveles de homoplasia, tanto en niveles bajos como altos de la clasificación, por lo que sugieren solo incluir caracteres muy bien estudiados o solo mapear los caracteres morfológicos sobre filogenias moleculares. Por otro lado, la codificación de los caracteres puede afectar el resultado de los análisis filogenéticos. Si bien algunos caracteres se pueden codificar como caracteres binarios, en cambio otros necesitan esquemas más complejos de codificación; en estos casos una mala codificación puede provocar relaciones equivocadas (Hawkins et al., 1997).

Aunque los datos morfológicos presentan las desventajas de no ser abundantes y pudieran tener moderados niveles de homoplasia, excluirlos del análisis filogenético eliminaría información valiosa. Los caracteres morfológicos se obtienen, en la mayoría de los casos, de revisar todos los ejemplares conocidos de la especie, los

cuales pudieron haber sido colectados a lo largo de varios siglos. Las filogenias realizadas con datos morfológicos sirven de marco de referencia para evaluar las filogenias moleculares. Hay varios factores que pueden producir que las filogenias de datos moleculares no reflejen la filogenia de las especies, como los problemas asociados a la paralogía (ver sección *Duplicación y homología de genes*) y atracción de ramas largas, entre otros. Además, no hay que descartar problemas de contaminación o identificación errónea al obtener los datos moleculares (Wiens, 2004). La inclusión de datos morfológicos posibilita incluir fósiles en los estudios filogenéticos. Esto es importante ya que se ha determinado que la inclusión de fósiles puede facilitar resolver las relaciones de las especies actuales, en cambio al ser excluidos se pueden obtener relaciones equivocadas (Hermesen y Hendricks, 2008).

Por su parte, científicos de otros campos de la ciencia, como los bioinformáticos, se han dado cuenta que para desentrañar las sutilezas del funcionamiento del genoma es necesario conocer el fenoma, el cual ha sido definido como la información que describe el fenotipo de un organismo de acuerdo a su genética y factores ambientales a los que se ve sometido (Varki y Altheide, 2005). El conocimiento de datos fenotípicos comparativos es muy escaso, incluso en organismos que han sido estudiados por décadas. Es bien conocido que los especialistas suelen trabajar de forma aislada, lo que produce que el trabajo taxonómico avance lentamente. Los reportes de los taxónomos comúnmente son difíciles de interpretar para científicos de otras áreas de la ciencia. Se ha sugerido que el trabajo del taxónomo debe ser más colaborativo, facilitando el acceso de datos e imágenes de los ejemplares que están siendo estudiados para que haya retroalimentación entre especialistas (Johnson, 2012).

Con el fin de que la sistemática sea más interactiva y reunir datos fenómicos, se han desarrollado iniciativas globales como el Morphobank (<http://www.morphobank.org>). Este sitio de internet alberga proyectos en los que colaboran varios taxónomos interesados en un grupo particular, permitiendo almacenar información de los ejemplares estudiados como imágenes, datos de colecta y caracteres morfológicos que eventualmente serán integrados en matrices para estudios filogenéticos. Tal ambiente promueve la interacción entre los participantes de cada proyecto, pudiendo debatir sobre la mejor manera de interpretar los diferentes caracteres de los organismos estudiados. Con las primeras estimaciones filogenéticas utilizando datos fenómicos se deduce que este tipo de datos son muy importantes para resolver conflictos de filogenias obtenidas con datos moleculares. En el estudio de

O'Leary et al. (2013) los datos fenómicos resuelven relaciones de mamíferos placentarios que eran inestables con datos moleculares como la diversificación basal de estos organismos; las relaciones dentro del linaje de murciélagos, o la relación entre elefantes y sirenios.

Marcadores moleculares

De los primeros marcadores moleculares en sistemática fueron las cuantificaciones de anticuerpos en pruebas de inoculación cruzada de albumina entre especies de grandes simios y los humanos (Sarich y Wilson, 1967). Así continuaron técnicas que utilizaron hibridación de ADN y sitios de restricción (p.ej. Sibley et al., 1988; Palmer y Zamir, 1982; Delarbre et al., 1988). Con el advenimiento de la amplificación del ADN con el método de PCR se desarrollaron marcadores derivados de la amplificación de secuencias neutras, como los que utilizan oligonucleótidos arbitrarios AFLPs (Vos et al., 1995); o oligonucleótidos específicos SSR (Zietkiewicz et al. 1994).

Sin embargo, aunque estas técnicas produzcan regiones amplificadas del mismo tamaño, no se garantiza que sean homologas y por lo tanto no son comparables. Para tener más certeza de la homología de los datos es necesario conocer la secuencia de nucleótidos de la región amplificada. Por esto, en sistemática, los datos más comúnmente utilizados son aquellos obtenidos de secuenciación directa de genes y regiones no codificantes, tanto del núcleo, cloroplasto y mitocondria. Se ha reportado que las secuencias de núcleo en mamíferos pueden tener una tasa de evolución de 0.15 - 0.7% cada millón de años, las secuencias de plantas de este mismo genoma pueden tener una tasa de 0.0014 - 0.114%; mientras que en los cloroplastos la tasa de mutación son de 0.004 - 0.116%; las secuencias de la mitocondria en mamíferos son de las que presentan las más elevadas tasas de mutación de 2% por millón de años, mientras que en plantas las tasas en este genoma son las más lentas de 0.004 - 0.042% (Freeland, 2006).

Con procedimientos estándar de secuenciación por PCR se obtienen regiones entre 100 a 1000 pares de bases, información diminuta comparada con la totalidad del ADN de un organismo, pero con la cual ha sido posible inferir un gran número de filogenias de todo tipo de organismo, desde relaciones entre especies cercanamente emparentadas, hasta relaciones entre las principales ramas del árbol de la vida. Con información de este tipo a partir de secuencias de regiones ribosomales Woese (1987) confirma la hipótesis de Woese y Fox (1977) sobre las

principales divisiones de la diversidad biológica en tres agrupamientos principales, dos de organismos procariontes (eubacterias y archeobacterias) y otro de organismos eucariontes. Esta división fue posteriormente formalizada por este autor y colaboradores en 1990, que en la actualidad se conoce como los dominios Bacteria, Archaea y Eucarya; los cuales son reconocidos como las divisiones más profundas de la diversidad biológica (Figura 2).



Figura 2. Representación del árbol de la vida. Se indican alrededor los nombres de los Dominios, principales divisiones de la vida. Las ramas atravesadas representan procesos de endosimbiosis y transferencia horizontal de genes, como por ejemplo los que ocurrieron en los eucariontes, que dieron origen a las mitocondrias y cloroplastos.

Alineamiento y gaps

Las matrices de datos que incluyen las secuencias moleculares para los análisis filogenéticos tienen que ser sometidas a un proceso de alineamiento de las diferentes secuencias para asegurar la correspondencia de cada uno de los diferentes nucleótidos entre las secuencias incluidas en el análisis, lo cual es el establecimiento de la homología posicional. Este proceso se puede realizar por una inspección manual insertando gaps, es decir, espacios entre los nucleótidos, lo cual implica eventos de inserción y/o deleción de bases (también nombrados indels), hasta que las secuencias encajen lo más posible. Para esto se han desarrollado algoritmos como los comúnmente utilizados en Clustal (Larkin et al., 2007) y T-Coffee (Notredame et al., 2000). En matrices con gran cantidad de indels, algoritmos como DIALIGN-T (Morgenstern 2004) y MAFFT (Kato et al., 2002) logran producir mejores alineamientos. También se han desarrollado métodos de alineamiento

como MCALIGN (Keightley y Johnson, 2004) que toma en cuenta modelos paramétricos sobre las inserciones o deleciones propuestas.

Por otro lado, los gaps producidos en un alineamiento de secuencias pueden brindar información adicional que se puede utilizar en la elaboración de una filogenia. Una posibilidad para sacar provecho de esta información es codificar un gap como un quinto estado, esto es, un estado alternativo a los cuatro nucleótidos (Swofford et al., 1996). Sin embargo, cuando los gaps son de dos o más posiciones de largo se complica su interpretación, ya que un gap de varias posiciones podría ser interpretado como resultado de tantos pasos mutacionales como posiciones presentes en el gap, aunque en realidad el gap se haya producido por un sólo evento de deleción.

Otra posibilidad es concatenar los gaps al alineamiento como caracteres adicionales de presencia/ausencia (Simmons y Ochoterena, 2000). La codificación de los gaps en esta modalidad puede ser realizada por programas como GapCoder (Young y Healy, 2003) y SeqState (Müller, 2006). También se han desarrollado modelos paramétricos para tomarlos en cuenta en análisis probabilísticos (Miklo's et al. 2004).

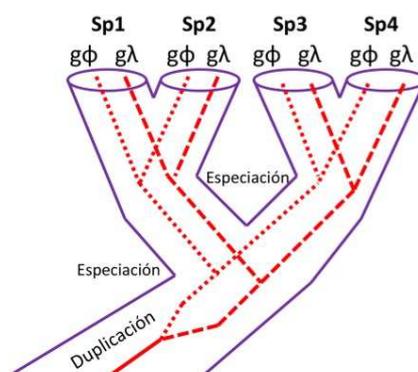


Figura 3. Representación de la paralogía de genes en un linaje de especies hipotéticas (Sp1 a Sp4). En la base del árbol, que representa la especie ancestral, se indica que antes del primer evento de especiación ocurrió un evento de duplicación de un gen formando dos linajes de genes, los cuales pasan a las especies hijas en dos eventos sucesivos de especiación. Todos los genes φ son ortólogos entre sí, lo mismo que los genes λ entre sí. En cambio la relación entre los genes φ y λ es de paralogía, ya que después del evento de duplicación han evolucionado independientemente.

Duplicación y homología de genes

Se ha determinado que los genes se pueden duplicar a partir de genes de copia única por medio de diferentes procesos celulares, como la duplicación cromosómica. El proceso de duplicación de genes parece ser bastante generalizado, se estima que en bacterias del 17 al 44 % de sus genes están duplicados y en eucariontes del 30 al 65 % (Zhang, 2003). Si en un linaje las diferentes copias del gen comienzan a cambiar a diferentes ritmos, se producen diferentes variantes del gen original que en conjunto son llamadas familias parálogas (Figura 3). Cuando estimamos filogenéticamente un grupo de organismo comparando copias de genes, que su evolución se corresponde a los eventos de especiación de estos organismos, se dice que se está realizando una comparación ortóloga. En cambio si se comparan copias de genes, que su evolución no se corresponde a los eventos de especiación de estos organismos, se dice que se está realizando una comparación paróloga (Figura 3). Es posible que al realizar un estudio filogenético entre especies se estén tomando secuencias parálogas sin saberlo, con lo que se estaría infiriendo equivocadamente la evolución de las especies, cuando en realidad solo se está obteniendo la filogenia del gen.

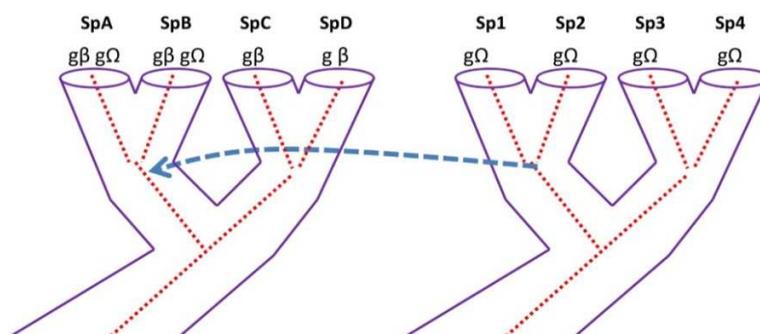


Figura 4. Representación de la xenología entre genes de dos linajes separados de especies. Las especies SpA y SpB comparten el gen $g\Omega$ que heredaron de un ancestro que sufrió una transferencia horizontal (endosimbiosis, plásmidos, hibridación, transposón, virus, etc.) a partir del gen $g\Omega$ de un linaje filogenéticamente lejano.

Otro proceso de evolución del ADN es la llamada transferencia horizontal de genes. En bacterias es un fenómeno frecuente la transmisión de genes que le confieren resistencia a los antibióticos (Gray y Fitch, 1983). Pero este fenómeno es más generalizado de lo que se pensaba (Doolittle, 1999) detectándose en organismos en todos los dominios de la vida, incluso afectando a los humanos (Salzberg et al.,

2011). Cuando en una estimación filogenética se incluyen sin saberlo secuencias que viene de una transferencia horizontal se dice que se está realizando una comparación xenóloga (Figura 4) y por lo tanto la inferencia obtenida distorsionaría en gran medida la interpretación de la evolución de las especies.

Genómica

Como se deduce de la sección anterior, la inferencia filogenética con secuencias de una región de ADN no es suficiente para determinar con confianza las relaciones entre taxa. Una forma de evitar esta complicación es incluir en un análisis filogenético varias secuencias de diferentes regiones del mismo orgánulo o combinar secuencias de diferentes orgánulos. En ese sentido, un análisis de toda la información genética de un organismo sería ideal para tener una mejor inferencia filogenética. Ahora se presenta una revolución en la obtención de datos moleculares con la posibilidad de secuenciar genomas completos. Los primeros genomas secuenciados se obtuvieron de organismos modelo como *Drosophila melanogaster* (Adams et al., 2000) *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000) y por supuesto el del humano (Venter et al., 2001), pero cada vez más genomas de organismos están siendo secuenciados (Delsuc et al., 2005; Strickler, 2012).

La obtención de estos datos genómicos se está facilitando con el desarrollo de secuenciadores de nueva generación o secuenciadores de secuenciación en paralelo, como la realizada por los equipos: 454 de Roche; GA, MiSeq, y HiSeq de Illumina; SOLiD e Ion Torrent de Life Technologies; RS Systems de Pacific Bioscience y Heliscope de Helicos Biosciences (El-Metwally, 2013), entre los que varía los tiempo de obtención de los datos, la cantidad de errores, y el costo (Kircher y Kelso, 2010; Türktas et al., 2015). La información en los datos genómicos no solo consiste de secuencias de nucleótidos, ya que es posible utilizar la información contenida en el orden de los genes, cambios en la posición de intrones, transposones o en las características de las regiones ricas en repeticiones de oligonucleótidos (Delsuc et al., 2005). Uno de los primeros retos que se tienen que afrontar con estas tecnologías de secuenciación es el manejar la gran cantidad de datos generados (terabytes). Por otro lado, los programas utilizados con los datos obtenidos por secuenciación convencional (Sanger) no pueden ser usados para manejar los datos crudos de la secuenciación masiva, por ejemplo, su calidad no puede ser revisada visualmente por el investigador por medio de la inspección de cromatogramas.

La generación de los datos por secuenciación masiva está sujeta a factores aleatorios, que generan abundantes errores en los datos (Lou et al., 2013). Eventos de poliploidia, paralogía, o heterocigosis complican el análisis de los datos generados por estas tecnologías, y se corre el riesgo de ser interpretados como errores de secuenciación y eventualmente ser omitidos (Hamilton et al., 2012). También es común que zonas repetitivas del genoma sean particularmente difíciles de secuenciar lo que produce una enorme cantidad de datos faltantes, fenómeno que es particularmente problemático en plantas (Hamilton et al., 2012). Para solventar algunos de estos problemas se han desarrollado varios programas para poder manipular los datos crudos generados de este tipo de secuenciación (p.ej. Ilie et al., 2011).

La información generada con estas nuevas tecnologías se puede utilizar para diversos propósitos como el descubrimiento de nuevos genes, detección de pérdida de genes, patrones de transcripción, etc. (p.ej. Bräutigam y Gowik, 2010; Steele 2011; Straub, 2011; Strickler, 2012; Nystedt et al., 2014). El tipo de estudios con datos genómicos que a este trabajo conciernen son los que tienen que ver con la generación de marcadores moleculares susceptibles de ser utilizados en la reconstrucción filogenética.

Por un lado, son relevantes los estudios que tratan de identificar marcadores con datos genómicos para incrementar la resolución en bajos niveles taxonómicos (Parks et al., 2009), que pueden ser utilizados para realizar filogenias de especies muy cercanas (p.ej. Heyduk et al., 2016; Nater et al., 2015; De Sousa et al., 2014). Otra aproximación que concierne para la perspectiva filogenética son los estudios con datos genómicos que se han realizado en estimaciones en niveles más inclusivos como: el inferir la historia evolutiva de las aves (Hackett et al., 2008), conocer la evolución de las algas para comprender el origen y diversidad de estos organismos eucariontes fotosintéticos (Kim et al., 2014), determinar la relación entre las plantas con semillas (Zhong et al., 2011), o resolver controversias en la evolución de los mayores linajes de los insectos que por mucho tiempo no habían sido resueltas (Misof et al., 2014). También, los datos genómicos han dado un gran impulso a la taxonomía de las bacterias (Chun y Rainey, 2014). Incluso con datos genómicos (Sicheritz-Pontén y Andersson, 2001) se ha confirmado la ya mencionada hipótesis de Woese (1987) de las relaciones principales de la vida en tres dominios.

Algoritmos de reconstrucción filogenética

Los algoritmos que se han utilizado para inferir árboles filogenéticos se pueden agrupar en los que utilizan los datos originales, en contraste con los que transforman los datos originales en distancias (índices de similitud) entre las especies. En los primeros, cada carácter es utilizado independientemente para construir y evaluar todos los árboles posibles, seleccionando el árbol que maximiza un criterio de optimización. Para matrices grandes puede requerirse de análisis que consumen excesivos recursos (computacional y tiempo). Los métodos de distancia, en cambio, son rápidos y computacionalmente relajados, ya que reducen los datos disponibles en un solo coeficiente con la información necesaria para reconstruir la filogenia, produciendo un solo árbol, aunque no necesariamente el que se ajusta mejor a los datos. Entre estos métodos se encuentran por ejemplo el UPGMA, y Neighbour-Joining, estos programas han sido utilizados por taxónomos numéricos, y todavía en la actualidad se reportan estudios que los utilizan en inferencias filogenéticas con datos moleculares.

Entre los algoritmos que no transforman los datos originales en distancias, destaca el de máxima parsimonia. La parsimonia en el contexto de la sistemática filogenética es aquella explicación que requiera el menor número de transformaciones. Consiste en evaluar todas las observaciones de manera simultánea, aceptando las agrupaciones que minimicen las hipótesis ad hoc, es decir, acepta aquellas agrupaciones que contengan el mayor grado de corroboración (Nixon y Ochoterena, 2001). En otras palabras, la parsimonia puede definirse como el criterio que maximiza el acierto en las hipótesis de la homología (Farris, 1983); de forma recíproca reduce y hace evidente la homoplasia (las similitudes no homólogas), llamada en otro contexto convergencia o analogía. En un análisis de parsimonia es común encontrar más de un árbol igualmente parsimonioso, razón por la cual surgen los métodos de consenso, como una forma de comparar todas las hipótesis y generar una que contenga los grupos que aparecen en todos, o en un porcentaje del universo de árboles.

El método de parsimonia ha permitido corroborar como monofiléticos los reinos al interior del dominio Eukarya, analizando una gran cantidad de taxones utilizando datos moleculares y morfológicos (Goloboff et al., 2009). El manejo de esta gran cantidad de información en una sola matriz de datos se da gracias a que se han

desarrollados programas con algoritmos eficientes como TNT (Goloboff et al., 2008) que de manera muy rápida pueden generar una hipótesis filogenética confiable.

Otros tipos de algoritmo que no transforman los datos originales son los que utilizan criterios probabilísticos, como son la máxima verosimilitud y el análisis bayesiano. La máxima verosimilitud evalúa la probabilidad de cambio en cada posición del alineamiento sobre cada una de las ramas de la filogenia, estos valores se conjugan para estimar la verosimilitud de cada árbol evaluado, valor que a su vez sirve para seleccionar un árbol filogenético que, de acuerdo a los datos y un modelo de evolución determinado, presente la función de probabilidad con un valor máximo. La máxima verosimilitud se considera un método robusto que es poco afectado por errores de muestreo, es un método que aun con pocos caracteres produce estimaciones con poca incertidumbre en comparación a otros métodos (Hillis et al., 1996). Sin embargo, su implementación requiere una gran demanda computacional que dependiendo del tamaño de la matriz, el análisis puede tardar días, en contraste con los minutos que tardaría un análisis de distancia para esa misma matriz. Generalmente, se obtiene una única hipótesis con el mejor valor de verosimilitud.

Otro método probabilístico es el análisis Bayesiano, el cual utiliza la probabilidad previa, que es la suposición del investigador, de acuerdo a su conocimiento previo del comportamiento del problema a evaluar. Esta probabilidad previa servirá para deducir, mediante el teorema de Bayes, el nivel de certeza del evento a evaluar, el cual es llamado probabilidad posterior. En el contexto filogenético la probabilidad posterior se determina al combinar la probabilidad previa de la filogenia contra su verosimilitud, evaluados a través de cada posición de la matriz de datos y de todas las combinaciones de largos de ramas de la filogenia (Huelsenbeck et al., 2001).

El realizar el cálculo directo de la probabilidad posterior con el teorema de Bayes implica la integración en tantas dimensiones como parámetros a evaluar, lo cual es muy complicado de realizar, por lo que se han desarrollado métodos para aproximarse a la probabilidad posterior por medio de muestreos aleatorios sobre un espacio virtual de parámetros, esto es llamado Cadena de Markov Monte Carlo (MCMC). Las cadenas suelen ser aplicadas por millones de generaciones y al final del análisis, a partir de los árboles que han sido seleccionados por la cadena, se hace un resumen (consenso) donde se indicará la probabilidad posterior de cada ramificación de la filogenia.

Actualmente, la gran cantidad de datos que se generan con la genómica pueden ser analizados sin problema para estimar filogenias con diversos programas existentes. En el contexto de la parsimonia ya se mencionaba al programa TNT (Goloboff et al., 2008) el cual es una gran alternativa para analizar un gran número de datos de forma rápida y eficiente. En el contexto probabilístico el programa RAxML (Stamatakis et al., 2005) realiza análisis de máxima verosimilitud de matrices grandes de datos. Sin embargo, los datos genómicos no solo implican tener una mayor cantidad de datos, también implica incluir diferentes regiones de estos datos que pueden tener diferente historia, como consecuencia de factores como transferencia horizontal de genes, diferenciación incompleta de linajes, hibridación, tasas altas de sustitución en linajes lejanamente relacionados, diferentes tasas de sustitución en diferentes regiones del gen, entre otros.

Se han propuesto alternativas para lidiar con estas dificultades, por ejemplo, se han sugerido la construcción de superárboles a partir de los árboles producidos por las regiones por separado, lo cual puede ser implementado en el programa PhySIC_IST (Scornavacca et al., 2008). Otra posibilidad es utilizar el programa BUCKy (Ané et al. 2007) para resumir los cladogramas producidos por las diferentes regiones analizadas indicando en cada rama la proporción de genes que apoyan cada clado. Una tercera posibilidad es analizar todas las regiones en conjunto y construir redes filogenéticas haciendo evidentes los eventos de hibridación y sorteo incompleto de linajes, como lo realizado por el programa STEM-hy (Kubatko et al., 2009). Diversos avances en la forma de analizar los datos genómicos se desarrollan en la actualidad y autores como Lemmon y Lemmon (2011) abordan consideraciones adicionales en el tratamiento de este tipo de datos.

Conclusiones y perspectivas

El objetivo de la sistemática es la reconstrucción de la historia evolutiva de la vida para elaborar la clasificación de los seres vivos. Esto no es tarea fácil ya que inferir la historia de la vida implicaría conocer, al menos, a los millones y millones de organismos extintos, algo imposible porque, aun con todos los fósiles que se hayan logrado conservar solo se tendría una parte minúscula de esa información. La información que realmente ayuda es aquella depositada en todas y cada una de las características de cada especie presente en la actualidad, principalmente características morfológicas y moleculares.

En el símil de que la historia de la vida es un árbol, reconstruir esa historia equivaldría a reconstruir las ramificaciones del árbol a partir de solo conocer las hojas más jóvenes, una tarea que podría parecer adivinatoria más que científica. Para lograr esta tarea se ha pasado por varios cambios en el marco teórico que ha abordado este problema. Los taxónomos "linneanos" realizaban su trabajo tomando en cuenta el criterio de similitud morfológica, sin estar conscientes que el criterio de similitud pudiera producir clasificaciones erróneas, ya que algunas características que derivan de diferentes estructuras ancestrales convergen en forma similar. Darwin entendió que solo las características que comparten ancestría común son la clave para clasificar. Por su parte Hennig precisó, que de las características que comparten ancestría común, solo las que son compartidas estrictamente por el ancestro y todos sus descendientes son las definitorias para clasificar. Estos avances conceptuales fueron lentamente aceptados después de la segunda mitad del siglo XX, primero con datos morfológicos y después con datos moleculares como los producidos por la secuenciación a baja escala que se desarrolló desde finales del siglo XX. Con esto se elaboraron inferencias que pudieron apoyar o cuestionar la mayoría de las clasificaciones realizadas en tiempos pasados o incluso establecieron nuevas relaciones no sospechadas anteriormente. A pesar de este gran avance, la mayoría de las inferencias realizadas tiene relaciones no resueltas en diferentes niveles de la filogenia y en algunos casos las inferencias obtenidas con diferentes datos producen resultados contradictorios.

El avance de la genómica ha propiciado un aumento descomunal de datos moleculares que comienzan a ser considerados como la respuesta para entender relaciones filogenéticas conflictivas. Pero antes de que se analicen estos datos se deben de hacer una serie de consideraciones. Por ejemplo, regiones del genoma con repeticiones de microsátelites, transposones y duplicaciones génicas enturbian la señal filogenética. Si no se toman en cuenta estos procesos se podrían inferir historias evolutivas erróneas.

Otra manera de dar más certeza a las filogenias de datos moleculares es contrastarlas con inferencias de datos morfológicos, independientemente o en forma conjunta. Por lo que iniciativas para seguir obteniendo datos morfológicos son muy importantes, como la llamada fenómica. Las primeras aplicaciones en sistemática de datos fenómicos indican que son una contraparte necesaria para apoyar o incluso modificar resultados con datos genómicos. Sin embargo, los datos fenómicos son mucho más difíciles de adquirir que los datos moleculares. Por este motivo, es recomendable que taxónomos interesados en estudiar el mismo taxón

se coordinen, para poder enriquecer bases de datos fenómicos en tiempos razonables y poder resolver problemas sistemáticos que no se han podido resolver con datos moleculares. Esto mismo se puede aplicar a otros campos de investigación “ómica” como la proteómica y transcriptómica que podrían aportar datos adicionales.

En este panorama inmediato, de abundancia de datos para realizar estudios sistemáticos, se pudiera llegar a pensar que determinando las similitudes entre especies a partir de una gran cantidad de datos seríamos capaces de inferir las relaciones de forma más exactas. Esta forma de pensar se ha utilizado anteriormente en la sistemática por M. Adanson en el siglo XVIII y por los taxónomos numéricos en el siglo XX, tendencias que ha demostrado tener deficiencias al estudiar la historia evolutiva, ya que mezclan similitudes aparentes con similitudes homologas. Por ello es importante continuar dando preferencias a los métodos de análisis que realizan explícitamente esta distinción.

Otros eventos que complican la inferencia filogenética son la hibridación, y el sorteo incompleto de genes. Esto produce historias evolutivas más complejas de lo que el paradigma del árbol filogenético nos permite solventar. Quizás esta avalancha de datos así como de formas de diagnosticar y analizar esta información produzca en la sistemática algo más que un cambio cuantitativo y en cambio se genere un cambio cuántico que escale a esta ciencia a un renovado marco conceptual que, como propone Doolittle (1999) nos dirija cada vez más a considerar los árboles reticulados para proponer clasificaciones.

Referencias

- Adams, M.D., S.E. Celniker, R.A Holt, C.A. Evans, J.D. Gocayne, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185–2195.
- Adanson, M. 1763. *Families des plantes*. Vincent. Paris. 640 pp.
- Ané, C., B. Larget, D.A. Baum, S.D. Smith, A. Rokas. 2007. Bayesian estimation of concordance among gene trees. *Mol. Biol. Evol.* 24:412–26.
- Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Ashlock, P.D. 1979. An evolutionary systematist's view of classification. *Syst. Zool.* 28: 441-450.
- Bauhin, C. 1671. *Pinax theatri botanici*. Arion. Basilea. 518 pp.

- Benoît, D. 2003. The roots of phylogeny: how did Haeckel build his trees? *Syst. Biol.* 52:515–527.
- Bräutigam, A. y U. Gowik. 2010. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol.* 12: 831-841.
- Chun, J. y F.A. Rainey. 2014. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. *Int. J. Syst. Evol. Microbiol.* 64: 316–324.
- Claßen-Bockhoff, R. 2001. Plant morphology: the historic concepts of Wilhelm Troll, Walter Zimmermann and Agnes Arber. *Ann. Bot.* 88: 1153-1172.
- Darwin, C. 1859. *On the origin of species*. Murray. London. 458 pp.
- De Queiroz, K. 1988. Systematics and the Darwinian revolution. *Philos. Sci.* 55: 238–259.
- Delarbre, C., Y. Kashi, P. Boursot, J.S. Beckmann, P. Kourilsky, F. Bonhomme y G. Gachelin. 1988. Phylogenetic distribution in the genus *Mus* of t-complex-specific DNA and protein markers: inferences on the origin of t-haplotypes. *Mol. Biol. Evol.* 5:120–133.
- Delsuc, F., H. Brinkmann y H. Philippe. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6:361–375.
- De Sousa, F., Y.J.K. Bertrand, S. Nylander, B. Oxelton, J.S. Eriksson, B.E. Pfeil. 2014. Phylogenetic properties of 50 nuclear loci in *Medicago* (Leguminosae) generated using multiplexed sequence capture and next-generation sequencing. *PloS ONE.* 9(10):e109704.
- Donoghue, M. J. y J.W. Kadereit. 1992. Walter Zimmermann and the growth of phylogenetic theory. *Syst. Biol.* 141:74-85.
- Doolittle, W.F. 1999. Phylogenetic classification and the universal tree. *Science.* 284: 2124–2129.
- El-Metwally S., T. Hamza, M. Zakaria y M. Helmy. 2013. Next-generation sequence assembly: four stages of data processing and computational challenges. *PLOS Comput. Biol.* 9 (12): e1003345.
- Farris, J. 1983. The logical basis of phylogenetic analysis. En: *Advances in Cladistics*. (Platnick, N.I., y V.A. Funk, eds.) Columbia University Press. New York. 7–36 pp.
- Fisher, R.A. 1930. *The genetical theory of natural selection*. The Clarendon Press. Oxford. 272 pp.
- Freeland, J.R. 2006. *Molecular Ecology*. John Wiley & Sons. England. 388 p.

- Goldschmidt, R. 1940. The material basis of evolution. Yale University Press. New Haven. 436 pp.
- Goloboff, P.A., J.S. Farris y K.C. Nixon 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24:774–786.
- Goloboff, P.A., S.A. Catalano, M.J. Mirande, C. Szumik, S. Arias, M. Källersjö, y J.S. Farris. 2009. Phylogenetic analysis of 73060 taxa corroborates major eukaryotic groups. *Cladistics* 25: 211–230.
- Gray, G.S. y W.M. Fitch. 1983. Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resistance gene from *Staphylococcus aureus*. *Molecular Biology and Evolution*, 1(1): 57–66.
- Hackett, S.J., y R.T. Kimball, S. Reddy, R.C.K. Bowie, E.L. Braun, et al. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768.
- Haeckel, E. 1894. Systematische phylogenie. Reimer. Berlin. 400 pp.
- Haldane, J.B.S. 1932. The causes of evolution. Princeton University Press. London. 235 pp.
- Hamilton, J.P. y C.R. Buell. 2012. Advances in plant genome sequencing. *Plant J.* 70:177–190.
- Hawkins, J.A., C.E. Hughes, R.W. Scotland .1997. Primary homology assessment, characters and character states. *Cladistics* 13:275–283.
- Hennig, W. 1966. Phylogenetic systematics. University of Illinois Press. Urbana. 263 pp.
- Hermesen, E.J. y J.R. Hendricks. 2008. W(h)ither fossils? Studying morphological character evolution in the age of molecular sequences. *Ann. Missouri Bot. Gard.* 95:72–110.
- Heyduk, K., D.W. Trapnell, C.F. Barrett, M.J. Leebens. 2016. Phylogenomic analyses of species relationships in the genus *Sabal* (Arecaceae) using targeted sequence capture. *Biol. J. Linnean Soc.* 117: 106-120.
- Hillis, D.M., C. Moritz, B.K. Mable. 1996. Molecular Systematics. Sinauer Associates. Sunderland. 655 pp.
- Huelsenbeck, J.P., F. Ronquist, R. Nielsen y J.P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314.
- Huxley, J.S. 1940. The new systematics. Clarendon Press. Oxford. 586 pp.
- Huxley, J.S. 1959. Clades and grades. En: *Function and taxonomic importance*, (Cain, A.J., ed.) Systematics Association. London. 21–22 pp.
- Ilie, L., F. Fazayeli y S. Ilie. 2011. HiTEC: accurate error correction in high-throughput sequencing data. *Bioinformatics* 27:295–302.

- Johnson, N.F. 2012. A collaborative, integrated and electronic future for taxonomy. *Invertebr. Syst.* 25:471–475.
- Katoh, K., K. Misawa, K. Kuma y T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30(14): 3059–3066.
- Keightley, P.D. y T. Johnson. 2004. MCALIGN: stochastic alignment of noncoding DNA sequences based on an evolutionary model of sequence evolution. *Genome Res.* 14:442–450.
- Kim, K.M., J.H. Park, D. Bhattacharya y H.S. Yoon. 2014. Applications of next-generation sequencing to unravelling the evolutionary history of algae. *Int. J. Syst. Evol. Micr.* 64: 333-345.
- Kircher, M. y J. Kelso. 2010. High-throughput DNA sequencing--concepts and limitations. *BioEssays* 32:524–536.
- Kubatko, L.S., B.C. Carstens y L. Knowles. 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics* 25:971–73.
- Larkin, M.A. G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
- Lemmon, E.M., y A.R. Lemmon. 2013. High-Throughput Genomic Data in Systematics and Phylogenetics. *Annu. Rev. Ecol. Syst.* 44:99–121.
- Linneo, C. 1753. *Species plantarum: exhibentes plantas rite cognitatas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas.* Holmiae-Impensis Laurentii Salvii. Estocolmo. 560 pp.
- Llorente-Bouquets, J. 1990. *La búsqueda del método natural.* Fondo de Cultura Económica. México. 155 pp.
- Lou, D.I., J.A. Hussmann, R.M. McBee, A. Acevedo, R. Andino, W.H. Press y S.L. Sawyer. 2013. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. *Proc. Natl. Acad. Sci. U.S.A.* 110:19872–19877.
- Mayr, E. 1969. *Principles of systematic zoology.* McGraw-Hill. New York. 428 pp.
- Mayr, E. 2001. The philosophical foundations of Darwinism. *Proc. Am. Philos. Soc.* 145: 488-495.
- Miklo's, I., G.A. Lunter y I. Holmes. 2004. A “long indel” model for evolutionary sequence alignment. *Mol. Biol. Evol.* 21:529–540.

- Misof, B., S. Liu, K. Meusemann, R.S. Peters, A. Donath y C. Mayer, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346:763–767.
- Morgan, T.H. 1916. *A Critique of the Theory of Evolution*. Princeton University Press. New Jersey. 197 pp.
- Morgenstern, B. 2004. DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Res.* 32:W33–W36.
- Morrone, J.J. 2000. *El lenguaje de la cladística*. UNAM-Facultad de Ciencias. México. 109 pp.
- Muller, H.J. 1927. Artificial transmutation of the gene. *Science* 66: 84–87.
- Müller, K. 2006. Incorporating information from length-mutational events into phylogenetic analysis. *Mol. Phylogenet. Evol.* 38:667-676.
- Nater, A., R. Burri, T. Kawakami, L. Smeds y H. Ellegren. 2015. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Syst. Biol.* 64:1000–1017.
- Nixon, C.K. y H. Ochoterena. 2001. Taxonomía tradicional, cladística y construcción de hipótesis filogenéticas. En: *Enfoques contemporáneos para el estudio de la biodiversidad*. (Hernández, H.M., A.N. García-Aldrete, F. Álvarez y M. Ulloa. eds.) UNAM-Instituto de Biología. México. 15-37 pp.
- Notredame, C., D.G. Higgins y J. Heringa. 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 302:205–217.
- Nystedt, B., N.R. Street, A. Wetterbom, A. Zuccolo, Y.C. Lin, D.G. Scofield, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497:579–584.
- O’Leary, M.A., J.I. Bloch, J.J. Flynn, T.J. Gaudin, A. Giallombardo, N.P. Giannini, et al. 2013. The Placental Mammal Ancestor and the Post-K-Pg Radiation of Placentals. *Science* 339:662–667.
- Palmer, J.D. y D. Zamir. 1982. Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. *Proc. Natl. Acad. Sci. U.S.A.* 79: 5006–5010.
- Parks, M., R. Cronn y A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology*, 7:1–17.
- Salzberg, S.L., O. White, J. Peterson, y J.A. Eisen. 2001. Microbial Genes in the Human Genome: Lateral Transfer or Gene Loss ?. *Science* 292: 1903–1906.
- Sanderson, M.J. y M.J. Donoghue. 1989. Patterns of variation in levels of homoplasy. *Evolution* 43(8): 1781-1795.

- Sarich, V.M. y A.C. Wilson. 1967. Immunological time scale for hominid evolution. *Science* 158: 1200–1203.
- Scornavacca, C., V. Berry, V. Lefort, E.J.P. Douzery y V. Ranwez. 2008. PhySIC_IST: cleaning source trees to infer more informative supertrees. *BMC Bioinformatics* 9: 413.
- Scotland, R.W., R.G. Olmstead, J.R. Bennett. 2003. Phylogeny reconstruction: the role of morphology. *Syst. Biol.* 52: 539–548.
- Sibley, C.G., J.E. Ahlquist y B.L. Monroe. 1988. A classification of the living birds of the world based on DNA-DNA hybridization studies. *Auk* 105:409–423.
- Sicheritz-Pontén, T. y S.G.E. Andersson. 2001. A phylogenomic approach to microbial evolution. *Nucleic Acids Res.* 29:545–552.
- Simmons, M.P. y H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analysis. *Syst. Biol.* 49:369–381.
- Sokal, R.R. y P.H.A. Sneath. 1963. *Principles of numerical taxonomy*. WH Freeman and Company. San Francisco. 359 pp.
- Stamatakis, A., T. Ludwig y H. Meier. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–63.
- Steele, R.P. y C.J. Pires. 2011. Biodiversity assessment: state of the art techniques in phylogenomics and species identification. *Am. J. Bot.* 98:415-425.
- Stevens, P.F. 1984. Metaphors and typology in the development of botanical systematics 1690-1960, or the art of putting new wine in old bottles. *Taxon* 33:169- 211.
- Straub, S.C., M. Fishbein, T. Livshultz, Z. Foster, M. Parks, R. Weitemier, R.C. Cron y A. Liston. 2011. Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 2:211-232.
- Strickler, S.R., A. Bombarely, L.A. Mueller. 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species. *Am. J. Bot.* 99: 257-266.
- Sturtevant, A.H. 2001. *A History of Genetics*. Cold Spring Harbor Laboratory Press. New York. 174 pp.
- Swofford, D.L., G.J. Olsen, P.J. Waddell y D.M. Hillis. 1996. Phylogenetic inference. P. 407–514. En: *Molecular systematics*, (Hillis D.M., C. Moritz y B.K. Mable, eds.) Sinauer Associates. Sunderland.

- Türktaş, M., K. Yücebilgili-Kurtoğlu, G. Dorado, B. Zhang, P. Hernandez y T. Ünver. 2015. Sequencing of plant genomes – a review. *Turk. J. Agric. For.* 39:361–376.
- Varki, A., y T.K. Altheide. 2005. Comparing the human and chimpanzee genomes: Searching for needles in a haystack. *Genome Res.* 15:1746–1758.
- Venter, J.C., M.D. Adams, E.W. Myers, et al. 2001. The sequence of the human genome. *Science* 291:1304–1351.
- Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van-de-Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper y M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407-4414.
- Wiens, J.J. 2004. The role of morphological data in phylogeny reconstruction. *Syst. Biol.* 53(4): 653-661.
- Woese, C.R. y G.E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* 74: 5088-5090.
- Woese, C.R. 1987. Bacterial evolution. *Microbiol. Rev.* 51:221–271.
- Woese, C.R., O. Kandler y M.L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87: 4576-4579.
- Wright, S. 1930 The Genetical Theory of Natural Selection: a review. *J. Hered.* 21:340-356.
- Young, N.D. y J. Healy. 2003. GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics.* 4:1-6.
- Zhang, J. 2003. Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, 18(6): 292–298.
- Zhong, B., O. Deusch, V.V. Goremykin, D. Penny, P.J. Biggs, R.A. Atherton, et al. 2011. Systematic error in seed plant phylogenomics. *Genome Biol. Evol.* 3:1340–1348.
- Zietkiewicz, E., A. Rafalski y D. Labuda. 1994. Genome fingerprinting by simple sequence repeat (SRR)-anchored polymerase chain reaction. *Genomics* 20: 176-183.
- Zimmermann, W. 1931. Arbeitsweise der botanischen Phylogenetik und anderer Gruppierungswissenschaften. P. 941-1053. En: Abderhalden E, ed. *Handbuch der biologischen Arbeitsmethoden.* Urban & Schwarzenberg. Berlin.