



RIDAA
Repositorio Institucional
Digital de Acceso Abierto de la
Universidad Nacional de Quilmes



Universidad
Nacional
de Quilmes

Roffé, Ariel

El estatus fáctico de la cladística : aportes desde una reconstrucción estructuralista



Esta obra está bajo una Licencia Creative Commons Argentina.
Atribución - No Comercial - Sin Obra Derivada 2.5
<https://creativecommons.org/licenses/by-nc-nd/2.5/ar/>

Documento descargado de RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes de la Universidad Nacional de Quilmes

Cita recomendada:

Roffé, A. (2020). *El estatus fáctico de la cladística: aportes desde una reconstrucción estructuralista*. *Metatheoria*, 11(1), 53-72. Disponible en RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes <http://ridaa.unq.edu.ar/handle/20.500.11807/2577>

Puede encontrar éste y otros documentos en: <https://ridaa.unq.edu.ar>

El estatus fáctico de la cladística: aportes desde una reconstrucción estructuralista*

The Empirical Status of Cladistics: Contributions from a Structuralist Reconstruction

Ariel Jonathan Roffé^{†‡}

Resumen

En el presente trabajo se analiza la controversia existente en el marco de la sistemática biológica en torno al estatus de la cladística. La utilización del método de parsimonia para la reconstrucción filogenética ha sido defendida apelando a un principio metodológico de simplicidad, así como a principios fácticos externos a la propia sistemática. Se propone aquí un nuevo tipo de enfoque, que consiste en considerarla una teoría fáctica, siendo su aplicación justificada por su éxito empírico. Para defender este punto se brindará una reconstrucción formal estructuralista de la cladística, que ayudará a clarificar cuáles son sus afirmaciones fácticas centrales. A su vez, se muestra cómo estos principios fácticos han sido utilizados para contrastar a tal teoría utilizando filogenias experimentales.

Palabras clave: cladística - estructuralismo metateórico - parsimonia

Abstract

The present work analyzes the controversy within biological systematics regarding the status of cladistics. The use of the parsimony method for phylogenetic reconstruction has been defended by appealing to a methodological principle of simplicity, as well as to empirical principles that external to systematics. I propose new kind of approach, which consists in considering it an empirical theory, thus justifying its application by its empirical success. To defend this point, a formal structuralist reconstruction of cladistics will be provided, which will help clarify what its central empirical statements are. At the same time, I show how these principles have been used to test that theory using experimental phylogenies.

Keywords: cladistics - metatheoretical structuralism - parsimony

* Recibido: 30 de julio de 2020. Aceptado con revisiones: 15 de septiembre de 2020.

[†] Centro de Filosofía e Historia de la Ciencia (CEFHIC), Universidad Nacional de Quilmes (UNQ)/Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)/Universidad de Buenos Aires (UBA), Argentina. Para contactar al autor, por favor, escribir a: ariroffe@hotmail.com.

[‡] El presente trabajo ha sido financiado por los proyectos PUNQ 1401/15 (Universidad Nacional de Quilmes, Argentina), UNTREF 32/15 255 (Universidad Nacional Tres de Febrero, Argentina), PICT-2014-1741 (ANPCyT, Argentina), PICT 2018-3454 (ANPCyT, Argentina) y UBACyT 20020170200106BA (Universidad de Buenos Aires, Argentina).

Metatheoria 11(1)(2020): 53-72. ISSN 1853-2322. eISSN 1853-2330.

© Editorial de la Universidad Nacional de Tres de Febrero.

© Editorial de la Universidad Nacional de Quilmes.

Publicado en la República Argentina.

1. Introducción

Una de las grandes teorías propuestas por Darwin afirma que todos los seres vivos actualmente comparten un único ancestro común. Con esto, Darwin no pretendía explicar la existencia de adaptaciones (el *explanandum* de la teoría de la selección natural, ver Ginnobili 2009, 2018) sino la de rasgos homólogos, esto es, de rasgos estructuralmente similares en especies distintas (a veces muy lejanas), que pueden incluso tener funciones distintas. Así, según Darwin, la historia de la vida en la tierra puede representarse de manera arbórea.

Recién a mediados del siglo XX se desarrollaron herramientas teóricas y computacionales que permitieron desentrañar detalladamente la forma de ese árbol de una manera precisa y sistemática. En particular, dentro de esos métodos de inferencia filogenética, tuvo particular éxito el programa cladista (iniciado por el entomólogo alemán Willi Hennig). Dicho brevemente (ya que lo desarrollaré con mayor precisión en la sección siguiente), el método cladista de inferencia filogenética permite obtener un árbol a partir de un conjunto de especies y una distribución de homologías eligiendo, de entre todos los árboles posibles, a aquel que requiere postular la menor cantidad de cambios evolutivos posible. De ese modo, a la base de este método parece haber una suposición de que la evolución actúa de manera parsimoniosa, prefiriendo conservar a los caracteres en su estado actual e introduciendo la menor cantidad de cambios posible. Así, se explicaría la mayor cantidad de homologías posible a partir de la herencia del rasgo desde ancestros comunes, minimizando la cantidad de convergencias. Por estos motivos, a dicho método se lo conoce comúnmente como el *método de parsimonia*.

Una pregunta metateórica recurrente entre los/as filósofos/as y biólogos/as que reflexionan sobre este método es, precisamente, cómo justificar su aplicación. En particular, la pregunta es por qué deberíamos asumir que la evolución es parsimoniosa, en lugar de creer que las convergencias son más frecuentes en la historia evolutiva. Las respuestas que se han dado a esta pregunta van desde considerar que su justificación descansa sobre un principio *metodológico* de simplicidad/parsimonia, hasta hacerla reposar sobre principios *empíricos* pero externos a la propia teoría (p.e. lo que sabemos acerca de la evolución desde otros ámbitos, como ser, lo que la genética molecular nos dice acerca de las tasas de sustitución).

En el presente trabajo me propongo brindar un tipo nuevo de justificación de la cladística. Defenderé, en primer lugar, que la cladística puede ser pensada como una *teoría* empírica y no meramente como un método (aunque la componente metodológica existe, ya que las teorías suelen implicar metodologías en sus condiciones de aplicación). En consecuencia, sostendré, en segundo lugar, que su aplicabilidad (a su dominio pretendido) se justifica del mismo modo que el de cualquier otra teoría empírica: a partir de su éxito empírico. Un factor que complica la atribución de éxito empírico a la cladística es que, en la mayor parte de sus aplicaciones, los eventos evolutivos relevantes se encuentran en el pasado remoto, lo cual vuelve imposible una contrastación. Sin embargo, mostraré que existen casos históricos de filogenias conocidas independientemente (incluso generadas en el laboratorio) en donde tales contrastaciones fueron llevadas a cabo.

Para lograr todo esto presentaré una reconstrucción formal de la cladística (una versión previa, no publicada, puede encontrarse en mi tesis doctoral, Roffé 2020b), realizada utilizando las herramientas del estructuralismo metateórico (Balzer, Moulines & Sneed [1987] 2012). Esto aportará motivos para defender la primera de mis dos tesis (que la cladística es una teoría), al mostrar que es un cuerpo de conocimiento estructurado de modo similar a otras teorías. Por supuesto que el hecho que algo sea representable con una reconstrucción estructuralista no lo vuelve inmediatamente una teoría empírica, pero en conjunción con las consideraciones acerca de la naturaleza fáctica de sus leyes, brindará un argumento fuerte para defender tanto su carácter de teoría como su justificación fáctica. Cabe notar, además, que la reconstrucción tiene valor por sí misma, y que puede ser aplicada a otras controversias en torno a la cladística (refiero, nuevamente, a la tesis doctoral para algunos ejemplos).

En adelante, procederé como sigue. En la sección 2 se hace una introducción informal, primero al origen común darwiniano (que es importante para comprender el *explanandum* y algunas de las

asunciones de la cladística) y luego a la cladística propiamente dicha. La sección 3 presenta mi reconstrucción estructuralista de dicha teoría. La sección 4 retorna al punto de la justificación de la cladística a partir de la reconstrucción. Por último, en la sección 5, se extraen algunas conclusiones.

2. Presentación informal de la teoría

2.1. El origen común darwiniano

Si bien la selección natural es la innovación conceptual de Darwin más conocida, y aquella con la que más frecuentemente se lo asocia, no fue la única que este autor propuso, ni la que consideraba la más importante. En efecto, Darwin también propuso que toda la rica diversidad de los seres vivos que observamos en la actualidad deriva de un único ancestro común. De ese modo, la historia de la vida en la tierra podría representarse por medio de un diagrama con forma de árbol: partiendo de una única raíz (una única especie ancestral), diversas ramas fueron surgiendo en diferentes direcciones (i.e. la misma especie original dio lugar a diferentes especies descendientes), las cuales a su vez continuaron ramificándose, etc.

Para comprender los motivos por los que Darwin postula el origen común es conveniente comenzar exponiendo lo que la tesis pretende explicar, es decir, aquello de lo que pretende dar cuenta. Lo que el origen común explica no son los rasgos adaptativos, sino los rasgos homólogos. En palabras de Darwin:

Hemos visto que los miembros de una misma clase, independientemente de sus costumbres, se parecen en el plan general de su organización. Esta semejanza se expresa frecuentemente por el término *unidad de tipo* o diciendo que las diversas partes y órganos son homólogos en las distintas especies de la clase. Todo el asunto se comprende con la denominación general de Morfología. Es ésta una de las partes más interesantes de la historia natural, y casi puede decirse que es su verdadera esencia. ¿Qué puede haber más curioso que el que la mano del hombre, hecha para coger; la del topo, hecha para minar; la pata del caballo, la aleta de la marsopa y el ala del murciélago, estén todas construidas según el mismo patrón y encierren huesos semejantes en las mismas posiciones relativas? (Darwin [1872] 1992, pp. 569-570, énfasis en original).

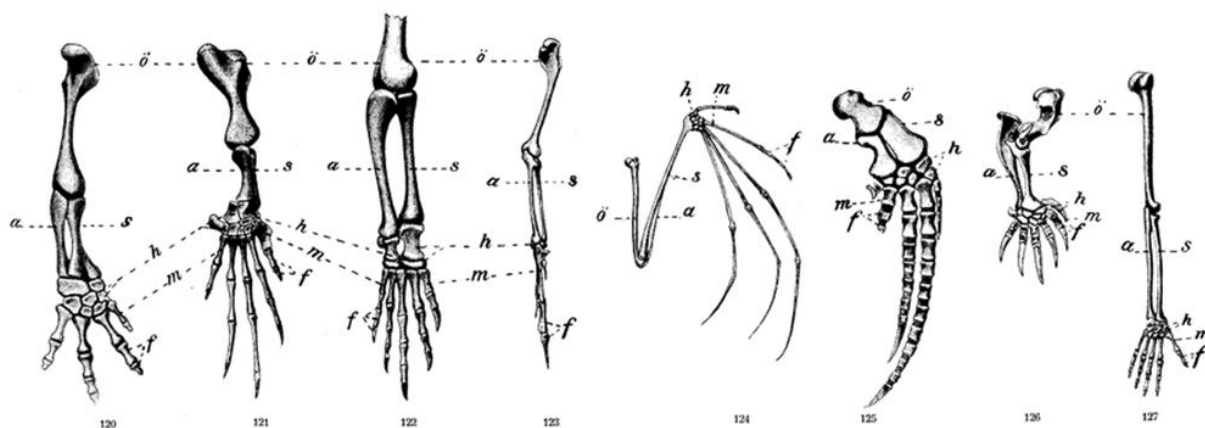


Figura 1. Homología en las extremidades de distintas especies de vertebrados: salamandra (120), tortuga marina (121), cocodrilo (122), ave (123), murciélago (124), ballena (125), topo (126) y humano (127). Fuente: https://commons.wikimedia.org/wiki/File:Arm_skeleton_comparative_NF_0102.5-2.png

El concepto de homología y los métodos para reconocerlas habían sido propuestos previamente por Geoffroy Saint-Hilaire (1830, quien usaba el término “analogías filosóficas” para referirse a ellas; para más sobre la obra de Saint-Hilaire, ver Caponi 2015) y luego teorizado por Richard Owen (1848, 1849), entre otros. Para Owen, la presencia de este tipo de similitudes estructurales (p.e. en las extremidades de los vertebrados) se explicaba apelando a que estaban construidas a partir del mismo *arquetipo*. Lo que Darwin hizo fue encarnar al arquetipo de Owen: este sería un ancestro. De ese modo, la homología entre la extremidad de los vertebrados se explicaría no por compartir algún plano de construcción abstracto,

sino porque el ancestro de todos ellos poseía una versión de tal extremidad. Según las necesidades impuestas por el ambiente la extremidad se habría ido modificando de distinto modo en cada rama del árbol evolutivo, preservando, sin embargo, la estructura general (Darwin [1872] 1992, pp. 571-572; ver Blanco 2012, para un tratamiento estructuralista de la teoría del origen común de Darwin; para discusiones sobre el concepto de homología en el marco de la cladística, ver Brady 1985, Pearson 2010, 2018, Roffé 2020a, Roffé, Ginnobili & Blanco 2018, entre muchos otros).¹

El reconocimiento de que el origen común explica la presencia de homologías tenía, para Darwin, profundas implicancias sobre la taxonomía. Las implicancias apuntaban, según este pensador, no tanto al modo de construir una taxonomía, sino más bien al modo de entenderla o interpretarla. Lo que los naturalistas habían estado haciendo, sin saberlo, era agrupar a los organismos con un criterio genealógico (Darwin [1872] 1992, pp. 551-552).

Sin embargo, la conexión entre el reconocimiento de la ancestría común y la idea de una taxonomía basada en las relaciones filogenéticas (en lugar de otros factores, como la semejanza morfológica) no es una de implicación lógica. En el período en el que Darwin publicó las sucesivas ediciones de *El origen* no se contaba aún con un método sistemático para identificar las relaciones filogenéticas a partir de las distribuciones “observadas” de homologías. De ese modo, algunos de sus colegas (p.e. Huxley) aceptaron la primera propuesta (la ancestría común) pero rechazaron la segunda (la de basar la taxonomía en las relaciones de ancestría), sobre la base de que la descendencia “no era cognoscible con suficiente certeza como para permitir que ese tipo de consideraciones tengan lugar en algo tan importante como la taxonomía” (Hull 1988, p. 98, traducción propia).²

Un factor que vuelve problemática la reconstrucción del árbol de la vida a partir de las distribuciones de homologías es que éstas raramente son perfectamente compatibles entre sí en casos interesantes. Por ejemplo, considérese el siguiente fragmento de una matriz de datos de arañas dada en Ramírez (2003):

Taxón / Carácter	Fila de ojos anterior	Mancha negra longitudinal ventral en el abdomen	Denticulos en la porción prolateral de C2
(1) <i>Monapia tandil</i>	Recta	Ausente	Presente
(2) <i>Monapia fierro</i>	Recta	Presente	Presente
(3) <i>Monapia carolina</i>	Recta	Presente	Ausente
(4) <i>Monapia dilaticollis</i>	Recurvada	Ausente	Ausente

Tabla 1. Fragmento de una matriz de datos real.

Nótese que el segundo carácter entra en conflicto con el tercero, ya que sugiere agrupar a (2) con (3) en lugar de con (1). Es decir, (asumiendo que las ausencias son estados ancestrales en ambos caracteres) el segundo carácter sugiere que *Monapia fierro* y *Monapia carolina* comparten un ancestro que no es un ancestro de *Monapia tandil*, mientras que el tercero sugiere que *Monapia fierro* y *Monapia tandil* comparten un ancestro que no es un ancestro de *Monapia carolina*. La pregunta se torna, entonces, en cómo acomodar la información incompatible provista por los distintos caracteres a fin de establecer cuál es el patrón de diversificación que realmente ocurrió.

¹ Cabe aclarar sin embargo que, dada la evidencia con la que contaba, Darwin fue prudente en su postulación de un único origen común. Lo limitó con seguridad a cuatro o cinco especies dentro de los animales y a un número similar para las plantas, y solo especuló acerca de un único origen común entre todos los organismos (Darwin, [1872] 1992, pp. 631-632).

² Hull (1988, p. 373) relata las posteriores discusiones, completamente anacrónicas, acerca de si debía considerarse a Darwin un cladista o un evolucionista en taxonomía.

2.2. La cladística

La aparición de criterios y herramientas sistemáticas que permitieran resolver de manera adecuada estos conflictos tuvo que esperar a la segunda mitad del siglo XX con el surgimiento de la cladística (entre otras metodologías de reconstrucción filogenética). El nacimiento del programa cladista puede datarse en 1950 con la publicación del libro *Grundzüge einer Theorie der phylogenetischen Systematik* del entomólogo alemán Willi Hennig. En tal obra, Hennig propuso algunos de los pilares de lo que sería el programa de investigación cladista.³ Entre esos pilares figura la idea de representar únicamente la relación de hermandad entre grupos en las clasificaciones, así como un método para inferir esa relación. Como puede apreciarse, la cladística como programa de investigación o paradigma en sistemática involucra tesis o propuestas en diferentes planos. En particular, deben distinguirse las tesis o propuestas en taxonomía de los métodos de inferencia filogenética (que serán mi foco de atención en adelante). Esta distinción entre la propuesta taxonómica y el método de inferencia filogenética es trazada por algunos de los rivales de la escuela cladista en taxonomía, ya que en ocasiones desean aceptar al método de inferencia filogenética como legítimo, pero rechazar que las clasificaciones deban estar basadas únicamente en información extraída a partir de sus productos (p.e. ver Ashlock 1974, Aubert 2015, pp. 2-3).⁴

La solución de la cladística al problema de la información incompatible mencionado anteriormente consiste en tomar muchos caracteres en consideración (cientos o miles) y preservar al agrupamiento (i.e. inferir como la historia evolutiva correcta a aquella) que mejor es soportada por la evidencia total. Por esto se entiende preservar el árbol que requiere postular el menor número de cambios entre estados en total, considerando todos los caracteres. Veamos este punto con mayor detalle. Aunque no del todo completa, la presentación que sigue bastará para mis propósitos. Para presentaciones históricas más completas de la teoría pueden verse Hull (1988) y Rieppel (2016). En cuanto a su contenido, pueden consultarse diversos manuales en la literatura, como ser, Kitching *et al.* (1998), Nelson y Platnick (1981), Wiley y Lieberman (2011), entre otros.

Dado un conjunto finito de especies hay también un número finito de posibles patrones de diversificación dicotómicos. Estos patrones son usualmente representados como árboles, llamados cladogramas, en donde las ramas representan la relación de ancestría inmediata y los nodos internos representan ancestros hipotéticos. En los cladogramas los nodos internos se dividen siempre en dos, y no ocurre nunca una mezcla de linajes (i.e. dos nodos internos que apuntan al mismo nodo posterior, como podría ocurrir en casos de especiación por hibridación). Por ejemplo, la figura siguiente ilustra a 6 de los 15 posibles árboles distintos para 4 taxa (t_1 - t_4):

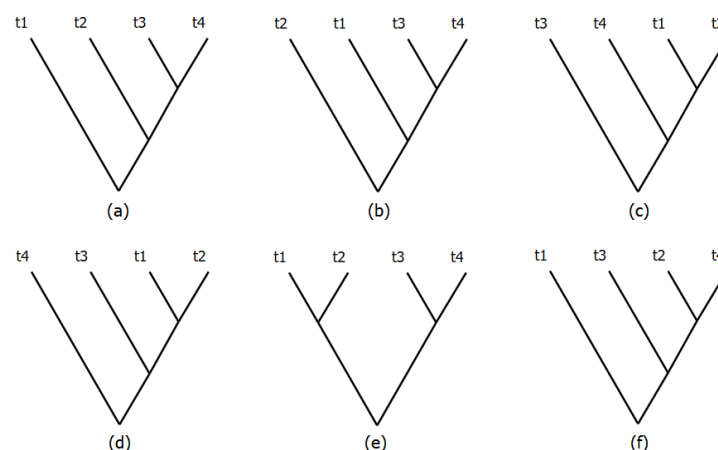


Figura 2. Seis posibles cladogramas para cuatro taxones.

³ Cabe notar que Hennig nunca utilizó el término “cladista”, sino que tal denominación es posterior y se debe a Mayr (1965), quien era miembro de una escuela rival. También es notable que el texto de Hennig generó, originalmente, muy poca repercusión, y que su metodología se extendió recién a mediados de los años sesenta de la mano del ictiólogo estadounidense Gareth Nelson y su grupo de investigación del *American Museum of Natural History* (ver Hull 1988, para más sobre los inicios del programa cladista).

⁴ Pese a ser conceptualmente independientes, Hull (1988, p. 256) ilustra cómo el hecho de proponer un único método, claro y aplicable, para reconocer las relaciones filogenéticas entre las especies fue instrumental para la victoria del programa cladista en taxonomía.

Cabe notar que el número de cladogramas posibles crece de manera extremadamente rápida con el aumento del número de taxa. Por ejemplo, con 5 taxa el número de cladogramas asciende ya a 106, con 9 taxa a 2.027.025, y con 20 taxa el número se eleva a 8.200.794.532.637.891.559.375 cladogramas (una derivación de la fórmula para realizar este cálculo puede encontrarse en Felsenstein 1978).⁵ El análisis cladista debe elegir a uno de todos aquellos árboles como aquel que representa a la historia filogenética real. Es decir, el resultado de la aplicación es una propuesta (hipótesis) acerca de la historia. Esta elección se realiza comparando a los árboles según un criterio de optimalidad.

Para elegir un árbol, cada carácter de la matriz de datos es mapeado en cada árbol para ver cuántos cambios evolutivos deben postularse para explicar su distribución observada. Más precisamente, optimizar un carácter sobre un árbol significa asignar sus estados a los nodos internos del árbol de modo tal de minimizar el número de transformaciones entre estados. Por ejemplo, considérese la siguiente matriz de datos (i.e. distribución de homologías C_1 - C_3 para los taxones t_1 - t_4):

	C_1	C_2	C_3
t_1	0	1	0
t_2	0	0	0
t_3	1	0	0
t_4	1	1	1

Tabla 2. Matriz de datos de ejemplo, que contiene 4 taxa y 3 caracteres. Los números 0 y 1 de las celdas representarían estados alternativos para los caracteres (véase el ejemplo de la tabla 2).

Tómese el árbol (a) de la figura 2. En él, los estados del carácter C_1 pueden asignarse a los nodos internos de 8 modos (para los nodos terminales, i.e., los taxa bajo estudio, el estado del carácter está dado en la matriz de datos, de modo que son iguales en todas las asignaciones). La siguiente figura muestra 3 de las 8 asignaciones posibles:

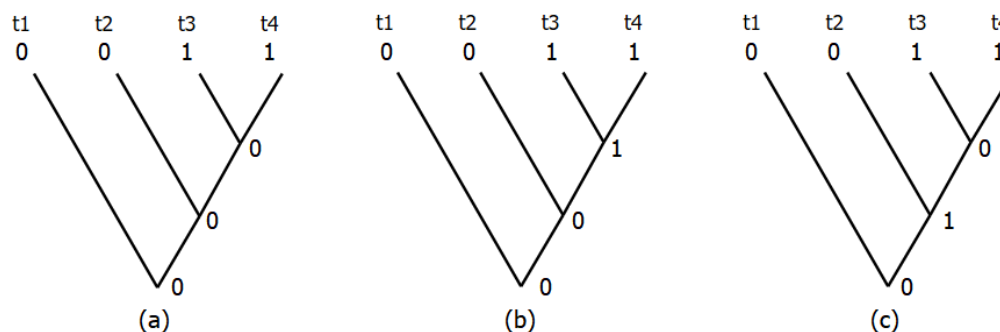


Figura 3. Posibles asignaciones del carácter C_1 en el árbol (a).

Nótese que la asignación (b) requiere postular un solo cambio evolutivo para acomodar al carácter C_1 en este árbol, mientras que puede chequearse fácilmente que el resto de las asignaciones requiere al menos dos (por ejemplo, la asignación (a) requiere dos cambios, desde al ancestro de t_3 y t_4 hacia cada uno de sus descendientes). De ese modo, queda establecido que el carácter C_1 puede ser acomodado en este árbol postulando un solo cambio.

El *largo* del árbol (a) es la sumatoria del mínimo número de cambios necesario para acomodar cada uno de los caracteres de la matriz a ese árbol. Así, si se realiza este mismo procedimiento con el resto de los caracteres de la matriz, puede verse que el árbol (a) posee un largo de 4 (el carácter C_2 se acomoda con un mínimo de 2 cambios, mientras que C_3 con 1 cambio). La cladística busca encontrar al árbol o

⁵ Por este motivo, los análisis reales suelen utilizar métodos “heurísticos”, que dan una solución aproximada al problema de encontrar el árbol óptimo, véase a continuación para una descripción del criterio de optimalidad.

cladograma de menor largo (i.e. aquel que requiere, en total, el menor número de cambios evolutivos para explicar las distribuciones observadas de caracteres).

Si hay más de un árbol con el mismo largo mínimo, la cladística no se pronuncia respecto de cuál de todos ellos representa al patrón de diversificación real de las especies. En cambio, solo afirma que ese patrón debe estar entre los árboles de largo mínimo. Aun así, de algunos conjuntos de árboles óptimos puede extraerse información relevante. Por ejemplo, si los árboles (a), (b) y (e) fuesen los únicos tres cladogramas óptimos (i.e. de largo mínimo)⁶ entonces, si bien no sabríamos la forma exacta del patrón de diversificación real, sabríamos que el grupo (t_3 , t_4) estaría presente en él, y por tanto que ese es un grupo monofilético (que incluye a *todos* y *solo* los descendientes de algún ancestro común). En cambio, se suspendería el juicio acerca de las relaciones entre ese grupo y los taxones t_1 y t_2 .⁷

Existen muchas otras complicaciones adicionales en las que no entraré aquí. Por mencionar un ejemplo, hasta el momento se asumió que toda transformación de un estado a otro en cualquier carácter tiene el mismo costo o peso (i.e. toda transformación suma 1 al largo del cladograma). Sin embargo, este no tiene por qué ser el caso. En ocasiones, se desea que una transformación en un carácter sea más costosa que una transformación en otro carácter, p.e. que sea más costoso ganar o perder una estructura morfológica compleja que una mancha de color superficial. Por otra parte, incluso dentro de un carácter, algunas transformaciones pueden establecerse como más costosas que otras, p.e. puede quererse que una transversión cueste más que una transición en un carácter de ADN, y/o que una inserción o delección cueste más que una sustitución (reflejando el hecho de que las primeras ocurren con menor frecuencia en la naturaleza que las segundas). El modo usual de lidiar con esto es tener una matriz de costo asociada a cada carácter, que indica el costo (en términos de cuánto añaden al largo del cladograma) de cada cambio de un estado a otro.

Esta metodología brinda un criterio para resolver los conflictos entre patrones anidados incompatibles de homologías, como el que se ilustró anteriormente. Sencillamente se prefieren aquellos patrones que están confirmados por la mayor cantidad de otros patrones. A este método de inferencia filogenética se lo ha llamado el método de *parsimonia* porque asume (o *prima facie* parece asumir, esto será discutido más adelante) que la explicación más simple (la que requiere postular menos eventos evolutivos de transformación de caracteres) es la correcta. Retornaremos al punto acerca la justificación de la utilización de métodos cladísticos para la reconstrucción filogenética en la sección 4. Como se dijo en la introducción, ello se hará a bajo la lente de una reconstrucción formal, que se presenta a continuación en la sección 3.

3. Una reconstrucción estructuralista de la cladística

Existe actualmente un campo de estudio dedicado al estudio matemático de la filogenética, en el que participan tanto biólogos como matemáticos (ver Brower 2000, Farris 1970, 1974, Martin, Blackburn & Wiley 2010, Semple & Steel 2003, Steel 2016, para algunos ejemplos). La presente reconstrucción construye sobre lo edificado en esa área. Sin embargo, la matematización de la cladística llevada a cabo allí fue elaborada sin un aparato conceptual metateórico explícito por detrás, que permita entender a la cladística y a diversas características suyas como análogos a otras teorías en otros ámbitos de la ciencia.

Por otra parte, los objetivos de esos autores tienen usualmente más que ver con probar teoremas y establecer resultados de complejidad algorítmica que con precisar el lenguaje para identificar las aserciones fácticas de la teoría y discutir cuestiones conceptuales/metateóricas que la rodean. Los objetivos de la formalización en este trabajo son más cercanos a los segundos que a los primeros (aunque, por supuesto, algunos teoremas podrán probarse a partir de los axiomas identificados). Por ello, para la reconstrucción dada en esta sección se utilizarán las herramientas brindadas por el estructuralismo

⁶ Esto puede no ser posible matemáticamente, ya que los tres son enraizamientos del mismo grafo no dirigido. No entraré en ese tipo de complicaciones aquí, ya que no afecta el punto conceptual ilustrado.

⁷ Un modo de representar a la información compartida por un conjunto de árboles es por medio del llamado *árbol de consenso estricto*, ver Wiley & Lieberman (2011), pp. 193-194.

metateórico (Balzer, Moulines & Sneed [1987] 2012), la cual ya fue aplicada exitosamente tanto a otras teorías biológicas (p.e. Balzer & Lorenzano 2000, Bernabé 2018, Díaz & Lorenzano 2017, Ginnobili 2010, Lorenzano 2002, Méndez & Casanueva 2006, entre muchos otros) como de otras áreas de la ciencia (p.e. Abreu 2014, Alleva, Díez & Federico 2017, Caamaño 2009, Carman 2015, Falguera & Donato-Rodríguez 2016, Gonzalo & Balzer 2012, Lastiri 2012, C. Lorenzano 2002, Sneed 1971, entre muchos otros). Esto permitirá, en la sección siguiente, resaltar diversas similitudes entre la cladística y esas otras teorías, así como discutir cuestiones conceptuales/metateóricas con mayor claridad.

Por último, cabe mencionar que reconstruiré aquí sólo los axiomas impropios y las leyes fundamentales (modelos parciales y modelos) de la teoría, ya que ello será suficiente para mis propósitos, dejando el tratamiento de las condiciones de ligadura, las especializaciones y los modelos parciales para otra ocasión (puede verse Roffé 2020b para ello; sí tendré algo que decir sobre la teoriedad en la sección siguiente).

3.1. Axiomas impropios

En esta subsección se darán los axiomas impropios de la reconstrucción de la cladística (**CLAD** en adelante). Lo haré primero de manera más didáctica pero menos ordenada y sistemática, brindando explicaciones informales de las fórmulas dadas. Al final de la subsección se da una presentación estructuralista más estándar.

El dominio de aplicación de la teoría (el conjunto de taxa cuyas relaciones de cercanía filogenética se desea inferir) puede ser caracterizado formalmente a partir de un conjunto T :

$T (= \{t_1, t_2, t_3, \dots, t_n\})$ es un conjunto finito que contiene al menos tres elementos.

Los árboles filogenéticos o cladogramas, con todos los posibles patrones de diversificación entre estos taxa, estarán contenidos en otro conjunto A :

$A = \{A_i / A_i \text{ es un árbol filogenético}\}$

Cada árbol filogenético A_i será representado formalmente como un grafo (una versión de los grafos basados en nodos [*node-based trees*] de Martin *et al.* 2010). Habrá para ello un conjunto finito de nodos N , que incluye a los nodos terminales e internos —por lo que $T \subset N$. A su vez, llamaré I al conjunto de los nodos internos ($I \stackrel{\text{def}}{=} N - T$). Dado que todo cladograma para un mismo conjunto de terminales tiene el mismo número de nodos internos (ver Steel 2016, p. 10, lema 1.4) no será necesario definir un conjunto de nodos internos diferente para cada cladograma. De ese modo, cada cladograma $A_i \in A$ puede definirse del siguiente modo. A_i es un cladograma sii $A_i = \langle N, D(A_i) \rangle$, tal que:

$$(i) D(A_i) \subseteq N^2$$

$D(A_i)$ representa a la relación de descendencia entre nodos según A_i . $\langle n_1, n_2 \rangle \in D(A_i)$ significa que, según el cladograma A_i , n_2 es un descendiente inmediato de n_1 . Que esto sea una relación significa que todo A_i será un grafo *dirigido* sobre el conjunto de vértices N .

El modo de dibujar un cladograma A_i a partir de su especificación formal es el usual en teoría de grafos: se representará a los nodos como puntos, y a los ejes (la relación de descendencia entre dos nodos) como líneas que los unen. Seguiré la convención biológica de representar a los ejes como teniendo una dirección hacia arriba. Por ejemplo, el cladograma A_x

$$N = \{t_1, t_2, t_3, t_4, n_1, n_2, n_3\}$$

$$D(A_x) = \{\langle n_1, n_2 \rangle, \langle n_1, t_1 \rangle, \langle n_2, n_3 \rangle, \langle n_2, t_4 \rangle, \langle n_3, t_2 \rangle, \langle n_3, t_3 \rangle\}$$

puede dibujarse del siguiente modo:

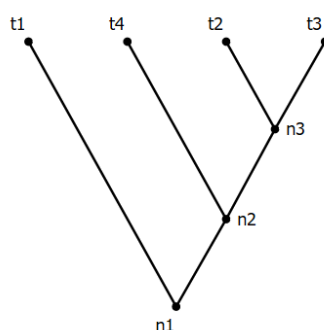


Figura 4. Ilustración del árbol filogenético (qua grafo) especificado anteriormente.

Un camino (en inglés *path*) entre dos nodos n_x y n_y es una secuencia de nodos $\langle n_x, n_{x1}, n_{x2}, \dots, n_y \rangle$ tal que para todo par de nodos consecutivos n_{xi} y n_{xi+1} , $\langle n_{xi}, n_{xi+1} \rangle \in D(A_i)$. Por ejemplo, en el árbol recién dado, $\langle n_1, n_2, t_4 \rangle$ es un camino. Con esta terminología en mente, es posible (y necesario) poner algunas restricciones adicionales sobre $D(A_i)$, ya que no todo grafo dirigido sobre N constituirá un cladograma (se dan aquí las versiones informales de estos axiomas impropios, las especificaciones formales se darán más adelante, en la presentación estructuralista):

- (ii) $D(A_i)$ es acíclica. Esto es, no hay ningún camino que salga desde un nodo y termine en el mismo nodo.
- (iii) $D(A_i)$ es enraizada. Esto es, existe un único nodo (llamado el nodo raíz del árbol) que no es el descendiente de ningún otro nodo, y tal que existe un camino desde él hacia todo otro nodo del árbol.
- (iv) Cada nodo interno tiene exactamente dos descendientes.
- (v) Los taxa T bajo consideración son nodos terminales (no tienen descendientes).
- (vi) $D(A_i)$ no tiene fusiones (*mergings*), i.e. dos nodos distintos no pueden tener el mismo descendiente (como ocurriría, por ejemplo, en la especiación por hibridación).

Por otro lado, nótese que lo que nos interesa es el patrón de diversificación de los taxa terminales. Los nodos internos pueden intercambiarse entre sí en un árbol sin modificar este patrón. Por ejemplo, los siguientes dos grafos representan el mismo escenario evolutivo:

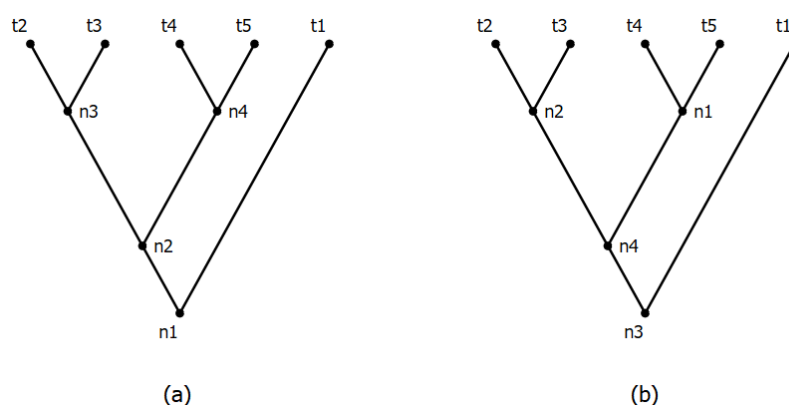


Figura 5. Ejemplo de dos cladogramas isomórficos.

Diremos, pues, que estos dos cladogramas son isomórficos (ver Roffé 2020b, pp. 57-58, Steel 2016, p. 9, para una caracterización formal de la noción de isomorfismo) y asumiremos en adelante que el conjunto A contiene al conjunto de los cladogramas *hasta el isomorfismo* (p.e. considerando como el mismo árbol a los dos de arriba).

El dominio de los caracteres constará de un conjunto finito y no-vacío $C = \{C_1, \dots, C_m\}$. Cada carácter C_i será a su vez un conjunto de estados —i.e. $C_i = \{c_{i,1}, c_{i,2}, \dots, c_{i,j}\}$. Adoptaré la convención usual de codificar

los estados para cada carácter por medio de números naturales. Otro punto importante es que los caracteres no comparten estados, es decir, se establecerá la restricción de que $\cap C = \emptyset$.

Adicionalmente, tendremos una serie de funciones de asignación para mapear estados de los caracteres a nodos. La función d_T asignará estados a los nodos terminales (está dada como input del análisis cladista), es decir: $d_T: T \times C \rightarrow UC$. En otras palabras, los conceptos T , C y d_T representan formalmente la matriz de datos.

Contaremos además con un conjunto $\delta = \{d_1, \dots, d_n\}$ de funciones de asignación completas, es decir, que mapean estados de caracteres a todo miembro de N , incluyendo a los nodos internos (i.e. para toda $d_i \in \delta$, $d_i: N \times C \rightarrow UC$). Para ambos tipos de funciones vale el requisito de que $d_T(t, C_i) \in C_i$ y $d_i(n, C_i) \in C_i$; es decir, dado un nodo y un carácter, ambas devuelven un estado de ese carácter y no de otro. δ es un conjunto de funciones y no una función particular ya que el estado de los nodos internos no está dado como un *input* en una aplicación de la teoría; en cambio, tal asignación puede realizarse de diversos modos. Sin embargo, todas las posibles asignaciones tienen que respetar la asignación (sí dada) de los nodos terminales. Por lo tanto, un requisito adicional para los miembros de δ es que:

$$\forall d_i \in \delta, \forall t \in T, \forall C_i \in C (d_i(t, C_i) = d_T(t, C_i))$$

Un ejemplo de una función d_T de asignación para un conjunto T de 4 taxa terminales ($T = \{t_1, t_2, t_3, t_4\}$), dados 2 caracteres con 2 estados cada uno ($C_1 = \{c_{1,1}, c_{1,2}\}$ y $C_2 = \{c_{2,1}, c_{2,2}\}$) sería la siguiente:

$$\begin{array}{ll} d_T(t_1, C_1) = c_{1,1} & d_T(t_1, C_2) = c_{2,1} \\ d_T(t_2, C_1) = c_{1,1} & d_T(t_2, C_2) = c_{2,2} \\ d_T(t_3, C_1) = c_{1,2} & d_T(t_3, C_2) = c_{2,2} \\ d_T(t_4, C_1) = c_{1,1} & d_T(t_4, C_2) = c_{2,2} \end{array}$$

Esta información representa formalmente a la siguiente matriz de datos:

Taxón/Carácter	C_1	C_2
t_1	1	1
t_2	1	2
t_3	2	2
t_4	1	2

Tabla 3. Matriz de datos representada por la función d_T de arriba.

Una función de asignación completa d_x basada en d_T podría verse así:

$$\begin{array}{ll} d_x(t_1, C_1) = c_{1,1} & d_x(t_1, C_2) = c_{2,1} \\ d_x(t_2, C_1) = c_{1,1} & d_x(t_2, C_2) = c_{2,2} \\ d_x(t_3, C_1) = c_{1,2} & d_x(t_3, C_2) = c_{2,2} \\ d_x(t_4, C_1) = c_{1,1} & d_x(t_4, C_2) = c_{2,2} \\ d_x(n_1, C_1) = c_{1,2} & d_x(n_1, C_2) = c_{2,2} \\ d_x(n_2, C_1) = c_{1,2} & d_x(n_2, C_2) = c_{2,1} \\ d_x(n_3, C_1) = c_{1,1} & d_x(n_3, C_2) = c_{2,1} \end{array}$$

Esta función de asignación podría luego ser dibujada sobre dos árboles distintos del siguiente modo:

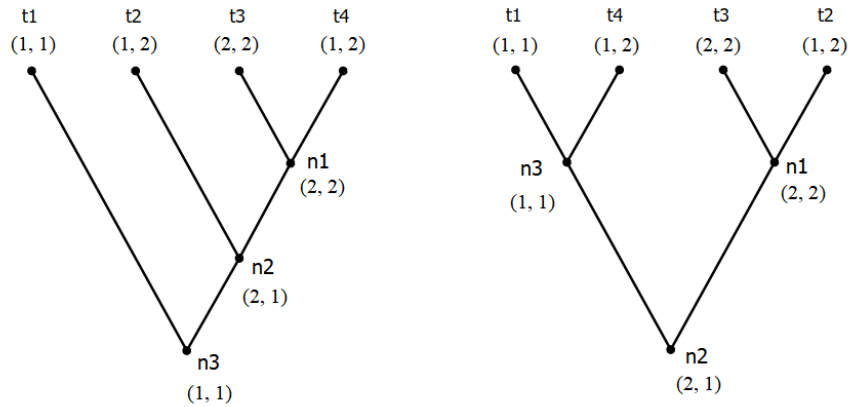


Figura 6. Función de asignación d_x dibujada sobre dos árboles.

Cabe notar que esta no será la asignación óptima para ninguno de los dos árboles (i.e. la que minimiza el largo), sino que, en ambos casos, esta será otra (u otras) función(es) del conjunto δ .

Las matrices de costo (que indican el costo de pasar de un estado a otro de un carácter) serán representadas por una única función $\$$, que toma como argumentos dos estados de un carácter y devuelve un número natural. Formalmente:⁸

$$\$: UC \times UC \rightarrow \mathbb{N}$$

Así, el costo evolutivo de pasar del estado $c_{i,j}$ al estado $c_{i,k}$ en el carácter C_i quedará representado por $\$(c_{i,j}, c_{i,k})$. En el tipo de aplicaciones que consideraré en esta contribución, el costo de mantenerse en el mismo estado es siempre 0. Sin embargo, a fin de mantenerme en un plano de mayor generalidad, no introduciré este requisito como axioma.

A continuación, se definen dos funciones binarias, w y l_1 , que devuelven el peso de un eje y el largo de un cladograma, respectivamente, bajo una asignación. Informalmente, el peso de un eje es igual a la suma de los costos de las transformaciones que ocurren entre los nodos que ese eje une. Formalmente:

$$w(\langle n_j, n_k \rangle, d_x) = \sum_{c_i \in C} \$(d_x(n_j, c_i), d_x(n_k, c_i))$$

De modo un poco más intuitivo, si $\$(c_{i,j}, c_{i,k})$ representa el costo de pasar del estado j al estado k en el carácter C_i , $\$(d_x(n_j, C_i), d_x(n_k, C_i))$ representará el costo de pasar de un nodo a otro para ese carácter, bajo una asignación d_x . w simplemente suma tal costo para todos los caracteres presentes en C .

El largo de un cladograma A_x bajo una asignación d_m es simplemente la suma de los pesos de todos sus ejes. Es decir:

$$l_1(A_i, D_m) = \sum_{\langle x, y \rangle \in D(A_i)} w(\langle x, y \rangle, d_m)$$

El largo a secas de un cladograma (l_2) quedará definido como el mínimo largo de ese árbol bajo toda asignación posible. Es decir:

$$l_2(A_i) = \min(\{l_1(A_i, d_j) / d_j \in \delta\})$$

El conjunto de los árboles óptimos AO estará definido como el conjunto de los árboles de largo₂ mínimo. Formalmente:

$$AO = \{A_i \in A / l_2(A_i) \in \min(\{l_2(A_k) / A_k \in A\})\}$$

⁸ Nótese que así caracterizados el dominio y el codominio, esta será una función parcial ya que habrá argumentos (p.e. dos estados que no pertenecen al mismo carácter) para los cuales no habrá ningún valor asignado. Podría modificarse el dominio y codominio para que $\$$ sea total, pero ello complejizaría el axioma. Por simplicidad se lo deja de esta forma.

Por otra parte, será útil para lo que sigue contar con la noción de asignación óptima para un árbol. Para ello, puede introducirse una función δ_O , definida del siguiente modo:

$$\delta_O(A_i) = \{d_j \in \delta / l_1(A_i, d_j) = l_2(A_i)\}$$

Es decir, dado un árbol, δ_O devuelve el conjunto de las asignaciones de estados a nodos que arrojan el mínimo largo para ese árbol.

Todo lo anterior permite definir, ahora sí usando el formato usual en una reconstrucción estructuralista, la clase de los modelos potenciales de **CLAD**.

M_p(CLAD): $x = \langle A, C, I, N, T, AO, A_R, \delta, \delta_O, \$, d_R, d_T, l_1, l_2, w \rangle$ es una cladística potencial ($x \in M_p(\text{CLAD})$) syss:

- (1) T es un conjunto finito tal que $|T| > 3$ [taxa terminales]
- (2) N es un conjunto finito tal que $T \subset N$ [nodos de los cladogramas]
- (3) $I = N - T$ [nodos internos]
- (4) $A = \{A_i / A_i = \langle N, D(A_i) \rangle\}$ que satisfacen los siguientes requisitos hasta el isomorfismo:
 - (i) $D(A_i) \subseteq N^2$ [relación binaria de descendencia inmediata]
 - (ii) $\nexists n_{x1} \in N, \exists n_{x2}, \dots, n_{xj} \in N$ tal que $\langle n_{x1}, n_{x2} \rangle \in D(A_i)$ y ... y $\langle n_{xj-1}, n_{xj} \rangle \in D(A_i)$ y $\langle n_{xj}, n_{x1} \rangle \in D(A_i)$ [$D(A_i)$ es acíclica]
 - (iii) $\exists! n_x \in N$ tal que $\forall n_y \in N$ (si $n_y \neq n_x, \langle n_y, n_x \rangle \notin D(A_i)$ y $\exists n_{x1}, \dots, n_{xj} \in N$ ($\langle n_x, n_{x1} \rangle \in D(A_i)$ y $\langle n_{x1}, n_{x2} \rangle \in D(A_i)$ y ... y $\langle n_{xj}, n_y \rangle \in D(A_i)$)) [$D(A_i)$ es enraizada]⁹
 - (iv) $\forall n \in I, \exists n_{x1}, n_{x2} \in N$ tal que $(n_{x1} \neq n_{x2} \text{ y } \langle n, n_{x1} \rangle \in D(A_i) \text{ y } \langle n_x, n_{x2} \rangle \in D(A_i) \text{ y } \forall n_{x3} \in N \text{ (si } n_{x1} \neq n_{x3} \neq n_{x2} \text{ entonces } \langle n_x, n_{x3} \rangle \notin D(A_i)))$ [Cada nodo interno tiene exactamente dos descendientes]
 - (v) $\forall t \in T (\nexists n \in N \text{ tal que } \langle t, n \rangle \in D(A_i))$ [Los taxa bajo consideración son nodos terminales]
 - (vi) $\forall n \in N \nexists n_x \in N \exists n_y \in N (\langle n_x, n \rangle \in D(A_i) \text{ y } \langle n_y, n \rangle \in D(A_i) \text{ y } n_x \neq n_y)$ [No ocurren fusiones]
- (5) $A_R \in A$ [A_R es un árbol particular, el árbol real, ver la sección siguiente]
- (6) $C = \{C_1, \dots, C_m\}$ es un conjunto finito, no-vacío tal que cada $C_i = \{c_{i,1}, c_{i,2}, \dots, c_{i,j}\}$ es un conjunto finito, no vacío [caracteres y estados]
- (7) $\cap C = \emptyset$ [Los caracteres no comparten estados]
- (8) $d_T: T \times C \rightarrow UC$, tal que $d_T(t, C_i) \in C_i$ [asignación para terminales]
- (9) $\delta = \{d_1, \dots, d_n\}$ es un conjunto finito, no-vacío [de funciones de asignación completas], tal que $\forall d_i \in \delta (d_i: N \times C \rightarrow UC \text{ y } d_i(n, C_i) \in C_i)$
- (10) $d_R \in \delta$ [d_R es una asignación particular, véase la sección siguiente]
- (11) $\forall d_i \in \delta, \forall t \in T, \forall C_i \in C (d_i(t, C_i) = d_T(t, C_i))$ [las asignaciones completas respetan las asignaciones para terminales]
- (12) $\$: UC \times UC \rightarrow \mathbb{N}$ [función parcial de costos]
- (13) $w: N^2 \times \delta \rightarrow \mathbb{N}$, tal que $w(\langle n_j, n_k \rangle, d_x) = \sum_{c_i \in C} \$ (d_x(n_j, c_i), d_x(n_k, c_i))$ [peso de un eje]
- (14) $l_1: A \times \delta \rightarrow \mathbb{N}$, tal que $l_1(A_i, D_m) = \sum_{\langle x, y \rangle \in D(A_i)} w(\langle x, y \rangle, d_m)$ [largo de un cladograma bajo una asignación]
- (15) $l_2: A \rightarrow \mathbb{N}$ tal que $l_2(A_i) = \min(\{l_1(A_i, d_j) / d_j \in \delta\})$ [largo de un cladograma]
- (16) $AO = \{A_i / l_2(A_i) \in \min(\{l_2(A_k) / A_k \in A\})\}$ [árboles óptimos]
- (17) $\delta_O: A \rightarrow \delta$, tal que $\delta_O(A_i) = \{d_j \in \delta / l_1(A_i, d_j) = l_2(A_i)\}$ [asignaciones óptimas para un árbol]

⁹ Con $\exists! x \varphi(x)$ utilizado para la notación lógica usual de “existe un y solo un individuo x que satisface φ ” (i.e. $\exists! x \varphi(x) \stackrel{\text{def}}{=} \exists x \varphi(x) \wedge \forall z \text{ si } \varphi(z) \text{ entonces } z = x$).

3.2. Leyes fundamentales

Recuérdese que lo que la cladística busca inferir es el árbol filogenético que dio lugar (y que explica) la distribución observada de caracteres y estados. Para hacerlo postula no solo un árbol como óptimo, sino también una asignación de estados para los ancestros hipotéticos (los nodos internos). Una aplicación de esta teoría será considerada exitosa cuando: (i) el árbol inferido como el óptimo coincide con el árbol real (o bien, si hay muchos árboles óptimos, cuando el real está entre los óptimos); y (ii) la distribución real de los caracteres de los ancestros coincide con la optimización de los caracteres sobre aquel árbol inferido (si hay más de una optimización posible, cuando la real es alguna de ellas). Una pregunta que podría surgir en este punto es cómo es posible determinar cuáles son el árbol y la distribución de caracteres reales independientemente de esta ley (a fin de poder contrastarla). Volveremos a esta cuestión más adelante.

Para poder expresar las afirmaciones fácticas (i) y (ii) recién mencionadas, es necesario introducir a la reconstrucción los conceptos de árbol real y distribución de caracteres real. Lo haré simplemente a través de dos constantes de individuo, A_R y d_R , bajo el único requisito formal de que $A_R \in A$ y $d_R \in \delta_O$ (estas afirmaciones ya están presentes en los axiomas impropios dados arriba). De ese modo, (i) y (ii) pueden expresarse formalmente simplemente diciendo que:

M(CLAD): x es una cladística actual ($x \in M(CLAD)$) syss:

- (1) $x \in M_p(CLAD)$
- (2) $A_R \in AO$
- (3) $d_R \in \delta_O(A_R)$

Esto expresa precisamente lo que se deseaba, es decir, que (2) el patrón de diversificación real de las especies está entre los árboles óptimos; y (3) la historia real de evolución de los caracteres está entre las asignaciones de estados óptimas para el árbol real, según la cladística.

Es interesante notar que estos dos axiomas (tomados conjuntamente) exhiben dos de los síntomas típicos de las leyes fundamentales (Lorenzano 2014-2015).¹⁰ Por un lado, valen en todas las aplicaciones intencionales de la teoría (en esto se distinguirían de las leyes especiales, que pretenden aplicarse solo a algunas aplicaciones de la teoría). Toda aplicación exitosa de la cladística será una en la que los árboles óptimos contienen al árbol real, y en la que la optimización de los caracteres identifica correctamente los estados de los ancestros. Si alguna de ellas falla, entonces la aplicación de la cladística sería considerada no-exitosa.¹¹

Por otra parte, otro síntoma usual que exhiben las leyes fundamentales es su carácter sinóptico o arracimado (suelen contener a todos, o a la mayoría, de los conceptos de la teoría). Es menos claro que los axiomas de arriba exhiban este síntoma. Éstos parecen, a primera vista, incluir solo unos pocos conceptos de la teoría. Nótese, sin embargo, que AO y δ_O son conceptos definidos. Por ejemplo, AO está definido a partir de A y l_2 . A su vez, l_2 está caracterizado a partir de A , δ y l_1 ; l_1 de otros conceptos, etc. Si ilustramos gráficamente cuáles conceptos están definidos/caracterizados a partir de cuáles otros, obtenemos algo como lo siguiente:

¹⁰ Otros de los síntomas usuales (el carácter cuasi-vacuo o empíricamente irrestricto, el papel sistematizador y la fuerza modal) se encuentran discutidos en Roffé (2020b), pp. 79-80, en relación con las especializaciones de la teoría (no discutidas aquí).

¹¹ El caso de la segunda ((3) en la presentación estructuralista) es un poco más debatible. En los inicios del programa cladista el énfasis estaba puesto más sobre la identificación de la topología correcta que sobre las reconstrucciones de estados ancestrales. Sin embargo, actualmente, el segundo de estos propósitos es considerado tan importante como el primero. Adicionalmente, las contrastaciones explícitas de la cladística midieron el grado de ajuste de las asignaciones de estados a ancestros a los estados reales de esos ancestros (ver la sección siguiente).

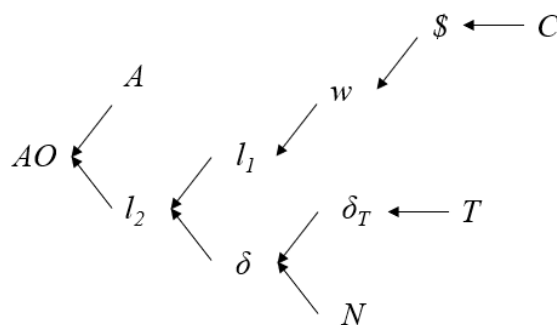


Figura 7. Grafo ilustrando qué conceptos están definidos o caracterizados a partir de cuáles otros. Algunas flechas fueron omitidas para mejorar la legibilidad (p.e. l_1 y l_2 también incluyen a A en su caracterización, etc.)

Si se realiza este ejercicio con todos los conceptos que figuran en las dos leyes puede verse que, en realidad, los dos axiomas dados contienen (conjuntamente) la gran mayoría de los conceptos de la teoría. Es decir, si bien puede sostenerse que, estrictamente hablando, las leyes no exhiben el requisito de ser arracimadas —ya que éste es un requisito sintáctico, y no hay aquí una única ley que contenga explícitamente a (casi) todos los conceptos de la teoría— sí lo exhiben en espíritu. Esto se debe a que tomados conjuntamente los axiomas propios que determinan la clase de los modelos, establecen conexiones/restricciones en las interpretaciones para todos los términos de la teoría.¹²

4. El estatus fáctico de la cladística

Es posible volver ahora a la pregunta mencionada en la sección introductoria. Recuérdese que el problema consiste en el estatus fáctico de la cladística, y en consecuencia, en la justificación que puede ofrecerse para la utilización de métodos cladísticos en la inferencia filogenética. En esta sección sostendré, a la luz de la reconstrucción presentada en la sección anterior, que la cladística es una teoría fáctica, cuyas afirmaciones centrales (sus leyes fundamentales) tienen el mismo estatus que cualquier otra ley de cualquier otra teoría fáctica. Así, la justificación para utilizarla no es distinto al de esas otras teorías: no es ni más ni menos que su éxito empírico aplicativo. Pero antes de defender este punto en mayor detalle es conveniente examinar algunas de las justificaciones que se han dado en la literatura para la utilización de la cladística.

En primer lugar, estaría la idea de que la justificación de la cladística está basada en la simplicidad. Por detrás de ello podría haber un principio metafísico *a priori* de que la realidad es simple (una posición que, en mi conocimiento, nadie sostuvo), o bien un principio epistemológico de que las explicaciones más simples son preferibles debido a nuestras limitaciones cognitivas (ver Sober 1988, cap. 2).

Esta posición no parece muy razonable. Independientemente de que no es claro por qué un árbol que requiere postular más convergencias sería menos simple que otro que requiera menos (no estaría claro qué noción de simplicidad está en juego), puede afirmarse lo siguiente. La apelación a la simplicidad en ciencias se realiza en general cuando las dos hipótesis en comparación son empíricamente equivalentes. En ese caso, un posible factor de decisión entre ellas es su simplicidad (como sea que se la entienda). Sin embargo, a pesar de que todos los árboles (bajo alguna asignación de caracteres) son capaces de dar cuenta de la distribución observada de homologías, no son empíricamente equivalentes en cuanto a la historia evolutiva que postulan para dar cuenta de ella. Si bien tal historia evolutiva (el patrón de diversificación) no suele ser “observable”, ya que los eventos de especiación relevantes suelen

¹² De todos modos, cabe notar que Moulines (1991, pp. 233-234) discute casos de teorías en donde ocurren cosas similares —p.e. en la mecánica relativista del continuo y en la electrodinámica, en las reconstrucciones ofrecidas por Bartelborth (1988)— en donde las diversas leyes fundamentales “no parecen poder reformularse como leyes sinópticas de manera plausible y natural” (Moulines 1991, p. 234). Sin embargo, habría una diferencia aquí, ya que en el presente caso (incluso tomadas conjuntamente) las leyes contienen/restringen a todos los conceptos a través de conceptos definidos —no conteniendo *explícitamente* todo concepto de la teoría—. Podría argumentarse, sin embargo, que, al ser eliminables los conceptos definidos, las leyes podrían reformularse (de modo mucho más engorroso) para contener explícitamente todo concepto de la teoría.

estar en el pasado remoto, sí existen casos en donde es posible acceder a ese patrón independientemente de la cladística (ver más abajo).

Otra posición posible consiste en sostener que las leyes fundamentales están justificadas *teóricamente*, dado lo que sabemos acerca de la evolución a través de teorías independientes, pero no empíricamente (i.e. no las podemos testear directamente). Por ejemplo, en el ámbito de la genética, sabemos que las tasas de sustitución de nucleótidos en poblaciones suelen ser bajas, y por lo tanto que es menos probable que una especie ancestral con cierto estado para un carácter (p.e. un nucleótido en una posición del genoma) haya dado lugar a dos especies con un estado diferente al original, y en donde tal estado haya surgido dos veces independientemente, en lugar de una vez en algún otro ancestro común (posterior) de ambas. Así, por ejemplo, algunos autores y autoras consideran que aplicar la cladística es hacer una inferencia a la mejor explicación (Quinn 2016) o que el método de parsimonia es el que minimiza las hipótesis *ad hoc* de homoplasia, dado que la hipótesis nula o explicación por *default* de la similitud homológica es la herencia del rasgo a partir de un ancestro común (Farris 1983).

No es mi intención aquí negar que consideraciones provenientes de otras teorías aportan plausibilidad a la utilización de métodos cladísticos. Sin embargo, creo que puede decirse algo más, que las consideraciones “externas” de ese tipo no son *todo* lo que puede decirse a favor de la cladística. La posición que defenderé aquí es que las leyes fundamentales recién presentadas son leyes fácticas, que poseen el mismo estatus (fáctico) que cualquier otra ley fundamental de cualquier otra teoría (p.e. $F = m \times a$ en la mecánica clásica). Es decir, que se las debe aceptar o rechazar por motivos empíricos, según el éxito aplicativo que tengan, y que la cladística no tiene nada de peculiar en esto. La reconstrucción presentada en la sección anterior aporta de entrada razones fuertes para considerar a la cladística una teoría: es un cuerpo de conocimiento estructurado de manera semejante a otras teorías (o al menos que puede ser representado como tal) y que hace una serie de afirmaciones fácticas generales acerca de un conjunto de fenómenos.

El mayor obstáculo para aceptar esta posición (y una razón por la que algunos autores sostuvieron alguna de las anteriores) parecería ser la duda sobre el estatus fáctico de lo que reconocí como las leyes fundamentales. En particular, pareciera ser que un test empírico de la cladística requeriría conocimiento independiente del patrón de diversificación real para establecer si este efectivamente se encuentra entre los que la cladística determina como los árboles óptimos (y lo mismo con la distribución histórica de caracteres). Sin embargo, en la mayoría de los casos, la historia real no es cognoscible por haber ocurrido hace mucho tiempo.

Existen, sin embargo, casos en donde sí es posible obtener esa información. Algunos son los estudios llevados a cabo en el área de la filogenética experimental. Un ejemplo de este tipo de enfoque se encuentra en Hillis *et al.* (1992).¹³ El objetivo de estos autores no era solamente testear la cladística, sino más bien realizar un experimento crucial que pudiese decidir entre diversos métodos de inferencia filogenética (la cladística entre ellos). Para ello, generaron una filogenia en el laboratorio, que usaron para comparar cuál de los métodos la reconstruía mejor. Dado que generar una filogenia con un grado interesante de diferenciación entre los individuos requiere del paso de muchas generaciones los autores usaron un virus (el bacteriófago T7) que tiene generaciones muy cortas, elevando además las tasas de mutación al criarlos en un ambiente con mutágenos. Así, partiendo de una población original, fueron dividiéndola en dos cada ciertos intervalos de tiempo, para obtener la siguiente topología (que conocían previamente a utilizar los métodos):

¹³ Otro modo de tener un conocimiento independiente de la filogenia sería a través de simulaciones (Huelsenbeck 1995), pero es más dudoso que pueda hablarse de un test fáctico de la cladística en tal caso (Hillis *et al.* 1992, pp. 589-590, 1993, p. 90)

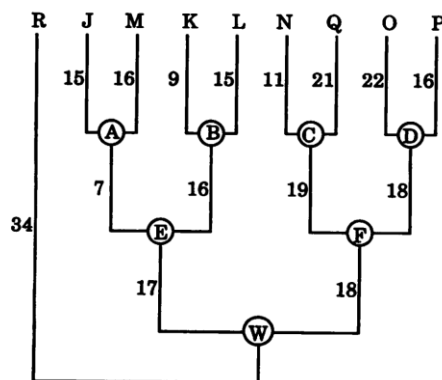


Figura 8. Filogenia conocida del bacteriófago T7, tomada de Hillis et al. (1992), p. 590.

Luego secuenciaron parte de los genomas y los usaron como matriz de datos para los diferentes métodos. Nótese que con 8 taxa terminales hay 135.135 topologías posibles, con lo cual la probabilidad de dar con la correcta por azar es mínima.¹⁴

Los resultados fueron los siguientes. De los métodos utilizados (5 en total, incluyendo parsimonia) todos dieron con la topología correcta. En este sentido, el experimento fracasó *qua* experimento crucial,¹⁵ pero fue exitoso *qua* contrastación de la cladística. En cuanto a la segunda ley fundamental, la cladística logró reconstruir correctamente 1369 de los 1404 estados de los nodos ancestrales (un 97,5%, aproximadamente), incorrectamente 18 (1,3%) y ambiguamente 20 (1,4%) —no había comparación aquí con los otros métodos, ya que la cladística era el único que permitía reconstruir los estados de los ancestros.¹⁶

Como puede notarse, además, en esta contrastación, lo que Hillis *et al.* midieron a fin de establecer el éxito o fracaso aplicativo de la cladística fueron precisamente las dos afirmaciones fácticas que constituyen los axiomas propios de **M(CLAD)** —lo cual confirma la adecuación de la reconstrucción. Es decir, se observó, por un lado, que la topología (el cladograma) de menor largo coincidiera con el patrón de diversificación real A_R (ahora conocido independientemente), y que la optimización de los caracteres utilizados sobre ese árbol coincidiese con la distribución de los caracteres de los ancestros (representado por d_R en la reconstrucción).

Adicionalmente, lo que esto muestra, nuevamente en jerga estructuralista, es que los conceptos A_R y d_R son **CLAD**-no-teóricos, esto es, que pueden determinarse independientemente de la propia cladística. Como sostienen Roffé, Ginnobili y Blanco (2018, p. 5; también en el contexto de debates en torno a la cladística) que un concepto sea **T**-no-teórico para una teoría **T** no significa que pueda ser determinado independientemente de **T** en toda aplicación de **T**. En el presente caso, en la mayoría de las aplicaciones estándar de la cladística no tenemos posibilidad de determinar independientemente a A_R y d_R . Sin embargo, ello no implica que dichos conceptos sean **CLAD**-teóricos, ya que hay casos (como los ilustrados arriba) en donde sí se da tal determinación independiente. Como también afirman estos autores, esto es lo que hace que esos conceptos pertenezcan a la “base contrastacional” de **CLAD**, ya que contrastar una teoría consiste en determinar un mismo concepto de manera **T**-teórica y **T**-no-teórica a la vez, y analizar la coincidencia o no entre los resultados de ambas determinaciones. Esto es precisamente lo que hacen experimentos como el de Hillis *et al.*, inferir teóricamente un cladograma óptimo a partir de las leyes de **CLAD** y obtenerlo **CLAD**-no-teóricamente al manipular experimentalmente la cladogénesis en las poblaciones de virus.

¹⁴ El noveno, R, es un *outgroup*, utilizado para enraizar el árbol, pero sin formar parte de los taxa bajo estudio.

¹⁵ Sober (1993) objetó que el experimento no servía para comparar métodos ya que la filogenia que diseñaron los autores (p.e. una con largos de rama idénticos) era “fácil” para todas las metodologías testeadas —esto era conocido a partir de simulaciones. Los autores respondieron que este experimento era solo una primera aproximación a lo que podía hacerse con filogenias experimentales, y que una contrastación más exhaustiva debía probar con otras topologías (Hillis *et al.* 1993, p. 91).

¹⁶ Por supuesto, los 18 casos de estimaciones incorrectas no implican que la contrastación haya fracasado rotundamente, dado que todos estamos dispuestos a aceptar el éxito aplicativo de una teoría tolerando algún margen de error.

En suma, la hipótesis de que la evolución procede parsimoniosamente, haciendo que el curso real de la filogenia sea el de menor largo, es una hipótesis fáctica, que puede contrastarse cuando se tiene acceso independiente al patrón de diversificación real. Esto último ocurre en casos como los de filogenias experimentales. Hay que aclarar, sin embargo, que a pesar de las declaraciones grandilocuentes iniciales de Hillis *et al.* (1992) la filogenética experimental sigue siendo un campo relativamente chico, no habiendo demasiados trabajos publicados con este tipo de enfoque. Según Oakley (2009), parte importante de la razón estriba en que generar una filogenia experimental consume mucho más tiempo y recursos que hacer una simulación. Adicionalmente, la discusión actual en torno al mejor método de inferencia filogenético ha disminuido en intensidad. En cambio, la mayoría de los sistemáticos emplean varios métodos a la vez, considerando como más sólidos a los resultados (p.e. a los grupos) que se obtienen en todos ellos. Otra razón por la que las filogenias experimentales no se popularizaron es que los tiempos generacionales requeridos para hacer análisis interesantes obligan a usar siempre el mismo tipo de organismos (virus, y quizás bacterias).

A partir de todo lo anterior, es posible concluir que en la gran mayoría de los casos de aplicación reales la cladística no es contrastada, sino que se la usa para retrodecir la historia evolutiva. En esto difiere sólo en grado de otras teorías, pero no hay una diferencia cualitativa sustancial. Como se dijo, las teorías no son contrastadas en todas sus aplicaciones. Su contrastación exitosa en ciertas aplicaciones brinda confianza inductiva en que las predicciones o retrodicciones de la teoría en lugares nuevos (en los que no fue contrastada aún) son correctas. Por ese motivo, en los casos reales de aplicación de la cladística en donde no hay posibilidad de contrastación independiente, los sistemáticos se contentan con los resultados del análisis dado (con el/los árbol/es óptimo/s obtenido/s) sin un chequeo independiente ulterior de sus resultados.¹⁷ En términos estructuralistas, determinan sólo **CLAD**-teóricamente al árbol real, sin realizar la determinación **CLAD**-no-teórica que permitiría contrastar a la teoría en esa aplicación.

5. Conclusiones

En el presente trabajo se discutió acerca del estatus de la cladística, defendiendo que posee un estatus fáctico similar al de otras teorías científicas. Para ello, se llevaron a cabo distintas tareas.

En primer lugar, en la sección 2, se introdujeron informalmente primero la ancestría común darwiniana (lo cual era relevante para comprender el *explanandum* y algunas de las asunciones de la cladística) y luego la cladística propiamente dicha, distinguiendo sus componentes taxonómicos de los filogenéticos (y centrándose en estos últimos). Se introdujeron las nociones de matriz de datos (caracteres y estados, y su asignación a taxa terminales), cladograma, función de asignación, largo de un cladograma, y árboles óptimos.

En la sección 3 se presentó una reconstrucción formal estructuralista de la cladística. Se presentaron en particular las clases de modelos potenciales y de modelos, ilustrando ampliamente con ejemplos y discutiendo, en el caso de los modelos, en qué medida los axiomas propios exhiben dos de los síntomas típicos de las leyes fundamentales. Como se dijo, esta reconstrucción tiene valor por sí misma, tanto desde el punto de vista de la extensión del dominio de aplicaciones exitosas del programa estructuralista a un nuevo ámbito, como resultando en una formulación que exhibe de manera clara el lenguaje, la estructura y las aserciones fácticas centrales de una porción de la biología actual. Por otra parte, se ilustró con la discusión posterior cómo contar con una reconstrucción tal puede generar nuevas hipótesis metateóricas y contribuir a resolver debates que ocurren al seno de la propia disciplina objeto. La reconstrucción puede ser aplicada a otros debates internos a la propia sistemática, tarea comenzada en Roffé (2020b) y a explorar en otras publicaciones.

¹⁷ Si hay, sin embargo, un chequeo que tiene que ver con la coherencia interna con otras aplicaciones de la misma teoría. Por ejemplo, dos árboles óptimos para un mismo conjunto de taxa construidos a partir de matrices de datos distintas (p.e. una utilizando rasgos morfológicos y otra moleculares) deberían ser compatibles entre sí. En una reconstrucción estructuralista, este tipo de coherencia se expresaría en las condiciones de ligadura. Las condiciones de ligadura de **CLAD** no fueron tratadas aquí por motivos de espacio y pertinencia; puede verse Roffé (2020b) para un tratamiento.

La sección 4 se adentró en la discusión específica acerca del estatus de la cladística. Se expusieron primero las posiciones previas que habían sido adoptadas en este debate, tanto las que la justifican a partir de un principio metodológico de simplicidad, como las que apelan a una justificación empírica pero externa a la propia teoría. Se defendió luego su estatus de teoría fáctica, prestando especial atención al modo como se la puede contrastar (y fue contrastada) con filogenias conocidas de manera independiente (en filogenias experimentales). Con ello pretendo no sólo haber presentado una posición nueva en el debate sino haber ilustrado el modo en el que el análisis conceptual riguroso (auxiliado por herramientas metateóricas formales) puede ser aplicado a controversias metateóricas de manera fructífera.

Bibliografía

- Abreu, C. (2014), “Análisis estructuralista de la teoría de la anomía”, *Metatheoria* 4(2): 9-22.
- Alleva, K., Díez, J. y L. Federico (2017), “Models, Theory Structure and Mechanisms in Biochemistry: The Case of Allostereism”, *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 63: 1-14.
- Ashlock, P. D. (1974), “The Uses of Cladistics”, *Annual Review of Ecology and Systematics* 5(1): 81-99.
- Aubert, D. (2015), “A Formal Analysis of Phylogenetic Terminology: Towards a Reconsideration of the Current Paradigm in Systematics”, *Phytoneuron* 66: 1-54.
- Balzer, W. y P. Lorenzano (2000), “The Logical Structure of Classical Genetics”, *Journal for General Philosophy of Science* 31(2): 243-266.
- Balzer, W., Moulines, C. U. y J. D. Sneed ([1987] 2012). *Una arquitectónica para la ciencia. El programa estructuralista*, Bernal: Universidad Nacional de Quilmes.
- Bartelborth, T. (1988), *Eine logische Rekonstruktion der klassischen Elektrodynamik*, Bern: Peter Lang.
- Bernabé, F. N. (2018), “Cerebros, hormonas y filosofía de la ciencia: La polémica de los cerebros tipo a la luz de la elucidación de la Hipótesis Organizacional Activacional”, en Chibeni, S., Zaterka, L., Ahumada, J., Letzen, D., Silva, C. C., Al-Chueyr Pereira Martins, L. y A. P. Oliveira Pereira de Moraes Brito (eds.), *Filosofía e Historia de la Ciencia en el Cono Sur: Selección de trabajos del X Encuentro de la Asociación de Filosofía e Historia de la Ciencia del Cono Sur*, Córdoba: Universidad Nacional de Córdoba, pp. 138-148.
- Blanco, D. (2012), “Primera aproximación estructuralista a la Teoría del Origen en Común”, *Ágora* 31(2): 171-194.
- Brady, R. (1985), “On the Independence of Systematics”, *Cladistics* 1(2): 113-126.
- Brower, A. V. Z. (2000), “Evolution is Not a Necessary Assumption of Cladistics”, *Cladistics* 16: 143-154.
- Caamaño, M. (2009), “A Structural Analysis of the Phlogiston Case”, *Erkenntnis* 70(3): 331-364.
- Caponi, G. (2015), “El impacto de la filosofía anatómica de Étienne Geoffroy Saint-Hilaire en el desarrollo de la historia natural”, *Gavagai - Revista Interdisciplinaria de Humanidades* 2(2): 9-31.
- Carman, C. C. (2015), “El Sistema de Epíclidos y Deferentes de Ptolomeo: Una reconstrucción”, *Metatheoria* 6(1): 47-72.
- Darwin, C. (1992), *El origen de las especies* (versión castellana de José P. Marco de la 6ª ed. inglesa), Barcelona: Planeta-Agostini.
- Díaz, M. y P. Lorenzano (2017), “La red teórica de la dinámica de poblaciones”, *Scientiae Studia* 15(2): 307-342.
- Falguera, J. L. y X. Donato-Rodríguez (2016), “Flogisto versus oxígeno: Una nueva reconstrucción y su fundamentación histórica”, *Crítica: Revista Hispanoamericana de Filosofía* 48(142): 87-116.
- Farris, J. S. (1970), “Methods for Computing Wagner Trees”, *Systematic Zoology* 19(1): 83-92.
- Farris, J. S. (1974), “Formal Definitions of Paraphyly and Polyphyly”, *Systematic Zoology* 23: 548-554.

- Farris, J. S. (1983), "The Logical Basis of Phylogenetic Analysis", en Platnick, N. y V. A. Funk (eds.), *Advances in Cladistics: Vol. II*, New York: Columbia University Press, pp. 7-36.
- Felsenstein, J. (1978), "The Number of Evolutionary Trees", *Systematic Zoology* 27(1): 27-33.
- Ginnobili, S. (2009), "Adaptación y función", *Ludus Vitalis* 17(31): 3-24.
- Ginnobili, S. (2010), "La teoría de la selección natural darwiniana", *Theoria* 25(1): 37-58.
- Ginnobili, S. (2018), *La teoría de la selección natural. Una exploración metacientífica*, Bernal: Universidad Nacional de Quilmes.
- Gonzalo, A. y W. Balzer (2012), "A Reconstruction of the 'Classical' Linguistic Transformational Theory CLT", *Metatheoria* 2(2): 25-49.
- Hennig, W. (1950), *Grundzüge einer Theorie der phylogenetischen Systematik*, Berlin: Deutscher Zentralverlag.
- Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R. y I. J. Molineux (1992), "Experimental Phylogenetics: Generation of a Known Phylogeny", *Science* 255(5044): 589-592.
- Hillis, D. M., Bull, J. J., White, M. E., Badgett, M. R. y I. J. Molineux (1993), "Experimental Approaches to Phylogenetic Analysis", *Systematic Biology* 42(1): 90-92.
- Huelsenbeck, J. P. (1995), "Performance of Phylogenetic Methods in Simulation", *Systematic Biology* 44(1): 17-48.
- Hull, D. (1988), *Science as a Process*, Chicago: The University of Chicago Press.
- Kitching, I. J., Forey, P. L., Humphries, C. J. y D. M. Williams (1998), *Cladistics: The Theory and Practice of Parsimony Analysis*, Oxford/New York: Oxford University Press.
- Lastiri, M. (2012), "Aplicaciones intencionales de la mecánica cuántica", *Ágora: Papeles de Filosofía* 31(2): 271-285.
- Lorenzano, C. (2002). "Una reconstrucción estructuralista de la bioquímica", en Díez, J. A. y P. Lorenzano (eds.), *Desarrollos actuales de la metateoría estructuralista: Problemas y discusiones*, Bernal: Universidad Nacional de Quilmes/Universidad Autónoma de Zacatecas/Universidad Rovira i Virgili, pp. 209-230.
- Lorenzano, P. (2002), "La teoría del gen y la red teórica de la genética", en Díez, J. A. y P. Lorenzano (eds.), *Desarrollos actuales de la metateoría estructuralista: Problemas y discusiones*, Bernal: Universidad Nacional de Quilmes/Universidad Autónoma de Zacatecas/Universidad Rovira i Virgili, pp. 285-330.
- Lorenzano, P. (2014-2015), "Principios-guía y leyes fundamentales en la metateoría estructuralista", *Cuadernos del Sur* 43-44: 35-74.
- Martin, J., Blackburn, D. y E. O. Wiley (2010), "Are Node-Based and Stem-Based Clades Equivalent? Insights from Graph Theory", *PLOS Currents Tree of Life* 2, RRN1196, doi: 10.1371/currents.RRN1196.
- Mayr, E. (1965), "Numerical Phenetics and Taxonomic Theory", *Systematic Zoology* 14(2): 73-97.
- Mendez, D. y M. Casanueva (2006), "A Reconstruction of Darwin's Pangenesis in a Graph Format", en Ernst, G. y K.-G. Niebergall (eds.), *Philosophie der Wissenschaft – Wissenschaft der Philosophie – Festschrift für C. Ulises Moulines zum 60. Geburtstag*, Paderborn: Mentis, pp. 157-164.
- Moulines, C. U. (1991), *Pluralidad y recursión*, Madrid: Alianza Universidad.
- Nelson, G. y N. Platnick (1981), *Systematics and Biogeography, Cladistics and Vicariance*, New York: Columbia University Press.
- Oakley, T. (2009), "A Critique of Experimental Phylogenetics", en Garland, T. y M. R. Rose (eds.), *Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments*, California: University of California Press, pp. 659-669.
- Owen, R. (1848), *On the Archetype and Homologies of the Vertebrate Skeleton*, London: John Van Voorst.
- Owen, R. (1849), *On the Nature of Limbs*, London: John Van Voorst.
- Pearson, C. H. (2010), "Pattern Cladism, Homology, and Theory-Neutrality", *History and Philosophy of the Life Sciences* 32(4): 475-492.
- Pearson, C. H. (2018), "Theoricity and Homology: A Reply to Roffé, Ginnobili, and Blanco", *History and Philosophy of the Life Sciences* 40: 42.

- Quinn, A. (2016), “Phylogenetic Inference to the Best Explanation and the Bad Lot Argument”, *Synthese* 193(9): 3025-3039.
- Ramírez, M. J. (2003), “The Spider Subfamily Amaurobioidinae (Araneae, Anyphaenidae): A phylogenetic revision at the generic level”, *Bulletin of the AMNH*, 277: 1-262.
- Rieppel, O. (2016), *Phylogenetic Systematics: Haeckel to Hennig*, Boca Raton: CRC Press.
- Roffé, A. J. (2020a), “Dynamic Homology and Circularity in Cladistic Analysis”, *Biology & Philosophy* 35: 21.
- Roffé, A. J. (2020b), *Contrastando reconstrucciones con herramientas computacionales: Una aplicación a la cladística*, Tesis doctoral, Buenos Aires: Universidad de Buenos Aires.
- Roffé, A. J., Ginnobili, S. y D. Blanco (2018), “Theoricity, Observation and Homology: A Response to Pearson”, *History and Philosophy of the Life Sciences* 40: 42.
- Saint-Hilaire, É. G. (1830), *Principes de philosophie zoologique*, Paris: Pichon et Didier.
- Semple, C. y M. Steel (2003), *Phylogenetics*, Oxford: Oxford University Press.
- Sneed, J. D. (1971), *The Logical Structure of Mathematical Physics*, Dordrecht: Reidel.
- Sober, E. (1988), *Reconstructing the Past: Parsimony, Evolution, and Inference*, Cambridge, MA/London: MIT Press.
- Sober, E. (1993), “Experimental Tests of Phylogenetic Inference Methods”, *Systematic Biology* 42(1): 85-89.
- Steel, M. (2016), *Phylogeny: Discrete and Random Processes in Evolution*, Philadelphia: SIAM-Society for Industrial and Applied Mathematics.
- Wiley, E. O. y B. S. Lieberman (2011), *Phylogenetics: Theory and Practice of Phylogenetic Systematics*, Hoboken, NJ: John Wiley & Sons.