

## Introducción

La filogenética es una rama de la biología cuyo fin último es establecer las relaciones evolutivas entre genes, caracteres y/u organismos (Stavrinides & Ochman, 2009). En ese sentido un árbol filogenético es una representación gráfica de una hipótesis evolutiva entre taxones, genes o caracteres. Los árboles filogenéticos también reciben el nombre de filogenias, además de ser hipótesis evolutivas, son una gran herramienta de la sistemática ya que permiten organizar, estructurar y clasificar el conocimiento de la diversidad biológica (Baum, 2008).

Existen distintos tipos de árboles filogenéticos, los primeros surgen a partir del trabajo del entomólogo Hennig (1950,1966), aunque la historia de los árboles filogenéticos es más compleja (Morrone, 2000), estos se conocen como árboles de la escuela cladista o cladogramas los cuales se construyen a partir de una matriz de caracteres de T columnas que representan los taxones a comparar y C filas que corresponden a los caracteres, cada carácter puede tener más de un estado y se hace una codificación de estos. Al final se construyen todos los diagramas de ramificación que impliquen el menor número de pasos (Staton, 2015). A medida de que las técnicas de secuenciación comenzaron a ser más asequibles se incrementó el uso de datos moleculares para estudios filogenéticos, sin embargo el método cladista no era el mejor para este tipo de datos, pues caía en un error conocido como Atracción de Rama Larga (Staton, 2015), entonces para las filogenias basadas en secuencias se comenzaron a utilizar árboles basados en distancia y máxima verosimilitud.

En este trabajo nos centraremos en los métodos basados en distancia, una vez que se tienen alineadas las secuencias ya sea de DNA o de proteínas, se aplica una métrica que en esencia lo que mide es la similitud entre dos secuencias, comparando una a una de sus bases (Stavrinides & Ochman, 2009), a esta se le debe aplicar una corrección utilizando un modelo de sustitución nucleotídica, estos describen una tasa de cambio de las mutaciones fijas entre las secuencias, lo ideal es escoger el modelo que se adapta mejor ya que sino se pueden producir sesgos en la inferencia filogenética. Por ejemplo Jukes-Cantor es un modelo que asume una misma tasa de cambio entre todos los nucleótidos, además de una misma frecuencia para cada uno, sin embargo se sabe que las transiciones, cambios entre dos pirimidinas o dos purinas, son más frecuentes que las transversiones, cambios entre pirimidinas y purinas, dados estos hechos surge el modelo de Kimura que asume distintas tasas de cambios para las transversiones y las transiciones, más modelos han surgido a partir de los primeros (Arenas, 2015), entre ellos el de Tamura, que toma en cuenta el exceso de transiciones, frecuencias desiguales de nucleótidos, y la variación de la tasa de sustitución de distintos sitios (Tamura & Nei, 1993), uno de los más complejos es el modelo general del tiempo reversible, que introduce distintas tasas para cada cambio y una frecuencia distinta para cada nucleótido, lo novedoso es que asume una proporción de sitios invariables y una tasa de variación a través de los sitios (Arenas, 2015).

Una vez hecha la corrección, se emplean distintos algoritmos para la generación del árbol a partir de una matriz de distancias. Uno de los más conocidos es UPGMA (Unweighted Pair Group Method with Arithmetic Mean), los dos taxones con mayor similitud entre sí se

agrupan, y se recalcula la matriz considerando a los dos taxones agrupados como si fuera uno solo, así el proceso se repite iterativamente hasta que todos los taxones hayan sido agrupados y añadidos al árbol (Stavrinides & Ochman, 2009). Neighbor Joining (NJ) es otro de los más utilizados, se define vecino a aquellas Unidades Taxonómicas Operacionales (OTUs) que están conectadas por un nodo interno en un árbol no enraizado. Se empieza con un árbol “estrella” porque no hay jerarquía inicialmente entre los taxones, cualquier par de OTUs pueden conectarse entre sí, formando un nuevo nodo interno en el árbol que conecta ese par con el resto de los taxones, para elegir al par de OTUs hay  $N(N-1)/2$  formas de hacerlo y entre esas se elige aquella que da la suma más pequeña de las longitudes de las ramas, el procedimiento continúa hasta que se forman  $N-3$  ramas internas (Saitou & Nei, 1987).

En el presente trabajo haremos la reconstrucción de la filogenia del género *Ambystoma* que pertenece a la familia Ambystomatidae. La composición de la familia sigue estando en debate (Frost, 2023). Hay autores que proponen que la familia está compuesta por la subfamilia Dicamptodontidae y Rhyacotritoninae (Tihen, 1958), mientras otros sostienen que Dicamptodon es el grupo hermano de Ambystomatidae (Sever, 1992). En 2009 Zhang y Wake publicaron una filogenia de las salamandras utilizando genoma mitocondrial, y proponen a Ambystomatidae y Dicamptodontidae como familias separadas, y argumentando que el registro fósil del género Dicamptodon del paleoceno difiere notablemente morfológicamente del de *Ambystoma*, además de que las larvas del primer grupo son perennes adaptadas a la corriente y las de *Ambystoma* son generalmente de corta duración y están adaptadas a aguas estancadas (Zhang & Wake, 2009). En ese sentido, para el presente trabajo, tomaremos en cuenta a la familia Ambystomatidae compuesta únicamente por el género *Ambystoma*.

El género *Ambystoma* está compuesto por 33 especies cuya distribución abarca desde el suroeste de Alaska y sur de Canadá hasta el Altiplano Mexicano (Semarnat, 2018). Se conocen también como ajolotes y/o salamandras Topo (Mole salamanders) debido a que la mayoría pasan la mayoría del año bajo tierra y salen para alimentarse y aparearse, se distribuyen en distintos hábitats como arroyos, montañas, y presas (Amphibia Web) Durante su fase larvaria deben estar en hábitats acuáticos, algunos son paedomórficos, es decir que retienen características de individuos juveniles (Ávila, 2021).

Hasta ahora se han publicado algunos artículos con filogenias de *Ambystoma* obtenidas por distintos métodos. En 1984, Shaffer fue el primero en construir un árbol filogenético del género *Ambystoma*, este se hizo con datos moleculares, específicamente utilizando aloenzimas, a través de un árbol de distancia UPGMA, tiempo después implementó el método de máxima verosimilitud (Shaffer, 1993). Posteriormente en 2004 se publicó otra filogenia utilizando un conjunto de genes del genoma mitocondrial de 5 especies de *Ambystoma*, utilizando análisis bayesianos (Samuels, 2005). En 2006, se publicó un trabajo de grado donde se crea una filogenia a partir de caracteres morfológicos categóricos y continuos de las larvas, utilizando un software (Mesquite) que a través de una búsqueda heurística construye árboles que se re-arreglan hasta que se encuentra el mejor ajuste utilizando el criterio de parsimonia (Cole, 2007), el cual establece que la hipótesis más probable es la que requiere menos pasos evolutivos.

Para la reconstrucción de la filogenia, utilizaremos el genoma mitocondrial (gm) de 10 especies de *Ambystoma*, *A. talpoideum*, *A. bishopi*, *A. laterale*, *A. californiense*, *A. tigrinum*, *A. dumerilii*, *A. andersoni*, *A. mexicanum*, *A. barbouri*, *A. texanum* y como outgroup, organismo que debe ser cercano al grupo pero no dentro de él, utilizaremos a *S. lacertina*. Decidimos utilizar el genoma mitocondrial completo dado que en un mayor número de genes se puede reflejar una mayor variación genética (Samuels, 2005), además es conveniente porque el tamaño del gm es constante entre estas especies.

## Pregunta de Investigación

¿Qué combinación de modelo de sustitución nucleotídica y árbol de distancia nos arroja la construcción de la filogenia más viable del género *Ambystoma* dado el criterio de parsimonia?

## Objetivo

El objetivo del presente trabajo es crear árboles filogenéticos del género *Ambystoma* utilizando distintos modelos de sustitución nucleotídica y distintos tipos de árboles de distancia para evaluar con el criterio de parsimonia cuál podría ser la hipótesis más viable.

Objetivos particulares:

- Conocer las diferencias que surgen de acuerdo a los distintos modelos y métricas empleados en la construcción del árbol
- Evaluar cual es la hipótesis más viable utilizando el criterio de parsimonia

## Métodos

### 1. Alineación de Secuencias.

#### a. Needleman-Wunsch (Pares de Secuencias)

Es un algoritmo de programación dinámica. Los algoritmos de programación dinámica suelen ser usados para resolver problemas de optimización (dependiendo del problema, las respuestas correctas son las mínimas o máximas). Funcionan construyendo la solución paso por paso. En nuestro caso para cada prefijo de las cadenas está buscando el score más alto que puede obtener, dependiendo de cuál de las celdas adyacentes haya sido la mayor y considerando si la decisión actual será que los caracteres harán match, mismatch o será un insert/delete. Hace esto hasta tener las dos secuencias completas. Desde la última celda de la matriz reconstruimos el camino de las “mejores” decisiones que cada paso tomó.

Si la entrada son dos secuencias, de longitud  $n$  y  $m$ , la complejidad de tiempo es de  $O(nm)$ . La justificación de esto es que el paso más complicado es que tenemos que llenar una matriz de  $(n+1)(m+1)$ . Además de las secuencias para la entrada también necesitamos la puntuación con la que vamos a evaluar el match, mismatch y los insert/delete. Aunque la implementación puede recibir cualquier número para estos parámetros, nosotros usamos 1, -1 y -1, respectivamente.

### b. Star-Alignment (Múltiples Secuencias)

Calcular el mejor alineamiento (según la SP-score) para múltiples secuencias es un problema NP-completo (Wang & Jiang, 1994). En la práctica, para resolverlo, un enfoque es también utilizar programación dinámica con una o más heurísticas. Tenemos como entrada múltiples secuencias y las puntuaciones de match, mismatch e in/del. Empezamos buscando la secuencia que se parezca más a todas las demás, esta será el centro de la estrella. Alineamos todas las demás (con algún algoritmo que lo haga por pares) con el centro.

En este caso, la implementación del algoritmo no fue hecha por nosotros. El [repositorio](#) original del proyecto menciona también una ligera optimización al algoritmo original, además de ciertos parámetros para match, mismatch e in/del. En esto último, nosotros decidimos usar los mismos que usamos en el algoritmo anterior. La razón de usar un repositorio y no una biblioteca de Python fue porque las que encontramos no eran descriptivas (además de mencionar programación dinámica) sobre qué algoritmo usaban.

## 2. Cálculo de distancias.

### a. Proporción de distancias observadas.

El más sencillo de todos. Lo único que hacemos es contar el número de diferencias (como si fuéramos a sacar la distancia de Hamming) y con eso sacar la proporción con la longitud de la secuencia. Mientras más se acerque a uno quiere decir que hay más sitios en los que las secuencias difieren, y mientras más se acerque a cero significa que son más parecidas.

### b. Jukes-Cantor (JK69).

Una vez que calculamos la proporción de diferencias observadas. Como contar las diferencias observadas no es suficiente lo que hacemos para corregirlo es usar ciertos modelos probabilísticos. En el modelo evolutivo de Jukes-Cantor cada sustitución tiene la misma probabilidad, lo que lo hace el más simple de todos. La fórmula es la siguiente (Cristianini & Hahn, 2006):

$$K = -\frac{3}{4} \ln\left(1 - \frac{4}{3}d\right)$$

donde d son las diferencias observadas.

### c. Motoo Kimura (K71).

Es una adaptación de Jukes-Cantor, solo que aquí distinguimos de las transversiones y las transiciones, a diferencia del anterior. Si dejamos de hacer la diferenciación entonces nos daría la fórmula anterior. Esta es (Cristianini & Hahn, 2006) la siguiente:

$$K = -\frac{1}{2} \ln(1 - 2P - Q) - \frac{1}{4} \ln(1 - 2Q)$$

donde Q es el número de transversiones y P el de transiciones

### d. Tamura (T92).

Este modelo sólo es válido para nucleótidos. Además de los parámetros anteriores agregamos 2 extra: en cada secuencia vamos a contar también su contenido de Guaninas y Citosinas. La fórmula es la siguiente (John Rot Laboratory):

$$K = -C \ln\left(1 - \frac{P}{C} - Q\right) - 0.5(1 - C) \ln(1 - 2Q)$$

donde P y Q son transiciones y transversiones, respectivamente. Y la C viene del resultado de  $GCsec1 + GCsec2 - 2 * GCsec1 * GCsec2$ . Aquí GSsecN es el contenido de GC en la secuencia N.

### 3. Generación del árbol.

#### a. Neighbor Joining.

Es un algoritmo iterativo, en cada paso vamos uniendo dos vértices. Iniciamos con una gráfica estrella, todos nuestros taxa están unidos a un nuevo nodo. Si tenemos más de dos nodos vamos a seguir los siguientes pasos (Cristianini & Hahn, 2006):

- I. Calcular R, que es un vector del mismo tamaño que nuestro número de secuencias. Donde  $R_i = \sum_{j=1}^n d(i, j)$ , donde d hace referencia a nuestra tabla de distancias original.
- II. Con R podemos calcular una nueva matriz, M, del mismo tamaño que la de distancias, donde la entrada  $M(i, j) = (n - 2)d(i, j) - R_i - R_j$ , donde n es el número de nodos. El valor mínimo de esta matriz son los nodos que vamos a unir. En la gráfica separamos los dos nodos con el valor más bajo, digamos a y b, del centro de la estrella y los unimos a un nuevo nodo, V. Agregamos V a nuestra lista de nodos por unir (de donde sacamos a los vértices a y b).
- III. Para el paso iterativo calculamos de nuevo nuestra matriz de distancias. Ahora en lugar de tener una matriz de nxn tenemos una de (n-1)x(n-1). Los valores de la matriz original que no correspondan a los nodos a y b los vamos a conservar en esta nueva matriz. Las que no las calculamos de la siguiente manera:  

$$d(V, k) = .5 (d(i, k) + d(j, k) - d(i, j))$$
- IV. Calculamos las longitudes de las ramas de a-V y de b-V.  

$$B_a = (0.5 d(a, b) + \frac{R_a - R_b}{2(n-2)})$$

$$B_b = (0.5 d(a, b) + \frac{R_b - R_a}{2(n-2)})$$
- V. Repetimos los pasos 1 a 4 hasta tener exactamente dos nodos. Digamos que son i,j. Los unimos con una longitud de rama de  $d(i, j)$

#### b. UPGMA.

Para UPGMA usamos pasos muy parecidos a NJ, las únicas diferencias son que para el paso 1 y 2 no necesitamos recalcular una nueva matriz, usamos el menor valor de la original. Además de eso, la forma de calcular las ramas de los árboles es diferente. Sin embargo, fuera de las diferencias sobre los cálculos siguen los mismos pasos. La implementación usada corresponde a la biblioteca [pylogeny](#).

### 4. Análisis de parsimonia.

Para el análisis de parsimonia con R (phangorn) necesitamos un árbol y las secuencias con las que estamos trabajando. Lo que regresa es el score de parsimonia, que es el mínimo de pasos que son necesarios para explicar el árbol.

## Resultados

### 1. Tablas de Distancias

En las siguientes tablas los encabezados son los siguientes: *A. andersoni*, *A. barbouri*, *A. bishopi*, *A. californiense*, *A. dumerilii*, *A. laterale*, *A. mexicanum*, *A. talpoideum*, *A. texanum*, *A. tigrinum*, *S. lacertina*.

#### a. Naive

|      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|
| 0    | .106 | .12  | .065 | .033 | .106 | .008 | .154 | .103 | .047 | .251 |
| .106 | 0    | .121 | .108 | .105 | .111 | .105 | .156 | .063 | .102 | .256 |
| .12  | .121 | 0    | .122 | .119 | .124 | .119 | .163 | .117 | .115 | .255 |
| .065 | .108 | .122 | 0    | .063 | .111 | .064 | .156 | .108 | .06  | .255 |
| .033 | .105 | .119 | .063 | 0    | .107 | .032 | .152 | .101 | .048 | .251 |
| .106 | .111 | .124 | .111 | .107 | 0    | .105 | .153 | .109 | .101 | .253 |
| .008 | .105 | .119 | .064 | .032 | .105 | 0    | .152 | .102 | .046 | .25  |
| .154 | .156 | .163 | .156 | .152 | .153 | .152 | 0    | .157 | .151 | .256 |
| .103 | .063 | .117 | .108 | .101 | .109 | .102 | .157 | 0    | .101 | .253 |
| .047 | .102 | .115 | .06  | .048 | .101 | .046 | .151 | .101 | 0    | .248 |
| .251 | .256 | .255 | .255 | .251 | .253 | .25  | .256 | .253 | .248 | 0    |

Tabla 1.1 Matriz de distancias generada contando las diferencias con naive.

#### b. JK69

|      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|
| 0    | .114 | .13  | .068 | .034 | .114 | .008 | .172 | .111 | .048 | .306 |
| .114 | 0    | .131 | .117 | .113 | .12  | .113 | .175 | .065 | .11  | .312 |
| .13  | .131 | 0    | .134 | .13  | .135 | .129 | .184 | .127 | .125 | .312 |
| .068 | .117 | .134 | 0    | .066 | .12  | .067 | .175 | .116 | .062 | .311 |
| .034 | .113 | .13  | .066 | 0    | .116 | .033 | .17  | .109 | .049 | .305 |
| .114 | .12  | .135 | .12  | .116 | 0    | .114 | .171 | .117 | .108 | .309 |
| .008 | .113 | .129 | .067 | .033 | .114 | 0    | .17  | .11  | .047 | .304 |
| .172 | .175 | .184 | .175 | .17  | .171 | .17  | 0    | .176 | .169 | .314 |
| .111 | .065 | .127 | .116 | .109 | .117 | .11  | .176 | 0    | .108 | .309 |
| .48  | .11  | .125 | .062 | .049 | .108 | .047 | .169 | .108 | 0    | .302 |
| .306 | .312 | .312 | .311 | .035 | .309 | .304 | .314 | .309 | .302 | 0    |

Tabla 1.2 Matriz de distancias generada con el modelo de JK69.

c. K71

|      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|
| 0    | .98  | .11  | .062 | .032 | .097 | .007 | .128 | .095 | .045 | .178 |
| .98  | 0    | .112 | .1   | .099 | .105 | .097 | .132 | .061 | .094 | .184 |
| .11  | .112 | 0    | .111 | .11  | .113 | .109 | .139 | .108 | .105 | .179 |
| .062 | .1   | .111 | 0    | .061 | .103 | .062 | .13  | .1   | .057 | .183 |
| .32  | .099 | .11  | .061 | 0    | .1   | .031 | .127 | .094 | .046 | .177 |
| .097 | .105 | .113 | .103 | .1   | 0    | .096 | .129 | .101 | .094 | .178 |
| .007 | .097 | .109 | .062 | .031 | .096 | 0    | .126 | .094 | .043 | .177 |
| .128 | .132 | .139 | .13  | .127 | .129 | .126 | 0    | .131 | .127 | .178 |
| .095 | .061 | .108 | .1   | .094 | .101 | .094 | .131 | 0    | .091 | .181 |
| .045 | .094 | .105 | .057 | .046 | .094 | .043 | .127 | .091 | 0    | .176 |
| .178 | .184 | .179 | .183 | .177 | .178 | .177 | .178 | .181 | .176 | 0    |

Tabla 1.3 Matriz de distancias generada con el modelo K71.

d. T92

|      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|
| 0    | .099 | .111 | .062 | .032 | .098 | .007 | .129 | .096 | .045 | .179 |
| .99  | 0    | .113 | .1   | .099 | .105 | .097 | .133 | .062 | .094 | .185 |
| .111 | .113 | 0    | .112 | .111 | .114 | .109 | .14  | .108 | .106 | .18  |
| .062 | .1   | .112 | 0    | .061 | .104 | .062 | .131 | .1   | .057 | .184 |
| .032 | .099 | .111 | .061 | 0    | .101 | .031 | .127 | .094 | .046 | .178 |
| .098 | .105 | .114 | .104 | .101 | 0    | .097 | .13  | .102 | .094 | .179 |
| .007 | .097 | .109 | .062 | .031 | .097 | 0    | .127 | .095 | .043 | .178 |
| .129 | .133 | .14  | .131 | .127 | .13  | .127 | 0    | .132 | .127 | .179 |
| .096 | .062 | .108 | .1   | .094 | .102 | .095 | .132 | 0    | .092 | .182 |
| .045 | .094 | .106 | .057 | .046 | .094 | .043 | .127 | .092 | 0    | .178 |
| .179 | .185 | .18  | .184 | .178 | .179 | .178 | .179 | .182 | .178 | 0    |

Tabla 1.4 Matriz de distancias generada con el modelo T92.

- 2. Árboles
  - a. Neighbor Joining
    - i. Naive

### a. Neighbor Joining

- i. Naive

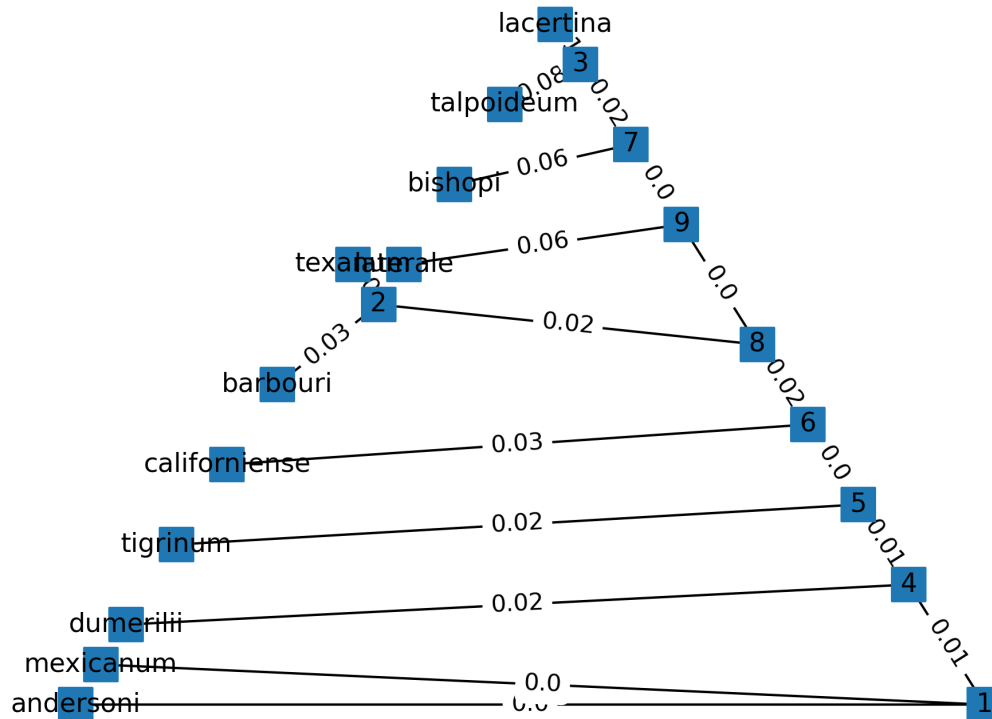


Imagen 1.1 Árbol generado con Neighbor Joining de la matriz Naive.

- ii. Jukes-Cantor

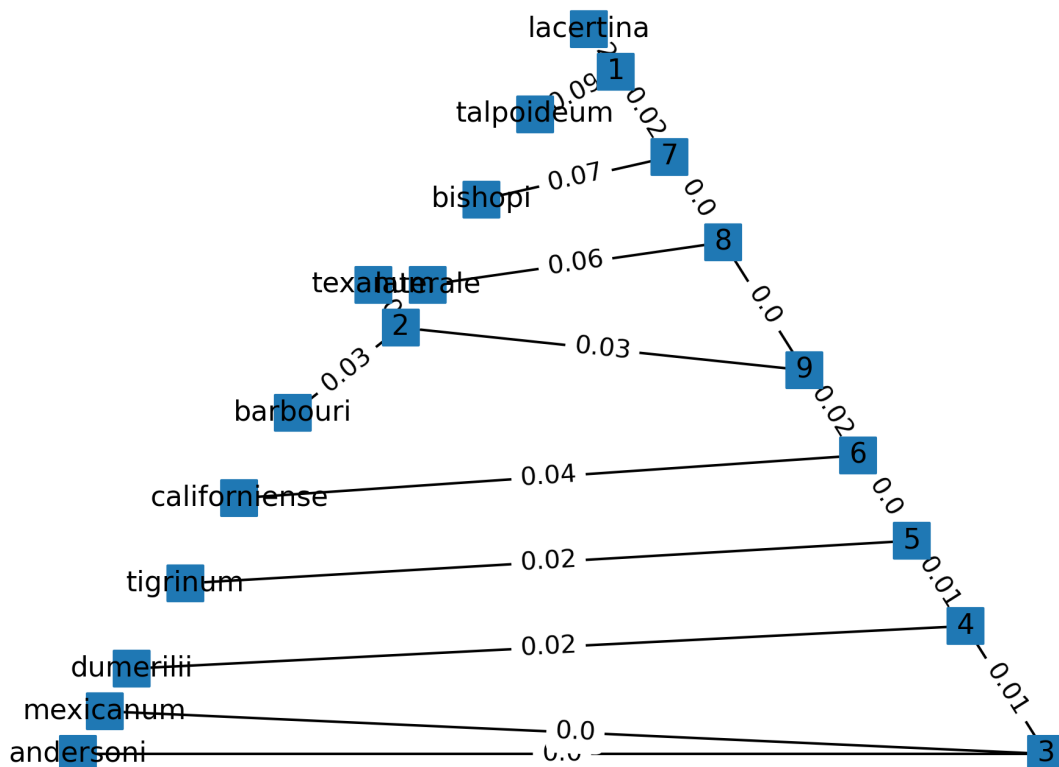


Imagen 1.2 Árbol generado con Neighbor Joining de la matriz JK.



iii. Motoo Kimura

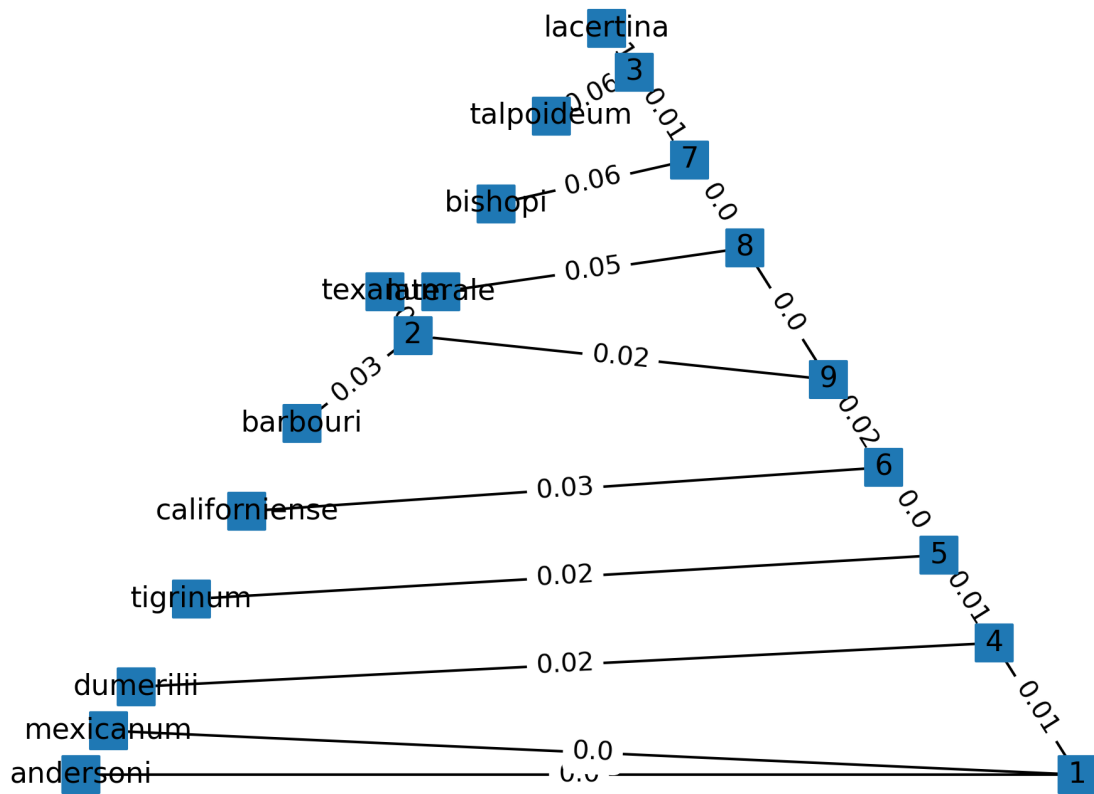


Imagen 1.3 Árbol generado con Neighbor Joining de la matriz MK.

iv. Tamura

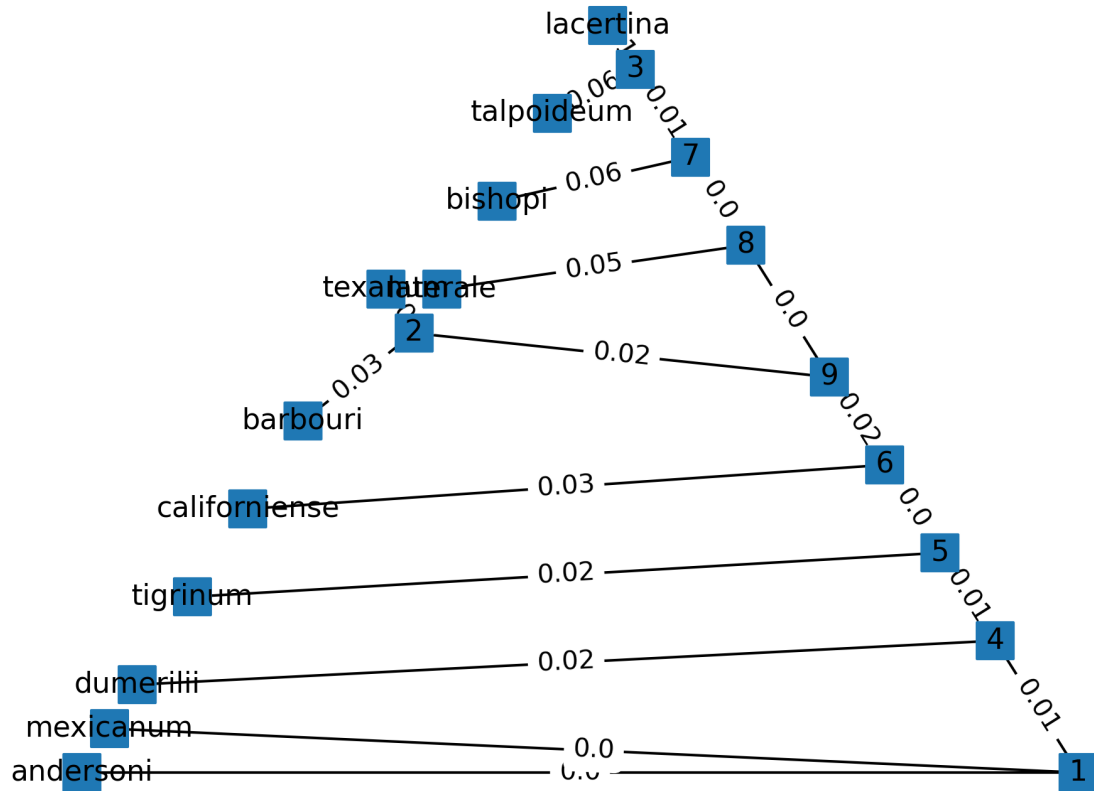


Imagen 1.4 Árbol generado con Neighbor Joining de la matriz Tamura.

b. UPGMA

Las siguientes tablas fueron generadas con [Interactive Tree Of Life \(iTOL\)](#), versión 5.

i. Naive

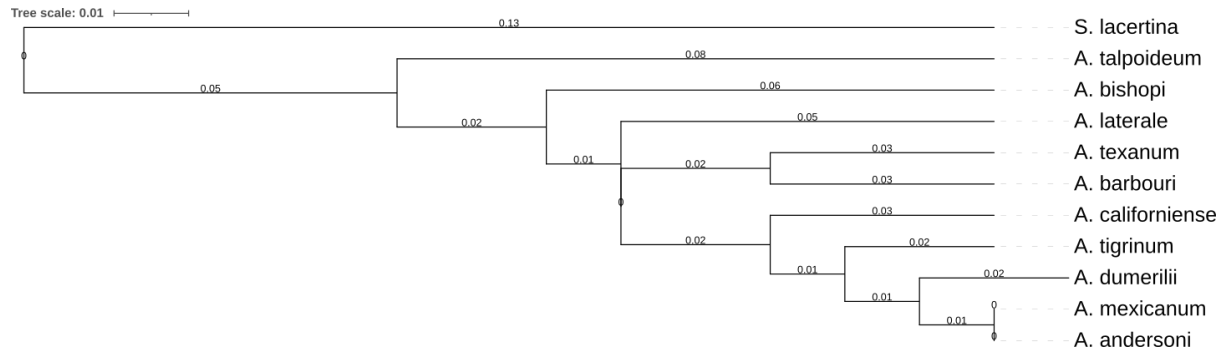


Imagen 2.1 Árbol generado con UPGMA de la matriz Naive.

ii. Jukes-Cantor

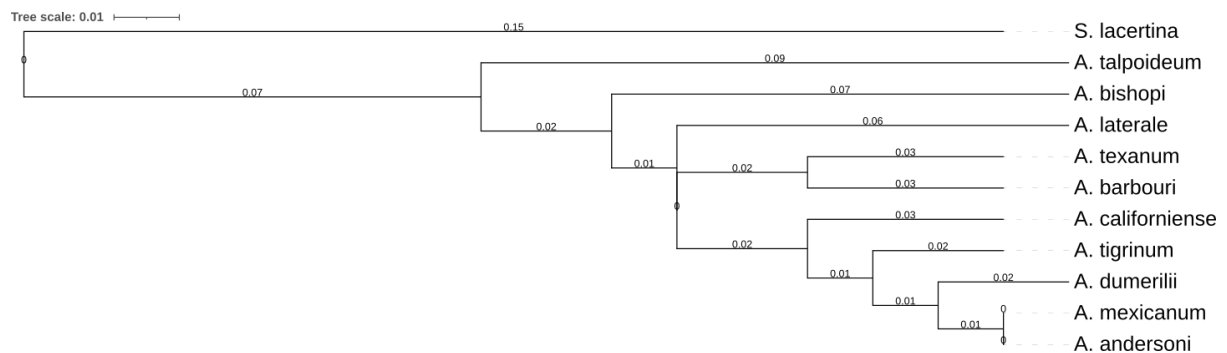


Imagen 2.2 Árbol generado con UPGMA de la matriz JK.

iii. Motoo Kimura

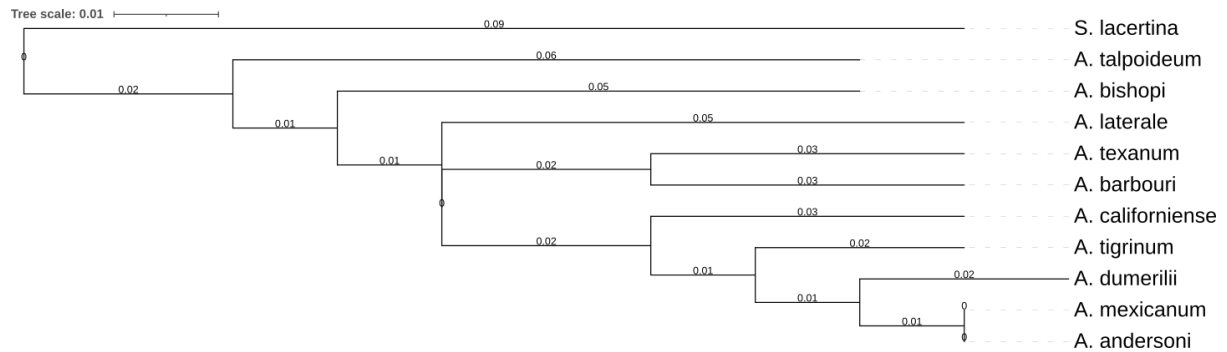


Imagen 2.3 Árbol generado con UPGMA de la matriz MK.

iv. Tamura

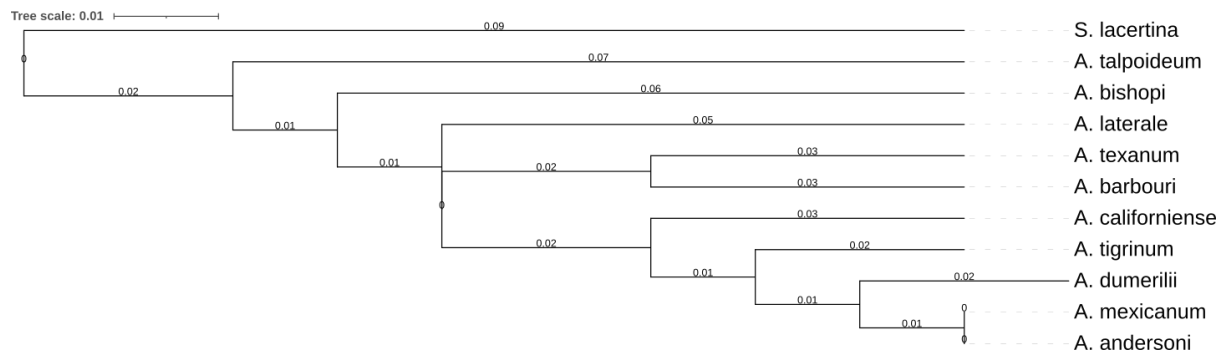


Imagen 2.4 Árbol generado con UPGMA de la matriz Tamura.

## Discusión

Además del alto costo computacional de los algoritmos (el tiempo de ejecución para los árboles que resultaron de las alineaciones de Needleman-Wunsch fue de aproximadamente 10 horas, y de MSA con la heurística de la estrella la salida tardó al momento de subir esto lleva 16 horas), nos hubiera gustado experimentar con más modelos (en particular GTR) y métodos para generar árboles no basados en las matrices. Todos los árboles generados, aunque varían en sus distancias, tienen la misma topología, por lo que del lado de comparación de algoritmos no es un ejercicio particularmente ilustrativo. Tiene más mérito pensándolo como un ejercicio sobre la fiabilidad de los algoritmos bioinformáticos. Y a pesar de que no fue posible conocer el score de parsimonia debido a que carecíamos de la matriz de alineación múltiple, el hecho de que todas las topologías sean las mismas, nos ayuda a generar un árbol consenso.

Observando los árboles obtenidos con respecto a los previamente publicados, nos damos cuenta que dos de ellos, el de Shaffer y el de Cole, no son muy comparables puesto que utilizan muchas especies que no se incluyen en nuestro análisis y varias de las que sí utilizamos no las incluyen, pero en el de Shaffer, al igual que en el nuestro, *A. dummerillii* y *A. andersoni*, se encuentran relativamente cercanos, mientras que en el árbol de Cole, basado en caracteres morfológicos, *A. mexicanum* y *A. dummerillii* mantienen la misma relación que el nuestro, además es interesante que hay 12 taxones que no incluimos en nuestro árbol pero en este se encuentran entre *A. mexicanum* y *A. andersoni*, estos dos en nuestra hipótesis se consideran grupos hermanos, por lo que sería interesante incluirlos a los otros 12 en un análisis de gm y constatar si se mantiene la misma topología del árbol de caracteres morfológicos. Por otro lado, el árbol generado a partir del análisis de varios genes de gm, coincide topológicamente de forma idéntica con nuestro árbol, solo que nosotros consideramos 5 miembros más de *Ambystoma*, pero en esencia es un subconjunto de nuestro árbol filogenético.

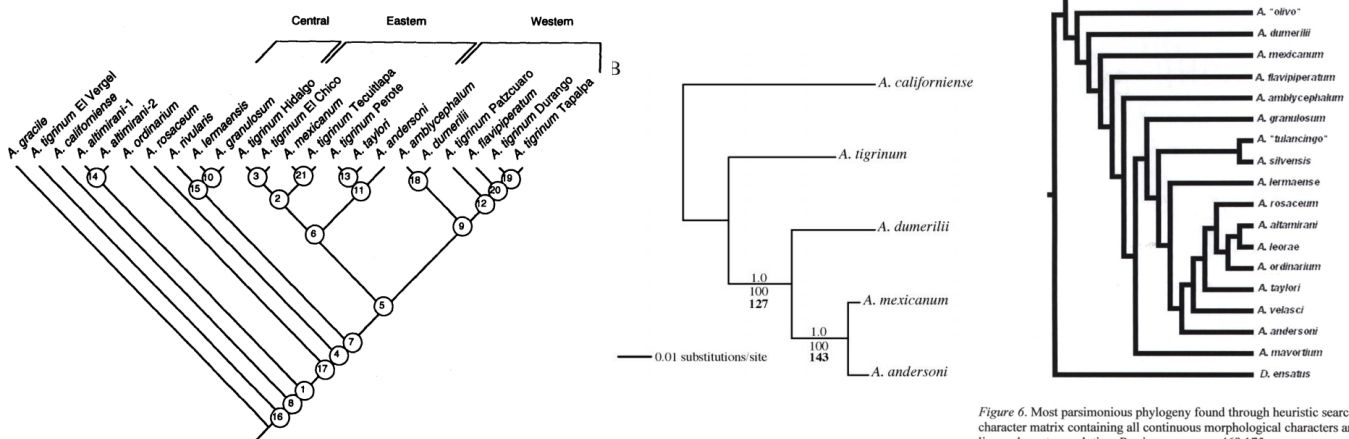


Figure 6. Most parsimonious phylogeny found through heuristic search of character matrix containing all continuous morphological characters and based on linear character evolution. Parsimony score = 468.175.

Imagen 3. Árboles previamente publicados por otros autores. A) árbol basado en datos moleculares de aloenzimas (Shaffer,1993), B) árbol basado en genes de gm (Samuels 2005), C) árbol basado en caracteres morfológicos (Cole, 2007).

Por otra parte, logramos vislumbrar un patrón filogeográfico, de manera que a grandes rasgos, entre más cercanos topológicamente se encuentran en el árbol dos especies, más cercanas están geográficamente, distribuyéndose en el sur de Norteamérica los taxones de

ramas más cortas y en el norte los de mayores ramas. Observamos las distribuciones de las especies analizadas que se encuentran en la base de datos Global Biodiversity Information Facility ([GBIF](https://www.gbif.org/)), hay que tomar en cuenta que es necesario curar los datos y filtrar los registros que no estén verificados pero en términos generales así se muestran sus distribuciones. También hay que considerar ciertos casos especiales como es el taxón *A. tigrinum* que anteriormente se consideraba una sola especie, y después pasó a nombrarse como el complejo tigrinum, que en realidad eran por lo menos 4 especies, entonces habría que esclarecer la distribución de cada una de estas.

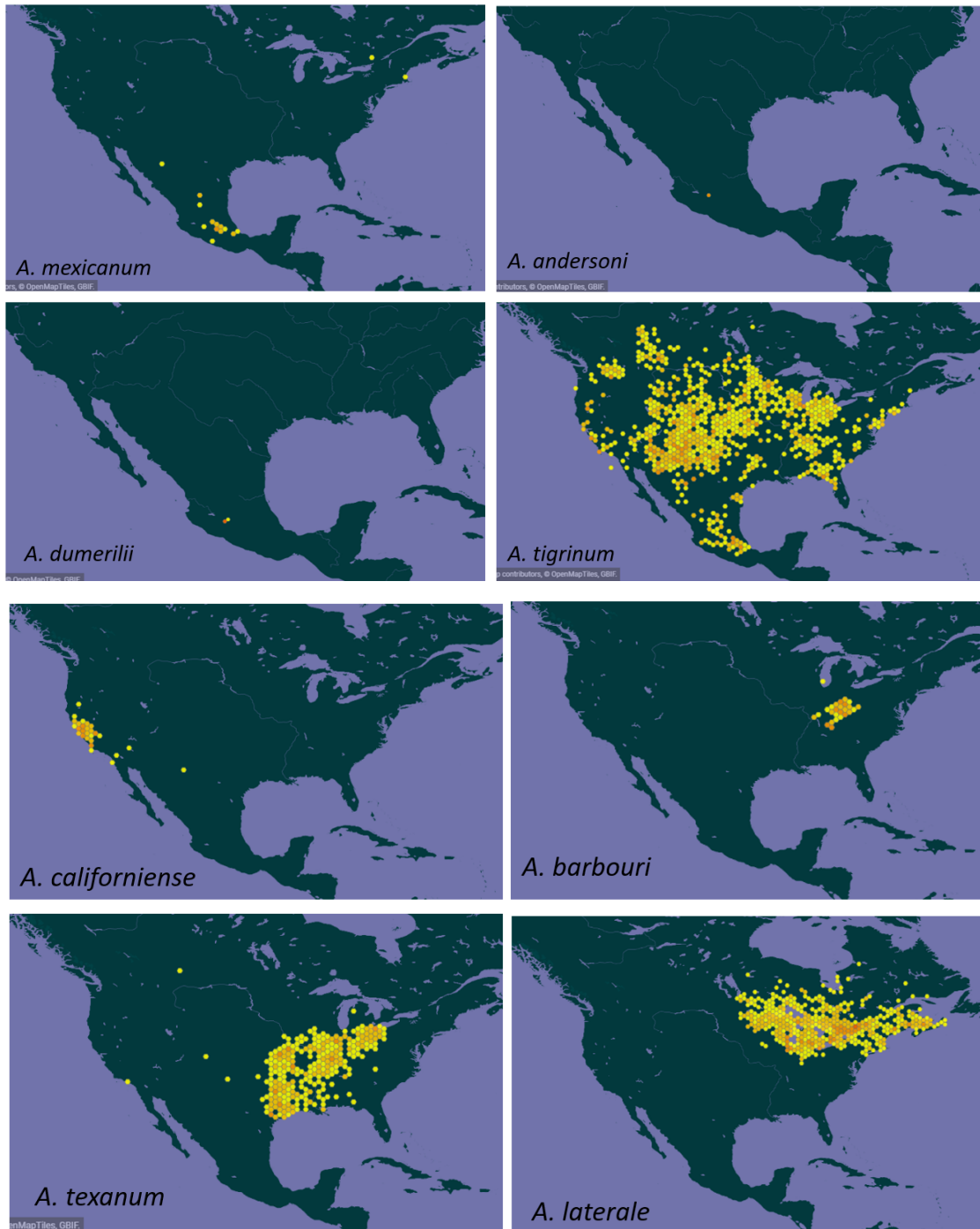




Imagen 4 Distribución de cada una de las especies de *Ambystoma* analizadas tomadas de GBIF

## Conclusión

Para poder llevar un análisis más profundo de los métodos y algoritmos utilizados, además de agregar más modelos para calcular las distancias, y otro tipo de métodos para construir los árboles, como máxima verosimilitud, sería útil usar otros datos o agregar más especies, ya que las topologías son prácticamente iguales en todos, solo cambian las distancias. Esto nos habla sobre lo confiable que es nuestra reconstrucción de la filogenia, pero no nos es muy útil para poder discernir cómo afectan las herramientas escogidas. Quizá con otro tipo de datos también podríamos comparar cómo afecta acoplar el registro fósil a nuestros cálculos.

Por otra parte, los datos que escogimos nos permitieron identificar un patrón entre la cercanía filogenética y la distribución geográfica de las especies.

## Referencias

- Arenas M. (2015). Trends in substitution models of molecular evolution. *Frontiers in genetics*, 6, 319. <https://doi.org/10.3389/fgene.2015.00319>
- Ávila, V. González, T. González, A. Trejo, M. (2021) El género *Ambystoma* en México ¿Qué son los ajolotes? *CIENCIA ergo-sum*, Revista Científica Multidisciplinaria de Prospectiva, vol. 28, núm. 2, 1, 2021 <https://doi.org/ces.v28n2a10>
- Baum, D. (2008). Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. *Nature Education* 1(1):190. Recuperado en mayo de 2022 de <https://www-nature-com.pbidi.unam.mx:2443/scitable/topicpage/reading-a-phylogenetic-tree-the-meaning-of-41956/>
- Cristianini, N., & Hahn, M. W. (2006). *Introduction to Computational Genomics: A Case Studies Approach*. Cambridge University Press.
- Cole, L. (2007). A phylogeny of Mexican *Ambystoma* Salamanders (Caudata: Ambystomatidae) from Larval Characters. Tesis para el grado de maestro en

ciencias. Charles E. Schmidt College of Science. Recuperado en junio de 2023 de <https://fau.digital.flvc.org/islandora/object/fau%3A34314/datastream/OBJ/view>

Frost, Darrel R. (2023). Amphibian Species of the World: an Online Reference. Version 6.1 (Date of access). Electronic Database accessible at <https://amphibiansoftheworld.amnh.org/index.php>. American Museum of Natural History, New York, USA. doi.org/10.5531/db.vz.0001. Recuperado en mayo de 2023 de <https://amphibiansoftheworld.amnh.org/Amphibia/Caudata/Ambystomatidae>

John Roth Laboratory. (2005, May 27). Distances. Recuperado Mayo, 2023. <https://rothlab.ucdavis.edu/genhelp/distances.html>

Letunic I and Bork P (2021) Nucleic Acids Res doi: [10.1093/nar/gkab301](https://doi.org/10.1093/nar/gkab301) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation

Morrone, J. (2000). Los árboles filogenéticos de Darwin (1859) a Hennig (1950). Museo; no. 14. Recuperado en mayo de 2022 de <http://sedici.unlp.edu.ar/handle/10915/49425>

"Pylogeny." Github, Diciembre, 2021. <https://github.com/pylogeny/cluster>

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution, 4(4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>

Salarian, Zahra. "Bioinformatics-Mini-Projects." Github, Agosto 2022, Irán. <https://github.com/zahrasalarian/Bioinformatics-Mini-Projects>

Samuels, A. K., Weisrock, D. W., Smith, J. J., France, K. J., Walker, J. A., Putta, S., & Voss, S. R. (2005). Transcriptional and phylogenetic analysis of five complete ambystomatid salamander mitochondrial genomes. Gene, 349, 43–53. <https://doi.org/10.1016/j.gene.2004.12.037>

SEMARNAT, 2018. Programa de Acción para la Conservación de las Especies *Ambystoma* spp, SEMARNAT/CONANP, México (Año de edición 2018). Recuperado en mayo de 2023 de [https://www.gob.mx/cms/uploads/attachment/file/444128/PACE\\_Ambystoma2.pdf](https://www.gob.mx/cms/uploads/attachment/file/444128/PACE_Ambystoma2.pdf)

Sever, D. M. 1992. Comparative anatomy and phylogeny of the cloacae of salamanders (Amphibia: Caudata). VI. Ambystomatidae and Dicamptodontidae. Journal of Morphology 212: 305–322

Shaffer, B. (1993) Phylogenetics of Model Organisms: The Laboratory Axolotl, *Ambystoma Mexicanum*, Systematic Biology, Volume 42, Issue 4, Pages 508–522, <https://doi.org/10.1093/sysbio/42.4.508>

Staton, Joseph L. (2015) "Understanding phylogenies: Constructing and Interpreting Phylogenetic trees", Journal of the South Carolina Academy of Science: Vol. 13: Iss. 1,

Stavrinides, J., & Ochman, H. (2009). Phylogenetic Methods. Encyclopedia of Microbiology, 247–260 .[https:// doi:10.1016/b978-012373944-5.00272-8](https://doi.org/10.1016/b978-012373944-5.00272-8)

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular biology and evolution, 10(3), 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>

Tihen, J. A. 1958. Comments on the osteology and phylogeny of ambystomatid salamanders. Bulletin of the Florida Museum of Natural History. Biological Sciences 3: 1–50.

Wang, L., & Jiang, T. (1994). On the complexity of multiple sequence alignment. Journal of Computational Biology, 1(4), 337–348. <https://doi.org/10.1089/cmb.1994.1.337>

Zhang, P., & Wake, D. B. (2009). Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. Molecular phylogenetics and evolution, 53(2), 492–508. <https://doi.org/10.1016/j.ympev.2009.07.010>