



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
FACULTAD DE CIENCIAS

Variación de especies: Mamut Lanudo
(*Mammuthus primigenius*)

GENÓMICA COMPUTACIONAL

15 de Diciembre del 2022

GUERRERO IZQUIERDO CECILIA

GUTIÉRREZ PRIETO LUISA ELENA



Índice

1. Abstract	3
2. Introducción	3
3. Pregunta de investigación	5
4. Hipótesis	5
5. Objetivo General	5
6. Objetivos Particulares	6
7. Métodos	6
8. Resultados	8
9. Discusión	13
10. Conclusión	13
11. Referencias	14

1. Abstract

Las diferencias genéticas entre especies son responsables de muchas de las diferencias en el comportamiento, morfología y fisiología que observamos entre especies. La variación entre especies también puede comenzar a informarnos sobre las relaciones entre éstas porque las especies más estrechamente relacionadas tendrán en promedio secuencias de ADN más similares, el examen de estas secuencias puede informarnos sobre cómo ha procedido la evolución durante millones de años. Si ahora queremos saber cuán lejanos son los mamuts (*Mammuthus primigenius*) con los elefantes actuales; africano, *Loxodonta africana* y dos indios, *Elephas maximus* y *Elephantulus sp. VB001*, necesitamos comparar secuencias entre especies. En el presente artículo realizamos un alineamiento de ciertos nucleótidos del mtDNA completo de los organismos anteriormente mencionados. Posteriormente, se analizan las distancias entre las secuencias por medio de diversos recursos como árboles filogenéticos, matrices de distancias de Jukes-Cantor y p-distance.

2. Introducción

La variación en el ADN se acumula a través de mutaciones, que son errores cometidos por la maquinaria celular que luego se codifican en el genoma. También en la configuración exacta de las secuencias de ADN puede introducirse mediante recombinación (cuando el organismo es diploide, es decir, tiene dos copias de cada cromosoma); este fenómeno es consecuencia de la reproducción sexual. El análisis de variación, a través de la comparación de dos o más secuencias, puede proporcionarnos una gran cantidad de información sobre varios aspectos de la estructura, función e historia genómicas en cualquier especie. [1]. En ocasiones, para el estudio de estas variaciones se emplea un árbol filogenético o filogenia, los cuales son útiles para organizar el conocimiento de la diversidad biológica, para estructurar clasificaciones y para proporcionar información sobre eventos que ocurrieron durante la evolución.

Mammuthus, *Elephas* y *Loxodonta* (familia *Elephantidae*, subfamilia *Elephantinae*) son géneros estrechamente relacionados que evolucionaron en el Plioceno africano. Pequeñas poblaciones aisladas del mamut lanudo *M. primigenius* sobrevivieron hasta mediados del Holoceno. Los análisis morfológicos han arrojado filogenias contradictorias para *M. primigenius*, *Elephas maximus* y *Loxodonta africana*. Los caracteres dentales sugieren una relación más estrecha entre *M. primigenius* y *E. maximus*, la morfología de la punta del tronco apoya una agrupación de *M. primigenius* y *L. africana*, y los caracteres inmunológicos y de estructura del pelo no pudieron resolver con seguridad la filogenia de estos tres taxones. [4]

Los mamuts tenían muchas adaptaciones anatómicas que minimizaban la pérdida de calor en su entorno hostil, como pelaje espeso, orejas pequeñas y colas pequeñas (en comparación con los elefantes modernos), y una gruesa capa de grasa debajo de la piel para reducir la pérdida de calor y posiblemente servir como fuente de calor o depósito de grasa para el invierno. [5] Las musarañas elefante (también llamadas sengis, orden *Macroscelidea*) son mamíferos insectívoros de cuerpo pequeño que se distribuyen en África. Actualmente se reconocen quince especies, de las cuales 9 se encuentran en la subregión del sur de África. En 2008, se publicó la primera secuenciación del genoma completo (WGS) (< 1) de un mamut lanudo. [6]

Por otra parte, la distancia genética entre dos secuencias homólogas se define como el número de sustituciones que se han acumulado entre ellas desde que se separaron de un ancestro común. Los nucleótidos se pueden dividir en purinas y pirimidinas. Las sustituciones dentro del mismo grupo se llaman "transiciones", las que se dan entre grupos se denominan "transversiones". [1] P-distance o distancia de Hamming o grado de divergencia: Consiste en contar la frecuencia relativa de cambios entre dos secuencias de un alineamiento múltiple. Si tenemos un alineamiento de longitud N con n diferencias la p-distance se define como:

$$p = n/N$$

[3]

En 1969, Thomas Jukes y Charles Cantor propusieron un modelo probabilístico para corregir el número observado de diferencias para tener en cuenta la posibilidad de sustituciones múltiples. [1] Asume que la probabilidad de mutar un nucleótido por otro es independiente de la posición de dicho nucleótido y del nucleótido en sí: La probabilidad de cambiar A por C , G o T es idéntica, $\frac{a}{3}$, de igual forma para C , G y T . A continuación mostramos la famosa corrección de Jukes-Cantor (JC);

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

Esto dice que el verdadero número de sustituciones por sitio entre dos secuencias (K) se puede estimar a partir de la fracción observada de sitios que difieren (d). [3]

3. Pregunta de investigación

¿Cuál de los siguientes organismos: Elefante asiático (*Elephas maximus*), Elefante africano de sabana (*Loxodonta africana*) y musarañas elefantes (*Elephantulus sp*) parece estar más estrechamente relacionado con el mamut lanudo (*Mammuthus primigenius*)?

4. Hipótesis

Con la pregunta de investigación se plantea lo siguiente: El genoma del Elefante asiático será el más parecido al genoma del Mamut lanudo.

5. Objetivo General

Identificar la secuencia con menor distancia respecto a la secuencia del mamut lanudo (*M. primigenius*).

6. Objetivos Particulares

Calcular la distancia entre estas secuencias, pero primero se deben eliminar las regiones hipervariables (HVR) y luego hacer la alineación global de las mismas. A continuación se describen estos pasos:

- Mantener en las secuencias sólo los siguientes nucleótidos:

Mamut Lanudo: 1 a 15421

Elefante Africano: 1 a 15417

Elefante Asiático: 1 a 15419

- Hacer el alineamiento múltiple de las secuencias.
- Calcular la distancia entre las secuencias alineadas.

7. Métodos

En esta investigación, analizamos las secuencias del mamut lanudo (*Mammuthus primigenius*), las de los elefantes actuales (*Loxodonta africana* y *Elephas maximus*) y la musaraña elefante (*Elephantulus sp. VB001*). Usando algoritmos de MUSCLE, identificamos las distancias mediante p-distance y Jukes-Cantor para determinar qué organismo es el más parecido genómicamente al mamut lanudo. En el proceso, realizamos el árbol filogenético de estos organismos. Los genomas analizados fueron excluidos de las regiones hipervariables (HVR), así que comparamos específicamente los nucleótidos: mamut; 1 a 15421, elefante africano; 1 al 15417 y del elefante indú y musaraña del 1 al 15419.

Por lo dicho anteriormente, se siguieron los siguientes pasos:

1. *Selección de las secuencias biológicas a analizar*

Como se ha mencionado anteriormente, analizaremos las secuencias de los mamuts, *M. primi-*

genius, y los elefantes actuales; africano, *L. africana* y dos indios, *E. maximus* y el mamífero *Elephantulus sp. VB001*.

2. *Obtención de datos*

Las secuencias de los genomas completos en archivo fasta fueron extraídas del *NCBI* (National Center for Biotechnology Information), <https://www.ncbi.nlm.nih.gov/>, específicamente la secuencia de *Mammuthus primigenius* se obtuvo de https://www.ncbi.nlm.nih.gov/nuccore/NC_007596.2/ (13 de diciembre del 2022, fecha último acceso) (número de acceso de la secuencia: NC_007596.2). La secuencia de *Loxodonta africana* se obtuvo de https://www.ncbi.nlm.nih.gov/nuccore/NC_000934.1/ (13 de diciembre del 2022, fecha último acceso) (número de acceso de la secuencia: NC_000934.1). La secuencia de *Elephas maximus* se obtuvo de https://www.ncbi.nlm.nih.gov/nuccore/NC_005129.2/ (13 de diciembre del 2022, fecha último acceso) (número de acceso de la secuencia: NC_005129.2). La secuencia de *Elephas maximus* se obtuvo de https://www.ncbi.nlm.nih.gov/nuccore/NC_005129.2/ (13 de diciembre del 2022, fecha último acceso) (número de acceso de la secuencia: NC_004921.1).

3. *Alineamiento Múltiple de Secuencias*

El alineamiento múltiple de las secuencias se realizó en MUSCLE version 11.

4. *Selección del modelo estadístico de evolución molecular de las correspondientes secuencias*

Elegimos la p-distance, distancia de Hamming o grado de divergencia y el modelo de Jukes-Cantor.

5. *Representación gráfica: Construcción de Árboles Filogenéticos*

El árbol filogenético se desarrolló a partir de los alineamientos obtenidos por MUSCLE.

8. Resultados

■ *Alineamiento Múltiple de Secuencias*



Fig. 1: Alineamiento de las secuencias con MUSCLE (primeros nucleótidos)

■ *Calculo de distancias:* Para el calculo de las distancias se utilizo pairwise distance, y bootstrap.

1. Modelo de sustitución de nucleótidos **Jukes-Cantor**

	1	2	3	4
1. ref NC 007596.2 :1-15421 Mammuthus primigenius mitochondrion complete genome		0.0018	0.0054	0.0017
2. ref NC 000934.1 :1-15417 Loxodonta africana mitochondrion complete genome	0.0511		0.0054	0.0018
3. ref NC 004921.1 :1-15419 Elephantulus sp. VB001 mitochondrion complete genome	0.3213	0.3266		0.0054
4. ref NC 005129.2 :1-15419 Elephas maximus mitochondrion complete genome	0.0471	0.0519	0.3212	

Fig. 2: Matriz de distancias


```

Description
Analysis =====
  Analysis                               = =====
  Scope                                 = Pairs of taxa
  Estimate Variance                     = =====
  Variance Estimation Method            = Bootstrap method
  No. of Bootstrap Replications          = 1000
  Substitution Model                    = =====
  Substitutions Type                    = Nucleotide
  Model/Method                          = Jukes-Cantor model
  Substitutions to Include              = All
  Rates and Patterns                    = =====
  Rates among Sites                     = Uniform Rates
  Pattern among Lineages                = Same (Homogeneous)
  Data Subset to Use                    = =====
  Gaps/Missing Data Treatment           = Pairwise deletion
  Select Codon Positions                 = 1st,2nd,3rd,Non-Coding

[1] #ref|NC_007596.2|:1-15421_Mammuthus_primigenius_mitochondrion_complete_genome
[2] #ref|NC_000934.1|:1-15417_Loxodonta_africana_mitochondrion_complete_genome
[3] #ref|NC_004921.1|:1-15419_Elephantulus_sp._VB001_mitochondrion_complete_genome
[4] #ref|NC_005129.2|:1-15419_Elephas_maximus_mitochondrion_complete_genome

[      1      2      3      4 ]
[1]
[2] 0.0511
[3] 0.3213 0.3266
[4] 0.0471 0.0519 0.3212

```

Fig. 3: Summary. Matriz de distancias utilizando Jukes-Cantor

2. Modelo de sustitución de nucleótidos **p-distance**

	1	2	3	4
1. ref NC_007596.2 :1-15421 Mammuthus primigenius mitochondrion complete genome		0.0017	0.0035	0.0016
2. ref NC_000934.1 :1-15417 Loxodonta africana mitochondrion complete genome	0.0494		0.0035	0.0018
3. ref NC_004921.1 :1-15419 Elephantulus sp. VB001 mitochondrion complete genome	0.2613	0.2648		0.0035
4. ref NC_005129.2 :1-15419 Elephas maximus mitochondrion complete genome	0.0457	0.0501	0.2613	

Fig. 4: Matriz de distancias

```

Description
Analysis =====
  Analysis                               = =====
  Scope                                 = Pairs of taxa
  Estimate Variance                     = =====
  Variance Estimation Method           = Bootstrap method
  No. of Bootstrap Replications         = 1000
  Substitution Model                   = =====
  Substitutions Type                   = Nucleotide
  Model/Method                         = p-distance
  Substitutions to Include              = d: Transitions + Transversions
  Rates and Patterns                   = =====
  Rates among Sites                    = Uniform Rates
  Pattern among Lineages                = Same (Homogeneous)
  Data Subset to Use                   = =====
  Gaps/Missing Data Treatment           = Pairwise deletion
  Select Codon Positions                = 1st,2nd,3rd,Non-Coding

[1] #ref|NC_007596.2|:1-15421_Mammuthus_primigenius_mitochondrion_complete_genome
[2] #ref|NC_000934.1|:1-15417_Loxodonta_africana_mitochondrion_complete_genome
[3] #ref|NC_004921.1|:1-15419_Elephantulus_sp._VB001_mitochondrion_complete_genome
[4] #ref|NC_005129.2|:1-15419_Elephas_maximus_mitochondrion_complete_genome

[      1      2      3      4 ]
[1]
[2] 0.0494
[3] 0.2613 0.2648
[4] 0.0457 0.0501 0.2613

```

Fig. 5: Summary. Matriz de distancias utilizando p-distance

- Árbol filogenético.

1. Método de construcción de árboles **Neighbour Joining** y modelo de sustitución **Jukes-Cantor**



Fig. 6: Árbol Original: NJ



Fig. 7: Árbol Bootstrap Consensus: NJ

2.

3. Método de construcción de árboles **UPGMA** y modelo de sustitución **p-distance**



Fig. 8: Árbol Original: UPGMA



Fig. 9: Árbol Bootstrap Consensus: UPGMA

9. Discusión

En general, en ambas matrices se obtuvieron los mismos resultados, es decir, existe poca distancia (diferencia de nucleótidos) entre el mamut lanudo (*Mammuthus primigenius*), el elefante asiático (*Elephas maximus*) y el elefante africano (*Loxodonta africana*) siendo este último el menos relacionado con los otros dos, por otra parte, la musaraña (*elephantulus sp*) está ligeramente relacionado con respecto a las anteriores especies, sin embargo, en el árbol filogenético bootstrap consensus con el modelo de sustitución de nucleótidos Jukes-Cantor y el método de construcción de árboles Neighbour Joining se mantiene de igual manera que el mamut lanudo y el elefante asiático son los más parecidos genéticamente, pero resalta que la musaraña tiene un ancestro en común con las anteriores dos especies, mientras que el elefante africano es el más alejado, por otra parte, el árbol original con los mismos parámetros (Jukes-Cantor y Neighbour Joining), indica que el mamut y el elefante asiático son los más cercanos genéticamente, además de que ambas especies con el elefante africano tienen un mismo ancestro en común, mientras que la musaraña es la especie que presenta mayor distancia, sin embargo, el valor del nodo de soporte en ambos árboles es de 51, esto implica que el nodo no tiene un buen soporte. Por otro lado, aplicando el modelo de sustitución de nucleótidos de p-distance y UPGMA, tenemos (de igual manera) los mismos resultados en ambos árboles (tanto en árbol original como bootstrap) lo que se obtuvo en el árbol original utilizando Neighbour Joining y Jukes-Cantor, además cabe resaltar que el valor de soporte fue 96 para ambos árboles.

10. Conclusión

Por los resultados obtenidos en ambas matrices, se puede decir que el mamut lanudo y el elefante asiático son los más parecidos genéticamente. Recientemente se supo que el mamut lanudo y el elefante asiático se separaron después de ramificarse a partir del elefante africano, ya que anteriormente (antes del 2008) se consideraba un problema que no tenía solución debido a que no existía la codificación completa del genoma del mamut.

Por otra parte, podemos descartar el Árbol bootstrap consensus que utiliza Neighbour Joining y Jukes-Cantor, ya que si tomamos en cuenta las matrices de distancias existe una contradicción, además de que su nodo de soporte es bajo, por lo cual este árbol no tiene mucho sentido. Podemos decir que el mejor árbol filogenético para representar la variación genética del mamut es UPGMA con p-distance, asimismo de que está sustentando por bootstrap

11. Referencias

Referencias

- [1] Cristianini, N. & Hahn, M. W. (2006). Introduction to Computational Genomics: A Case Studies Approach. Cambridge University Press.
- [2] Merino, E. (2011, 28 enero). Multiple sequence alignments. pdf. Recuperado 12 de diciembre de 2022, de https://www.ccg.unam.mx/~vinuesa/tlem/docs/Multiple_Sequence_Alignments.pdf
- [3] Pérez Hurtado, I. & Dpto. Ciencias de la Computacion e Inteligencia Artificial Universidad de Sevilla. (2014). TECNICAS INTELIGENTES EN BIOINFORMÁTICA: Evolución Molecular y Filogenia. pdf. Recuperado 12 de diciembre de 2022, de <https://www.cs.us.es/~marper/docencia/bioinformatics/temas/ignacio-2.pdf>
- [4] Rogaev E., Moliaka Y., Malyarchuk B., Kondrashov F., Derenko M., Chumakov I., Grigorenko A. (2006 Feb 7). Complete Mitochondrial Genome and Phylogeny of Pleistocene Mammoth *Mammuthus primigenius*. PLOS Biologue <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040073>
- [5] Smith, S. D., Kawash, J. K., Karaikos, S., Biluck, I., & Grigoriev, A. (2017). Evolutionary adaptation revealed by comparative genome analysis of woolly mammoths and elephants. DNA Research, 24(4), 359–369. <https://doi.org/10.1093/dnares/dsx007>



- [6] Smith, H. A., Robinson, T. J., Watson, J., & van Vuuren, B. J. (2008). A new species of elephant-shrew (Afrotheria: Macroscelidea: Elephantulus) from South Africa. *Journal of Mammalogy*, 89(5), 1257–1268. <https://doi.org/10.1644/07-mamm-a-254.1>