

Genómica Computacional

Proyecto: Alineamiento de secuencias e información mutua entre el gen mitocondrial humano y animales modelo

Sofía Hernández Galván
Cecilia Elizabeth Servín Ornelas
Eduardo Alexis Valencia Dorantes

14 de diciembre de 2022

Introducción

El ADN mitocondrial animal es un genoma extracromosómico pequeño, típicamente de aproximadamente 16 kb de tamaño. Con pocas excepciones, todos los genomas mitocondriales animales contienen los mismos 37 genes: dos para ARNr, 13 para proteínas y 22 para ARNt. El estudio de estos genomas sirve como modelo para la evolución del genoma. Además, la comparación de los arreglos de genes mitocondriales animales se ha convertido en un medio muy poderoso para inferir relaciones evolutivas antiguas, ya que los reordenamientos parecen ser eventos únicos, generalmente raros, que es poco probable que surjan de forma independiente en linajes evolutivos separados.

Hay dos especies de gorilas en el mundo: el gorila oriental y el gorila occidental. El gorila de montaña es una subespecie del gorila oriental. Compartimos alrededor del 98 % de nuestro ADN con gorilas. Esto significa que la exposición a enfermedades humanas, incluso un resfriado, puede tener un impacto potencialmente perjudicial en los gorilas, ya que son muy similares genéticamente a nosotros, pero no han desarrollado las inmunidades necesarias.

Los bonobos y los chimpancés se parecen mucho y ambos comparten el 98,7 % de su ADN con los humanos, lo que convierte a las dos especies en nuestros parientes vivos más cercanos. Los bonobos suelen ser un poco más pequeños, delgados y oscuros que los chimpancés. Su sociedad también es diferente: los grupos bonobo tienden a ser más pacíficos y están dirigidos por mujeres. También mantienen relaciones y dirimen conflictos a través del sexo.

El cerdo (*Sus scrofa*) es un miembro de los artiodáctilos, o mamíferos de pezuña hendida, que son un clado evolutivo distinto de los primates y roedores. Los cerdos existen tanto en poblaciones salvajes como domesticadas que tienen distintos fenotipos y cariotipos. *Sus scrofa* es un organismo modelo importante para la investigación en salud debido a los paralelismos con los humanos. Los cerdos son omnívoros y su fisiología digestiva es similar a la de los humanos. También existen similitudes entre los humanos y los cerdos en la función renal, la estructura vascular y la frecuencia respiratoria. Los cerdos también son importantes desde el punto de vista agrícola, ya que la carne de cerdo es una fuente líder de proteínas en todo el mundo.

El hurón es una especie de mamífero de la familia Mustelidae. En América del Norte, el hurón se ha convertido en una opción cada vez más importante como mascota doméstica, con más de cinco millones solo en los Estados Unidos. La legalidad de la propiedad de hurones varía según la ubicación. En Nueva Zelanda y algunos otros países, se aplican restricciones debido al daño causado a la fauna nativa por colonias salvajes de híbridos turón-hurón. Es un organismo modelo importante para múltiples enfermedades respiratorias humanas. Se considera el 'estándar de oro' para modelar la infección y transmisión del virus de la influenza humana.

Pregunta de investigación

¿Existe una forma de medir la similitud genética entre humanos y las especies modelo más utilizadas en investigación científica?

Hipótesis

Al escoger al gorilla, bonobo, puerco, y hurón se espera que la información mutua nos arroje valores “ideales” que aseguren que la diferencia de ADN sea mínima.

Objetivo general

Identificar las diferencias y similitudes entre el genoma mitocondrial humano y el genoma mitocondrial de 4 diferentes animales seleccionados.

Objetivos particulares

1. Obtener las secuencias alineadas del genoma mitocondrial humano y cada uno de los genomas mitocondriales de los animales seleccionados.
2. Comparar las secuencias alineadas del genoma mitocondrial humano y el de cada uno de los cuatro animales seleccionados usando la Información Mutua.
3. Determinar una métrica para definir el grado de similitud entre el genoma mitocondrial humano y el de los animales seleccionados de interés.

Métodología

Selección de genes y especies

Para detectar similitudes entre el *Homo sapiens* y organismos modelo se requiere de genes claves que nos ayuden a realizar esto de forma no exhaustiva, es por ello que se optó por el uso del genoma mitocondrial, ya que éste sirve para encontrar parentesco entre grupos de individuos.

En cuanto a las especies, se escogieron cuatro en particular:

- *Gorilla beringei graueri* (Primates: Hominidae)
- *Pan paniscus* (Primates: Hominidae)
- *Sus scrofa* (Artiodactyla: Suidae)
- *Mustela eversmanii* (Carnivora: Mustelidae)

El motivo de esta elección radica en las similitudes de los organismos con el humano, por ejemplo, el sistema respiratorio de estos cuatro es muy similar al del ser humano.

Obtención de datos

Se hizo uso de la base de datos GenBank para obtener los archivos FASTA del genoma mitocondrial completo del humano, así como las cadenas completas del genoma mitocondrial de los cuatro animales que seleccionamos (<https://www.ncbi.nlm.nih.gov/nuccore/KF914213>, https://www.ncbi.nlm.nih.gov/nuccore/NC_001644, <https://www.ncbi.nlm.nih.gov/nuccore/4220565>, <https://www.ncbi.nlm.nih.gov/nuccore/KT224449>).

Alineamiento de secuencias

El alineamiento de las secuencias se realizó con el software EMBOSS Stretcher (https://www.ebi.ac.uk/Tools/psa/emboss_stretcher/), lo cual dió como resultado un óptimo alineamiento global mediante una modificación de la programación dinámica clásica; así, se unificó el tamaño de los genes mitocondriales de interés.

Algoritmos implementados.

Entropía mutua de Shannon.

La ecuación de entropía de Shannon proporciona una forma de estimar el número mínimo promedio de bits necesarios para codificar una cadena de símbolos, en función de la frecuencia de los símbolos.

En nuestro caso la entropía mutua nos daría un índice de correlación, que sería nuestra métrica para medir la similitud entre especies.

Información mutua.

La información mutua se calcula entre dos variables y mide la reducción de la incertidumbre de una variable dado un valor conocido de la otra variable.

Resultados

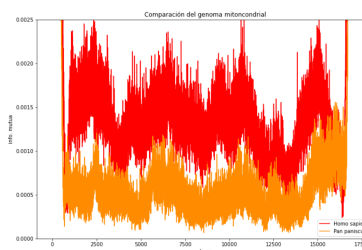
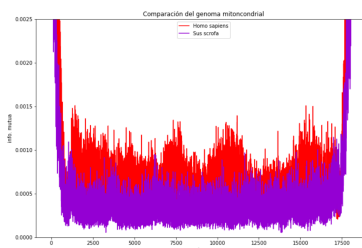
Para comparar la similitud entre cada par de secuencias (humano-animal) usamos el algoritmo de la Entropía de Shannon, del cual obtuvimos la entropía de cada secuencia, además de la entropía conjunta; las cuales necesitamos para la implementación del algoritmo de Información Mutua. Este último nos generó tablas de correlación con los datos relevantes para nuestro proyecto, y para una mayor interpretación de estos, hicimos las gráficas de comparación del genoma mitocondrial, observando que las diferencias son mínimas, por no decir casi nulas, no obstante, hay un margen de error que diferencia cada genoma y lo identifican como único.

Observamos que el orden de la correlación entre el genoma mitocondrial humano vs. otras especies, fue de la siguiente forma: chimpancé (33.25), gorila (33.13), cerdo (31.49) y hurón (31.37).

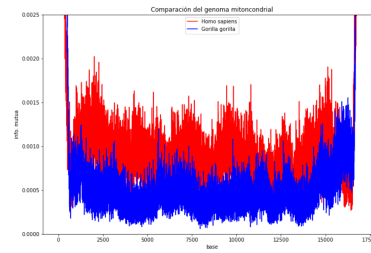
En las gráficas de información mutua se sobrelapa más la I.M. del humano v.s.hurón, así como la gráfica humano vs. cerdo. Pero la forma se parece más a la de los simios.

	Humano	Cerdo
Humano	24.995947	31.494874
Cerdo	31.494874	10.893617

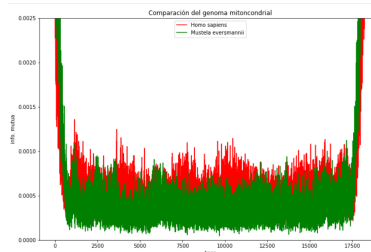
	Humano	Chimpancé
Humano	25.761513	33.250145
Chimpancé	33.250145	11.691745



	Humano	Gorila
Humano	25.821988	33.131755
Gorila	33.131755	11.525212



	Humano	Hurón
Humano	25.019634	31.369896
Hurón	31.369896	10.763415



Conclusiones

El algoritmo implementado en este proyecto proporciona una visualización simple entre la secuencia del genoma mitocondrial del ser humano y el de organismos modelo. Además, gracias a los métodos utilizados, sabemos que podemos utilizar la entropía de Shannon, los coeficientes de correlación y la información mutua para obtener medidas que nos indiquen la diferencia o similitudes entre ambos genomas.

Referencias

- [1] Boore JL. Animal mitochondrial genomes. Nucleic Acids Res. 1999 Apr 15;27(8):1767-80. doi: 10.1093/nar/27.8.1767. PMID: 10101183; PMCID: PMC148383.
- [2] Brownlee J. Information Gain and Mutual Information for Machine Learning. Recuperado el 14 de diciembre de 2022 de <https://machinelearningmastery.com/information-gain-and-mutual-information/#:~:text=Mutual%20information%20is%20calculated%20between,one%20random%20variable%20given%20another>.
- [3] Fox, James G.; Marini, Robert P. (2014). Biology and Diseases of the Ferret (3rd ed.). Ames, Iowa: John Wiley Sons. p. 5. ISBN 978-1-118-78273-6. OCLC 863695703. Archived from the original on 2022-01-05. Retrieved 2021-12-31.
- [4] NCBI. Sus scrofa. Recuperado el 14 de diciembre de 2022 de <https://www.ncbi.nlm.nih.gov/genome?term=sus%20scrofa%20%5B0rganism%5D&cmd=DetailsSearch>
- [5] Shannon Entropy. Recuperado el 14 de diciembre de 2022 de [http://bearcave.com/misl/misl_tech/wavelets/compression/shannon.html#:~:text=The%20Shannon%20entropy%20equation%20provides,10%20or%20loge\)%3A](http://bearcave.com/misl/misl_tech/wavelets/compression/shannon.html#:~:text=The%20Shannon%20entropy%20equation%20provides,10%20or%20loge)%3A)
- [6] WWF. Bonobo. Recuperado el 14 de diciembre de 2022 de <https://www.worldwildlife.org/species/bonobo>
- [7] WWF. TOP 10 FACTS ABOUT MOUNTAIN GORILLAS. Recuperado el 14 de diciembre de 2022 de <https://www.wwf.org.uk/learn/fascinating-facts/gorillas>