



UNIVERSIDADE DA CORUÑA

Grao en Bioloxía

Memoria do Traballo de Fin de Grao

Traballo de investigación: **Tráfico de ADN entre el genoma mitocondrial y el nuclear: estudio de las inserciones de Numts en la filogenia de los primates.**

Traballo de investigación: **Tráfico de ADN entre o xenoma mitocondrial e o nuclear: estudo das insercións de Numts na filoxenia dos primates.**

Research work: **DNA traffic between mitochondrial and nuclear genomes: study of Numts insertions in primate phylogeny.**

Xabier Roldán Rodríguez

Septiembre, 2021

Director Académico: Horacio Naveira Fachal



Tráfico de DNA entre el genoma mitocondrial y el nuclear: estudio de las inserciones de Numts en la filogenia de los primates



ÍNDICE

RESUMEN	5
ABSTRACT	5
PALABRAS CLAVE	6
1. INTRODUCCIÓN	7
2. MATERIAL Y MÉTODOS	10
2.1 Herramientas bioinformáticas	11
2.1.1 Ensembl	11
2.1.2 Navegador genómico USCS	11
2.1.3 BLAT	11
2.2 Procedimiento	12
3. RESULTADOS	13
3.1 Cariotipos	13
3.1.1 Cariotipo humano	13
3.1.2 Cariotipo chimpancé	13
3.1.3 Cariotipo de gorila	13
3.1.4 Cariotipo de orangután	13
3.1.5 Cariotipo de macaco	13
3.2 Sintenias	14
3.2.1 Sintenia cromosoma 7	14
3.2.2 Sintenia cromosoma 13	15
3.2.3 Sintenia cromosoma 21	16
3.3 Numts	17
3.3.1 Numts en el ser humano	17
3.3.2 Numts en el chimpancé	18
3.3.3 Numts en gorila	19
3.3.4 Numts en orangután	20
3.3.5 Numts en macaco	20
3.3.6 Comparación del número de Numts	21
3.4 Estudio de Numts en humanos	22
3.4.1 Inserción de Numt en el cromosoma 13	22



3.4.2	Inserción de Numt en el cromosoma 7	24
3.4.3	Inserción de Numt en el cromosoma 21	26
4.	DISCUSIÓN Y CONCLUSIONES	28
5.	BIBLIOGRAFÍA	30
6.	ANEXOS.	32



RESUMEN

El genoma humano se encuentra sujeto a las fuerzas evolutivas que dan forma a su arquitectura. En concreto, la integración de fragmentos del genoma mitocondrial en cromosomas nucleares (Numts) moldea esos genomas nucleares. Se han definido Numts en diversas familias de eucariotas, como la de los homínidos. Como consecuencia del proceso continuado de generación de Numts, estos han sido considerados buenos marcadores para llevar a cabo los estudios filogenéticos, así como loci potencialmente informativos para reconstruir su historia evolutiva.

Para llevar a cabo este estudio, hemos seleccionado cinco especies de primates (humano, chimpancé, orangután, gorila y macaco) en las que estimaremos el número de Numts presentes en cada genoma, y que usaremos para realizar un análisis comparativo entre las secuencias de sus Numts, así como entre las secuencias mitocondriales correspondientes a estos. Además, obtendremos las relaciones de sintenia entre sus cromosomas. De los Numts identificados en la especie humana, elegiremos tres como casos de estudio y determinaremos si también se encuentran presentes en los genomas de las otras cuatro especies (copias ortólogas), para poder estimar su antigüedad mínima. También estimaremos el grado de identidad nucleotídica de cada inserción de Numt con la secuencia mitocondrial de su especie para saber el porcentaje de similitud entre la inserción de Numt y la secuencia mitocondrial contemporánea correspondiente.

ABSTRACT

The human genome is subject to the evolutionary forces that shape its architecture. Specifically, the integration of fragments of the mitochondrial genome into nuclear chromosomes (Numts) shapes those nuclear genomes. Numbers have been defined in various families of eukaryotes, such as that of the hominids. As a consequence of the continuous process of generation of Numts, they have been good marker results to carry out phylogenetic studies, as well as potentially informative loci to reconstruct their evolutionary history.

To carry out this study, we have selected five species of primates (human, chimpanzee, orangutan, gorilla and macaque) in which we will estimate the number of Numts present in each genome, and that we will use to perform a comparative analysis between the sequences. of their Numts, as well as between the mitochondrial sequences corresponding to. In addition, we will obtain the synteny relationships between their chromosomes. Of the Numts identified in the human species, we will



choose three as case studies and determine if they are also present in the genomes of the other four species (orthologous copies), in order to estimate their minimum age. We will also estimate the degree of nucleotide identity of each Numt insert with the mitochondrial sequence of its species to know the percentage of similarity between the Numt insert and the corresponding contemporary mitochondrial sequence.

PALABRAS CLAVE

BLAT, Ensembl, evolución, mutación, NHEJ, transposición.

OBJETIVOS

El objetivo principal de este Trabajo de Fin de Grado es realizar un estudio sobre el origen y la evolución de los Numts a lo largo de la línea evolutiva de los primates a través del análisis comparativo de las secuencias de cada uno de los Numts estimados para cada una de las cinco especies de primates seleccionadas (humano, chimpancé, gorila, orangután y macaco); así como las secuencias mitocondriales que corresponden a cada uno de ellos. Para esto elegimos tres Numts del ser humano que muestren un elevado grado de identidad nucleotídica con la secuencia mitocondrial contemporánea.



1. INTRODUCCIÓN

La Real Academia Española (RAE) define genoma como la “*secuencia de nucleótidos que constituye el ADN de un individuo o de una especie*”¹. En concreto, en células eucariotas animales esta información genética reside en el ADN nuclear (ADNn) y en el ADN mitocondrial (ADNmt), ubicados, como su nombre indica, en el núcleo y en las mitocondrias, respectivamente.

El ADNn es una molécula lineal de doble cadena que contiene miles de genes, y se encuentra organizado en cromosomas. Por el contrario, el ADNmt es una molécula circular, cerrada y de doble cadena, con un tamaño aproximado en los seres humanos de 16569 pares de bases (pb). De sus 37 genes, 22 codifican para ARNt, 2 para ARNr y 13 para ARNm. Las dos cadenas complementarias tienen un peso molecular muy distinto ya que poseen una proporción diferente de guanina y citosina; originando así una cadena pesada o H (*heavy*) rica en bases púricas que codifica 28 de sus genes, y la complementaria rica en bases pirimidínicas, conocida como cadena ligera o L (*light*) que solo codifica nueve genes (*Figura 1*)².

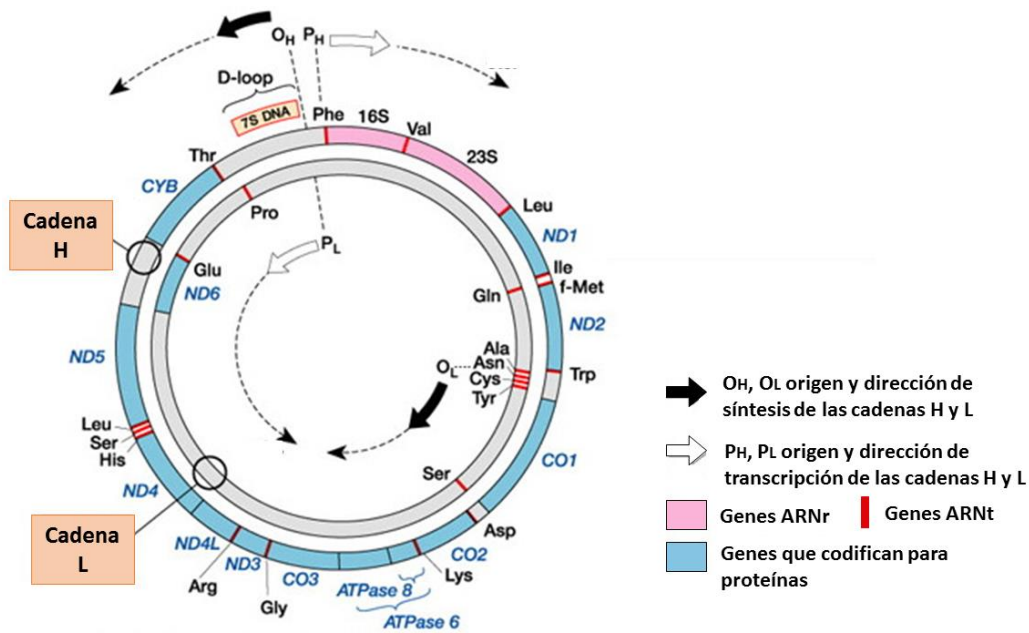


Figura 1. Mapa genético del ADN mitocondrial. Se representan la cadena ligera y pesada. Las regiones codificantes en distintos colores y las no codificantes en gris².



La reparación del ADNmt es un tema que no está muy estudiado, incluso algunas vías de reparación no están del todo claras. Lo que sí está claro es que poseen estos mecanismos de reparación que son muy diferentes a los de la reparación del ADNn. En las mitocondrias el mecanismo de reparación más común es la reparación por escisión de bases (BER) que se utiliza para evitar las bases oxidadas. No se ha esclarecido todavía si poseen reparación de errores de apareamiento (MMR) y de ruptura de doble cadena del ADN (DSB), aunque algunas de las proteínas encargadas de estas reparaciones han sido encontradas dentro de la mitocondria ³.

Las mitocondrias son remanentes de un proceso de endosimbiosis alfa-proteobacteriana⁴, tras el cual el genoma mitocondrial ancestral sufrió una reducción pronunciada de su tamaño debido a la transferencia horizontal de genes de la mitocondria al núcleo⁵. En 1994, se acuñó por primera vez el acrónimo Numts (*Nuclear Mitochondrial Segments*) para designar estas secuencias de origen mitocondrial presentes en el ADNn, que habían sido descubiertos en la década de 1980 en *Neurospora crassa* ^{6,7}. Años después se confirmó que los fragmentos de ADNmt se insertaban en el núcleo durante el proceso de reparación de roturas de doble cadena a través de la maquinaria de unión de extremos no homólogos, con o sin el requisito de microhomología^{5,8,9}. Una vez implantados en el genoma nuclear, la mayoría de estos fragmentos pierden su función original y se convierten en pseudogenes sujetos a la tasa de evolución nuclear, que es mucho más baja que la mitocondrial ^{10,11}. Estas inserciones surgen debido a la acción de agentes endógenos y exógenos como la radiación ionizante y el envejecimiento, y dependen estrictamente de la tasa de roturas de doble cadena en el ADN nuclear^{12,13}.

A lo largo de la vida de la célula, esta información genética puede sufrir distintos daños a causa de diversos factores. Como consecuencia, las células eucariotas han desarrollado distintos mecanismos especializados para poder detectar y reparar estas lesiones.

Uno de los daños más graves, frecuentes y difíciles de reparar que pueden tener lugar, es la rotura de la doble cadena del ADN (DSB) como resultado de la exposición a agentes exógenos, como puede ser la radiación ionizante y ciertas sustancias químicas, a causa de procesos endógenos, incluidas la replicación y reparación del ADN. Esta ruptura ocurre cuando las cadenas principales de azúcar de ambas cadenas se rompen suficientemente cerca como para no permitir la asociación de las bases de Watson-Crick, originando la liberación de dos extremos de ADN ^{14,15}. En estas células existen dos vías principales de reparación de DSB, que funcionan de maneras diferentes pero complementarias, HR y NHEJ (Non-Homologous End Joining) ¹⁵.



La recombinación de homólogos (HR), es el modo predominante de reparación de DSB en levaduras e implica una reparación precisa a través de una secuencia homóloga extendida y no dañada como plantilla, como se puede observar en la *Figura 2* (cadena azul). Por el contrario, el mecanismo de unión de extremos no homólogos (NHEJ), es un proceso propenso a errores en el que a menudo se eliminan o agregan nucleótidos; y el más importante en la reparación de DSB en organismos superiores, como los mamíferos. Esta reparación no exacta de roturas, ha impulsado la evolución cromosómica debido a la introducción de cambios estructurales ¹⁶.

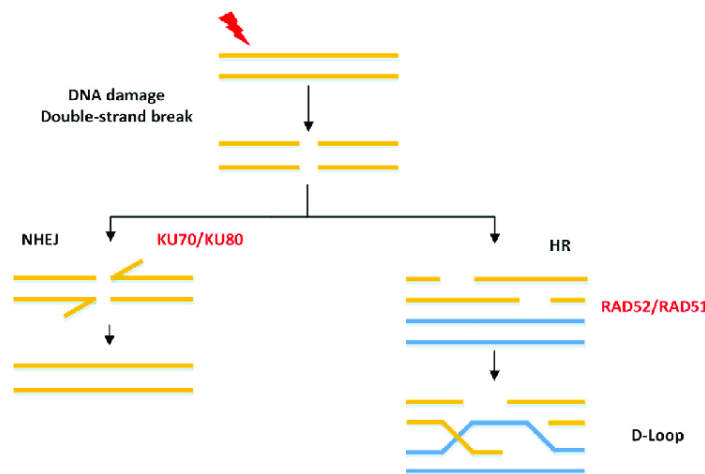


Figura 2. Vías de reparación de rotura de doble hebra del ADN (DSB) mediante recombinación homóloga (HR) y unión de extremos no homólogos (NHEJ)¹⁷.

Los Numts se capturan pasivamente en los genomas y no tienen actividad de transposición en el genoma nuclear ¹⁸. Como se cree que la mayoría derivan de eventos de inserción única, pueden considerarse como loci libres de homoplasia molecular, lo que los hace buenos candidatos a marcadores filogenéticos para conocer la evolución histórica de un taxón. Dado que si la aparición de una inserción de Numt en un locus específico se encuentra en más de un taxón, significa que poseen un ancestro común. La presencia o ausencia de Numts en loci específicos servirían para determinar el orden de ramificación filogenética de diferentes especies. Se han descrito Numts en varias especies eucariotas, incluido el *Homo sapiens* y otros primates. Por ello, la comparación de la presencia o la ausencia de la inserción de Numt entre los genomas de humanos, chimpancés y otros primates ha permitido de hecho la reconstrucción de una filogenia precisa de primates¹⁹ (*Figura 3*).

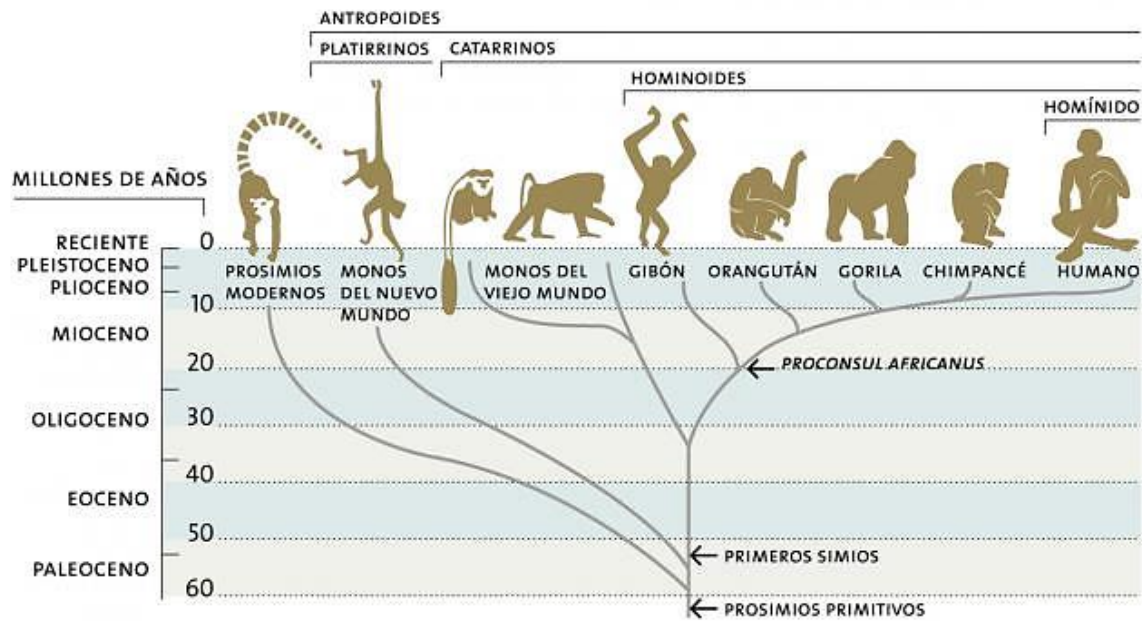


Figura 3. Imagen de la filogenia de los primates estándar. Se ha de tener en cuenta que una de las especies estudiadas es el macaco, que en la imagen se corresponde con la rama de monos del viejo mundo ²⁰.

2. MATERIAL Y MÉTODOS

Como se ha explicado anteriormente, en este TFG se pretende realizar un análisis filogenético de los primates mediante el estudio de los Numts. En este caso, examinaremos tres Numts del genoma humano para realizar una comparación con los genomas de los chimpancés, gorilas, orangutanes y macacos. Para ello, nos ayudaremos de varias herramientas bioinformáticas mediante las cuales podremos comparar los distintos ADN.

Ha de tenerse en cuenta que, para realizar este trabajo, los genomas estudiados fueron obtenidos con los siguientes ensamblajes específicos:

- Genoma humano: GCA_000001405.28
- Genoma del chimpancé: GCA_000001515.5
- Genoma del gorila: GCA_000151905.3
- Genoma del orangután: GCA_000003175.1
- Genoma del macaco: GCA_003339765.3



2.1 Herramientas bioinformáticas

2.1.1 Ensembl

La página web de Ensembl es un buscador de genomas de vertebrados que ayuda a la investigación en genómica comparada, evolución, variación de secuencias y regulación transcripcional. Este navegador permite calcular múltiples alineaciones, predecir las funciones reguladoras y recopilar datos de enfermedades. Entre sus herramientas encontramos: BLAST, BLAT, BioMart y Variant Effect Predictor ²¹.

2.1.2 Navegador genómico UCSC ²²

Este navegador genómico fue creado por la Universidad de California, Santa Cruz por Jim Kent y David Haussler, y ofrece acceso en línea a una base de datos de secuencias genómicas y datos de anotación para una amplia variedad de organismos. El navegador también tiene muchas herramientas para visualizar, comparar y analizar conjuntos de datos genómicos tanto disponibles públicamente como generados por el usuario, alineando secuencias y cargando datos del usuario ²³.

2.1.3 BLAT

BLAT (Blast-like alignment tool) es un programa de alineamiento, basado en un algoritmo que localiza las zonas más similares a la secuencia propuesta, dentro de un genoma en concreto. Este programa está especializado en secuencias con un alto grado de identidad entre ellas, lo que favorece la velocidad de este navegador. Dentro del Navegador genómico UCSC, esta herramienta está diseñada para encontrar rápidamente secuencias de una similitud mayor o igual al 95%, en tramos de una longitud de 25 pares de bases por lo que no registraría alineamientos de secuencias más cortas y divergentes. Los resultados se muestran de forma que aparece: la puntuación, el inicio y el final de la secuencia consultada, el tamaño de la secuencia coincidente, la identidad en tanto por ciento, las coordenadas genómicas y la orientación de la cadena ²⁴.



2.2 Procedimiento

Para poder lograr los objetivos planteados con anterioridad, se seguirán los siguientes pasos:

1. Obtener la secuencia mitocondrial de cada una de las cinco especies en la página de Ensembl
2. Obtener el listado completo de los Numts de las cinco especies. Para ello, hacemos uso de la herramienta BLAT del instituto genómico de la UCSC.
3. Seleccionar 3 Numts del listado anterior de la especie humana, haciendo uso de dos criterios: una longitud de la inserción de Numt baja y una identidad superior al 98%, para facilitar el estudio.
4. Obtener la imagen ``Side by side alignment`` de la inserción de Numt seleccionado.
5. Calcular la antigüedad de la inserción de Numt seleccionada. Para ello, en la página de Ensembl seleccionamos las otras cuatro especies a comparar, e introducimos las coordenadas del genoma mitocondrial del que procede la inserción de Numt que estamos estudiando. Así obtenemos el alineamiento de las secuencias mitocondriales de las cinco especies. En este observaremos verticalmente las posiciones coincidentes de las cuatro especies con la posición en la que discrepaba el genoma nuclear y mitocondrial humano en la imagen ``Side by side alignment``.



3. RESULTADOS

Las sintenias hacen referencia a la conservación de fragmentos de cromosomas en comparación con otras especies. Las utilizaremos para comparar los cariotipos de las 5 especies. En concreto, realizaremos la sintenia de los cromosomas 7, 13 y 21 del ser humano en relación al chimpancé, al gorila, al orangután y al macaco.

3.1 *Cariotipos*

El cariotipo es la colección de cromosomas de un individuo. También se denomina así a la técnica de laboratorio que produce una imagen de los cromosomas de un individuo y se puede usar para detectar enfermedades genéticas en fetos o niños. A continuación, observaremos los cariotipos de los primates que se van a estudiar.

3.1.1 *Cariotipo humano*

En el *Anexo 1* se observa un pictograma del cariotipo humano, en el que se distinguen 22 cromosomas autosómicos, dos cromosomas sexuales y el genoma mitocondrial.

3.1.2 *Cariotipo chimpancé*

En el *Anexo 2* se muestra una imagen del cariotipo de chimpancé, donde se pueden ver 23 autosomas, un cromosoma sexual y un cromosoma mitocondrial. Si lo comparamos con el del humano veremos que el número de autosomas y cromosomas sexuales es la misma.

3.1.3 *Cariotipo de gorila*

En el *Anexo 3* se representa el cariotipo de gorila, donde se pueden observar 23 autosomas y un cromosoma sexual. Además, se observa el cromosoma mitocondrial.

3.1.4 *Cariotipo de orangután*

En el *Anexo 4* se observa un pictograma el cariotipo de chimpancé, donde se pueden observar 23 autosomas y un cromosoma sexual. Además, se observa el cromosoma mitocondrial.

3.1.5 *Cariotipo de macaco*

En el *Anexo 5* se aprecia una imagen el cariotipo del macaco, donde se pueden observar 20 autosomas, dos cromosomas sexuales y un cromosoma mitocondrial. Además, se observa el cromosoma mitocondrial.



3.2 Sintenias

3.2.1 Sintenia cromosoma 7

A continuación, en la *Figura 4* se pueden observar las relaciones de sintenia del cromosoma 7 humano con los de las otras cuatro especies que se están estudiando. Se observa como el cromosoma 7 del humano estaría representado en gran parte en el cromosoma 7 del chimpancé, a excepción de una zona muy pequeña que estaría representada en el cromosoma 16. En el caso del gorila sucede algo similar, y es que la gran mayoría del cromosoma 7 humano se encontraría representado en el cromosoma 7 de gorila, a excepción de una región pequeña que se correspondería con el cromosoma 2A. En cuanto al orangután, se puede observar cómo en prácticamente su totalidad corresponde al cromosoma 7 a excepción de una región del brazo largo del cromosoma 7 que correspondería con un pequeño fragmento del cromosoma 1. Por último, en relación con el macaco, podemos comentar que en su gran parte corresponde con el cromosoma 3 de macaco a excepción de dos pequeñas zonas que corresponde con fragmentos del cromosoma 16 y 1.

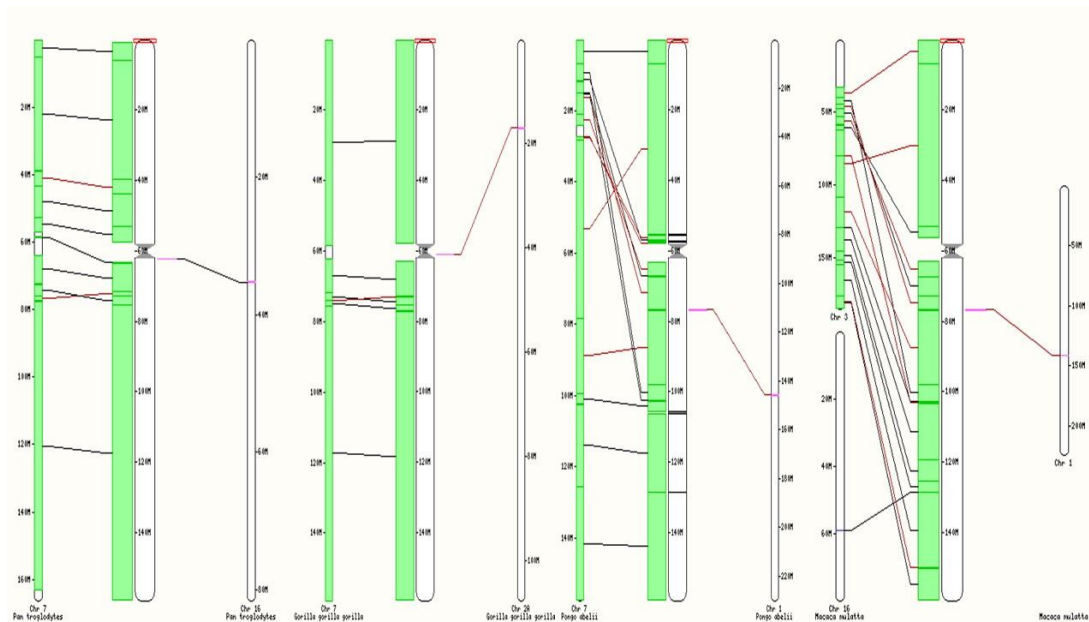


Figura 4. Sintenia del cromosoma 7 del humano con respecto a chimpancé (primera figura), gorila (segunda figura), orangután (tercera figura) y macaco (cuarta figura).



3.2.2 Sintenia cromosoma 13

En la *Figura 5* podemos observar la delación de sintenia del cromosoma 13 en relación a los cuatro primates en los que se centra el estudio. En la sintenia del chimpancé se puede observar como el brazo largo del cromosoma 13 de humano se relaciona con prácticamente todo el cromosoma 13 de chimpancé; y sólo una pequeña parte de la región telomérica corresponde a una pequeña región del cromosoma 28. La sintenia del orangután se puede apreciar como también el brazo largo del cromosoma 13 de humano corresponde al cromosoma 13 de orangután casi por completo, y como sucedía en el chimpancé una pequeña región telomérica corresponde, esta vez, con una pequeña porción del cromosoma 21. En tercer lugar, la sintenia entre el cromosoma 13 del humano con el orangután, podemos ver como el brazo largo del cromosoma 13 del humano se corresponde con casi la totalidad del cromosoma 13 de orangután. Por último, la sintenia entre el cromosoma 13 y el macaco, tiene como resultado que el brazo largo del cromosoma 13 de humano se corresponde a la totalidad del cromosoma 17 de macaco; se observa también como un fragmento de la región telomérica corresponde a una región de cromosoma 10 y otra pequeña zona del telómero corresponde con una pequeña parte del cromosoma 3 de macaco.

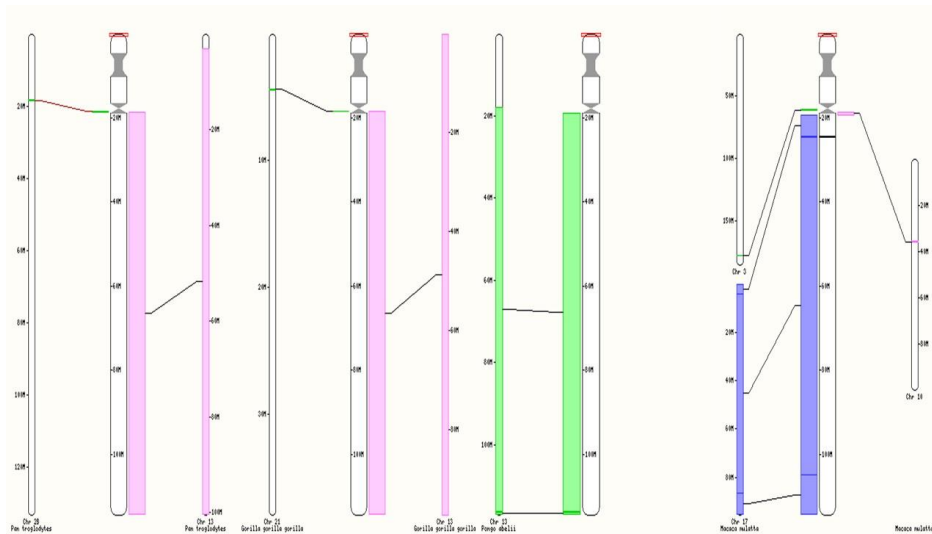


Figura 5. Sintenia del cromosoma 13 del ser humano en relación al chimpancé (primera figura), gorila (segunda figura), orangután (tercera figura) y macaco (cuarta figura).



3.2.3 Sintenia cromosoma 21.

La relación de sintenia del cromosoma 21 es más complicada que las anteriores estudiadas, como se puede observar en la *Figura 6*. En primer lugar, la sintenia entre este cromosoma humano y el chimpancé se resuelve con que el brazo largo del cromosoma 21 del humano se corresponde con cromosoma 21 de chimpancé, a excepción una zona justo antes del telómero que se corresponde con una zona del cromosoma 18 del chimpancé. En cuanto al brazo corto del cromosoma 21 del humano, tiene varias correspondencias con los cromosomas del chimpancé, en concreto, tiene cuatro con el cromosoma 7, dos con el cromosoma 4, uno con el cromosoma 9, dos con el cromosoma 21 y una con el cromosoma 18.

Por otra parte, la sintenia entre el cromosoma 21 del humano y el gorila es todavía más complicada que con el chimpancé; a pesar de esta complicación destaca la correspondencia del brazo largo del cromosoma 21 del ser humano con más de $\frac{3}{4}$ del cromosoma 21 de gorila. Además de esta coincidencia, el cromosoma 21 de humano y de gorila tienen otras cinco coincidencias mucho más pequeñas en tamaño que la anteriormente mencionada. Estas coincidencias se dan en los cromosomas 2B, 5, 9, 3, 18 y 4, estos dos últimos con dos coincidencias cada uno de ellos.

En tercer lugar, comentamos la sintenia entre el cromosoma humano estudiado y el orangután, el cual posee una coincidencia que ocupa $\frac{3}{4}$ partes de su cromosoma 21 con el brazo largo del cromosoma 21 del humano. En el cromosoma 21 del orangután existen otras tres coincidencias más pequeñas; a parte de estas correspondencias existe otras dos, una en el cromosoma 7 y otra en el cromosoma 4.

Para concluir, estudiamos la sintenia entre el macaco y el cromosoma 21 del humano. En esta sintenia destaca la coincidencia entre el brazo largo del cromosoma 21 del ser humano, con una porción del cromosoma 3 del macaco; en el que existen otras cinco coincidencias. Además, en el cromosoma 5 de macaco hay tres coincidencias más y en el cromosoma 10 observamos ocho coincidencias.

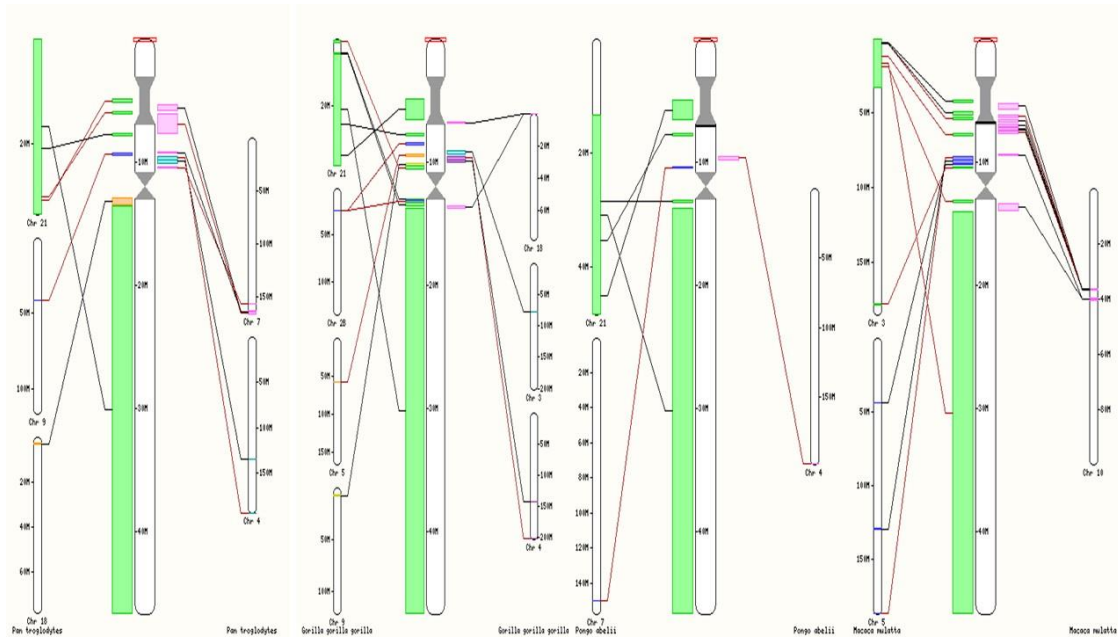


Figura 6. Sintenia entre el cromosoma 21 y el chimpancé (superior izquierda), gorila (superior derecha), orangután (inferior izquierda) y el macaco (inferior derecha).

3.3 Numts

Los Numts aparecen en cantidades diferentes dependiendo de la especie. Dos procesos pueden contribuir a esta diferencia: la tasa de integración y los procesos de duplicación o eliminación¹⁸.

3.3.1 Numts en el ser humano

Se identificó un total de 128 inserciones, con un mínimo de identidad de 75.1% dentro del cromosoma 5 en la cadena forward y con una longitud de 1354 pares de bases; y un máximo de identidad de 99,3% en el cromosoma 13, en la cadena reverse y con una longitud de 256 pares de bases (*Tabla 1*). La inserción más larga en cuanto a número de pares de bases posee 15259 pares y se encuentra situada en el cromosoma 2, con una identidad de 80.5% y un score de 3713.



ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	YourSeq	577	1054	2604	16569	84.5%	chrX	+	143430381	143432006	1626
browser details	YourSeq	574	6165	7496	16569	81.3%	chr1	-	106804419	106805780	1362
browser details	YourSeq	555	5139	6955	16569	79.4%	chr9	-	80563586	80565400	1815
browser details	YourSeq	546	12220	13537	16569	81.3%	chr16	-	20721085	20722401	1317
browser details	YourSeq	509	9395	11100	16569	79.0%	chr6	+	153667510	153669231	1722
browser details	YourSeq	508	14357	15910	16569	75.1%	chr5	+	86189888	8628341	1354
browser details	YourSeq	478	5403	6461	16569	80.7%	chr8	+	133755345	133756402	1058
browser details	YourSeq	475	9880	11115	16569	75.7%	chr9	+	92109986	92111213	1228
browser details	YourSeq	453	7045	7824	16569	79.5%	chr5	-	5396101	5396877	777
browser details	YourSeq	447	5165	6154	16569	85.4%	chr8	-	110934283	110935290	1008
browser details	YourSeq	437	6642	7511	16569	79.5%	chr7	-	69331762	69332632	871
browser details	YourSeq	272	1053	1502	16569	89.4%	chr14_GL000009v2_random	+	1923	2374	452
browser details	YourSeq	269	987	1507	16569	90.0%	chr20	-	30744402	30744928	527
browser details	YourSeq	264	2417	2844	16569	85.4%	chr10	+	19746746	19747170	425
browser details	YourSeq	260	987	1507	16569	89.7%	chr22	-	12135357	12135884	528
browser details	YourSeq	252	982	1237	16569	99.3%	chr13	-	109424125	109424380	256
browser details	YourSeq	249	687	953	16569	95.9%	chrX	-	126472467	126472732	266
browser details	YourSeq	244	6741	7012	16569	94.9%	chr2	-	49229628	49229899	272
browser details	YourSeq	230	1925	2218	16569	89.7%	chr3	+	160947654	160947947	294
browser details	YourSeq	226	537	998	16569	83.7%	chr2	+	125811722	125812539	818

Tabla 1. BLAT, donde el alineamiento se efectuó con el genoma humano y la secuencia del ADN mitocondrial humano como secuencia query. Se observan subrayadas las inserciones con menor y mayor identidad respectivamente. Los hits están ordenados de mayor a menor con respecto a su score. Obtenida de la web de Universidad California Santa Cruz.

3.3.2 Numts en el chimpancé

En la Tabla 2 se comparan los Numts del chimpancé. Cotejamos una secuencia de ADN mitocondrial tipo obtenida de Ensembl con todo el genoma del chimpancé. En este caso, al realizar el BLAT, se observan un total de 139 inserciones en el ADN nuclear. El máximo de identidad se da en el cromosoma 1, siendo de 100% en la cadena reverse con una longitud de 151 de pares de bases. Sin embargo, el mínimo de identidad es de 75,5%, en el cromosoma 5 y con una longitud de 1105 pares de bases.

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	YourSeq	601	236	1988	16554	81.6%	chrX	+	138651485	138653308	1824
browser details	YourSeq	593	5870	7061	16554	79.7%	chr7	-	65543681	65545503	1823
browser details	YourSeq	568	2949	4144	16554	76.2%	chr3	+	105124859	105126040	1182
browser details	YourSeq	563	4368	5646	16554	80.8%	chr9	+	47824402	47825664	1263
browser details	YourSeq	563	14240	15344	16554	75.5%	chr5	+	8520135	8521239	1105
browser details	YourSeq	557	11110	12452	16554	78.4%	chr15	-	18537352	18538689	1338
browser details	YourSeq	541	4368	5691	16554	82.5%	chr2B	+	98402545	98403878	1334
browser details	YourSeq	534	11229	13329	16554	78.5%	chr10	-	36554492	36556589	2098
browser details	YourSeq	530	14009	15192	16554	77.2%	chr1	-	212892086	212893266	1181
browser details	YourSeq	211	141	367	16554	95.1%	chr5	-	35270307	35270530	224
browser details	YourSeq	194	2578	2796	16554	92.7%	chr2A	+	81549502	81549710	209
browser details	YourSeq	187	1515	1724	16554	95.7%	chr12	+	12543444	12543654	211
browser details	YourSeq	184	12157	12344	16554	99.0%	chr2B	-	105694322	105694509	188
browser details	YourSeq	177	3795	3979	16554	97.9%	chr13	-	93253089	93253273	185
browser details	YourSeq	171	1063	1237	16554	98.9%	chr5	+	110039693	110039867	175
browser details	YourSeq	151	2368	2518	16554	100.0%	chr1	-	94504232	94504382	151
browser details	YourSeq	148	9982	10141	16554	96.3%	chr2B	+	99549227	99549386	160
browser details	YourSeq	143	1841	1993	16554	96.8%	chr7	+	45581332	45581484	153
browser details	YourSeq	142	14543	14688	16554	98.7%	chr4	+	183425631	183425776	146

Tabla 2. BLAT obtenido del navegador genómico de la página web de la Universidad de California, Santa Cruz. Los hits están ordenados de mayor a menor en relación a su score.



3.3.3 Numts en gorila

El tercer BLAT que realizamos será el del gorila (*Tabla 3*), siguiendo el mismo método de comparación entre el ADN mitocondrial y el nuclear. Realizado el BLAT, obtenemos 125 inserciones entre las cuales, la mayor identidad es de un 100% y se produce 4 veces, dos en el cromosoma 9, una de ellas en la cadena forward y la otra en la reverse. Las otras dos inserciones con mayor identidad de producen en lo que la página denomina “chrUN”, que significa que contiene contigs de clones que no se pueden colocar con seguridad en un cromosoma específico. Para los archivos chrN_random y chrUn_random, básicamente concatenamos todos los contigs en pseudocromosomas cortos. Las coordenadas de estos son bastante arbitrarias, aunque las posiciones relativas de las coordenadas son buenas dentro de un contig. La inserción que se produce en la cadena reverse del cromosoma 9 tiene una longitud de 563 pares de bases, mientras que la que se produce en la cadena forward tiene solamente 120. Las ocurridas en el chrUN suceden en la cadena reverse, una tiene una longitud de 378 pares de bases y la otra 538. En contraposición con la máxima identidad, tenemos la mínima, la cual es de 75,2% y está situada en el cromosoma 2, en la cadena forward y tiene una longitud total de 3814 pares de bases.

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	YourSeq	814	4544	6127	16412	77.6%	chr8	+	129629520	129631080	1561
browser details	YourSeq	767	13574	15444	16412	75.3%	chr16	-	66273817	66275681	1865
browser details	YourSeq	765	60	1485	16412	85.7%	chr2B	-	29824231	29825649	1419
browser details	YourSeq	724	11160	15038	16412	75.2%	chr2A	+	99126008	99129821	3814
browser details	YourSeq	724	1263	2727	16412	85.9%	chr14	+	64842487	64843959	1473
browser details	YourSeq	715	4423	6281	16412	81.4%	chr9	-	51682578	51684425	1848
browser details	YourSeq	715	8656	10315	16412	77.5%	chr2B	+	89431215	89432865	1651
browser details	YourSeq	702	1837	4661	16412	79.2%	chr10	+	19695357	19697205	1849
browser details	YourSeq	606	9180	10710	16412	81.2%	chr9	+	63259360	63261182	1823
browser details	YourSeq	605	13859	15323	16412	76.1%	chr17	-	82167646	82169109	1464
browser details	YourSeq	563	213	775	16412	100.0%	chr9	-	27205830	27206392	563
browser details	YourSeq	550	78	1108	16412	83.4%	chr11	+	38034263	38035289	1027
browser details	YourSeq	538	619	1156	16412	100.0%	chrUn_NW_022150917v1	-	22104	22641	538
browser details	YourSeq	520	4431	5636	16412	81.4%	chr9	-	49846316	49847488	1173
browser details	YourSeq	514	8973	10468	16412	77.3%	chr4	-	25858325	25859812	1488
browser details	YourSeq	503	8528	13672	16412	92.5%	chr9	+	27205498	27206597	1100
browser details	YourSeq	501	4525	5611	16412	88.7%	chr2B	-	41977870	41978954	1085
browser details	YourSeq	385	4930	5618	16412	83.7%	chr2B	+	98598830	98599550	721
browser details	YourSeq	378	1	378	16412	100.0%	chrUn_NW_022153807v1	-	1	378	378
browser details	YourSeq	371	3120	4304	16412	78.3%	chr12	+	43249616	43250794	1179
browser details	YourSeq	364	4588	5311	16412	81.7%	chr8	-	106756082	106756823	742
browser details	YourSeq	364	7705	8225	16412	85.0%	chr10	-	95292720	95293240	521
browser details	YourSeq	132	8958	9094	16412	98.6%	chr2B	+	76579778	76579964	187
browser details	YourSeq	129	8831	8965	16412	97.8%	chr8	+	72962300	72962434	135
browser details	YourSeq	129	14815	14985	16412	87.8%	chr6	+	131569860	131570030	171
browser details	YourSeq	120	662	781	16412	100.0%	chr9	+	12314781	12314900	120
browser details	YourSeq	112	6716	7013	16412	87.1%	chr11	+	29127178	29128125	948

Tabla 3. Podemos observar subrayados la inserción con menor identidad y las 4 inserciones con la mayor identidad. Los hits están ordenados de mayor a menor en relación a su score. BLAT obtenido a partir del navegador genómico de la Universidad de California, Santa Cruz.



3.3.4 Numts en orangután

Para la obtención de las inserciones dentro del genoma nuclear de orangután realizamos un BLAT, el cual usaremos el ensamblaje PPYG2 (Tabla 4). Obtenemos una cantidad total de 130 inserciones. Con relación a la identidad, tenemos un máximo de 99% en el cromosoma 17, en la cadena reverse y una longitud de 4634 pares de bases. El menor porcentaje de identidad es de 75,7% aparece en el cromosoma 5, en la cadena forward de 1340 pares de bases.

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	YourSeq	16499	1	16499	16499	100.0%	chrM	+	1	16499	16499
browser details	YourSeq	10302	3519	14975	16499	95.3%	chr4	+	137126998	137138455	11458
browser details	YourSeq	7073	32	8981	16499	89.9%	chr5	-	36728021	36736961	8941
browser details	YourSeq	6382	1	10568	16499	83.5%	chr17_NW_019937296v1_random	+	9707	20226	10520
browser details	YourSeq	5268	3559	9654	16499	93.3%	chr9	-	39878508	39884545	6038
browser details	YourSeq	5184	448	12072	16499	80.6%	chr10	-	38338894	38350075	11182
browser details	YourSeq	4962	11220	16498	16499	97.1%	chr17	-	13874018	13879291	5274
browser details	YourSeq	4542	1	4636	16499	99.0%	chr17	-	13869383	13874016	4634
browser details	YourSeq	772	4277	5640	16499	80.8%	chr28	-	41107177	41108536	1360
browser details	YourSeq	771	842	2462	16499	83.3%	chr3	-	101681420	101689374	7955
browser details	YourSeq	748	12289	13220	16499	90.2%	chr22	-	16423071	16424002	932
browser details	YourSeq	708	9005	11171	16499	82.2%	chrUn_NW_019937437v1	-	20570	22730	2161
browser details	YourSeq	706	974	2564	16499	82.0%	chr8	+	32015294	32016853	1560
browser details	YourSeq	693	4283	5385	16499	87.3%	chr6	-	110429593	110451211	21619
browser details	YourSeq	665	13889	15228	16499	75.7%	chr5	+	8590580	8591919	1340
browser details	YourSeq	642	5906	7079	16499	80.8%	chr5	-	95072553	95073749	1197
browser details	YourSeq	642	8702	10224	16499	77.4%	chr3	-	163663599	163665122	1524
browser details	YourSeq	640	1432	2740	16499	81.2%	chr14	+	65795676	65796988	1313
browser details	YourSeq	628	3380	4317	16499	87.5%	chr20	-	27768118	27769054	937

Tabla 4. BLAT realizado en el Navegador genómico UCSC. Las filas subrayadas corresponden a las inserciones con mayor y menor identidad. Los hits están ordenados de mayor a menor en relación a su score.

3.3.5 Numts en macaco

Para detectar los Numts en macaco usaremos la herramienta BLAT (Tabla 5). El ensamblaje usado del genoma de macaco será: GCA_003339765.3. Como resultado obtenemos 106 inserciones, con una identidad máxima de 100% y una mínima de 80,4%. La máxima se produce en el cromosoma 8, en la cadena reverse con una longitud de 4685 pares de bases y la mínima se da en el cromosoma 7 en la cadena forward con una longitud de 1601 pares de bases.



ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	YourSeq	5395	11140	16564	16564	99.8%	chr8	-	25076649	25082068	5420
browser details	YourSeq	5130	10373	16406	16564	92.9%	chr5	+	167556164	167562198	6035
browser details	YourSeq	5037	329	7453	16564	88.9%	chr7	-	13225071	13231761	6691
browser details	YourSeq	4967	5129	10520	16564	96.2%	chr3	-	139202848	139208227	5380
browser details	YourSeq	4775	4551	9872	16564	95.1%	chr1	-	118283481	118288796	5316
browser details	YourSeq	4664	23	4692	16564	100.0%	chr8	-	25071942	25076626	4685
browser details	YourSeq	4512	1	4836	16564	96.7%	chr13	-	27182630	27187471	4842
browser details	YourSeq	4224	10613	16564	16564	86.5%	chr7	+	13233352	13239297	5946
browser details	YourSeq	3963	1	4486	16564	94.6%	chr6	-	73961795	73966284	4490
browser details	YourSeq	883	15071	16563	16564	84.0%	chr19	-	27231444	27232932	1489
browser details	YourSeq	867	2365	3641	16564	86.6%	chr8	+	95461132	95462403	1272
browser details	YourSeq	846	292	1360	16564	92.0%	chr6	+	33587560	33588623	1064
browser details	YourSeq	845	707	2767	16564	84.7%	chr9	-	82435999	82438053	2055
browser details	YourSeq	843	1833	3437	16564	80.4%	chr7	+	146363100	146364700	1601
browser details	YourSeq	837	11006	12090	16564	88.7%	chr16	+	9378734	9379824	1091
browser details	YourSeq	789	15188	16564	16564	83.4%	chr4	-	88763155	88764530	1376
browser details	YourSeq	777	15330	16564	16564	89.8%	chr7	-	13232088	13233332	1245
browser details	YourSeq	679	1503	2966	16564	85.5%	chr2	+	96224735	96226164	1430
browser details	YourSeq	679	13666	14525	16564	90.4%	chr2	+	67451491	67452350	860

Tabla 5. BLAT realizado en el navegador genómico de la Universidad de Santa CRUZ, California. Destacados en color azul se encuentran las inserciones con máxima y mínima identidad.

3.3.6 Comparación del número de Numts

Para establecer una comparativa entre los Numts de cada especie, realizamos el estudio de dos relaciones. La primera, expresa la cantidad de Numts en relación con el número de Megabases (Tabla 6); y la segunda consiste en clasificar los Numts en relación a la cantidad de cromosomas totales (Tabla 7).

Como se puede apreciar en la Tabla 6, la especie que posee mayor número de Numts es el chimpancé seguido por el humano y el gorila, siendo el macaco el que menor cantidad tiene. Se observa una tendencia al aumento de Numts según avanzamos en la línea evolutiva de los primates.

Especie	Numts	Megabases	Numts/Megabases
Humano	128	3097	0,0413
Chimpancé	139	3231	0,0430
Gorila	125	3063	0,0408
Orangután	130	3447	0,0377
Macaco	106	2971	0,0357

Tabla 6. Tabla dónde se observa la relación Numts/Megabases de cada una de las cinco especies estudiadas.

Especie	Numts	Cromosomas	Numts/cromosomas
Humano	128	24	5,33
Chimpancé	139	25	5,56
Gorila	125	24	5,21
Orangután	130	24	5,42
Macaco	106	22	4,82

Tabla 7. Tabla dónde se aprecia la relación Numts/cromosomas de las cinco especies de primates que estudiamos.



Para discernir la posición filogenética de esta inserción en primates, realizaremos un alineamiento de la secuencia mitocondrial de las 5 especies de primates estudiadas. En la *Figura 8* se pueden observar las posiciones MT:1009 y MT:1018, ambas subrayadas en amarillo. En la posición MT:1009 se observa que solo existe una citosina en el ser humano, mientras que el resto de las especies estudiadas tienen una timina en dicha posición, lo que puede indicar que el cambio en esa secuencia se haya producido posterior a la divergencia de estas especies de primates, en un período relativamente corto de tiempo. Lo mismo ocurre con la posición MT: 1018 dentro del genoma mitocondrial de humano, la cual corresponde a una guanina mientras que en el resto de los primates es una adenina, y por tanto esta variación en el ADN mitocondrial se produjo después de la divergencia del linaje humano.

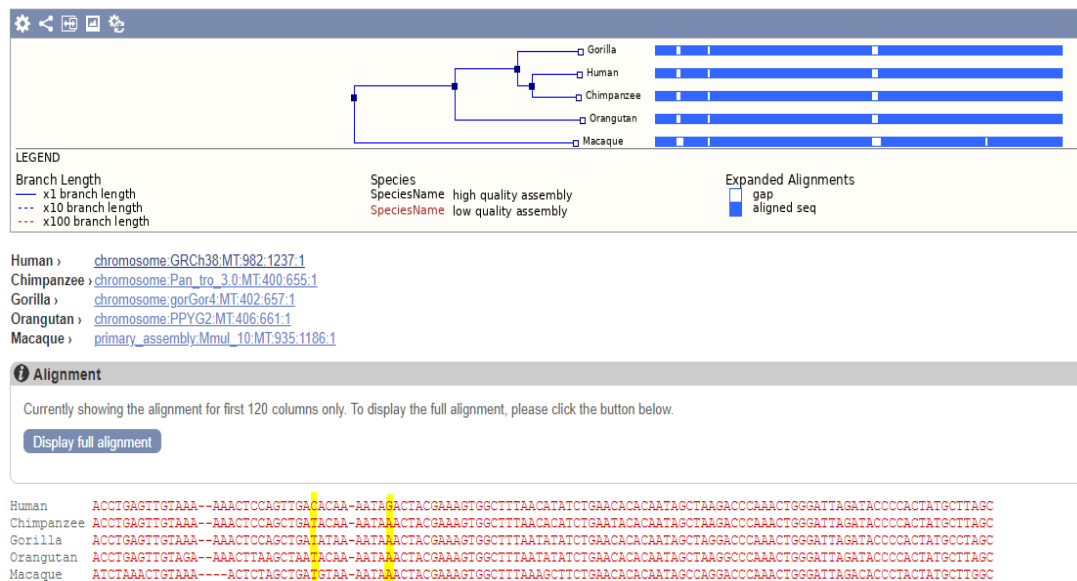


Figura 8. Filogenia y alineamiento de las secuencias del ADN mitocondrial de las cinco especies estudiadas correspondiente al intervalo MT: 982-1237 del genoma mitocondrial humano. Se resaltan en amarillo las posiciones divergentes con la inserción del Numt 13.

En la *Figura 9* podemos observar un alineamiento de la región que comprende la inserción del cromosoma 13 (± 100 pares de bases) con las secuencias nucleotídicas de las otras especies estudiadas. Como se puede apreciar, esta inserción de Numt es específica del genoma humano (no está presente en ninguna de las otras cuatro especies).



Para cotejar la antigüedad de esta inserción de Numt realizamos un alineamiento de un *cluster* de las cinco especies de primates (el humano, el chimpancé, el gorila, el orangután y el macaco) (*Figura 11*). Como se puede observar en esta figura, en la posición MT:1619 existe un dimorfismo entre la citosina y la timina, estando este último nucleótido, que es el que caracteriza al Numt 7, presente tan sólo en chimpancé y macaco.

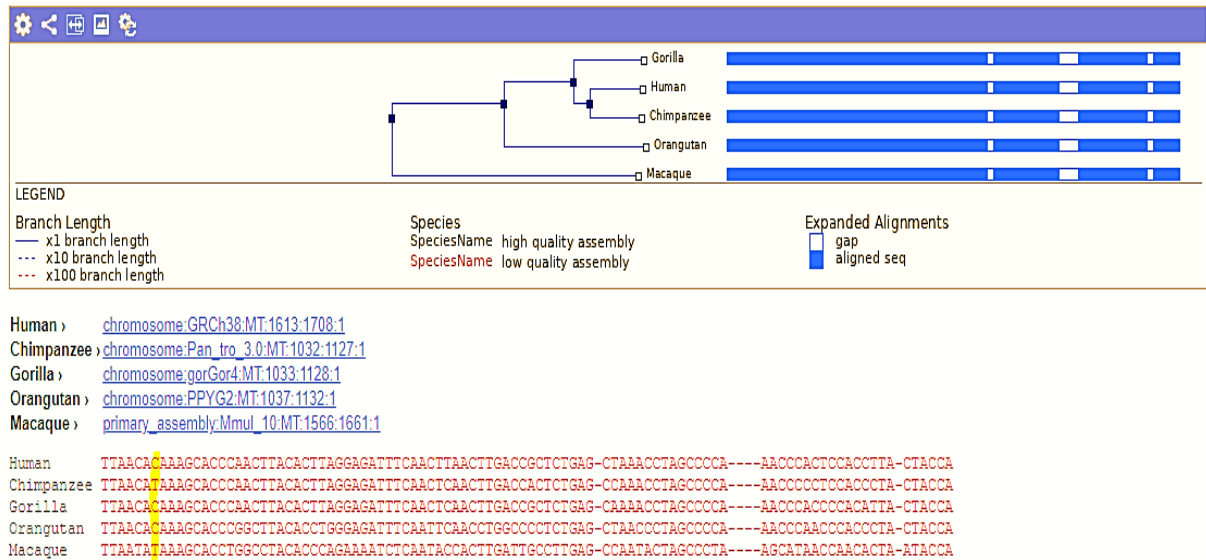


Figura 11. Filogenia y alineamiento de las secuencias del ADN mitocondrial de las cinco especies estudiadas correspondiente al intervalo MT:1613-1708. Se resalta en amarillo la posición divergente con la inserción del Numt 7.

En la *Figura 12* podemos observar un alineamiento de la región que comprende la inserción del cromosoma 7 (± 100 pares de bases) con las secuencias nucleotídicas de las otras especies estudiadas. De nuevo, esta inserción de Numt es específica del genoma del humano.

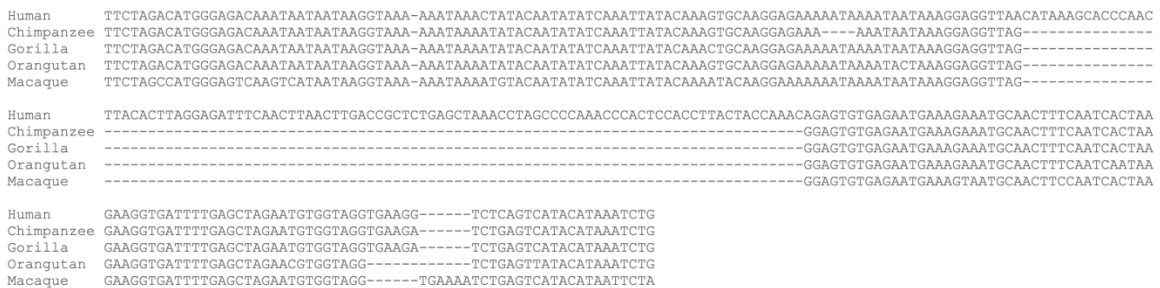


Figura 12. Alineamiento del intervalo 7:145997233-145997528 del genoma humano, que comprende la inserción de Numt del cromosoma 7, con el resto de primates estudiados.



Tráfico de DNA entre el genoma mitocondrial y el nuclear: estudio de las inserciones de Numts en la filogenia de los primates

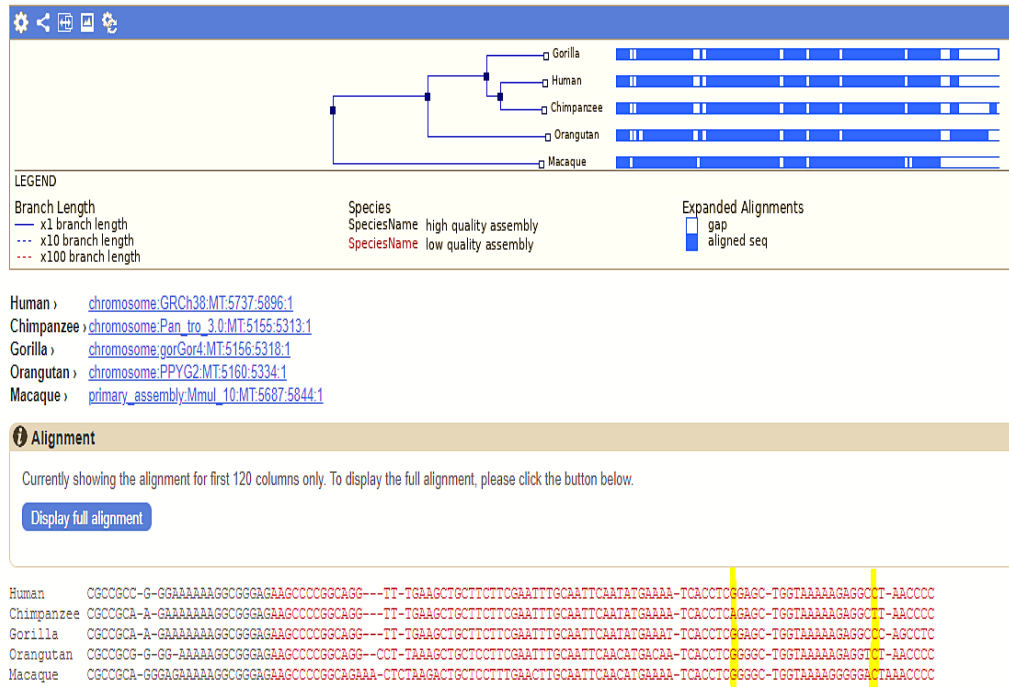


Figura 14. Filogenia y alineamiento de las secuencias del ADN mitocondrial de las cinco especies estudiadas correspondiente al intervalo MT: 5737-5896 del genoma mitocondrial humano. Se resaltan en amarillo las posiciones divergentes con la inserción del Numt 21.

En la *Figura 15* se muestra el alineamiento de la región que comprende la inserción del cromosoma 21 (± 100 pares de bases) con las secuencias nucleotídicas de las otras especies estudiadas. Una vez más se comprueba que esta inserción de Numt es específica del genoma humano.



```

Human      GAAAATTTTAAGGACCTTTT-----TCTTTTACTAATGCTTTGATTAGGATTCAGTCTGCATACGAAATTTGTCATGATTTAGCTG-----AAAC
Human      GAAAATTTTAAGGACCTTTT-----TCTTTTACTAATGCTTTGATTAGGATTCAGTCTGCATACAAATTTGCAAGATTTTAGCTGATACTCTTCAGGAAAC
Human      GAAAATTTTAAGGACCTTTT-----TCTTTTACTAATGCTTTGATTAGGATTCAGTCTGCATACAAATTTGCAAGATTTTAGCTGATACTCTTCAGGAAAC
Human      GAAAATTTTAAGGACCTTTT-----TCTTTTACTAATGCTTTGATTAGGATTCAGTCTGCATACAAATTTGCAAGATTTTAGCTGATACTCTTCAGGAAAC
Human      GAAAATTTTAAGGACCTTTTCATT-----TCTTTTACTAATGCTTTGATTAGGATTCAGTCTGCATACGAAATTTGTCATGATTTAGCTGATACTCATTTCAGGAAAC
Gorilla    GAAAATTTTAAGGACCTTTT-----TCTTTTACTAATGCTTTGATTAGGATTCAGTCTGCATACGAAATTTGTCATGATTTAGCTGATACTCATTTCAGGAAAC
Human      GAAAATTTTAAGGACCTTTT-----TCATT-----TCTTTTACTTACGCTTTGATTAGGATTCAGTCTGCATACGAAATTTGTCATGATTTAGCTGATACTCATTTCAGGAAAC
Chimpanzee GAAAATTTTAAGGACCTTTT-----TCATT-----TCTTTTACTTACGCTTTGATTAGGATTCAGTCTGCATACGAAATTTGTCATGATTTAGCTGATACTCATTTCAGGAAAC
Orangutan  GAAAATTTTAAGGACCTTTT-----TCATT-----TCTTTTACTTACGCTTTGATTAGGATTCAGTCTGCATACAAATTTGTCATGATTTAGCTGATACTCATTTCAGGAAAC

Human      TATTTTAATAGATCGCCCGGGAAAAAGCGGGGAGAAGCCCGGCAGGTTTGAAGCTGCTTCTCGAATTTGCAATTCAAATGAAAATCACCTCAGAGCTGGTAAAAAGAGGCTTAA
Human      TATTTTAATAG-----
Human      TATTTTAATAG-----
Human      TATTTTAATAG-----
Human      TATTTTAATAG-----
Gorilla    TATTTTAATAG-----
Human      GATTTTAATAG-----
Chimpanzee GATTTTAATAG-----
Orangutan  TATTTTAATAG-----

Human      CCCCTGTCCTTAGATTTACAGTCCAATGCTTCACTCAGCCATTTACCTCACCAGATCAATACCTTTCCTTATCTCTCAATGTCATTAGATATATACATTTTAACTTAATTTATAGAAC
Human      -----ATCATTACCCTTTCCTTATCTCTCAATGTCATTAGATATATACATTTTAACTTAATTTCTAGGAC
Human      -----ATCATTACCCTTTCCTTATCTCTCAATGTCATTAGATATATACATTTTAACTTAATTTCTAGGAC
Human      -----ATCATTACCCTTTCCTTATCTCTCAATGTCATTAGATATATACATTTTAACTTAATTTCTAGGAC
Human      -----ATCATTACCCTTTCCTTATCTCTCAATGTCATTAGATATATACATTTTAACTTAATTTCTAGGAC
Gorilla    -----ATCATTACCCTTTCCTTATCTCTCAATGTCATTAGATATATACATTTTAACTTAATTTCTAGGAC
Human      -----ATCATTACCCTTTCCTTATCTCTCAATGTCATTAGATATATACATTTTAACTTAATTTCTAGGAC
Chimpanzee -----ATCATTACCCTTTCCTTATCTCTCAATGTCATTAGATATATACATTTTAACTTAATTTCTAGGAC
Orangutan  -----ATCATTACCCTTTCCTTATCTCTCAATGTCATTAGATATATATTTTAACTTAATTTCTAGGAC

Human      AAATATACATATCTTTTTCTGATTATGTGCCTC
Human      AAATATACATATCTTTTTCTGATTATGTGCCTC
Human      AAATATACATATCTTTTTCTGATTATGTGCCTC
Human      AAATATACATATCTTTTTCTGATTATGTGCCTC
Gorilla    AAATATACATATCTTTTTCTGATTATGTGCCTC
Human      AAATATACATATCTTTTTCTGATTATGTGCCTC
Chimpanzee AAATATACATATCTTTTTCTGATTATGTGCCTC
Orangutan  AAATATATGTATCTTTTTCTGATTATGTGCCTC

```

Figura 15. Alineamiento del intervalo 21:10014755-100115114 del genoma humano, que comprende la inserción de Numt del cromosoma 21, con el resto de primates estudiados.

4. DISCUSIÓN Y CONCLUSIONES

Tras la revisión bibliográfica llevada a cabo para realizar este trabajo de fin de grado, hemos podido comprobar que son varios los artículos que realizan un análisis del número de Numts de varias especies, sin llegar al mismo resultado. Autores como Mourier et al., 2001, han encontrado 296 Numts en el genoma nuclear humano²⁵, Hazkani-Covo et al., 2003 han identificado 82 Numts²⁶. Las estimas más recientes, Dayama et al. 2020 señalan que el número de Numts en el genoma de referencia humano es de 767²⁷. Sin embargo, a través de nuestro protocolo de trabajo explicado anteriormente, tal y como se muestra en el apartado de resultados hemos obtenido, tan sólo, 128 Numts, debido a que el BLAT no registra las inserciones más antiguas (más divergentes).

Por otra parte, también existen disyuntivas del número de Numts entre las diferentes especies estudiadas. Dayama et al., 2020 aseguran que existe un aumento del número de Numts de orangutanes a chimpancés y una disminución en humanos, siendo el macaco la especie con mayor número de Numts respecto a las anteriores²⁸. Por el contrario, en nuestros resultados se puede comprobar que el



orden no corresponde al anteriormente citado, sino que, la especie con mayor número de inserciones sería el chimpancé, seguido del orangután, el humano y por último el macaco.

Atendiendo a los resultados obtenidos en la *Figura 9*, *Figura 12* y *Figura 15*, observamos que estas inserciones son específicas de la especie humana, ya que al realizar el alineamiento no se encuentran coincidencias en la secuencia nucleotídica en las otras 4 especies estudiadas. Lo que se puede ver en las imágenes de los alineamientos de los genomas mitocondriales es que la base presente en el Numt del ser humano y que diverge con la secuencia de su propio genoma mitocondrial, es la misma que se encuentra en el genoma mitocondrial de chimpancé. Este resultado nos puede conducir a pensar que estas inserciones se han producido en el momento en el que el ser humano y el chimpancé comenzaron a divergir.

Un tema que se ha estado repitiendo durante todo el trabajo es el caso de los dimorfismos, y es que nos gustaría comentar que ya sea entre especies o entre ADNmt y ADNn siempre existe un caso de dimorfismo en las bases de una cierta posición, siendo este dimorfismo producido por las dos mismas bases en cada posición estudiada. En el estudio de la bibliografía no hemos encontrado nada al respecto y en este estudio no hemos encontrado una conclusión clara sobre este tema ya que debería de hacerse con un estudio más profundo y numeroso.

Los resultados de este Trabajo de Fin de Grado nos llevan a pensar que los Numts han podido ser una de las causas de la especiación en los primates, como aseguran otros autores como Gunbin et al. 2017 y Calabrese et al. 2017^{29,30}.

Por último, teniendo en cuenta los datos aportados por los alineamientos de las secuencias mitocondriales en los cromosomas estudiados y sus correspondientes arboles filogenéticos máximos parsimoniosos, podemos suponer que el análisis de los Numts puede ser un buen procedimiento para la realización de una filogenia.



5. BIBLIOGRAFÍA

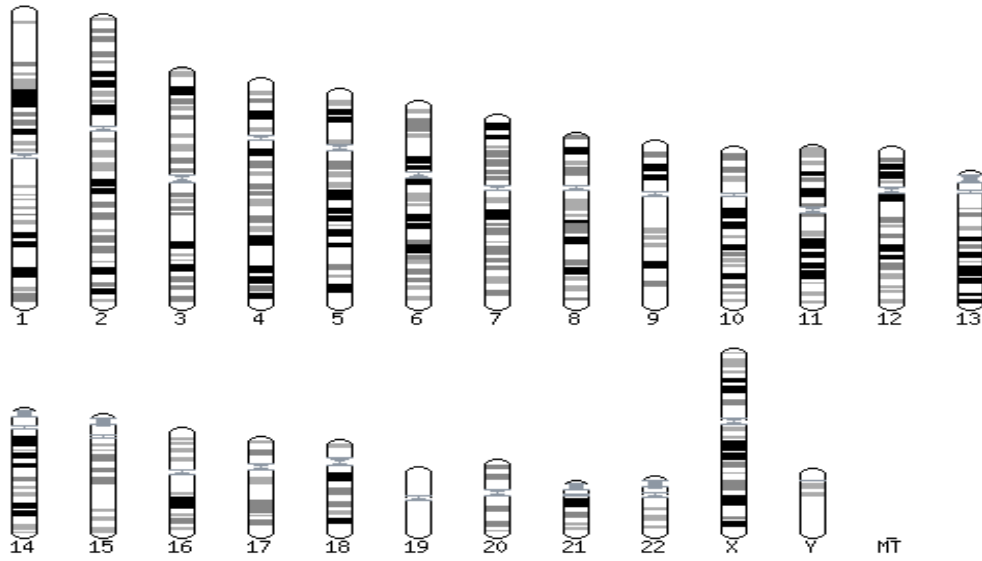
- (1) ASALE, R.-; RAE. genoma | Diccionario de la lengua española <https://dle.rae.es/genoma> (accessed 2021 -07 -19).
- (2) Álvarez Iglesias, V. *Estudio multidisciplinar de la variabilidad del ADN mitocondrial en poblaciones humanas*; Universidade de Santiago de Compostela. Servizo de Publicacións e Intercambio Científico, 2008.
- (3) Fontana, G. A.; Gahlon, H. L. Mechanisms of Replication and Repair in Mitochondrial DNA Deletion Formation. *Nucleic Acids Research* **2020**, *48* (20), 11244–11258. <https://doi.org/10.1093/nar/gkaa804>.
- (4) Margulis, L. Origin of eukaryotic cells: Evidence and research implications for a theory of the origin and evolution of microbial, plant, and animal cells on the Precambrian earth. *Zeitschrift für Allgemeine Mikrobiologie* **1970**, *13*, 186.
- (5) Calabrese, F. M.; Balacco, D. L.; Preste, R.; Diroma, M. A.; Forino, R.; Ventura, M.; Attimonelli, M. NumtS Colonization in Mammalian Genomes. *Scientific Reports* **2017**, *7* (1). <https://doi.org/10.1038/s41598-017-16750-2>.
- (6) Lopez, J. V.; Yuhki, N.; Masuda, R.; Modi, W.; O'Brien, S. J. Numt, a Recent Transfer and Tandem Amplification of Mitochondrial DNA to the Nuclear Genome of the Domestic Cat. *Journal Molecular Evolution* **1994**, *39* (2), 174–190. <https://doi.org/10.1007/BF00163806>.
- (7) Attimonelli, M.; Calabrese, F. M. Chapter 6 - Human Nuclear Mitochondrial Sequences (NumtS). In *The Human Mitochondrial Genome*; Gasparre, G., Porcelli, A. M., Eds.; Academic Press, 2020; pp 131–143. <https://doi.org/10.1016/B978-0-12-819656-4.00006-1>.
- (8) Blanchard, J. L.; Schmidt, G. W. Mitochondrial DNA Migration Events in Yeast and Humans: Integration by a Common End-Joining Mechanism and Alternative Perspectives on Nucleotide Substitution Patterns. *Molecular Biology and Evolution* **1996**, *13* (3), 537–548. <https://doi.org/10.1093/oxfordjournals.molbev.a025614>.
- (9) Ricchetti, M.; Fairhead, C.; Dujon, B. Mitochondrial DNA Repairs Double-Strand Breaks in Yeast Chromosomes. *Nature* **1999**, *402* (6757), 96–100. <https://doi.org/10.1038/47076>.
- (10) Thalmann, O.; Serre, D.; Hofreiter, M.; Lukas, D.; Eriksson, J.; Vigilant, L. Nuclear Insertions Help and Hinder Inference of the Evolutionary History of Gorilla MtDNA. *Molecular Ecology* **2005**, *14* (1), 179–188. <https://doi.org/10.1111/j.1365-294X.2004.02382.x>.
- (11) Fernández Domínguez, E. Polimorfismos de DNA mitocondrial en poblaciones antiguas de la cuenca mediterránea. Ph.D. Thesis, Universitat de Barcelona, 2005.
- (12) Gaziev, A. I.; Shaikhaev, G. O. Nuclear mitochondrial pseudogenes. *Molecular Biology (Mosk)* **2010**, *44* (3), 405–417.
- (13) Calabrese, F. M.; Simone, D.; Attimonelli, M. Primates and Mouse NumtS in the UCSC Genome Browser. *BMC Bioinformatics* **2012**, *13 Suppl 4*, S15. <https://doi.org/10.1186/1471-2105-13-S4-S15>.
- (14) Cannan, W. J.; Pederson, D. S. Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin. *Journal of Cellular Physiology* **2016**, *231* (1), 3–14. <https://doi.org/10.1002/jcp.25048>.
- (15) Bassing, C. H.; Alt, F. W. The Cellular Response to General and Programmed DNA Double Strand Breaks. *DNA Repair* **2004**, *3* (8), 781–796. <https://doi.org/10.1016/j.dnarep.2004.06.001>.
- (16) Hazkani-Covo, E.; Covo, S. Numt-Mediated Double-Strand Break Repair Mitigates Deletions during Primate Genome Evolution. *PLoS Genet* **2008**, *4* (10), e1000237. <https://doi.org/10.1371/journal.pgen.1000237>.
- (17) Ji, Q.; Mai, J.; Ding, Y.; Wei, Y.; Ledesma-Amaro, R.; Ji, X.-J. Improving the Homologous Recombination Efficiency of *Yarrowia Lipolytica* by Grafting Heterologous Component from



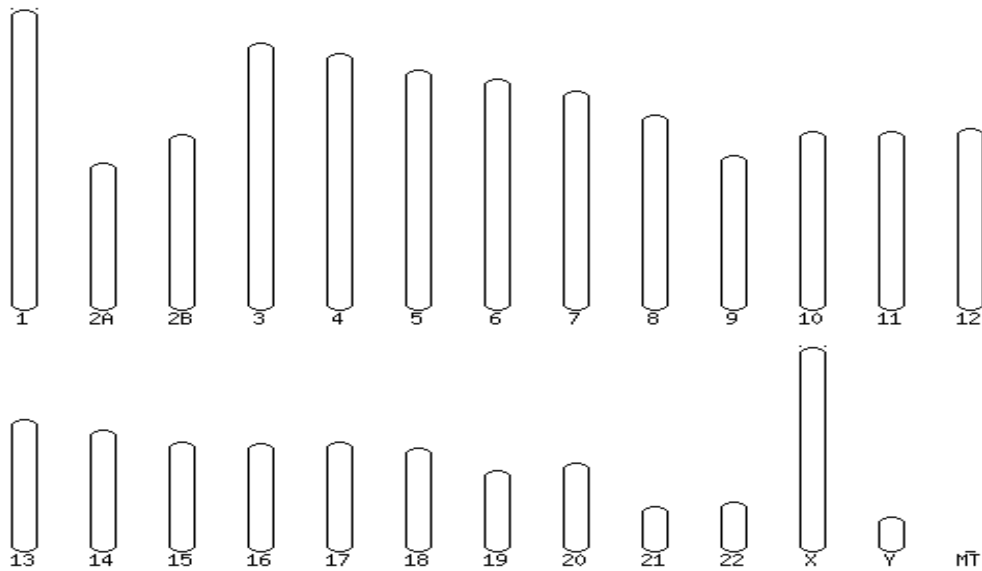
- Saccharomyces Cerevisiae. *Metabolic Engineering Communications* **2020**, *11*, e00152. <https://doi.org/10.1016/j.mec.2020.e00152>.
- (18) Hazkani-Covo, E. Mitochondrial Insertions into Primate Nuclear Genomes Suggest the Use of Numts as a Tool for Phylogeny. *Molecular Biology and Evolution* **2009**, *26* (10), 2175–2179. <https://doi.org/10.1093/molbev/msp131>.
- (19) Lang, M.; Sazzini, M.; Calabrese, F. M.; Simone, D.; Boattini, A.; Romeo, G.; Luiselli, D.; Attimonelli, M.; Gasparre, G. Polymorphic NumtS Trace Human Population Relationships. *Human Genetics* **2012**, *131* (5), 757–771.
- (20) Bazán, E. G. ¡DEJADME VIVIR!: Apuntes de Paleontología Humana: Registro Fósil y Esquema Evolutivo. ¡DEJADME VIVIR!, 2012.
- (21) Ensembl Genome Browser <https://m.ensembl.org/index.html> (accessed 2021 -09 -05).
- (22) UCSC Genome Browser Gateway <https://genome.ucsc.edu/cgi-bin/hgGateway> (accessed 2021 -09 -05).
- (23) Fujita, P. A.; Rhead, B.; Zweig, A. S.; Hinrichs, A. S.; Karolchik, D.; Cline, M. S.; Goldman, M.; Barber, G. P.; Clawson, H.; Coelho, A.; Diekhans, M.; Dreszer, T. R.; Gardine, B. M.; Harte, R. A.; Hillman-Jackson, J.; Hsu, F.; Kirkup, V.; Kuhn, R. M.; Learned, K.; Li, C. H.; Meyer, L. R.; Pohl, A.; Raney, B. J.; Rosenbloom, K. R.; Smith, K. E.; Haussler, D.; Kent, W. J. The UCSC Genome Browser Database: Update 2011. *Nucleic Acids Research* **2011**, *39* (Database issue), D876–D882. <https://doi.org/10.1093/nar/gkq963>.
- (24) Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Research*. **2002**, *12* (4), 656–664. <https://doi.org/10.1101/gr.229202>.
- (25) Mourier, T.; Hansen, A. J.; Willerslev, E.; Arctander, P. The Human Genome Project Reveals a Continuous Transfer of Large Mitochondrial Fragments to the Nucleus. *Molecular Biology and Evolution* **2001**, *18* (9), 1833–1837. <https://doi.org/10.1093/oxfordjournals.molbev.a003971>.
- (26) Hazkani-Covo, E.; Sorek, R.; Graur, D. Evolutionary Dynamics of Large Numts in the Human Genome: Rarity of Independent Insertions and Abundance of Post-Insertion Duplications. *Journal of Molecular Evolution* **2003**, *56* (2), 169–174. <https://doi.org/10.1007/s00239-002-2390-5>.
- (27) Dayama, G.; Zhou, W.; Prado-Martinez, J.; Marques-Bonet, T.; Mills, R. E. Characterization of Nuclear Mitochondrial Insertions in the Whole Genomes of Primates. *bioRxiv* **2020**, 2020.02.24.963504. <https://doi.org/10.1101/2020.02.24.963504>.
- (28) Dayama, G.; Zhou, W.; Prado-Martinez, J.; Marques-Bonet, T.; Mills, R. E. *Characterization of Nuclear Mitochondrial Insertions in the Whole Genomes of Primates*; 2020; p 2020.02.24.963504. <https://doi.org/10.1101/2020.02.24.963504>.
- (29) Gunbin, K.; Peshkin, L.; Popadin, K.; Annis, S.; Ackermann, R. R.; Khrapko, K. Integration of MtDNA Pseudogenes into the Nuclear Genome Coincides with Speciation of the Human Genus. A Hypothesis. *Mitochondrion* **2017**, *34*, 20–23. <https://doi.org/10.1016/j.mito.2016.12.001>.
- (30) Calabrese, F. M.; Balacco, D. L.; Preste, R.; Diroma, M. A.; Forino, R.; Ventura, M.; Attimonelli, M. NumtS Colonization in Mammalian Genomes. *Scientific Reports* **2017**, *7* (1). <https://doi.org/10.1038/s41598-017-16750-2>.
- (31) Ensembl Genome Browser <https://m.ensembl.org/index.html> (accessed 2021 -09 -05).



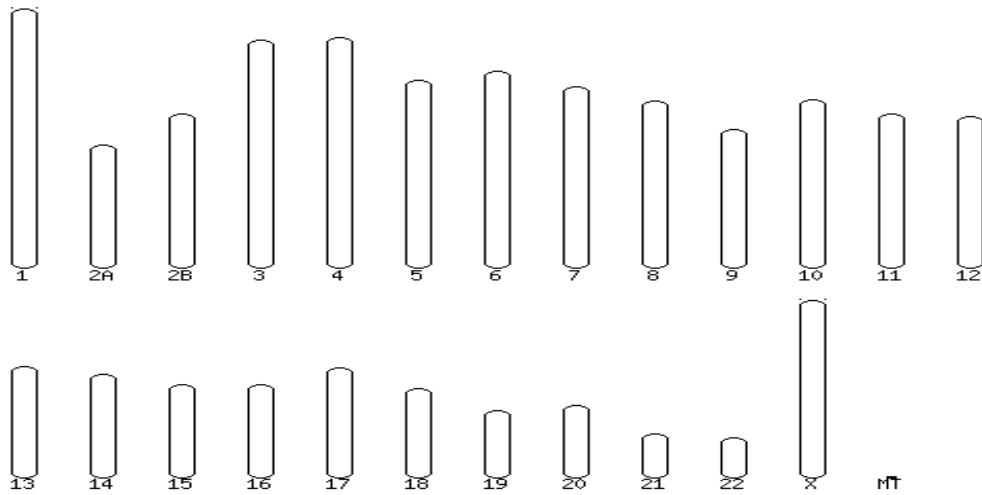
6. ANEXOS.



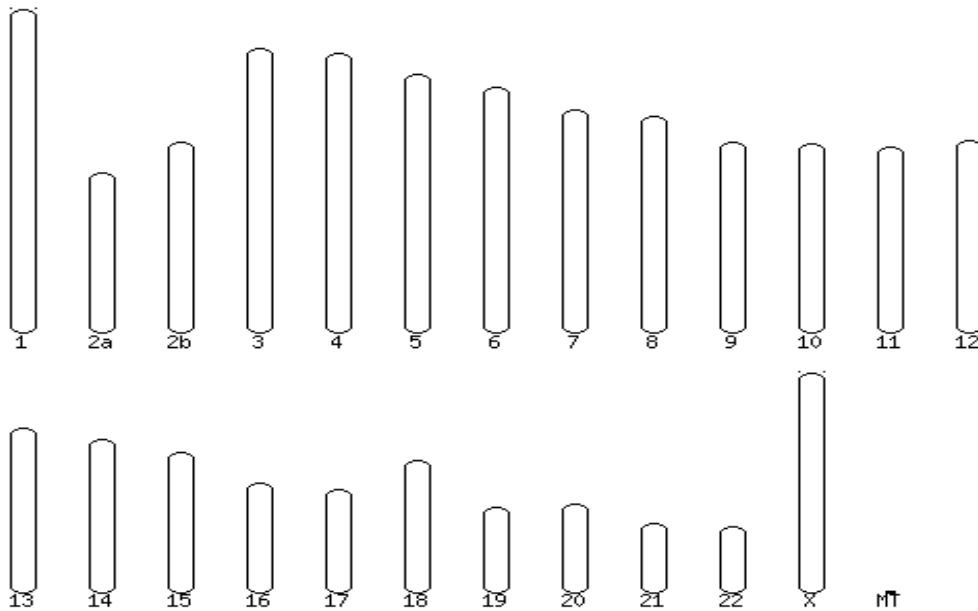
Anexo 1. Cariotipo humano. Obtenida del navegador genómico Ensembl, con el ensamblaje: GCA_000001405.28 ³¹



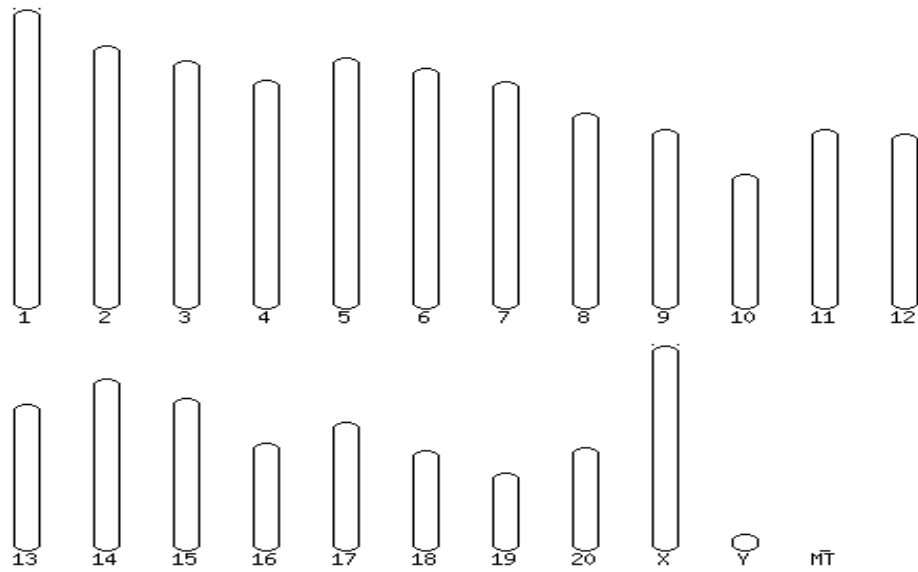
Anexo 2. Cariotipo de chimpancé. Obtenido de Ensembl, con el ensamblaje: GCA_000001515.5 ³¹.



Anexo 3. Cariotipo de gorila con el ensamblaje GCA_000151905.3. Obtenida del navegador genómico Ensembl³¹.



Anexo 4. Cariotipo de orangután con el ensamblaje: GCA_000003175.1. Obtenida del navegador genómico Ensembl³¹.



Anexo 5. Cariotipo de macaco con el ensamblaje GCA_003339765.3. Obtenida del navegador genómico Ensembl³¹.