



Histonas de Mamíferos

Montero Rasgado Dulce María¹, Mariana Jocelyn Robles Lara¹, Gisselle Ibarra Moreno¹ y
Diego Alonso Medina Amayo¹

1. Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad Universitaria, Coyoacán 04510, CDMX, México.

Introducción

El genoma eucarionte se caracteriza por encontrarse altamente empaquetado en una estructura de cromatina, este proceso es llamado compactación del DNA. La cromatina es el conjunto de DNA y proteínas asociadas, de entre estas proteínas, las histonas son las más abundantes. Ellas en asociación con el DNA, conforman la unidad básica estable de la cromatina, el nucleosoma (Lee, J. Y., & Orr-Weaver, 2001).

El nucleosoma es un octámero de histonas. Cada octámero está conformado por cuatro pares de histonas diferentes, H2A, H2B, H3 y H4, estas son las *histonas canónicas*. Además, hay una histona linker H1 que ayuda al correcto ensamblaje de los nucleosomas (Workman & Kingston, 1998; Lee, J. Y., & Orr-Weaver, 2001).

Debido a su importancia en el empaquetamiento del DNA, y por ende, en la estabilidad de la cromatina y la regulación génica a través de sus modificaciones post-traduccionales, las histonas son proteínas ampliamente conservadas entre los organismos eucariontes. Inclusive, por cada histona canónica, existen varios genes diferentes que la codifican y se encuentran en clusters. Sin embargo, a lo largo de la evolución, se han originado variantes de histonas con sus respectivos genes alejados de los clusters de las histonas canónicas. Dichas histonas, tienen características muy específicas que cambian por completo la estructura de la cromatina en estadio y tejido específico (Ausio J., 2006; Albig, W. *et al.*, 1997).

En mamíferos, los centrómeros presentan la característica presencia de la proteína del centrómero A (CENPA), que es una variante de la histona H3 canónica. CENPA es tiempo-específica de la división celular y es indispensable para conservar la integridad de la cromatina centromérica durante la intervención de los microtúbulos en el cinetocoro en la prometáfase hasta la anafase (Hu, H. *et al.*, 2011). Debido a su vital importancia durante la mitosis, CENPA es una variante de histonas altamente conservada entre mamíferos. Si se analiza su secuencia de aminoácidos entre distintos grupos de mamíferos, se espera que sus variaciones sean mínimas.

Pregunta de investigación

¿Qué tanto divergen las secuencias de aminoácidos de la histona H3 canónica respecto a su variante CENPA en algunas especies de mamíferos?

Hipótesis

La divergencia de las secuencias de aminoácidos de la proteína CENPA variará muy poco dentro de cualquier grupo de mamíferos.

Objetivo general

Identificar las diferencias en la secuencia de aminoácidos entre H3 y CENPA en 17 especies de mamíferos.

Objetivos particulares

1. Obtener secuencia de la proteína H3 canónica y su variante CENPA en 17 especies de mamíferos.
2. Comparar las secuencias de proteínas.
3. Obtener índices de divergencia y parentesco entre secuencias.
4. Obtener un árbol filogenético a partir de las secuencias de proteínas.

Métodos

Selección de especies

Debido a que se espera que las secuencias aminoácidos se encuentren bien conservadas y con poca divergencia, se seleccionaron 17 especies pertenecientes a dos super-órdenes de mamíferos, Laurasiatheria y Euarchontoglires:

1. *Equus caballus* (Caballo).
2. *Cebus imitator* (Capuchino cariblanco).
3. *Pan troglodytes* (Chimpancé).
4. *Loxodonta africana* (Elefante africano).
5. *Mirounga angustirostris* (Elefante marino)
6. *Halichoerus grypus* (Foca gris).
7. *Felis catus* (Gato doméstico).
8. *Cricetulus griseus* (Hámster chino)
9. *Homo Sapiens* (Humano).
10. *Mustela putorius* (Hurón).
11. *Sus scrofa* (Jabalí).
12. *Desmodus rotundus* (Murciélago vampiro).
13. *Pongo abelii* (Orangután).
14. *Ursus maritimus* (Oso polar).
15. *Mus musculus* (Ratón).
16. *Suricata suricatta* (Suricata).
17. *Vulpes lagopus* (Zorro ártico).

Una vez delimitado el grupo de mamíferos, se utilizó GenBank para descargar la secuencia de la proteína H3 canónica de *homo sapiens* y las secuencias de la histona variante CENPA del resto de especies de mamíferos incluyendo a *homo sapiens*.

Se alinearon las secuencias utilizando Clustal Omega.

Se obtuvo el árbol filogenético a partir de los alineamientos de las secuencias

Se obtuvo el índice de divergencias entre las secuencias de CENPA a través de la implementación de un script de Python (Ver repositorio GitHub).

Comparación de secuencia de H3 con las secuencias de CENPA

H3 HOMO SAPIENS	AKRVTMPKDIQLARRIGERA---	136
HOMO_SAPIENS	AGRVTLFPKDVQLARRIGLEEGLG	114
PAN TROGLODYTES	AGRVTLFPKDVQLARRIGLEEGLG	138
PONGO ABELII	AGRVTLFPKDVQLARRIGLEEGLG	140
DESMODUS ROTUNDUS	AGRVTLFPKDVQLARRIGIEEGLG	136
LOXODONTA AFRICANA	AGRVTLYPKDVQLARRIRIGIQEGLG	136
EQUUS CABALLUS	AGRVTLFPKDVQLARRIRIGIQEGLG	139
VULPES LAGOPUS	AGRVTLFPKDVQLARRIRIGIQEGLG	135
SUS SCROFA	AGRVTLFPKDIQLARRIRIGIQEGLG	131
CEBUS IMITATOR	AGRVTLFPKDVQLARRIRIGIEEGLG	138
SURICATA SURICATTA	AGRVTLFPKDVQLARRIRIGIQEGLG	137
HALICHOERUS GRYPUS	AGRVTLFPKDVQLARRIRIGIQEGLG	137
MIROUNGA ANGUSTIROSTRIS	AGRVTLFPKDVQLARRIRIGIQEGLG	137
FELIS CATUS	AGRVTLFPKDVQLARRIRIGIQEGLG	138
MUSTELA PUTORIUS	AGRVTLFPKDVQLARRIRIGIQEGLG	138
URSUS MARITIMUS	AGRVTLFPKDVQLARRIRIGIQEGLG	138
MUS MUSCULUS	AGRVTLFPKDIQLTRRIRIGFEGGLP	108
CRICETULUS GRISEUS	AGRVTFPKDIQLTRRIRIGIEGGLG	129
	* * * * * * * * * * * * * * * *	

Fig. 1. Alineamiento de secuencias de la histona H3 canónica de *Homo sapiens* respecto a su variante CENPA con todas las especies de mamíferos seleccionadas.

Comparación de las secuencias CENPA

HOMO SAPIENS	LLIRKLPFSRLA-----AEAFVLHFLFEDAYLLTLHAGRV	93
PAN_TROGLODYTES	LLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAEEAFVLHFLFEDAYLLTLHAGRV	117
PONGO ABELII	LLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAEEAFVLHFLFEDAYLLTLHAGRV	119
DESMODUS_ROTUNDUS	LLLRKTPFSRLAREICVKFTRGVDFNWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	115
LOXODONTA_AFRICANA	LLLRKRPFGLLAREVCAFFTRGVDFNWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	115
EQUUS_CABALLUS	LLLRKYPFSRLAREICITFTTRGVDFNWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	118
VULPES_LAGOPUS	LLLRKNPFSLRVREICVKFTRGVDFSWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	114
SUS_SCROFA	LLLRKNPFSLAREICVKFTRGVDFNWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	110
CEBUS_IMITATOR	LLLRKYPFSRLAREICVKFTRGVDFNWQAQALLALQEAEEAFVLHFLFEDAYLLTLHAGRV	117
SURICATA_SURICATTA	LLLRKSPFGLAREICVKFTRGVDFCWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	116
HALICHOERUS_GRYPUS	LLLRKSPFSRLAREICITFTTRGVDFSWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	116
MIROUNGA_ANGUSTIROSTRIS	LLLRKSPFSRLAREICITFTTRGVDFSWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	116
FELIS_CATUS	LLLRKSPFGLAREICVKFTRGVDFYWAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	117
MUSTELA_PUTORIUS	LLLRKSPFGLAREICIKFTTRGVDFSWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	117
URSUS_MARITIMUS	LLLRKSPFGLAREICVKFTRGVDFSWQAQALLALQEAEEAFVLHFLFEDAYLLSLHAGRV	117
MUS_MUSCULUS	LLFRKPPFSMVVREICEKFSRGVDFWQAQALLALQEAEEAFILHFLFEDAYLLSLHAGRV	87
CRICETULUS_GRISEUS	LLLRKLPFSRVVREICGKFTRGVDFCWQAQALLALQEAEEAFVLHFLFEDAYLLTLHAGRV	108
	:* ** : , *** ,***** ,*****	
HOMO SAPIENS	TLFPKDVQLARRIRGLEEGLG	114
PAN_TROGLDYTES	TLFPKDVQLARRIRGLEEGLG	138
PONGO ABELII	TLFPKDVQLARRIRGLEEGLG	140
DESMODUS_ROTUNDUS	TLFPKDVQLARRIRGLEEGLG	136
LOXODONTA_AFRICANA	TLYPKDVQLARRIRGIEEGLG	136
EQUUS_CABALLUS	TLFPKDVQLARRIRGIEEGLG	139
VULPES_LAGOPUS	TLFPKDVQLARRIRGIEEGLG	135
SUS_SCROFA	TLFPKDIQLARRIRGIEEGLG	131
CEBUS_IMITATOR	TLFPKDVQLARRIRGIEEGLG	138
SURICATA_SURICATTA	TLFPKDVQLARRIRGIEEGLG	137
HALICHOERUS_GRYPUS	TLFPKDVQLARRIRGIEEGLG	137
MIROUNGA_ANGUSTIROSTRIS	TLFPKDVQLARRIRGIEEGLG	137
FELIS_CATUS	TLFPKDVQLARRIRGIEEGLG	138
MUSTELA_PUTORIUS	TLFPKDVQLARRIRGIEEGLG	138
URSUS_MARITIMUS	TLFPKDVQLARRIRGIEEGLG	138
MUS_MUSCULUS	TLFPKDIQLTRIRIGFEGGLP	108
CRICETULUS_GRISEUS	TIFPKDIQLTRIRIGIEGGLG	129
	*:*****:*****:*	

Fig. 2. Alineación de secuencias de la histona variante CENPA respecto el resto de especies seleccionadas.

Diferencias entre las secuencias de aminoácidos entre H3 y CENPA

H3 HOMO SAPIENS	MARTKQTARKSTGGKAPRKQL---ATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQ	56
HOMO_SAPIENS	MGPRR---RSRKPEAPRRSPSPPTPTPGPSRRGPSLGASSHQHSRRR-QGWLKEIRKLQ	55
PAN_TROGLDYTES	MGPRR---RSRKPEAPRRRSPS---PTPGPSRRGPSLGASSHQHSRRR-QGWLKEIRKLQ	53
PONGO ABELII	MGPRR---RSRKPEAPRRRSPSPPTPTPGPSRRGPSLGASSHQHSRRR-QAWLKEIRKLQ	55
DESMODUS_ROTUNDUS	MGPRR---RKRKTETPRRRAASPT--PAAPRPAPSTGTSFRRR---RRHLILKEIKNLQ	51
LOXODONTA_AFRICANA	MGPRR---RRRKPEPLRRSPSPPT--PAPRR--PSGAPSRRRGQWRP-RVLKEIRKLQ	51
EQUUS_CABALLUS	MGPRR---RIRKPETPRRRAASPT--PGRPRPGPPSGASSRPSGRRRSTKVLREIRKLQ	54
VULPES_LAGOPUS	MGPR---RKPEVPKRRPASPA--PSAPRRGPSLGTSSRHGSRKH-LVLKEIRKLQ	50
SUS_SCROFA	MGPRR---RRRKPETPRRRPASPA--PAAPRRGPPL-----GPRRH-RVLREIRTLQ	46
CEBUS_IMITATOR	MGPRR---RSRKPETPRRRRASPT--PSPTRRGPSLGASSRPRGYRRQ-GWLKEIRKLQ	53
SURICATA_SURICATTA	MGPRR---RSRKPEVPRRRTASPT--PGTPRRGP-SGTPSRQRRPQRH-RVLKEIRKLQ	52
HALICHOERUS_GRYPUS	MGPRR---RIRKPEVPRRRAVSPA--PGPPRRGP-VGTPSRQRGPPRS-RVLREIRTLQ	52
MIROUNGA_ANGUSTIROSTRIS	MGPRR---RIRKPEVPRRRTVSPA--PGPPRRGP-LGTPSRQRGPPRS-RVLKEIRTLQ	52
FELIS_CATUS	MGPRR---RRRKPEVPKRRRTASPT--SPPPRRAPSLGASSRQRGPPRS-RVLKEIRKLQ	53
MUSTELA_PUTORIUS	MGPRR---RSRKAEPVRRRTASPA--AGPPRRAPSLGASSRQRGPPRS-RILKEIRKLQ	53
URSUS_MARITIMUS	MGPRR---RIRKPEVPRRRTASPA--PSPPPRRGPRLGAPSRQRGPPRL-RVLREIRMLQ	53
MUS_MUSCULUS	-----M-----DG---SQTLLRRQKFMWLKEIKTLQ	23
CRICETULUS_GRISEUS	MGPR-----RKPRTPRRRPSPVPGPS-----RR---SSRPGKRRKFLWLKEIKKLQ	44
	*:***:*	

H3_HOMO_SAPIENS	KSTELLIRKLPFQRLVREIAQDFKT--DLRFQSAAGALQEASEAYLVGLFEDTNLCAT	114
HOMO_SAPIENS	KSTHLLIRKLPFSRLA-----AEAFVLHFLFEDAYLLTLH	89
PAN_TROGLODYTES	KSTHLLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHFLFEDAYLLTLH	113
PONGO_ABELII	KSTHLLIRKLPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHFLFEDAYLLTLH	115
DESMODUS_ROTUNDUS	KTTHLLLRKTPFSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	111
LOXODONTA_AFRICANA	KSTDLLLRKTPFGLLAREVCAFFTRGVDFNWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	111
EQUUS_CABALLUS	KSTHLLLRKYPFPSRLAREICITFTRGVDFNWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	114
VULPES_LAGOPUS	KSTNLLLRKNPFPSRLVREICVKFTRGVDFSWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	110
SUS_SCROFA	KSTHLLLRKNPFQRLAREICVQFTRGVDFNWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	106
CEBUS_IMITATOR	KSTHLLLRKYPFPSRLAREICVKFTRGVDFNWQAQALLALQEAAEAFLVHFLFEDAYLLTLH	113
SURICATA_SURICATTA	KSTNLLLRKSPFQRLAREICVKFTRGVDFCWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	112
HALICHOERUS_GRYPUS	KSTALLIRKSPFSRLAREICITFTRGVDFSWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	112
MIROUNGA_ANGUSTIROSTRIS	KSTALLIRKSPFSRLAREICITFTRGVDFSWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	112
FELIS_CATUS	KSTDLLLRKSPFQRLAREICVKFTRGVDFYWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	113
MUSTELA_PUTORIUS	KSTGLLIRKSPFQRLAREICIKFTRGVDFSWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	113
URSUS_MARITIMUS	KSTGLLIRKSPFQRLAREICVKFTRGVDFSWQAQALLALQEAAEAFLVHFLFEDAYLLSLH	113
MUS_MUSCULUS	KSTDLLFRKPPFSMVVREICEKFSRGVDFWQAQALLALQEAAEAFLHFLFEDAYLLSLH	83
CRICETULUS_GRISEUS	RSTDLLLRKLPFSRVVREICGKFTRGVDLCWQAQALLALQEAAEAFLVHFLFEDAYLLTLH	104
	::* **:* ** ::	::*:*: ***** * ::*
H3_HOMO_SAPIENS	AKRVTIMPDKDIQLARRIRGERA--	136
HOMO_SAPIENS	AGRVTLFPKDVQLARRIRGLEEGLG	114
PAN_TROGLODYTES	AGRVTLFPKDVQLARRIRGLEEGLG	138
PONGO_ABELII	AGRVTLFPKDVQLARRIRGLEEGLG	140
DESMODUS_ROTUNDUS	AGRVTLFPKDVQLARRIRGLEEGLG	136
LOXODONTA_AFRICANA	AGRVTLFPKDVQLARRIRGLEEGLG	136
EQUUS_CABALLUS	AGRVTLFPKDVQLARRIRGLEEGLG	139
VULPES_LAGOPUS	AGRVTLFPKDVQLARRIRGLEEGLG	135
SUS_SCROFA	AGRVTLFPKDIQLARRIRGLEEGLG	131
CEBUS_IMITATOR	AGRVTLFPKDVQLARRIRGLEEGLG	138
SURICATA_SURICATTA	AGRVTLFPKDVQLARRIRGLEEGLG	137
HALICHOERUS_GRYPUS	AGRVTLFPKDVQLARRIRGLEEGLG	137
MIROUNGA_ANGUSTIROSTRIS	AGRVTLFPKDVQLARRIRGLEEGLG	137
FELIS_CATUS	AGRVTLFPKDVQLARRIRGLEEGLG	138
MUSTELA_PUTORIUS	AGRVTLFPKDVQLARRIRGLEEGLG	138
URSUS_MARITIMUS	AGRVTLFPKDVQLARRIRGLEEGLG	138
MUS_MUSCULUS	AGRVTLFPKDIQLTRRRIRFEGGLP	108
CRICETULUS_GRISEUS	AGRVTLFPKDIQLTRRRIRFEGGLG	129
	* ***: ***:***:*****	

Fig. 3. Alineamiento de las secuencias de aminoácidos entre la histona CENPA de todos los mamíferos seleccionados y la H3 canónica en *homo sapiens*. Los aminoácidos idénticos entre las secuencias CENPA y H3 se encuentran resaltadas en azul, en amarillo las variaciones homogéneas de CENPA y H3 en el resto de especies, y en rosa las variaciones únicas de CENPA en alguna especie en concreto.

Índice de similitud

Para facilitar el análisis y la discusión, creamos un programa que recibe los datos de Clustal Omega, a partir de los cuales comparamos las secuencias de todos los organismos a estudiar con la secuencia canónica del *homo sapiens*.

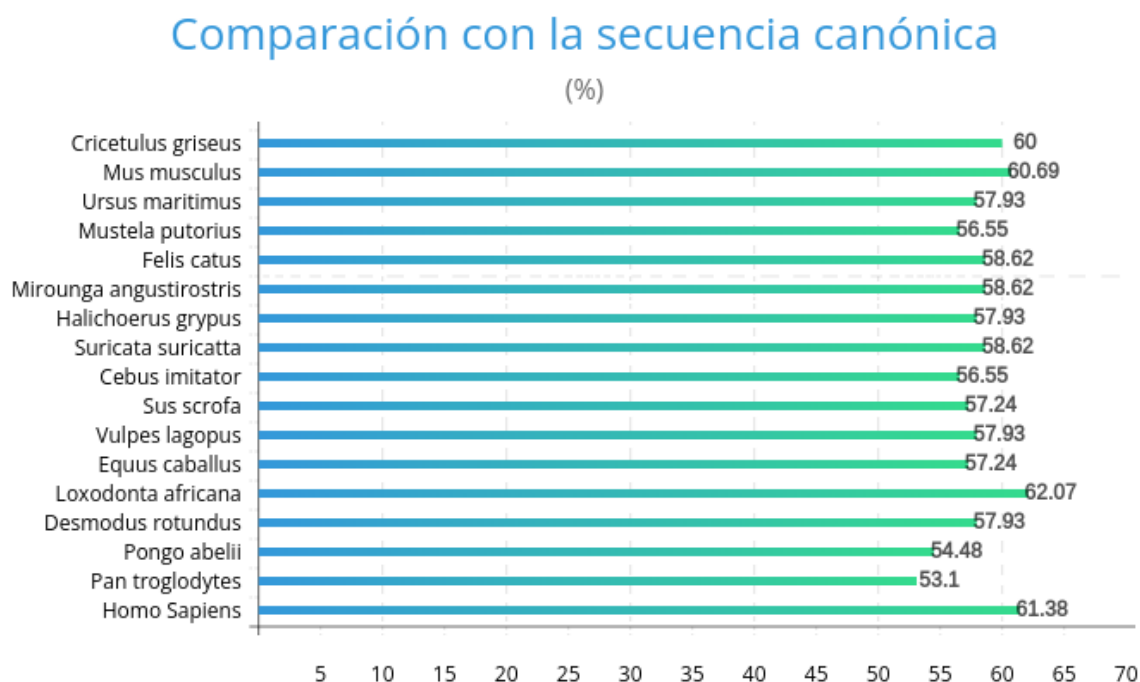


Fig. 4. Porcentaje de similitud entre las secuencias de la variante CENPA respecto a la histona H3 canónica de *Homo sapiens*.

Árbol filogenético

Se utilizó el árbol generado por Clustal Omega, este es generado mediante el método de distancia *BLOSUM62 score matrix* y el método de agrupación de uniones por vecinos (*Neighbor joining*) ya que es el método estándar para árboles de secuencias de DNA y proteínas.

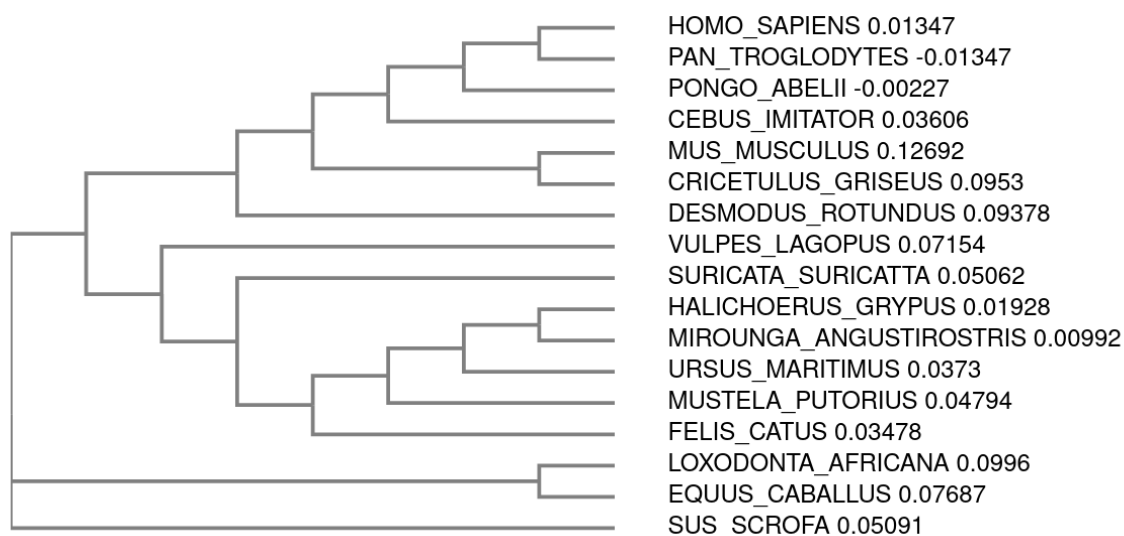


Fig. 5. Árbol filogenético obtenido de las secuencias de la proteína CENPA entre las 17 especies seleccionadas.

Discusión

En el presente trabajo, obtuvimos alineamientos de secuencias para comparar qué tanto cambia la secuencia de aminoácidos entre la H3 canónica y su variante CENPA (fig. 1 y 2). Como se esperaba, las secuencias variaban en unos cuantos aminoácidos entre las diferentes especies. Sin embargo, no varían aleatoriamente sino que siguen patrones similares (Fig. 3). Concretamente, las variaciones más abundantes se encontraban en el extremo amino terminal (N-terminal), ya que en ese sitio es donde ocurren la mayoría de las modificaciones post-traduccionales.

Además, el patrón de cambio se comienza a ver más claro hacia el final de la secuencia de aminoácidos (Fig. 3), lo cual sugiere la presencia de los dominios proteicos de CENPA. Dichos dominios le permiten a la proteína interactuar con otras proteínas o incluso con la secuencia de DNA. La integridad de dichos dominios de interacción es fundamental para el correcto funcionamiento y plegamiento de las proteínas, por lo que es de esperar que no varíen entre las especies seleccionadas, ya que las proteínas podrían ser inviables. Estos patrones hacia el final de la secuencia nos dan una idea de la ubicación de los dominios de interacción de CENPA.

Por otro lado, respecto al índice de divergencia calculado, se obtuvo que la secuencia de CENPA más parecida a la de la H3 canónica de *homo sapiens* es *Loxodonta africana* y *Mus musculus*, mientras que la más divergente es *Pan troglodytes*.

Finalmente, el árbol filogenético obtenido mostró que las secuencias de CENPA más relacionadas con *Homo sapiens* son *Pan troglodytes* y *Pongo abelii*. Esto concuerda con la filogenia más reciente en consenso, donde la mayoría de especies obedecen la clasificación de los super órdenes Laurasiatheria y Euarchontoglires.

Conclusiones

En conclusión, los alineamientos de secuencias y sus comparaciones nos permitieron tener una idea sobre la conformación de CENPA, sus dominios de interacción y cómo varían en cada especie respecto a la secuencia de la H3 canónica. El árbol filogenético obtenido con las secuencias de CENPA, se apega a la clasificación actual de los dos super-órdenes seleccionados en el presente trabajo.

Referencias

Lee, J. Y., & Orr-Weaver, T. L. (2001). Chromatin. *Encyclopedia of Genetics*, 340–343. doi:10.1006/rwgn.2001.0199

Workman, J.L. and Kingston, R.E. (1998) Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.*, 67, 545–579.

Ausio, J. (2006). Histone variants--the structure behind the function. *Briefings in Functional Genomics and Proteomics*, 5(3), 228–243. doi:10.1093/bfgp/ell020

Hu, H., Liu, Y., Wang, M., Fang, J., Huang, H., Yang, N., ... Xu, R.-M. (2011). Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes & Development*, 25(9), 901–906. doi:10.1101/gad.204511

Albig, W., Kioschis, P., Poustka, A., Meergans, K., & Doenecke, D. (1997). Human Histone Gene Organization: Nonregular Arrangement within a Large Cluster. *Genomics*, 40(2), 314–322. doi:10.1006/geno.1996.4592