

Análisis evolutivo del gen *PRKN* en el orden Primates

Yordy Stiven Cangrejo Useda

Trabajo de Grado para Optar al Título de Biólogo

Director

Francisco José Martínez Pérez

Doctor en Ciencias

Codirector

Lina María Vera Cala

Doctora en Epidemiología

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de Biología

Biología

Bucaramanga

2023

Dedicatoria

Sin duda, este trabajo, el lugar donde estoy y la persona que soy, todo, se lo debo a mi madre, quien durante toda mi vida me ha brindado su apoyo incondicional, sin esperar nada a cambio. El amor que he recibido de ella, el esfuerzo de su trabajo y el ejemplo de lucha que me ha dado, hoy se lo demuestro con este trabajo, que marca el inicio de una nueva etapa de mi vida, la cual dedicaré a retribuirle todo lo que ha depositado en mí...

A mi padre, quien en la adversidad ha estado a mi lado y me ha sabido guiar a su manera. A él, dedico también este proyecto como muestra que de a poco le compensaré los esfuerzos de su vida...

A mis padres, que han sido el motor de mi vida, el motivo de continuar adelante a pesar de los tropiezos. Es a ellos, a quienes hoy dedico este pequeño gran paso, como muestra de agradecimiento y respeto, y como promesa de que voy a continuar esforzándome tal y como ellos me lo han enseñado, para compensarles lo que se merecen.

A ustedes dos, muchísimas gracias.

Agradecimientos

A la Universidad Industrial de Santander, a la Facultad de Ciencias y a la Escuela de Biología, junto con sus administrativos y personal de trabajo.

A quien más que un director, ha sido un mentor de vida y me ha enseñado cómo debe ser un buen científico, quien muchas veces ha confiado más en mí que yo mismo y que me permitió trabajar con él durante todo este tiempo: al Dr. Francisco José Martínez Pérez.

A la Dra. Lina María Vera Cala y al Dr. Carlos Jaime Barrios Hernández por permitirme hacer parte del grupo de investigación Cómputo Avanzado y a Gran Escala, CAGE.

A mis compañeros, amigos del Laboratorio de Genómica Celular y Aplicada, quienes han formado parte de este proceso.

A mis amigos, quienes sabrán disculparme, pues el espacio es tan limitado y el agradecimiento es tan grande, que no se podría resumir en un par de oraciones. Pero espero se pueda demostrar lo mucho que significan para mí y lo agradecido que estoy con ustedes.

A mis amigos con los que inicié esta aventura y que, por motivos de cada uno, hoy nos encontramos distantes. Sin embargo, espero de todo corazón que se vuelvan los profesionales que desean. Infinitas gracias.

A esa persona que ocupó y ocupará un espacio en mi vida, y a quien espero la vida le retribuya lo que me dio.

A esa amiga que en poco tiempo se ha vuelto alguien importante en mi vida y me ha ayudado a sobrellevar este proceso, gracias.

A mi familia y a todas las personas que depositaron su apoyo en mí.

A ustedes lectores.

Tabla de Contenido

	Pág.
Introducción	12
2. Objetivos	15
2.1 Objetivo General	15
2.2 Objetivos Específicos.....	15
3. Competencias	16
3.1 Competencias cognitivas	16
3.2 Competencias actitudinales.....	16
4. Materiales y Métodos.....	17
4.1 Obtención y Caracterización de secuencias	17
4.2 Caracterización de la fase de intrones y su estado evolutivo.....	17
4.3 Caracterización de exones y marcos abiertos de lectura dentro de los intrones del gen <i>PRKN</i> de Primates.....	18
4.4 Análisis filogenético	18
5. Resultados	19
5.1 Organización estructural del gen <i>PRKN</i> en el orden Primates	19
5.2 Fases y estado de conservación de los intrones del gen <i>PRKN</i> en el orden Primates.....	27
5.3 Exones y Marcos Abiertos de Lectura dentro de los intrones del gen <i>PRKN</i> en el orden Primates	34
5.4 Análisis filogenético del gen <i>PRKN</i> en el orden Primates.....	36
6. Discusión.....	39
7. Conclusiones	45

8. Recomendaciones	46
Referencias Bibliográficas	47

Lista de Tablas

	Pág.
Tabla 1 Especies con la secuencia del gen <i>PRKN</i> completo	20
Tabla 2 Dominios proteicos presentes en los intrones del gen <i>PRKN</i> en primates	34

Lista de Figuras

	Pág.
Figura 1 Delimitación de exones e intrones del gen <i>PRKN</i> en <i>L. catta</i> , <i>M. murinus</i> y <i>C. jacchus</i>	21
Figura 2 Delimitación de exones e intrones del gen <i>PRKN</i> en la subfamilia Homininae	22
Figura 3 Organización de exones e intrones del gen <i>PRKN</i> en las especies <i>P. abelii</i> , <i>P. pygmaeus</i> , <i>S. syndactylus</i> y <i>N. leucogenys</i>	23
Figura 4 Organización de exones e intrones del gen <i>PRKN</i> en <i>M. fascicularis</i> , <i>M. mulatta</i> , <i>M.</i> <i>thibetana thibetana</i> , <i>R. roxellana</i> y <i>T. francoisi</i>	24
Figura 5 Dominios estructurales de Parkin.....	27
Figura 6 Caracterización de las fases de intrones de <i>C. jacchus</i> , <i>M. murinus</i> y <i>L. catta</i>	28
Figura 7 Caracterización de las fases de intrones en la subfamilia Homininae.....	30
Figura 8 Caracterización de las fases de intrones de <i>P. abelii</i> , <i>P. pygmaeus</i> , <i>S. syndactylus</i> y <i>N.</i> <i>leucogenys</i>	32
Figura 9 Caracterización de la fase de intrones de <i>M. fascicularis</i> , <i>M. mulatta</i> , <i>M. thibetana</i> <i>thibetana</i> , <i>R. roxellana</i> y <i>T. francoisi</i>	33
Figura 10 Sitios funcionales identificados en los intrones del gen <i>PRKN</i> en el orden Primates. 36	
Figura 11 Relación filogenética del gen <i>PRKN</i> en el orden Primates	38

Lista de Apéndices

pág.

Los Apéndices mencionados en este informe se encuentran en línea en el siguiente enlace:

https://drive.google.com/drive/folders/1Ki1kfC8L8oAa5of2-_6Ss_0zauhAvqia?usp=drive_link.

Glosario

ARNm: Ácido ribonucleico mensajero.

Estado evolutivo: se define de acuerdo con la presencia o ausencia de un intrón, producto de los distintos ARNm, a través de las secuencias genéticas.

Exón: región nucleotídica de una secuencia genética que codifica para un ARNm, el cual se traducirá a proteína.

Fase del intrón: la posición que puede ocupar un intrón dentro de la secuencia genética con base a los codones. Se definen tres fases: intrón fase 0, ubicado en medio de dos codones; intrón fase 1, ubicado entre el primer y segundo nucleótido de un codón; intrón fase 2, ubicado entre el segundo y tercer nucleótido de un codón.

Intrón: región nucleotídica de una secuencia genética que se encuentra en medio de exones y que puede o no, generar ARNm los cuales no se van a traducir.

ORF: Marco Abierto de Lectura por sus siglas en inglés “*Open Read Frame*”.

PRKN: gen *Parkin RBR E3 Ubiquitin Protein Ligase*.

Variante: secuencia de ARNm producto del procesamiento alternativo y predicciones bioinformáticas reportadas para cada especie en la base de datos GenBank.

Resumen

Título: Análisis evolutivo del gen *PRKN* en el orden Primates*

Autor: Yordy Stiven Cangrejo Useda**

Palabras Clave: Enfermedad de Parkinson, *PRKN*, Primates, Evolución, Fase del Intrón, Terapia génica.

Descripción: La Enfermedad de Parkinson es un trastorno neurodegenerativo que afecta a la población mundial. Se estima que el número de personas afectadas aumentará a los 17 millones para el 2040. Aunque no se conocen las características fisiológicas de la enfermedad, se ha establecido una relación con variantes y mutaciones del gen *PRKN*. El gen se caracteriza por tener un tamaño promedio de 1 200 000 pb. Sin embargo, el ARNm tiene un tamaño promedio de 4000 pb, lo cual hace que tan sólo el 0.3% del gen sean exones. La diferencia de tamaño entre exones e intrones predispone diversas preguntas acerca de su origen, así como la posibilidad de encontrar ARNm dentro de ellos, sin embargo, no se sabe el estado evolutivo de *PRKN* en Primates, ni las posibles implicaciones fisiológicas del mismo. Para esto, se descargaron del GenBank las secuencias del gen *PRKN* de Primates y se delimitaron exones e intrones a partir de sus variantes. Se caracterizó la estructura proteica y se identificó la fase de los intrones de *PRKN*. Se hizo una comparación de posibles ORFs y exones dentro de los intrones de *PRKN*, y se estableció una relación evolutiva a partir de un árbol filogenético. La organización estructural promedio del gen se compone por entre 11 y 12 exones, los cuales mantienen un patrón de conservación hacia el extremo 3'. Así mismo, las fases de los intrones se mantienen conservadas hacia esta misma región, lo que significa que son intrones antiguos. La estructura de la proteína se mostró con mayor variación. Se determinaron posibles ORFs dentro de los intrones de las secuencias, así como sitios funcionales, los cuales se mostraron conservados. La especie *C. jacchus* mostró una agrupación filogenética en un suborden diferente. Aunque el gen esté conservado, los diferentes arreglos exónicos y posibles ORFs promueven cambios en *PRKN*.

* Trabajo de Grado

** Facultad de Ciencias. Escuela de Biología. Biología. Director: Francisco José Martínez Pérez. Doctor en Ciencias. Codirector: Lina María Vera Cala. Doctora en Epidemiología

Abstract

Title: Evolutionary analysis of the *PRKN* gene in the order Primates*

Author(s): Yordy Stiven Cangrejo Useda**

Key Words: Parkinson Disease, *PRKN*, Primates, Evolution, Phase Intron, Genetic Therapy

Description: Parkinson's disease is a neurodegenerative disorder that affects the world's population. The number of people affected is expected to increase to 17 million by 2040. Although the physiological characteristics of the disease are not known, a relationship has been established with variants and mutations of the *PRKN* gene. The gene is characterized by an average size of 1 200 000 bp, however, the mRNA has an average size of 4000 bp, it means that only 0.3% of the gene are exons. The difference in size between exons and introns raises several questions about their origin, as well as the possibility of finding mRNAs within them. However, neither the evolutionary status of *PRKN* in primates nor the possible physiological implications are known. For this purpose, the sequences of the primate *PRKN* gene were downloaded from the GenBank, and exons and introns were delimited based on their variant. The protein structure was characterized and the phase of *PRKN* introns was identified. A comparison of possible ORFs and exons within *PRKN* introns was made and an evolutionary relationship was established from a phylogenetic tree. The average structural organization of the gene is composed of 11 to 12 exons, which maintain a conservation pattern towards the 3' end. Likewise, the phases of the introns remain conserved towards this same region, which means that they are ancient introns. The protein structure was shown to have more variation. Possible ORFs within the introns of the sequences were determined, as well as functional sites, which were shown to be conserved. The species *C. jacchus* showed phylogenetic clustering in a different suborder. Although the gene is conserved, different exonic arrangements and possible ORFs promote changes in *PRKN*.

* Bachelor thesis

** Science Faculty. School of Biology. Tutor: Francisco José Martínez Pérez. PhD. in Science. Cotutor: Lina María Vera Cala. PhD. in Epidemiology.

Introducción

La Enfermedad de Parkinson (EP), es un trastorno neurodegenerativo caracterizado por la pérdida de neuronas dopaminérgicas en la *substantia nigra pars compacta* (Vizziello et al., 2021). Los síntomas se evidencian principalmente en la vejez, con ralentización generalizada de los movimientos y síntomas de rigidez o temblor en reposo (Alexoudi et al., 2018). Los diagnósticos médicos para identificar la EP se basan principalmente en la Escala Unificada de Calificación de la Enfermedad de Parkinson de la Sociedad de Trastornos del Movimiento (UPDRS por sus siglas en inglés) (Liu et al., 2023) y en algunos marcadores moleculares en sangre como son: la α -sinucleína, la Intereucína-6 (IL6) y ADN mitocondrial (Borsche et al., 2020; Kluge et al., 2022). Sin embargo, estos elementos se emplean en una etapa avanzada de la enfermedad (Adler et al., 2014).

Aunque no se conocen las razones fisiológicas que promueven su desarrollo, se ha establecido una correlación molecular con el gen *Parkin RBR E3 Ubiquitin Protein Ligase* (*PRKN*), que participa en la homeostasis mitocondrial (Bayne & Trempe, 2019). *PRKN* es el segundo gen más grande del genoma humano, después del gen de la distrofina (Tanaka, 2020), con más de 1 380 000 pares de base (pb). Codifica para una proteína de 465 aminoácidos (Wang et al., 2021), es decir, su marco abierto de lectura (ORF), es de 1398 pb dentro de un ARN mensajero (ARNm) de 4178 pb; por lo tanto, el 0.3% del gen son exones. Diferentes estudios de la organización intrones-exones o mutaciones del gen *PRKN*, lo relacionan con la EP (Ahmad et al., 2023; Jensen et al., 2022; Seike et al., 2021). Pero, el conocimiento acerca de los intrones del

gen *PRKN* es muy limitado por lo que se requieren de estudios de ciencia básica para entender su relación con la EP, la mitocondria y otras enfermedades neurológicas (Leduc-Gaudet et al., 2022).

El gen *PRKN* fue descrito por primera vez en 1997 en el cromosoma 6q25.2-27, a partir de muestras de pacientes con Parkinsonismo juvenil (Matsumine et al., 1997). Posteriormente, se aisló la proteína a partir de muestras de cerebro y músculo esquelético, y se le asignó el nombre de “Parkin” (Kitada et al., 1998). Pertenece a la familia de proteínas RBR (por sus siglas en inglés: *ring between ring family*) debido a la estructura de sus dominios: un dominio similar a la ubiquitina (UBQ) en su extremo N-terminal, y múltiples dominios RING/dedos de Zinc en su extremo C-terminal separados por un dominio IBR (In-Between-RING), los cuales se encuentran presentes en diversos grupos taxonómicos, incluyendo algunos protistas (Marín et al., 2004).

Se estima que el dominio RBR de la familia, se originó una única vez en la historia evolutiva (Marín & Ferrúst, 2002), lo que ha permitido establecer un origen antiguo para los genes que pertenecen a esta familia en todo el grupo de los eucariotas e incluso, en algunos virus (Eisenhaber et al., 2007). Sin embargo, se sabe que cada subfamilia y, de hecho, cada gen, tienen sus propios mecanismos de regulación y activación (Spratt et al., 2014; Walden & Rittinger, 2018), así como participan en distintos procesos fisiológicos, como factores de transcripción de ciertos genes o en la mitofagia, que es el proceso celular en el que las mitocondrias se auto degradan y reciclan (Leduc-Gaudet et al., 2022; X. S. Wang et al., 2023); lo cual es un aspecto para destacar, si consideramos su origen evolutivo. *PRKN*, es un gen conservado presente únicamente en animales (Smit & Sixma, 2014), con ortólogos a nivel de taxones como *Caenorhabditis elegans*, *Drosophila melanogaster*, entre otros, incluido *Homo sapiens* (Von Coelln et al., 2004). Aunque

se ha estudiado el origen evolutivo de la subfamilia Parkin (Marín, 2009), poco se sabe de su estado en el orden de los Primates.

A través de la secuenciación de nueva generación (NGS), se han establecido genomas de diversas especies, entre ellos, los de diversos primates (Jasinska, 2020; L. F. Kuderna et al., 2020). Aunado a lo anterior, las herramientas bioinformáticas y los análisis *in silico* han logrado identificar las características de estos genomas (Uesaka et al., 2022), así como realizar predicciones para detectar cambios en la fisiología y metabolismo (Martin et al., 2021). Nuestro grupo de investigación Cómputo Avanzado a Gran Escala (CAGE), demostró que los intrones del gen *PRKN* humano contienen elementos de la transcripción y de la traducción. Por ejemplo, una repetición en tándem en el intrón 7 o exones cuya traducción genera dominios para la interacción con membranas o la fosforilación. Sin embargo, se desconoce si algunos de estos elementos podrían estar en el gen *PRKN* de otros primates, lo que sugeriría varias hipótesis a nivel evolutivo, su participación en el procesamiento del intrón o su traducción. Por lo tanto, es necesario la caracterización de los intrones del gen *PRKN* en otros primates, para determinar con base a su conservación la posibilidad de ser empleados como marcadores moleculares de la EP u otras, además de emplearse en estudios de ciencia básica.

2. Objetivos

2.1 Objetivo General

- ✓ Establecer la relación evolutiva de las regiones homólogas y variables de los intrones del gen *PRKN* en primates para establecer *in silico* su función celular

2.2 Objetivos Específicos

- ✓ Determinar *in silico* la presencia de secuencias específicas de unión para factores de la transcripción (SEUFT) y posibles exones en los intrones del gen *PRKN* de genomas de primates reportados en la base de datos GENE del NCBI.
- ✓ Identificar la posible función celular de los ARNm generadas por los exones determinados *in silico* dentro de los intrones del gen *PRKN* en primates.
- ✓ Establecer la relación de hermandad del gen *PRKN* en el orden Primates.

3. Competencias

3.1 Competencias cognitivas

- ✓ Genera bases de datos del gen *PRKN* de primates con base a la fidelidad de resultados de bases de datos públicas.
- ✓ Identifica intrones y exones del gen *PRKN* para las diferentes secuencias de primates publicadas en bases de datos públicas.
- ✓ Establece parámetros para realizar alineamientos múltiples entre las secuencias del gen *PRKN* de primates.
- ✓ Utiliza herramientas bioinformáticas para análisis evolutivos y moleculares del gen *PRKN* de primates.
- ✓ Comprende modelos evolutivos propuestos para explicar la historia evolutiva del gen *PRKN* de primates.

3.2 Competencias actitudinales

- ✓ Actualiza aprendizaje para el uso de bases de datos y herramientas bioinformáticas.
- ✓ Analiza resultados bioinformáticos para dar una explicación biológica de forma crítica.
- ✓ Desarrolla habilidades de escritura para redactar informes científicos de manera clara y concisa.
- ✓ Emplea valores éticos y bioéticos para el desarrollo y divulgación del conocimiento científico.

4. Materiales y Métodos

4.1 Obtención y Caracterización de secuencias

Las secuencias de los diferentes taxones de primates que reportaron al gen *PRKN* y sus ARNm (incluyendo sus variantes) hasta el 24 de abril del 2023, se obtuvieron de la base de datos Gene del Centro Nacional de Información Biotecnológica (NCBI) (Geer et al., 2010) por medio de las siguientes palabras claves: “PRKN” y “Primates”. Posteriormente, se seleccionaron únicamente aquellas que no presentaron regiones indeterminadas.

Para cada secuencia del gen, se hizo la delimitación de intrones y exones con un alineamiento entre el gen y los ARNm reportados a partir del programa lamassemble (Frith et al., 2021) a través de su servidor en línea (<https://mafft.cbrc.jp/alignment/server/index-rawreads.html>) con parámetros por defecto, y se corroboró manualmente. Las regiones del gen que presentaron homología con el ARNm se consideraron como exones; las demás regiones se identificaron como intrones. Estas posiciones se compararon con las reportadas por la base de datos Gene.

Así mismo, se hizo una caracterización de los dominios estructurales de la proteína PRKN, a partir de las secuencias de aminoácidos de cada variante, mediante el servidor en línea SMART (Simple Modular Architecture Research Tool) (Letunic et al., 2021).

4.2 Caracterización de la fase de intrones y su estado evolutivo

Con el fin de evidenciar el estado de conservación de los intrones del gen *PRKN* en el orden Primates, se describió la fase de cada intrón. Para ello, se realizó un alineamiento múltiple entre las secuencias de la proteína del gen *PRKN* descrita para primates, mediante el programa ClustalW (Larkin et al., 2007) en el servidor en línea Kalign (Lassmann, 2020), con los parámetros del modelo de pérdida de ADN (Martínez-Pérez et al., 2007). Posteriormente, se ubicó cada

aminoácido en su respectivo codón, dentro de los diferentes ARNm del gen *PRKN* para cada taxón y los criterios de fase se determinaron de acuerdo con lo propuesto por Patthy (1978) (Patthy, 1987).

Los intrones se clasificaron como recientes o antiguos, en base a los criterios propuestos por Martínez *et al.*, (2009) (Martínez-Pérez *et al.*, 2009). Los intrones que compartían una misma posición se consideraron como antiguos y los intrones presentes o ausentes dentro del alineamiento, se consideraron como recientes.

4.3 Caracterización de exones y marcos abiertos de lectura dentro de los intrones del gen *PRKN* de Primates

A partir del programa GENSCAN (Burge & Karlin, 1997) en su servidor en línea (<http://hollywood.mit.edu/GENSCAN.html>), se identificaron posibles Marcos Abiertos de Lectura (ORF) y exones dentro de los intrones del gen *PRKN* en primates. Así mismo, se obtuvieron las secuencias de aminoácidos de los posibles ORF, para las cuales se identificaron los posibles dominios mediante la base de datos PROSITE (Hulo *et al.*, 2006).

4.4 Análisis filogenético

Las diferentes secuencias del gen *PRKN* de primates se alinearon mediante la herramienta lamassemble (Frith *et al.*, 2021) a través de su servidor en línea (<https://mafft.cbrc.jp/alignment/server/index-rawreads.html>), con parámetros por defecto. Se empleó el gen *PRKN* de *Mus musculus* como grupo externo. Los análisis filogenéticos se realizaron a través del portal en línea CIPRES Science Gateway (Miller *et al.*, 2015). Para la reconstrucción filogenética se empleó una estimación acorde a máxima verosimilitud. El análisis de máxima verosimilitud se realizó mediante el programa IQ-Tree V.2.1.2 (Minh *et al.*, 2020), con un soporte de ramas basado en una aproximación de ultra Bootstrap (Hoang *et al.*, 2018) (1000 réplicas) y un

soporte de ramas individuales basado en SH-aLTR (Anisimova et al., 2011). El modelo de sustitución nucleotídica se calculó a partir del programa jModelTest 2 (Darriba et al., 2012). Se seleccionó GTR+G como modelo de sustitución a partir del índice de Akaike (AIC). Los demás parámetros se dejaron por defecto.

5. Resultados

5.1 Organización estructural del gen *PRKN* en el orden Primates

Se obtuvieron un total de 34 secuencias del gen *PRKN*, reportadas para diferentes taxones del orden Primates, las cuales se distribuyeron de la siguiente manera: la mayoría de las secuencias pertenecían al suborden Haplorrhini con 3 especies para la familia Hylobatidae, 6 para la familia Homínidae, 14 para la familia Cercopithecidae, 4 para la familia Cebidae y una especie para las familias Aotidae y Tarsiidae respectivamente. Para el suborden Strepsirrhini, se reportó únicamente una secuencia de el gen *PRKN* para las siguientes familias: Cheirogaleidae (*Microcebus murinus*), Indriidae (*Propithecus coquereli*), Lemuridae (*Lemur catta*), Galagidae (*Otolemur garnettii*) y Lorisidae (*Nycticebus coucang*) (Apéndice A). Del total de secuencias, se seleccionaron 16 (Tabla 1), las cuales no presentaban regiones indeterminadas.

El gen *PRKN* de las especies seleccionadas presentó un tamaño promedio de 1 300 000 pb y se encontró ubicado en los cromosomas 2, 3, 4 o 6 de acuerdo con la especie, excepto por *T. francoisi*, especie para la cual no se ha determinado su ubicación cromosomal (Tabla 1).

Tabla 1*Especies con la secuencia del gen PRKN completo*

Especie	Código de acceso	Cromosoma	Región	pb
<i>Callithrix jacchus</i>	NC_071445	4	163 974 244 ... 165 437 796	1 463 553
<i>Gorilla gorilla gorilla</i>	NC_073230	6	185 491 318 ... 186 870 493	1 379 176
<i>Homo sapiens</i>	NG_008289	6	161 347 417 ... 162 727 766	1 380 350
<i>Lemur catta</i> *	NC_059129	2	141 451 127 ... 142 564 438	1 113 312
<i>Macaca fascicularis</i>	NC_052258	4	158 647 209 ... 160 018 445	1 371 237
<i>Macaca mulatta</i>	NC_041757	4	8 030 146 ... 9 397 207	1 367 062
<i>Macaca thibetana tibetana</i>	NC_065581	4	159 441 106 ... 160 824 171	1 383 066
<i>Microcebus murinus</i> *	NC_033665	6	115 752 192 ... 115 934 925	182 734
<i>Nomascus leucogenys</i> *	NC_044383	3	147 100 804 ... 148 494 662	1 393 859
<i>Pan paniscus</i>	NC_073255	6	179 428 720 ... 180 815 536	1 386 817
<i>Pan troglodytes</i>	NC_072404	6	173 407 906 ... 174 819 435	1 411 530
<i>Pongo abelii</i> *	NC_071991	6	162 671 977 ... 164 054 063	1 382 087
<i>Pongo pygmaeus</i>	NC_072379	6	163 150 454 ... 164 529 929	1 379 476
<i>Rhinopithecus roxellana</i>	NC_044552	4	155 438 516 ... 156 834 937	1 396 422
<i>Symphalangus syndactylus</i>	NC_072424	2	147 448 104 ... 148 845 029	1 396 926
<i>Trachypithecus francoisi</i>	NW_02268146 0	Sin determinar	7 755 891 ... 9 160 787	1 404 897

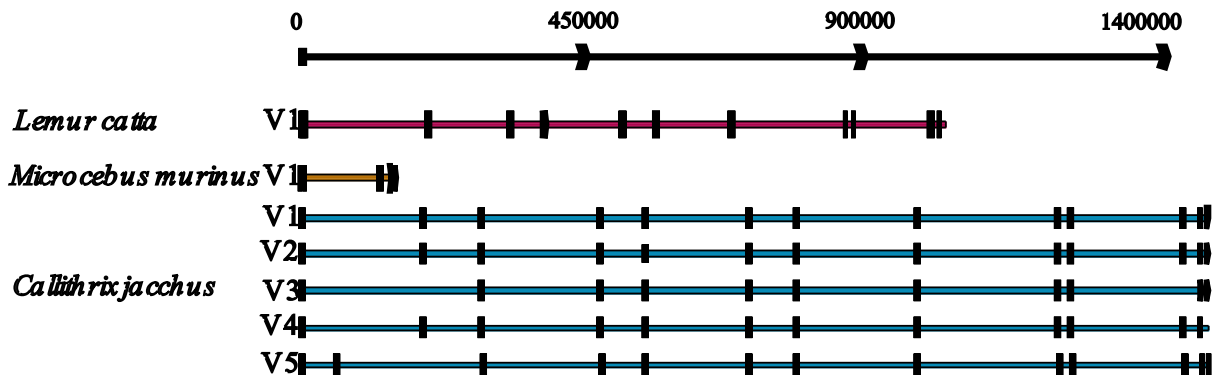
Nota. (*) especies para las cuales se reportó una sola variante.

El gen mostró una organización de intrones y exones relativamente conservada entre las especies seleccionadas del orden Primates (Figura 1-4). Las posiciones de los exones e intrones dentro de las secuencias correspondieron con las reportadas por la base de datos Gene del NCBI para cada variante del gen en los diferentes taxones (Apéndice B)

En promedio, las variantes reportaron entre 10 a 12 exones en total, no obstante, el gen *PRKN* de *M. murinus*, fue el de menor cantidad de exones a comparación los seleccionados con un total de 5 exones y 4 intrones (Figura 1).

Figura 1

Delimitación de exones e intrones del gen *PRKN* en *L. catta*, *M. murinus* y *C. jacchus*



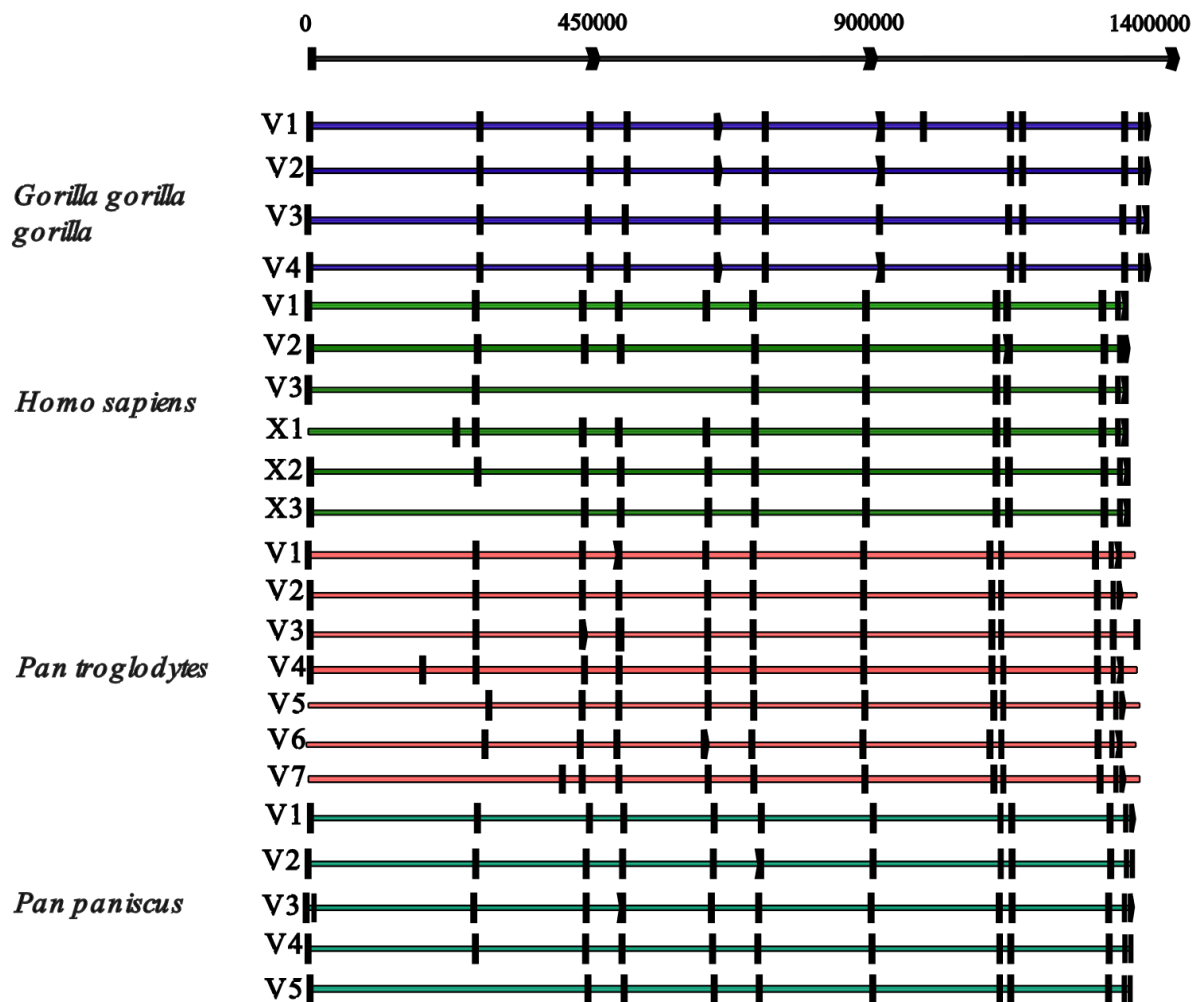
Nota. Organización de exones (barras verticales) e intrones (espacios entre barras verticales) en variantes de ARNm de *Lemur catta*, *Microcebus murinus* y *Callithrix jacchus*.

Las variantes de las especies indicadas en el gráfico mostraron tener una organización de exones e intrones independiente entre sí, como se puede observar en la especie *Callithrix jacchus*, en donde la disposición de exones de las variantes de ARNm cambia la cantidad de intrones de cada gen (Figura 1).

Así mismo, se evidencia que cada variante ARNm puede tener el mismo exón en distintas posiciones del gen, sin cambiar la estructura organizacional. Esto se puede observar, por ejemplo, en las variantes de ARNm de las especies de la subfamilia Hominae, y en general, para las especies que reportan dos o más variantes; donde la posición del exón se puede desplazar unos cientos pares de bases, u ocupar un espacio totalmente distinto, como el exón 1 de la variante 7 (V7), de *Pan troglodytes* (Figura 2).

Figura 2

Delimitación de exones e intrones del gen *PRKN* en la subfamilia Homininae



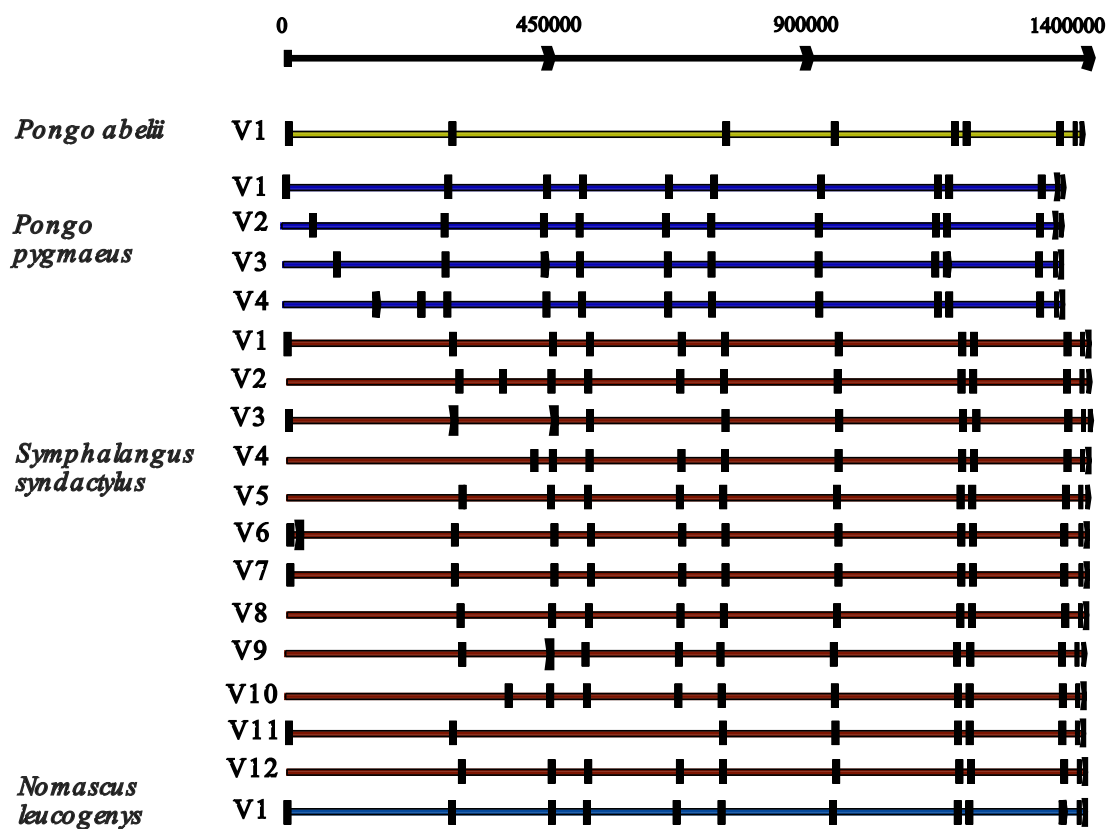
Nota. Organización de exones (barras verticales) e intrones (espacios entre barras verticales) en variantes de ARNm de *Gorilla gorilla gorilla*, *Homo sapiens*, *Pan troglodytes* y *Pan paniscus*.

Los eventos de aparición, delección o desplazamiento de exones se pueden evidenciar con mayor claridad conforme aumenta la cantidad de variantes. Tal es el caso de *Symphalangus syndactylus*, para la cual se reportan la mayor cantidad de variantes ARNm (12 variantes en total). Se puede observar la aparición de exones, como en la variante 6 (V6) (característica presente en

otras especies como: *C. jacchus* (V1, 2, 4 y 5), *G. gorilla gorilla* (V1 y V2), *M. thibetana thibetana* (V4), *P. paniscus* (V3), *P. troglodytes* (V4) y *P. pygmaeus* (V4)), así como una evidente falta de varios exones en la variante 11 (V11), que cambian la estructura del gen y el tamaño del ARNm en una misma especie (Figura 3).

Figura 3

Organización de exones e intrones del gen *PRKN* en las especies *P. abelii*, *P. pygmaeus*, *S. syndactylus* y *N. leucogenys*

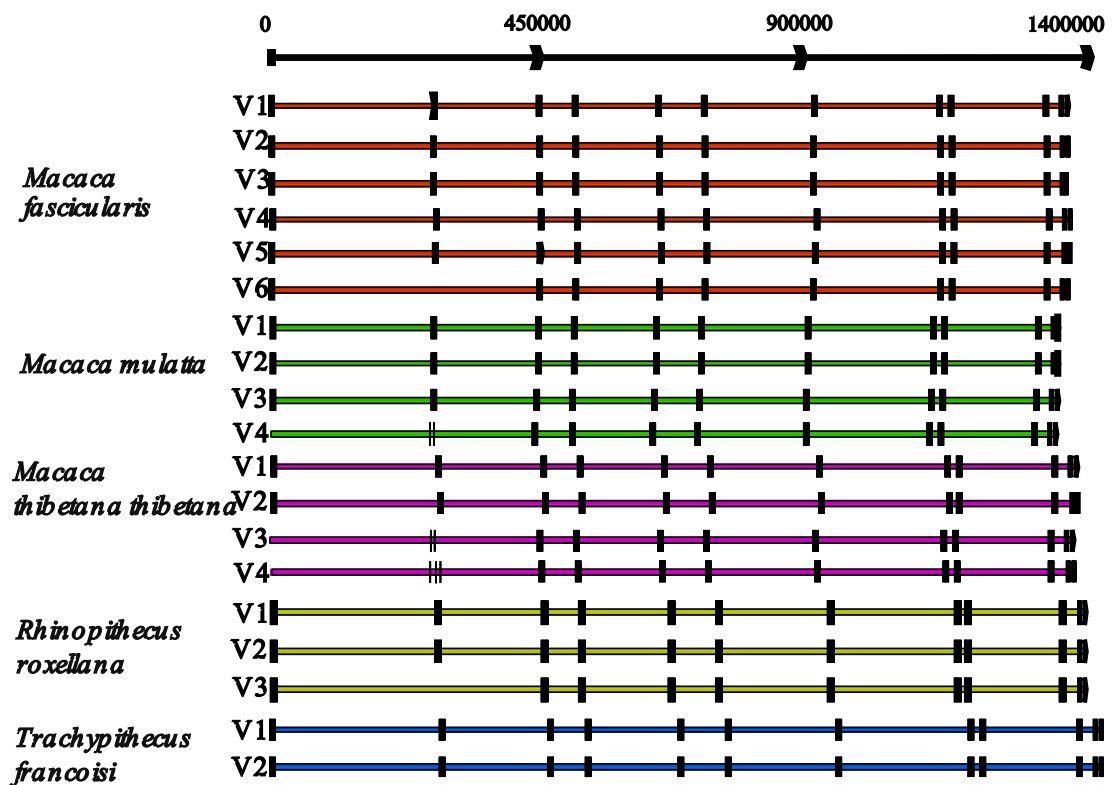


Nota. Organización de exones (barras verticales) e intrones (espacios entre barras verticales) en variantes de ARNm de *Pongo abelii*, *Pongo pygmaeus*, *Symphalangus syndactylus* y *Nomascus leucogenys*.

Es de resaltar que las diferencias en el tamaño del gen no se presentan únicamente entre especies de distintas familias, sino que también se da en especies del mismo género, como sucede con *Macaca* sp. (Figura 4). Sin embargo, a pesar de la diferencia en el tamaño, la organización de exones e intrones se mantiene casi conservada en las variantes reportadas para las especies de este género (Figura 4).

Figura 4

Organización de exones e intrones del gen *PRKN* en *M. fascicularis*, *M. mulatta*, *M. thibetana thibetana*, *R. roxellana* y *T. francoisi*



Nota. Organización de exones (barras verticales) e intrones (espacios entre barras verticales) en variantes de ARNm de *Macaca fascicularis*, *Macaca mulatta*, *Macaca thibetana thibetana*, *Rhinopithecus roxellana* y *Trachypithecus francoisi*.

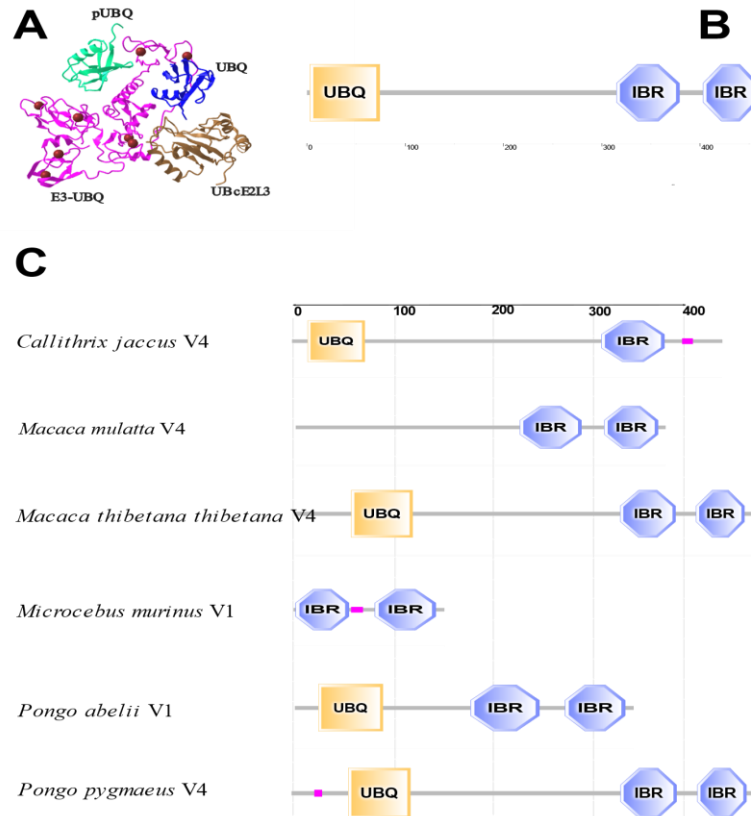
A pesar de las evidentes diferencias que existen entre las secuencias del gen *PRKN* de primates, se puede observar un patrón de conservación en la organización de exones e intrones, hacia las regiones media (entre 500 000 y 800 000 pb) y final (>1 200 000 pb) de la secuencia del gen, así como la presencia de un exón en los nucleótidos de inicio del gen. Esto se explicará con mayor detalle en el apartado de “Fases y estado de conservación de los intrones del gen *PRKN* en el orden Primates” en la sesión de Resultados.

Por otro lado, las secuencias de ARNm de cada variante corresponden con lo obtenido en la organización de exones e intrones del gen *PRKN*, es decir, a nivel de ARNm las variantes presentan diferencias en el tamaño y en la composición nucleotídica, lo cual es congruente con las estructuras del gen en los diferentes taxones del orden Primates. Sin embargo, en los ARNm también se pueden apreciar regiones conservadas entre las diferentes variantes (Apéndice C), lo cual refleja el patrón de conservación previamente descrito.

Aunado a lo anterior, la estructura de dominios codificados por dichos ARNm se muestra igualmente conservada entre las diferentes variantes, las cuales, en su mayoría, presentan un dominio de ubiquitina (UBQ) en su extremo N-terminal, junto con dos dominios IBR (*In-Between-Ring Fingers*) hacia el extremo C-terminal (Figura 5 A-B). Sin embargo, se pueden encontrar algunas modificaciones estructurales: la ausencia del dominio de ubiquitina, como se puede evidenciar en la V4 de *M. mulatta* y la V1 de *M. murinus* (Figura 5C). Así mismo, se pudo observar la ausencia de algún dominio IBR, como en la V4 de *C. jacchus*.

Como se mencionó anteriormente, la disposición de exones cambia el tamaño y la composición del ARNm, lo cual puede modificar la secuencia final de aminoácidos. Esto se puede observar en variantes de especies como *M. thibetana thibetana* o *P. pygmaeus*, cuyas variantes 4, respectivamente, mostraron tener mayor cantidad de pb en su ARNm (Apéndice B). Al hacer la

caracterización de dominios estructurales de estas variantes, se pudo observar un desplazamiento del sitio de ubiquitina con respecto a otras especies que presentaron este dominio (aproximadamente 50 aminoácidos, Figura 5C). Sin embargo, para otras especies el tamaño del ARNm no afectó la estructura de la proteína. Esto se puede corroborar con la V1 de *P. abelii*, ya que, aunque su ARNm tiene un tamaño promedio al de otras secuencias de *PRKN*, la secuencia de aminoácidos es más corta en comparación con otras especies (un total de 316 aminoácidos, en comparación con el promedio de 465, Figura 5C). Contrario a lo esperado, la estructura de la proteína no se ve afectada por el tamaño de su secuencia de aminoácidos, puesto que presenta los 3 dominios estructurales característicos de *PRKN*. Si bien, para esta especie la cantidad de aminoácidos no modificó la estructura de la proteína, no sucedió lo mismo con *M. murinus*, ya que el tener una secuencia del gen tan corta, la cantidad de aminoácidos cambió considerablemente en comparación con las demás especies (139 aminoácidos). A pesar de esto, se logran obtener los dos dominios IBR conservados en la mayoría de las variantes.

Figura 5*Dominios estructurales de Parkin*

Nota. Caracterización estructural de Parkin. (A) Complejo UbcH7-U con R0RBR Parkin y fosfoubiquitina. E3-UBQ: E3 Ubiquitina-proteína ligasa Parkin; UBcE2L3: Enzima conjugadora de ubiquitina E2 L3; UBQ: Ubiquitina; pUBQ: fosfoubiquitina. B-C: Dominios de Parkin obtenidos en SMART. UBQ: Ubiquitina; IBR: Dominio “*In Between Ring fingers*”. Los cuadros fucsias dentro de la secuencia indican dominios no reconocidos.

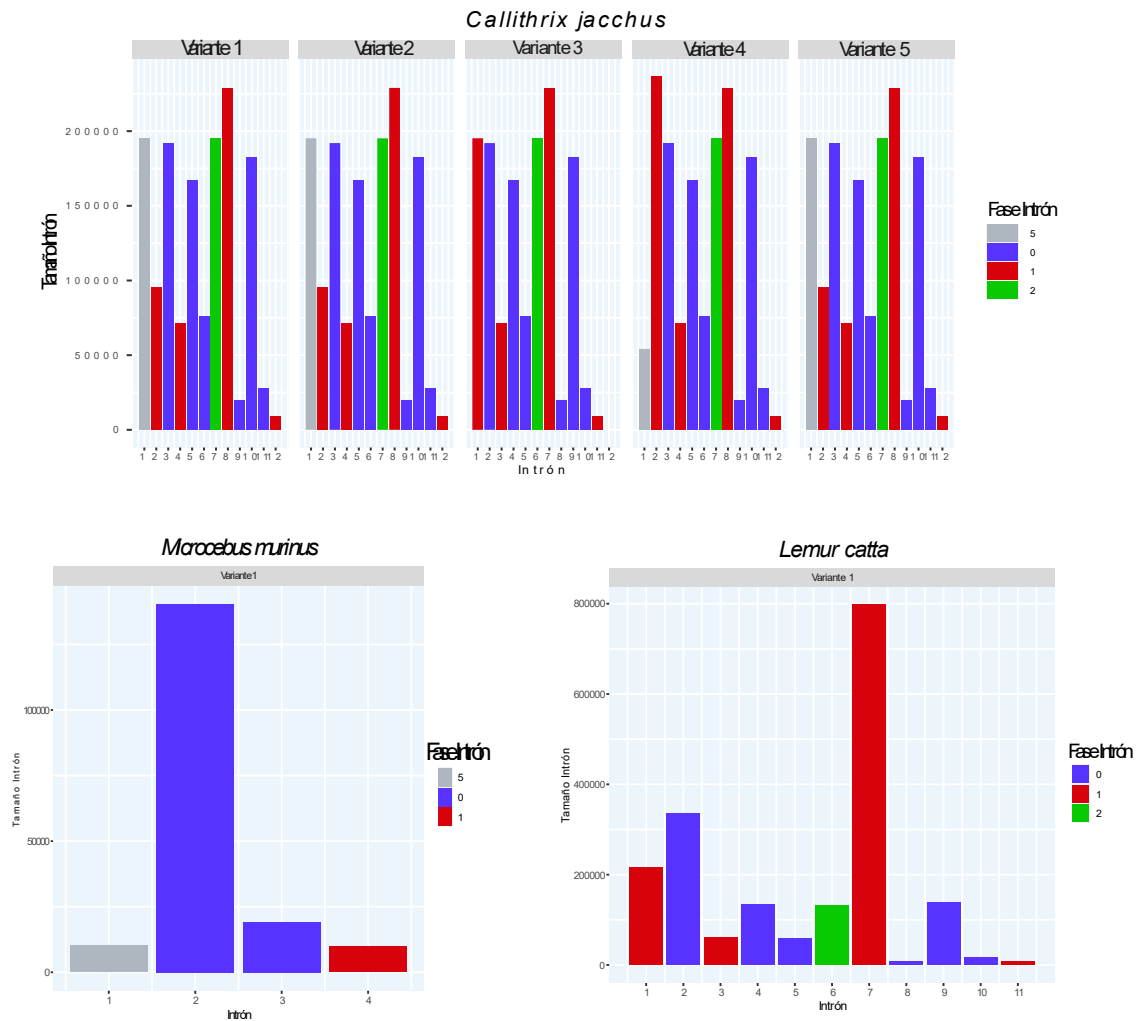
5.2 Fases y estado de conservación de los intrones del gen *PRKN* en el orden Primates

Como se mencionó en el apartado anterior, a pesar de las diferencias entre el tamaño, cantidad de exones e intrones, y organización del gen *PRKN* en el orden de los Primates, se puede observar un patrón de conservación. La caracterización de las fases de los intrones del gen *PRKN* nos permite ver dicho patrón, por ejemplo, un intrón de fase 2 presente en la región media-final

del gen en todas las variantes de todas las especies (la posición cambia de acuerdo con el número de exones), con excepción de *M. murinus* (Figura 6).

Figura 6

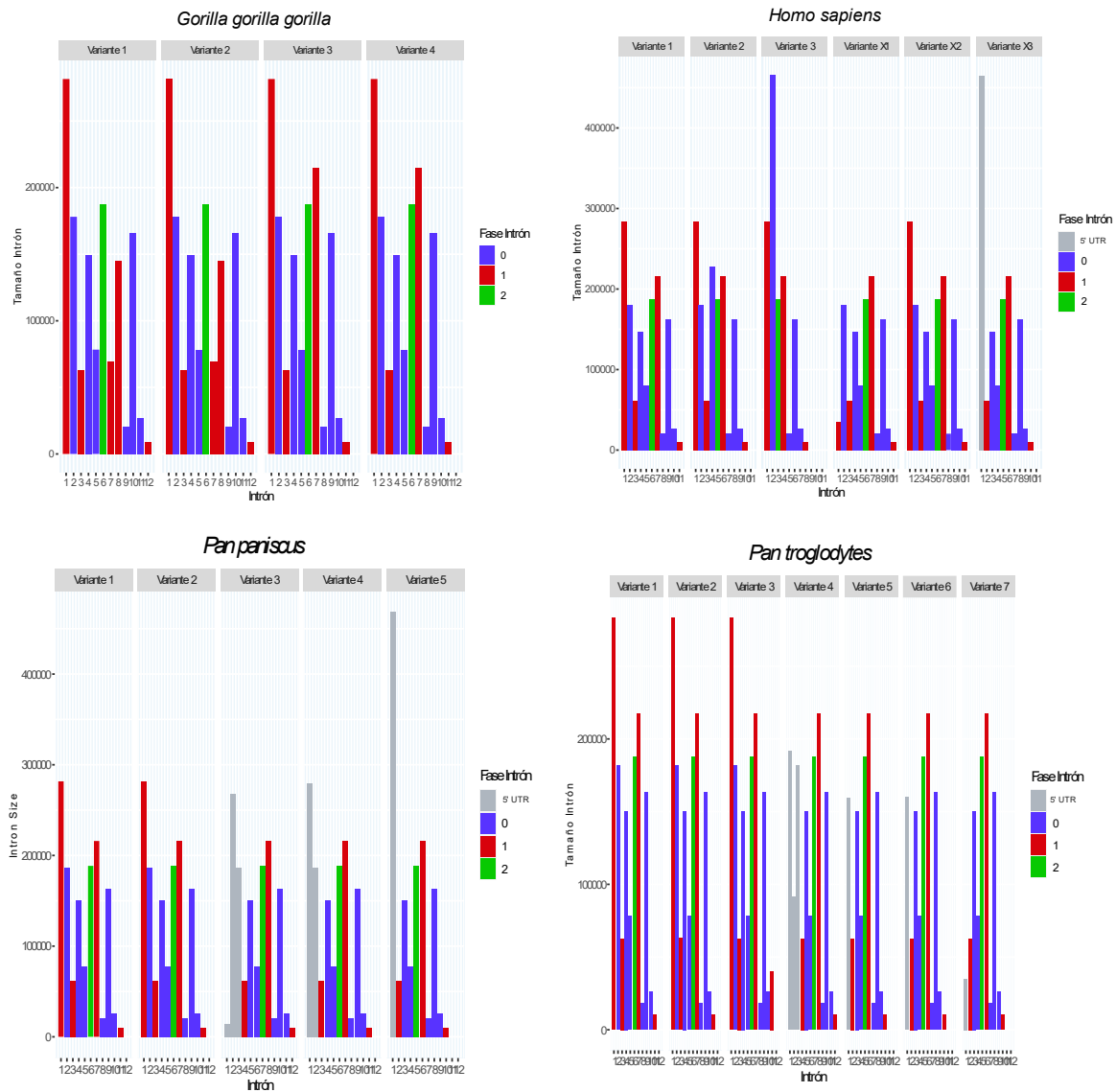
Caracterización de las fases de intrones de C. jacchus, M. murinus y L. catta



Nota. Descripción del tamaño y fase de los intrones delimitados por las variantes de ARNm en *PRKN* para *Callithrix jacchus*, *Microcebus murinus* y *Lemur catta*. Morado: fase 0; Rojo: fase 1; Verde: fase 2; Gris: región 5' no transcrita.

Así mismo, se puede observar que este intrón de fase 2 suele estar acompañado por 1 o 2 intrones de fase 0 en su extremo 5', y un intrón fase 1 hacia el extremo 3' (Figura 6-9). La presencia

de uno o dos intrones de fase 0 en el extremo 5' dependerá de la organización exón/intrón de la variante. Por ejemplo, en las variantes 3 y 5 de *H. sapiens*, se puede observar la presencia de un único intrón fase 0 en el extremo 5' (Figura 7), lo cual, si se compara con la organización de exones e intrones (Figura 2), se identifica la ausencia de lo que sería el intrón 5 (si se toma como referencia la V1) en dichas variantes. Esta característica se puede corroborar al compararla con la variante X3. A pesar de que esta variante carece de la delimitación de intrones en el extremo 5' que evidencian las demás variantes (es decir, la presencia de lo que sería el exón 2), conserva la estructura cercana del intrón de fase 2 y, por lo tanto, conserva los dos intrones fase 0 hacia el extremo 5' (Figura 7). La presencia de un único intrón fase 0 hacia el extremo 5' del intrón fase 2, se observa en otras especies como *P. abelii* y *S. symphalangus* (variante 1 y 11 respectivamente, Figura 8).

Figura 7*Caracterización de las fases de intrones en la subfamilia Homininae*

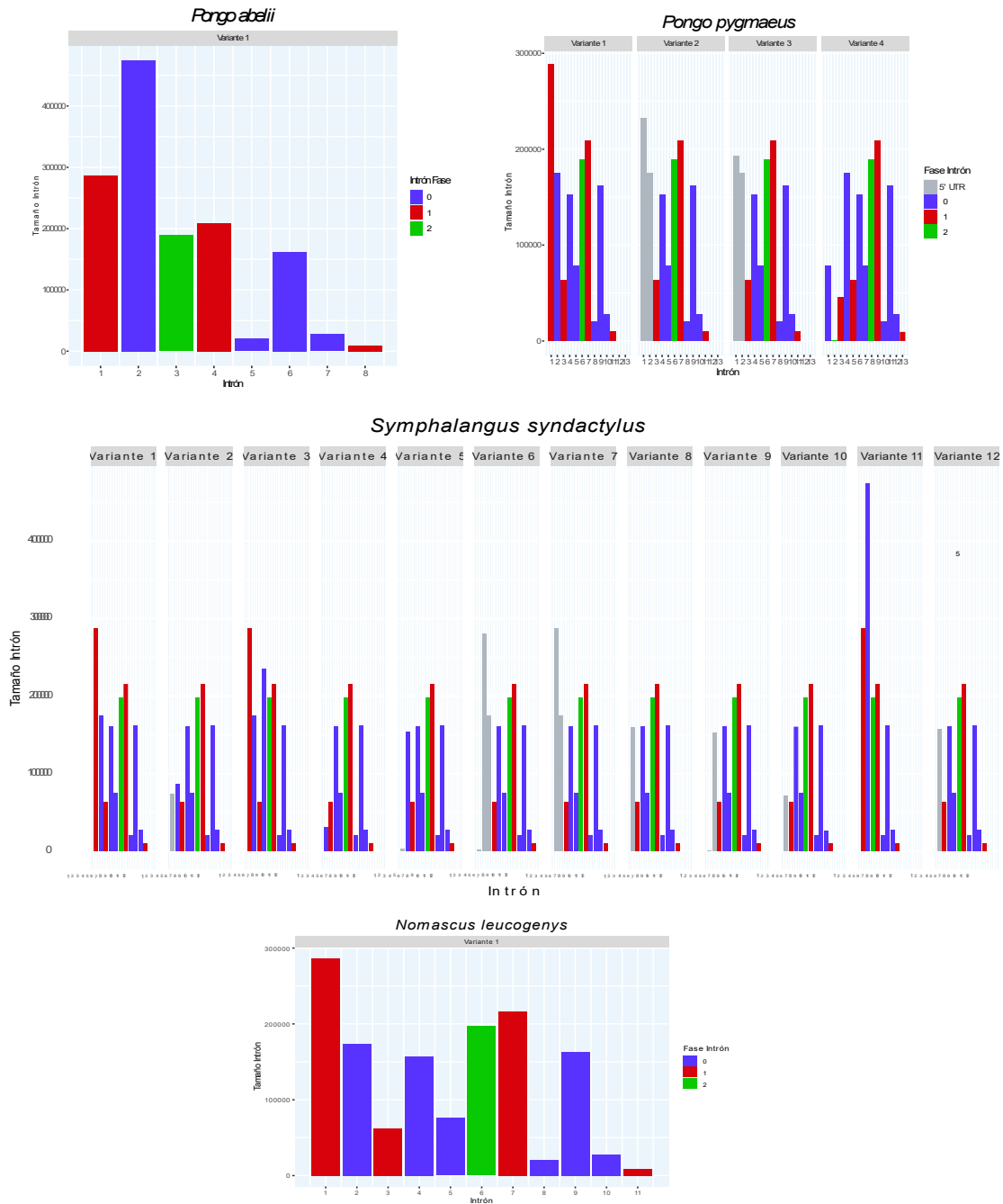
Nota. Descripción del tamaño y fase de los intrones delimitados por las variantes de ARNm en *PRKN* para *Gorilla gorilla gorilla*, *Homo sapiens*, *Pan paniscus* y *Pan troglodytes*. Morado: fase 0; Rojo: fase 1; Verde: fase 2; Gris: región 5' no transcrita.

Otra característica que se muestra constante tanto en la organización de exones e intrones, como en la caracterización de la fase de intrón, es la presencia de un exón en la posición inicial

del gen (entre 1 y 100 pb), así como un intrón de fase 0 en el extremo 5' de las secuencias. Cabe resaltar que, para algunas variantes, la fase del primer y segundo intrón se representan con color gris (en la leyenda figura 5' UTR). Esto es debido a que presentan una región 5' no traducida (5' UTR). A pesar de esto, se puede mantiene el intrón de fase 1, en la posición inmediata a esta región. Ejemplo de esto se puede ver en las en las especies *M. fascicularis* (V6), *M. mulatta* (V4), *M. thibetana thibetana* (V3) o *R. roxellana* (V3) (Figura 9).

Figura 8

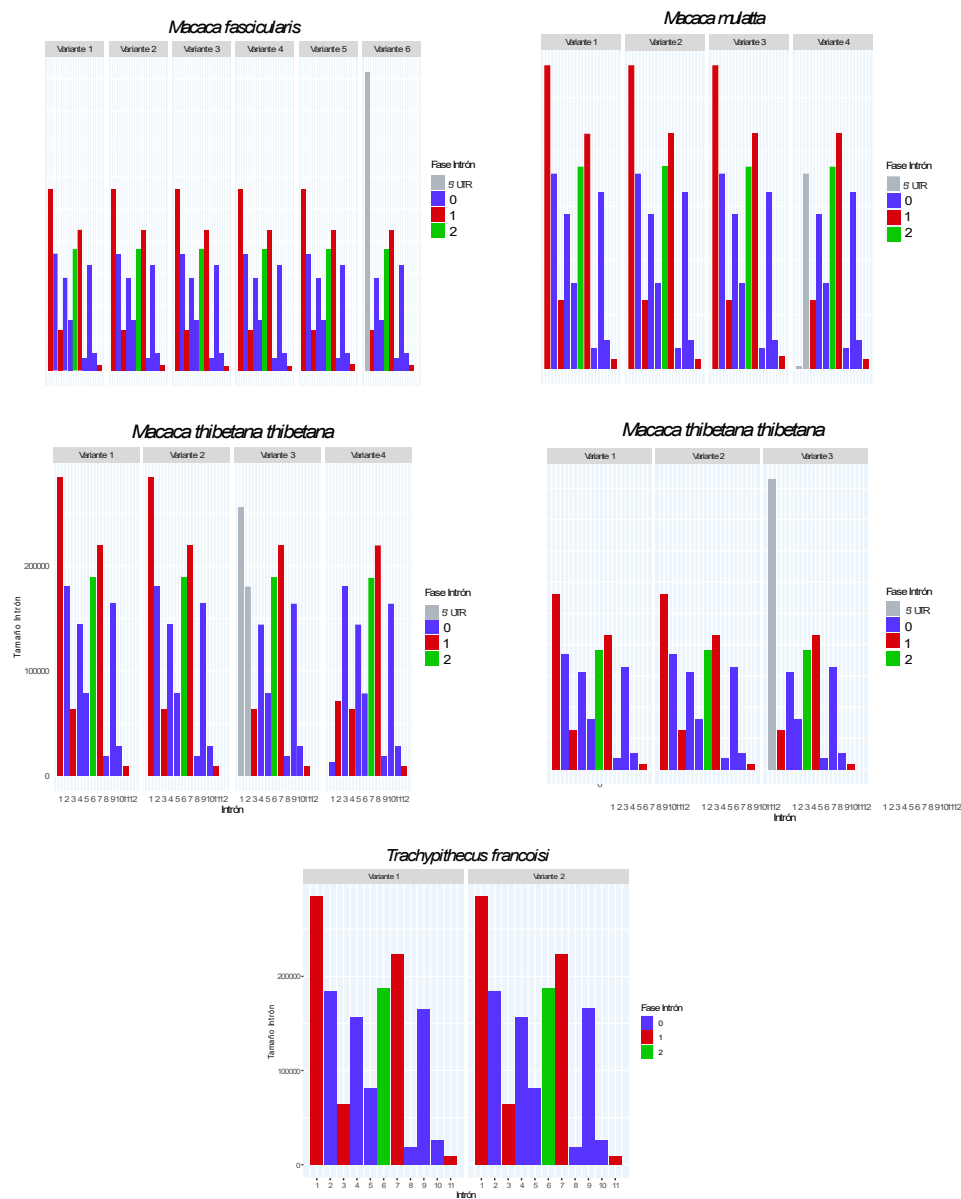
Caracterización de las fases de intrones de *P. abelii*, *P. pygmaeus*, *S. syndactylus* y *N. leucogenys*



Nota. Descripción del tamaño y fase de los intrones delimitados por las variantes de ARNm en *PRKN* para *Pongo abelii*, *Pongo pygmaeus*, *Symphalangus syndactylus* y *Nomascus leucogenys*. Morado: fase 0; Rojo: fase 1; Verde: fase 2; Gris: región 5' no transcrita.

Figura 9

Caracterización de la fase de intrones de M. fascicularis, M. mulatta, M. thibetana thibetana, R. roxellana y T. francoisi



Nota. Descripción del tamaño y fase de los intrones delimitados por las variantes de ARNm en *PRKN* para *Macaca fascicularis*, *Macaca mulatta*, *Macaca thibetana thibetana*, *Rhinopithecus*

roxellana y *Trachypithecus francoisi*. Morado: fase 0; Rojo: fase 1; Verde: fase 2; Gris: región 5' no transcrita.

5.3 Exones y Marcos Abiertos de Lectura dentro de los intrones del gen *PRKN* en el orden Primates

Mediante el programa GENSCAN se lograron obtener Marcos Abiertos de Lectura dentro de los intrones del gen *PRKN* en el orden Primates, tanto en sentido positivo (5' – 3'), como en la cadena contraria (3' – 5'). En total, se obtuvieron 1777 ORFs en sentido positivo, los cuales en su mayoría presentaron una estructura estándar de un marco abierto de lectura: una región promotora, un exón de inicio y final, exones internos y cola de poliadenilación (Apéndice D). En menor proporción se encontraron ORFs de un solo exón. Así mismo, se obtuvieron las secuencias de aminoácidos para cada ORF hallado.

A partir de la base de datos PROSITE, se identificaron los posibles dominios estructurales, sitios funcionales y familias de proteínas en la mayoría de ORFs predichos en GENSCAN (Apéndice E), excepto por algunas secuencias de algunos intrones ubicados en la región final de las secuencias del gen, entre los intrones 8 y 10 de algunas variantes de ARNm (Apéndice D y E). Los sitios funcionales que se encontraron con mayor frecuencia en los intrones del gen *PRKN*, se enlistan en la Tabla 2.

Tabla 2

Sitios funcionales presentes en los intrones del gen PRKN en primates

Sitio funcional	Actividad fisiológica	Referencia
Fosforilación de la proteína quinasa C	Fosforilación de residuos de serina en histonas H1 y H2B.	(Yamamoto et al., 1978)
Fosforilación de la proteína cinasa dependiente de AMPc y GMPc	Fosforilación de proteínas relacionadas en vasopresión.	(Waldmann et al., 1987)
N-miristoilación	Acilación de proteínas.	(Towler et al., 2003)

Fosforilación de la caseína quinasa C	Proliferación y diferenciación celular, y transducción hormonal.	(Pinna, 1990)
N-glicosilación	Plegamiento de proteínas, tráfico y transducción de señales.	(Hirata & Kizuka, 2021)
Amidación	Construcción de proteínas.	(Barak & Batra, 2021)

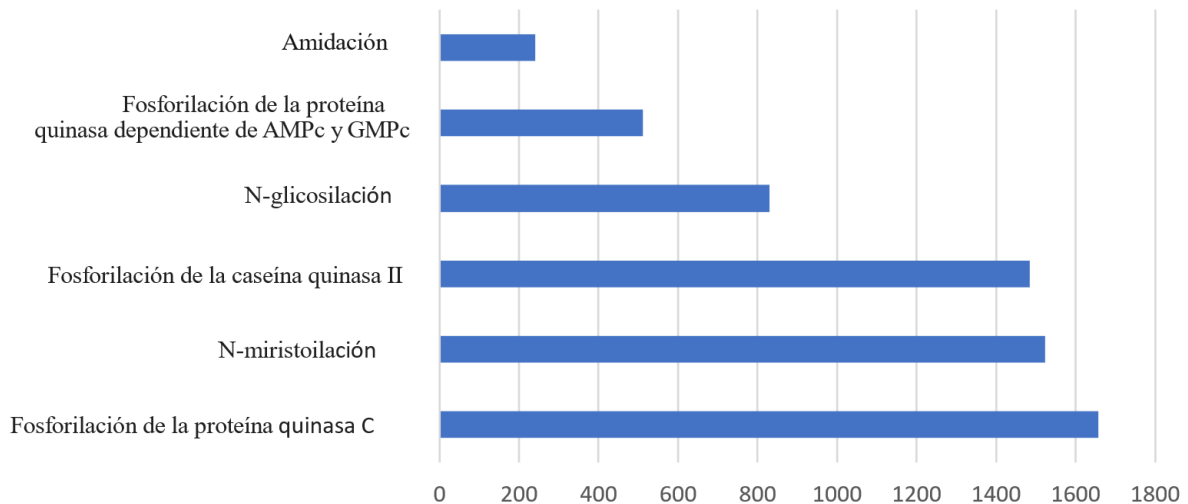
Nota. Se mencionan los posibles sitios funcionales determinados a partir de los posibles ORFs predichos por GENSCAN en los intrones del gen *PRKN*.

En promedio, el sitio de fosforilación de la proteína quinasa C fue el más frecuente entre las secuencias predichas por GENSCAN, encontrándose en 1658 ORFs (Figura 6). *M. murinus* fue la única especie para la cual no se encontró un sitio de amidación entre los ORFs de sus intrones. Cabe resaltar que las cadenas de aminoácidos obtenidas para cada ORF en sentido positivo podían presentar una o varias de las estructuras mencionadas anteriormente.

Así mismo, se identificaron dominios específicos para ciertas especies: Perfil de los lugares de fijación de las lipoproteínas de membrana procariotas, en *C. jacchus*; Perfil bipartito de la señal de localización nuclear y Sitio de hidroxilación del ácido aspártico y la asparagina, en *N. leucogenys*; Perfil del dominio similar a Ig, en *R. roxellana*; y señal de la familia MIP, en *T. francoisi* (Apéndice E).

Figura 10

Sitios funcionales identificados en los intrones del gen PRKN en el orden Primates



Nota. Las cantidades mostradas en la figura 10 hacen referencia a la cantidad de ORFs predichos que podrían presentar estos sitios, más no a la cantidad de sitios hallados, ya que en un ORF se encontraron varias posiciones que pudieran expresar estas estructuras.

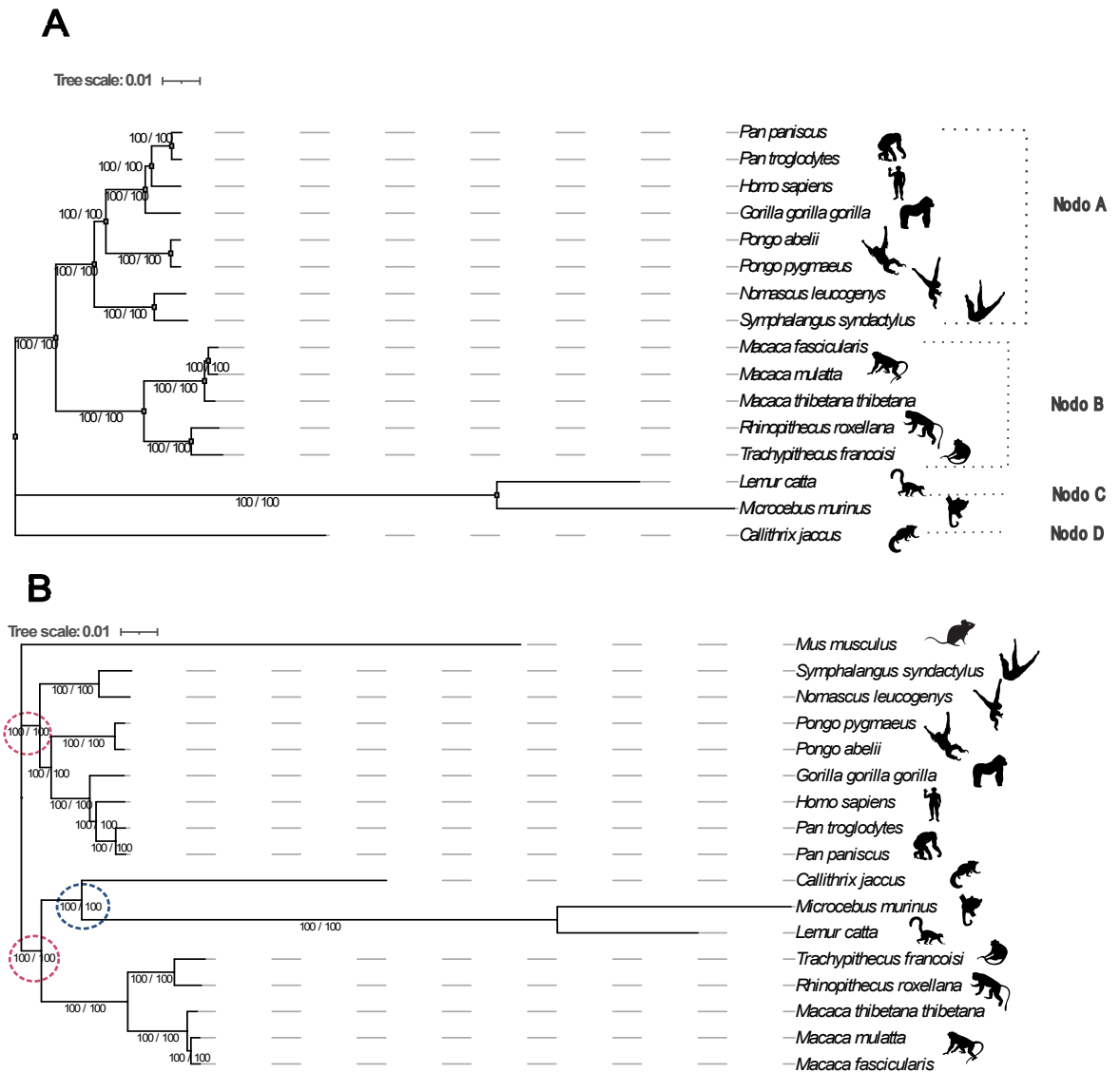
5.4 Análisis filogenético del gen *PRKN* en el orden Primates

Los análisis filogenéticos basados en máxima verosimilitud revelaron relaciones evolutivas y de hermandad para el gen *PRKN* en el orden primates, con un soporte de Bootstrap del 100% para cada rama (Figura 11A-B). Se pueden observar 4 grandes divisiones de hermandad: las especies *P. paniscus*, *P. troglodytes*, *G. gorilla gorilla* y *H. sapiens*, conforman un grupo hermano. El siguiente grupo está conformado por las especies *P. abelii* y *P. pygmaeus*, las cuales mantienen una hermandad con el primer grupo. Como grupo hermano de estos dos clados, se encuentran las especies *N. leucogenys* y *S. syndactylus*, relacionadas entre sí. Las especies nombradas conforman el primer nodo de hermandad (Nodo A).

El siguiente grupo, el cual se relaciona con el nodo previamente descrito está conformado por las especies del género *Macaca* (las cuales conforman un clado), junto con *R. roxellana* y *T.*

francoisi (relacionadas entre sí). Este sería el segundo nodo (Nodo B). El tercer nodo (Nodo C), comprende las especies *M. murinus* y *L. catta*, y finalmente, la especie *C. jacchus* la cual se observa como una rama independiente (Nodo D). A pesar de que los nodos internos están bien resueltos, el árbol presenta una politomía en los nodos más basales.

Al enraizar el árbol con *Mus musculus*, se resolvió la politomía presentada anteriormente. Se logró establecer una relación de hermandad entre *C. jacchus* con el clado *M. murinus-L. catta*. Este grupo a su vez se relacionó con el que sería el nodo B, descrito anteriormente. Finalmente, se recuperan dos grandes clados para el gen *PRKN* en el orden Primates.

Figura 11Relación filogenética del gen *PRKN* en el orden Primates

Nota. (A) árbol filogenético obtenido a partir de una aproximación de máxima verosimilitud, sin enraizar. (B) árbol enraizado con *Mus musculus* como grupo externo. Los círculos rojos denotan la politomía resuelta. El círculo azul indica la rama resuelta de *C. jacchus*. Las imágenes se obtuvieron de Phylopic (<https://www.phylopic.org/>).

6. Discusión

Con la implementación de herramientas de secuenciación masiva, se pudieron llevar a cabo proyectos que permitieron obtener el genoma completo de diferentes especies, por ejemplo, el Proyecto del Genoma Humano (Craig Venter et al., 2001) el cual cambió la manera de estudiar los genes en el siglo pasado y abrió una nueva ventana de información: el enfoque terapéutico se volvió un punto fuerte en la secuenciación de genomas completos, lo que permitió tener un mayor entendimiento a cerca de la fisiología de los genes en relación con diversas patologías (Prokop et al., 2018).

Hasta la fecha, se han podido caracterizar distintos genomas, como el de *Saccharomyces cerevisiae* (Goffeau et al., 1996), *Caenorhabditis elegans* (The C. elegans Sequencing Consortium, 1998), *Drosophila melanogaster* (Adams et al., 2000), así como en organismos cercanos a *H. sapiens*, por ejemplo, *Pan troglodytes* (Mikkelsen et al., 2005). Todo esto se ha logrado gracias al mejoramiento en las tecnologías para procesar datos masivos (Burian et al., 2021), así como la implementación de análisis *in silico*, que, a pesar de requerir su validación experimental, han permitido procesar de manera eficiente la información producto de secuenciaciones (Katsonis et al., 2022).

Esto último se pudo observar para el gen *PRKN* de los diferentes taxones del orden Primates que se evaluaron en este proyecto, en donde las variantes V1, V2 y V3 del gen *PRKN* de *H. sapiens*, son las únicas que tienen estado de “Revisado” en la Base de Datos GenBank (códigos de acceso del GenBank: NM_004562, NM_013987, NM_013988); mientras que las demás las variantes son predicciones que se obtuvieron a partir del programa Gnomon (Thibaud-Nissen et al., 2013).

Si bien, en las anotaciones *in silico* se pueden cometer errores de identificación o delimitación de variantes (Kress et al., 2023; Vihinen, 2023), no dejan de ser una herramienta útil en los análisis genéticos (Abugessaisa et al., 2017; Butkiewicz & Bush, 2016) así como en otras áreas de la ciencia (Diakou et al., 2022). Tal es el caso para el gen *PRKN* de humano, el cual ha sido utilizado en diversos análisis bioinformáticos e *in silico*, con el fin de realizar predicciones, encontrar variantes o establecer alternativas terapéuticas relacionadas con la Enfermedad de Parkinson (Augusto et al., 2020; Chaudhry et al., 2022; Conceição et al., 2017; Kondeva-Burdina et al., 2022).

Con lo anterior, se podría establecer la posibilidad de que las variantes que se han reportado en el GenBank correspondan a eventos de empalme alternativo, los cuales tengan implicaciones en cambios fisiológicos de los organismos. Es sería un reflejo de la versatilidad que tiene este gen en los diferentes organismos que lo presentan (Walden & Rittinger, 2018).

Con base a lo anterior, en esta pasantía se buscó indagar en el estado evolutivo del gen *PRKN* en el orden de los Primates, para poder establecer relaciones estructurales entre las diferentes especies, lo cual pueda reforzar el conocimiento actual en ciencia básica y a su vez, sea utilizado en estudios fisiológicos que permitan comprender mejor el desarrollo de la Enfermedad de Parkinson.

Al observar la organización del gen *PRKN* en los diferentes taxones del orden Primates, es evidente que existe un patrón de conservación el cual se sustenta en diferentes niveles. Primero, independientemente de la diferencia entre el tamaño de los genes de las especies, la ubicación de los distintos exones en base a los ARNm estimados, mantienen una posición similar.

Si comparamos la estructura de exones e intrones de las secuencias del gen con la fase de los intrones, podremos observar este mismo patrón de conservación. Antes se ha de definir la fase

de un intrón. De acuerdo a lo propuesto por Patthy en 1987 (Patthy, 1987), y en base a la teoría del “barajado de exones”, los intrones presentan 3 fases: la fase 0, en la cual un intrón se encuentra en medio de dos codones, sin interrumpir su transcripción; para el caso de las fases 1 y 2, el intrón se va a encontrar en medio de un codón pero en distintas posiciones: el intrón fase 1 estará entre el primer y segundo nucleótido de un codón, y el intrón fase 2 estará entre el segundo y tercer nucleótido de un codón (Patthy, 1987).

Con base a esto, existe un intrón conservado entre las especies trabajadas en esta pasantía, que corresponde al intrón fase 2 ubicado en la región media del gen. Al observar la delimitación de intrones y exones, se observa que, en esta misma región, todas las variantes de ARNm reportan un exón, el cual delimita el intrón de fase 2 en la región media de la secuencia del gen, a excepción de *M. murinus*. Según Martínez *et al.*, (2009), los intrones que se mantienen constantes en todas las variantes y especies se consideran como intrones antiguos (Martínez-Pérez *et al.*, 2009). Por lo tanto, el intrón de fase 2 caracterizado en este trabajo, se identifica como un intrón antiguo. Ahora bien, si se toma este intrón como punto de referencia, se puede observar que el patrón de conservación a nivel de las fases de los intrones, también se presenta hacia la región 5' y 3'.

Al observar la región 5', se puede notar que el intrón fase 2 suele estar acompañado por uno o dos intrones fase 1, y entre tres y dos intrones fase 0. Si se observa nuevamente la delimitación de exones e intrones, se podrá notar que, hacia esta región, las variantes reportan una mayor variación en cuanto a la posición de los exones, lo cual está a disposición del aumento o disminución en el número de exones. Sin embargo, es de notar que el primer intrón de todas las variantes, con excepción de *M. murinus*, es de fase 1. Es de resaltar que para algunas variantes se indica una región 5' no transcrita, sin embargo, cabe recordar que la definición de la fase está dada por la relación entre el intrón con los codones del ARNm, por lo tanto, estos intrones realmente no

están dentro de la delimitación de fases. Con base a esto, el primer intrón de todas las variantes también se considera como antiguo. Nótese que a diferencia del intrón de fase 2, para este intrón la posición de los exones cambia mucho más entre variantes, sin embargo, se observa el patrón de conservación.

Para el caso de los demás intrones de esta región, debido a que no presentan un patrón fijo ya que la posición de los exones afecta su presencia o usencia, se clasifican como intrones recientes. Finalmente, para los intrones hacia la región 3', se puede observar que tanto las fases como la estructura del gen se notan conservadas, por lo tanto, los intrones de esta región son antiguos. Si se tienen estas características en cuenta, se nota otro resultado sobresaliente del gen: la región más conservada en base a la posición de exones se halla hacia el extremo 3', mientras que la cantidad y posición de exones en el extremo 5' es más variante. Esto podría indicar un posible crecimiento del gen desde la región 3' hacia la región 5'.

Por otro lado, al caracterizar e identificar estructuras de proteínas, también se puede evidenciar el patrón de conservación, con algunas diferencias. Por ejemplo, la estructura de la proteína Parkin codificada por cada ARNm de cada especie, se encuentra conformada por el dominio de ubiquitina y dos dominios IBR en su mayoría. En los casos anteriores se había observado, que, si bien la posición y organización de exones cambiaba, las fases de los intrones, y por ende de los exones, se mantenía constante. Al comparar estos cambios con la proteína, podemos notar que esta sí se ve afectada por la organización exones-intrones. Por ejemplo, algunas variantes reportan la falta de uno de los dominios IBR, las cuales, al revisar la estructura del gen, se puede observar que los exones de la región final (>1 200 000 pb), se encuentran desplazados con respecto a otras variantes. De igual forma, al observar la región 5', en los primeros exones, se puede observar que hay variantes las cuales pierden el sitio de ubiquitina, o, este se desplaza

alrededor de 50 aminoácidos. A partir de este resultado, se debe resaltar el hecho de que, si bien la organización y el número de exones e intrones se mantiene conservado, la estructura del gen varía, lo cual se puede observar finalmente en la traducción a proteína. Si se tiene esto en cuenta, se podría establecer una relación con las distintas funciones que cumple este gen en los diferentes tejidos y órganos, así como su estrecha relación con la aparición de la EP (Barazzuol et al., 2020; Barodia et al., 2017; Kamienieva et al., 2023; Pinto et al., 2018).

Lo anterior también se puede observar con las predicciones obtenidas en PROSITE. Se evidencia que dentro de los intrones de las secuencias del gen *PRKN* de primates, pueden encontrarse distintos sitios funcionales, relacionados con varios procesos de regulación, estructura, transcripción, entre otros, que además se comparten entre especies. Es de resaltar que los sitios aquí mencionados no representan el número total hallado para todos los intrones, si no que hacen referencia a la cantidad de ORFs que pueden presentar estos sitios, ya que en los ORFs se pudieron encontrar varias regiones similares a los distintos sitios, lo cual estaba en función del tamaño de la secuencia.

Si se tiene en cuenta la función de Parkin, la caracterización de estos dominios, sitios funcionales y familias de proteínas, refuerzan la idea de que, de acuerdo con cada individuo, la expresión del gen puede cambiar, independientemente de si este se encuentra conservado incluso a nivel de orden, y, por lo tanto, las distintas combinaciones de exones que se puedan hallar en un único organismo están en función del tamaño del gen. Si se pudiera llevar este análisis a una fase experimental de laboratorio, muy probablemente se puedan validar los resultados obtenidos en este trabajo. Sin embargo, cabe recordar que las características del gen y su tamaño son variables que dificultan su validación en un laboratorio.

En el análisis filogenético se puede observar que la politomía del árbol A se resuelve a partir del grupo externo. Los grupos hermanos obtenidos se pueden comparar con la literatura: en el árbol sin enraizar se pueden distinguir 3 grandes ramas de las cuales, el nodo A y el nodo B, corresponden a un mismo suborden (Haplorrhini). Esto mismo se observa en el árbol enraizado. Es de resaltar que la subfamilia Homininae se pudo recuperar por completo en ambos árboles. En esta subfamilia se encuentra ubicada nuestra especie, *Homo sapiens*, la cual muestra una relación de hermandad más cercana al género *Pan*, resultado que concuerda con lo reportado por la literatura (Duda & Zrzavý, 2013) y una agrupación con el género *Gorilla*. Sin embargo, la especie *C. jacchus*, la cual hace parte del suborden Haplorrhini (Chen et al., 2023), presenta una agrupación no acorde a lo que se conoce actualmente de la taxonomía de primates, ya que esta especie se agrupa con *L. catta* y *M. murinus*, las cuales hacen parte del suborden Strepsirrhini (L. F. K. Kuderna et al., 2023).

Sin embargo, se puede notar la división en dos grandes ramas las cuales corresponderían al suborden Strepsirrhini y Haplorrhini, aunque es necesario resolver el estado de *C. jacchus*. Esto puede estar dado debido a que no se emplearon los otros miembros de la familia Cebidae, de la cual hace parte *C. jacchus*, debido a que las secuencias del gen presentaron regiones indeterminadas. Probablemente este grupo se pueda resolver si se emplean más miembros cercanos a *C. jacchus* que permitan dilucidar su estado. Por otro lado, durante la descripción de los resultados y en la discusión, se mencionó en gran medida que *M. murinus* presenta varias excepciones a lo obtenido en este proyecto. Esto puede estar dado en gran parte por el tamaño de su secuencia (182 734 pb) la cuál corresponde alrededor del 10% del tamaño promedio del gen *PRKN*. Así mismo, la caracterización de su genoma se encuentra en proceso de secuenciación, por lo cual presenta regiones aún por determinar.

Finalmente, durante esta pasantía se trabajó el estado evolutivo del gen *PRKN* en el orden Primates, sin embargo, se debe resaltar que la cantidad de taxones empleados no corresponde al total de especies de este orden. A pesar de esto, los resultados de esta pasantía demuestran la importancia de la implementación de herramientas bioinformáticas, análisis *in silico* y el acceso a base de datos públicas, para el estudio de los genes con interés en ciencia básica, así como en ciencia aplicada y terapias génicas. De igual modo, se establece un punto de partida para el estudio del gen *PRKN* en el orden de los Primates, con el fin de establecer relaciones evolutivas y fisiológicas que permitan comprender la aparición y desarrollo de la Enfermedad de Parkinson, así como en otras patologías que involucren dicho gen.

7. Conclusiones

La Enfermedad de Parkinson es un trastorno neurodegenerativo de gran interés debido a su prevalencia y características fisiológicas, las cuales están relacionadas principalmente con el gen *PRKN*. Diversos análisis han permitido establecer una relación entre *PRKN* y la EP, sin embargo, se requieren de estudios comparativos y análisis bioinformáticos que permitan esclarecer los procesos fisiológicos en los que se relaciona *PRKN* con diversos trastornos como la EP. Con base en lo anterior, se concluye que el gen *PRKN* se encuentra conservado entre diferentes grupos del orden Primates, principalmente hacia la región 3', lo cual se puede observar en la organización de exones e intrones de las diferentes variantes de ARNm. Así mismo, la caracterización de la fase de los intrones permite corroborar dicho patrón y a su vez, definir el estado evolutivo de los intrones, los cuales mostraron el patrón de conservación: los intrones antiguos se ubicaron hacia la región 3', mientras que hacia la región 5' se encuentran intrones recientes.

Por otro lado, los dominios estructurales de las proteínas generadas por los ARNm de las variantes presentan mayor variación a comparación de la estructura del gen. Por lo cual, a pesar de que el gen se encuentra conservado en ciertas regiones, la traducción a proteína puede variar de acuerdo con la organización de exones e intrones. Aunado a lo anterior, se demuestra de manera *in silico*, la presencia de Marcos Abiertos de Lectura y exones dentro de los intrones del gen *PRKN* en el orden primates, que, junto con lo anterior, permitirían explicar las diferentes características fisiológicas de *PRKN*, sin embargo, esto requiere de validación ya sea *in vitro* o *in vivo*.

Finalmente, los resultados obtenidos en esta pasantía demuestran la importancia de los análisis bioinformáticos en áreas de ciencia básica y ciencia aplicada. Así mismo, permiten cumplir con los objetivos propuestos en este trabajo.

8. Recomendaciones

Se deben validar y caracterizar de manera experimental las variantes de ARNm del gen *PRKN*, así como las distintas estructuras halladas dentro de los intrones. Así mismo, se deben implementar herramientas de secuenciación de tercera generación y bioinformáticas actuales para obtener una mayor cantidad de genomas que permitan establecer relaciones evolutivas con un enfoque terapéutico.

Referencias Bibliográficas

- Abugessaisa, I., Kasukawa, T., & Kawaji, H. (2017). Genome Annotation. *Methods in Molecular Biology* (Clifton, N.J.), 1525, 107–121. https://doi.org/10.1007/978-1-4939-6622-6_5.
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., ... Craig Venter, J. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185–2195. <https://doi.org/10.1126/science.287.5461.2185>.
- Adler, C. H., Beach, T. G., Hentz, J. G., Shill, H. A., Caviness, J. N., Driver-Dunckley, E., Sabbagh, M. N., Sue, L. I., Jacobson, S. A., Belden, C. M., & Dugger, B. N. (2014). Low clinical diagnostic accuracy of early vs advanced Parkinson disease: clinicopathologic study. *Neurology*, 83(5), 406–412. <https://doi.org/10.1212/wnl.0000000000000641>.
- Ahmad, A., Nkosi, D., & Iqbal, M. A. (2023). *PARK2* Microdeletion or Duplications Have Been Implicated in Different Neurological Disorders Including Early Onset Parkinson Disease. *Genes*, 14(3), 600. <https://doi.org/10.3390/genes14030600>.
- Alexoudi, A., Alexoudi, I., & Gatzonis, S. (2018). Parkinson's disease pathogenesis, evolution and alternative pathways: A review. *Revue Neurologique*, 174(10), 699–704. <https://doi.org/10.1016/J.neurol.2017.12.003>.
- Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C., & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*, 60(5), 685–699. <https://doi.org/10.1093/sysbio/syr041>.

- Augusto, R. L., Mendonça, I. P., de Albuquerque Rego, G. N., Pereira, D. D., da Penha Gonçalves, L. V., dos Santos, M. L., de Souza, R. F., Moreno, G. M. M., Cardoso, P. R. G., de Souza Andrade, D., da Silva-Júnior, J. C., Pereira, M. C., Peixoto, C. A., Medeiros-Linard, C. F. B., de Souza, I. A., & Andrade-da-Costa, B. L. da S. (2020). Purified anacardic acids exert multiple neuroprotective effects in pesticide model of Parkinson's disease: *in vivo* and *in silico* analysis. *IUBMB Life*, 72(8), 1765–1779. <https://doi.org/10.1002/iub.2304>.
- Barak, D. S., & Batra, S. (2021). Direct Access to Amides from Nitro-Compounds via Aminocarbonylation and Amidation Reactions: A Minireview. *Chemical Record* (New York, N.Y.), 21(12), 4059–4087. <https://doi.org/10.1002/tcr.202100224>.
- Barazzuol, L., Giamogante, F., Brini, M., & Cali, T. (2020). *PINK1/Parkin* Mediated Mitophagy, Ca²⁺ Signalling, and ER–Mitochondria Contacts in Parkinson's Disease. *International Journal of Molecular Sciences*, 21(5), 1772. <https://doi.org/10.3390/ijms21051772>.
- Barodia, S. K., Creed, R. B., & Goldberg, M. S. (2017). *Parkin* and *PINK1* functions in oxidative stress and neurodegeneration. *Brain Research Bulletin*, 133, 51–59. <https://doi.org/10.1016/j.brainresbull.2016.12.004>.
- Bayne, A. N., & Trempe, J. F. (2019). Mechanisms of PINK1, ubiquitin and Parkin interactions in mitochondrial quality control and beyond. *Cellular and Molecular Life Sciences: CMLS*, 76(23), 4589–4611. <https://doi.org/10.1007/s00018-019-03203-4>.
- Borsche, M., König, I. R., Delcambre, S., Petrucci, S., Balck, A., Bruggemann, N., Zimprich, A., Wasner, K., Pereira, S. L., Avenali, M., Deuschle, C., Badanjak, K., Ghelfi, J., Gasser, T., Kasten, M., Rosenstiel, P., Lohmann, K., Brockmann, K., Valente, E. M., ... Klein, C. (2020). Mitochondrial damage-associated inflammation highlights biomarkers in *PRKN/PINK1*

- parkinsonism. *Brain: A Journal of Neurology*, 143(10), 3041–3051.
<https://doi.org/10.1093/brain/awaa246>.
- Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology*, 268(1), 78–94. <https://doi.org/10.1006/jmbi.1997.0951>.
- Burian, A. N., Zhao, W., Lo, T. W., & Thurtle-Schmidt, D. M. (2021). Genome sequencing guide: An introductory toolbox to whole-genome analysis methods. *Biochemistry and Molecular Biology Education: A Bimonthly Publication of the International Union of Biochemistry and Molecular Biology*, 49(5), 815–825. <https://doi.org/10.1002/bmb.21561>.
- Butkiewicz, M., & Bush, W. S. (2016). *In Silico* Functional Annotation of Genomic Variation. *Current Protocols in Human Genetics*, 88, 6.15.1-6.15.17. <https://doi.org/10.1002/0471142905.hg0615S88>.
- Chaudhry, Z. L., Gamal, M., Ferhati, I., Warda, M., & Ahmed, B. Y. (2022). ER Stress in COVID-19 and Parkinson's Disease: *In Vitro* and *In Silico* Evidences. *Brain Sciences*, 12(4), 507. <https://doi.org/10.3390/brainsci12040507>.
- Chen, T., Garber, P. A., Zhang, L., Yang, L., & Fan, P. (2023). The pattern and drivers of taxonomic bias in global primate research. *Global Ecology and Conservation*, 46, 2351-9894. <https://doi.org/10.1016/J.gecco.2023.e02599>.
- Conceição, I. C., Rama, M. M., Oliveira, B., Café, C., Almeida, J., Mougá, S., Duque, F., Oliveira, G., & Vicente, A. M. (2017). Definition of a putative pathological region in *PARK2* associated with autism spectrum disorder through *in silico* analysis of its functional structure. *Psychiatric Genetics*, 27(2), 54–61. <https://doi.org/10.1097/ypg.000000000000159>.
- Craig Venter, J., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson,

- D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., ... Zhu, X. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), 1304–1351. <https://doi.org/10.1126/science.1058040>.
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 2012 9:8, 9(8), 772–772. <https://doi.org/10.1038/nmeth.2109>.
- Diakou, I., Papakonstantinou, E., Papageorgiou, L., Pierouli, K., Dragoumani, K., Spandidos, D. A., Bacopoulou, F., Chrousos, G. P., Eliopoulos, E., & Vlachakis, D. (2022). Novel computational pipelines in antiviral structure-based drug design (Review). *Biomedical Reports*, 17(6), 97. <https://doi.org/10.3892/br.2022.1580>.
- Duda, P., & Zrzavý, J. (2013). Evolution of life history and behavior in Hominidae: towards phylogenetic reconstruction of the chimpanzee-human last common ancestor. *Journal of Human Evolution*, 65(4), 424–446. <https://doi.org/10.1016/j.jhevol.2013.07.009>.
- Eisenhaber, B., Chumak, N., Eisenhaber, F., & Hauser, M. T. (2007). The ring between ring fingers (RBR) protein family. *Genome Biology*, 8(3), 209. <https://doi.org/10.1186/gb-2007-8-3-209>.
- Frith, M. C., Mitsuhashi, S., & Katoh, K. (2021). lamassemble: Multiple Alignment and Consensus Sequence of Long Reads. *Methods in Molecular Biology (Clifton, N.J.)*, 2231, 135–145. https://doi.org/10.1007/978-1-0716-1036-7_9.
- Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., & Bryant, S. H. (2010). The NCBI BioSystems database. *Nucleic Acids Research*, 38(suppl_1), D492–D496. <https://doi.org/10.1093/nar/gkp858>.
- Goffeau, A., Barrell, G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P.,

- Tettelin, H., & Oliver, S. G. (1996). Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287), 546–567. <https://doi.org/10.1126/science.274.5287.546>.
- Hirata, T., & Kizuka, Y. (2021). N-Glycosylation. *Advances in Experimental Medicine and Biology*, 1325, 3–24. https://doi.org/10.1007/978-3-030-70115-4_1.
- Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>.
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M., & Sigrist, C. J. A. (2006). The PROSITE database. *Nucleic Acids Research*, 34(Database issue), D227-D230. <https://doi.org/10.1093/nar/gkj063>.
- Jasinska, A. J. (2020). Resources for functional genomic studies of health and development in nonhuman primates. *American Journal of Physical Anthropology*, 171, 174-179. <https://doi.org/10.1002/ajpa.24051>.
- Jensen, I., Hendrich, C., Klietz, M., Berding, G., Höglinger, G. U., & Wegner, F. (2022). Case report: Early-onset Parkinson's disease with initial spastic paraparesis and hyperreflexia caused by compound heterozygous *PRKN*-gene exon 2 and 4 deletions. *Frontiers in Neurology*, 13, 969232. <https://doi.org/10.3389/fneur.2022.969232>.
- Kamienieva, I., Charzyńska, A., Duszyński, J., Malińska, D., & Szczepanowska, J. (2023). In search for mitochondrial biomarkers of Parkinson's disease: Findings in *parkin*-mutant human fibroblasts. *Biochimica et Biophysica Acta. Molecular Basis of Disease*, 1869(7), 166787. <https://doi.org/10.1016/j.bbadis.2023.166787>.

- Katsonis, P., Wilhelm, K., Williams, A., & Lichtarge, O. (2022). Genome interpretation using *in silico* predictors of variant impact. *Human Genetics*, 141(10), 1549–1577. <https://doi.org/10.1007/S00439-022-02457-6>.
- Kitada, T., Asakawa, S., Hattori, N., Matsumine, H., Yamamura, Y., Minoshima, S., Yokochi, M., Mizuno, Y., & Shimizu, N. (1998). Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature*, 392(6676), 605–608. <https://doi.org/10.1038/33416>.
- Kluge, A., Bunk, J., Schaeffer, E., Drobny, A., Xiang, W., Knacke, H., Bub, S., Lückstädt, W., Arnold, P., Lucius, R., Berg, D., & Zunke, F. (2022). Detection of neuron-derived pathological α -synuclein in blood. *Brain: A Journal of Neurology*, 145(9), 3058–3071. <https://doi.org/10.1093/brain/awac115>.
- Kondeva-Burdina, M., Mateev, E., Angelov, B., Tzankova, V., & Georgieva, M. (2022). *In Silico* Evaluation and *In Vitro* Determination of Neuroprotective and MAO-B Inhibitory Effects of Pyrrole-Based Hydrazones: A Therapeutic Approach to Parkinson's Disease. *Molecules* (Basel, Switzerland), 27(23), 8485. <https://doi.org/10.3390/molecules27238485>.
- Kress, A., Poch, O., Lecompte, O., & Thompson, J. D. (2023). Real or fake? Measuring the impact of protein annotation errors on estimates of domain gain and loss events. *Frontiers in Bioinformatics*, 3, 1178926. <https://doi.org/10.3389/fbinf.2023.1178926>.
- Kuderna, L. F., Esteller-Cucala, P., & Marques-Bonet, T. (2020). Branching out: what omics can tell us about primate evolution. *Current Opinion in Genetics & Development*, 62, 65–71. <https://doi.org/10.1016/j.gde.2020.06.006>.
- Kuderna, L. F. K., Gao, H., Janiak, M. C., Kuhlwilm, M., Orkin, J. D., Bataillon, T., Manu, S., Valenzuela, A., Bergman, J., Rousselle, M., Silva, F. E., Agueda, L., Blanc, J., Gut, M., de Vries, D., Goodhead, I., Harris, R. A., Raveendran, M., Jensen, A., ... Marques Bonet, T.

- (2023). A global catalog of whole-genome diversity from 233 primate species. *Science* (New York, N.Y.), 380(6648), 906–913. <https://doi.org/10.1126/science.abn7829>.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* (Oxford, England), 23(21), 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>.
- Lassmann, T. (2020). Kalign 3: multiple sequence alignment of large datasets. *Bioinformatics*, 36(6), 1928–1929. <https://doi.org/10.1093/bioinformatics/btz795>.
- Leduc-Gaudet, J. P., Hussain, S. N. A., & Gouspillou, G. (2022). *Parkin*: a potential target to promote healthy ageing. *The Journal of Physiology*, 600(15), 3405–3421. <https://doi.org/10.1113/jp282567>.
- Letunic, I., Khedkar, S., & Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Research*, 49(D1), D458–D460. <https://doi.org/10.1093/nar/gkaa937>.
- Liu, W., Lin, X., Chen, X., Wang, Q., Wang, X., Yang, B., Cai, N., Chen, R., Chen, G., & Lin, Y. (2023). Vision-based estimation of MDS-UPDRS scores for quantifying Parkinson’s disease tremor severity. *Medical Image Analysis*, 85, 102754. <https://doi.org/10.1016/j.media.2023.102754>.
- Marín, I. (2009). RBR ubiquitin ligases: Diversification and streamlining in animal lineages. *Journal of Molecular Evolution*, 69(1), 54–64. <https://doi.org/10.1007/S00239-009-9252-3>.
- Marín, I., & Ferrúst, A. (2002). Comparative genomics of the RBR family, including the Parkinson’s disease-related gene *parkin* and the genes of the ariadne subfamily. *Molecular Biology and Evolution*, 19(12), 2039–2050. <https://doi.org/10.1093/oxfordjournals.molbev.a004029>.

- Marín, I., Lucas, J. I., Gradilla, A. C., & Ferrús, A. (2004). Parkin and relatives: the RBR family of ubiquitin ligases. *Physiological Genomics*, 17(3), 253–263. <https://doi.org/10.1152/physiolgenomics.00226.2003>.
- Martin, N., Bergougnoux, A., Baatallah, N., Chevalier, B., Varilh, J., Baux, D., Costes, B., Fanen, P., Raynal, C., Sermet-Gaudelus, I., Girodon, E., Taulan-Cadars, M., & Hinzpeter, A. (2021). Exon identity influences splicing induced by exonic variants and *in silico* prediction efficacy. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society*, 20(3), 464–472. <https://doi.org/10.1016/J.JCF.2020.12.003>.
- Martínez-Pérez, F., Bendena, W. G., Chang, B. S. W., & Tobe, S. S. (2009). FGLamide Allatostatin genes in Arthropoda: introns early or late? *Peptides*, 30(7), 1241–1248. <https://doi.org/10.1016/j.peptides.2009.04.001>.
- Martínez-Pérez, F., Durán-Gutiérrez, D., Delaye, L., Becerra, A., Aguilar, G., & Zinker, S. (2007). Loss of DNA: a plausible molecular level explanation for crustacean neuropeptide gene evolution. *Peptides*, 28(1), 76–82. <https://doi.org/10.1016/j.peptides.2006.09.021>.
- Matsumine, H., Saito, M., Shimoda-Matsubayashi, S., Tanaka, H., Ishikawa, A., Nakagawa-Hattori, Y., Yokochi, M., Kobayashi, T., Igarashi, S., Takano, H., Sanpei, K., Koike, R., Mori, H., Kondo, T., Mizutani, Y., Schäffer, A. A., Yamamura, Y., Nakamura, S., Kuzuhara, S., ... Mizuno, Y. (1997). Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27. *American Journal of Human Genetics*, 60(3), 588–596.
- Mikkelsen, T. S., Hillier, L. W., Eichler, E. E., Zody, M. C., Jaffe, D. B., Yang, S. P., Enard, W., Hellmann, I., Lindblad-Toh, K., Altheide, T. K., Archidiacono, N., Bork, P., Butler, J., Chang, J. L., Cheng, Z., Chinwalla, A. T., Dejong, P., Delehaunty, K. D., Fronick, C. C., ...

- Waterston, R. H. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055), 69–87. <https://doi.org/10.1038/nature04072>.
- Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., Passarotti, M., Kaufman, S., & Oleary, M. A. (2015). A restful API for access to phylogenetic tools via the CIPRES Science Gateway. *Evolutionary Bioinformatics Online*, 11, 43–48. <https://doi.org/10.4137/ebo.s21501>.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., & Teeling, E. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Patthy, L. (1987). Intron-dependent evolution: preferred types of exons and introns. *FEBS Letters*, 214(1), 1–7. [https://doi.org/10.1016/0014-5793\(87\)80002-9](https://doi.org/10.1016/0014-5793(87)80002-9).
- Pinna, L. A. (1990). Casein kinase 2: an “eminence grise” in cellular regulation? *Biochimica et Biophysica Acta*, 1054(3), 267–284. [https://doi.org/10.1016/0167-4889\(90\)90098-x](https://doi.org/10.1016/0167-4889(90)90098-x).
- Pinto, M., Nissanka, N., & Moraes, C. T. (2018). Lack of *Parkin* anticipates the phenotype and affects mitochondrial morphology and mtDNA levels in a mouse model of Parkinson’s Disease. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 38(4), 1042–1053. <https://doi.org/10.1523/jneurosci.1384-17.2017>.
- Prokop, J. W., May, T., Strong, K., Bilinovich, S. M., Bupp, C., Rajasekaran, S., Worthey, E. A., & Lazar, J. (2018). Genome sequencing in the clinic: the past, present, and future of genomic medicine. *Physiological Genomics*, 50(8), 563–579. <https://doi.org/10.1152/physiolgenomics.00046.2018>.

- Seike, N., Yokoseki, A., Takeuchi, R., Saito, K., Miyahara, H., Miyashita, A., Ikeda, T., Aida, I., Nakajima, T., Kanazawa, M., Wakabayashi, M., Toyoshima, Y., Takahashi, H., Matsumoto, R., Toda, T., Onodera, O., Ishikawa, A., Ikeuchi, T., & Kakita, A. (2021). Genetic variations and neuropathologic features of patients with *PRKN* mutations. *Movement Disorders: Official Journal of the Movement Disorder Society*, 36(7), 1634–1643. <https://doi.org/10.1002/mds.28521>.
- Smit, J. J., & Sixma, T. K. (2014). RBR E3-ligases at work. *EMBO Reports*, 15(2), 142–154. <https://doi.org/10.1002/embr.201338166>.
- Spratt, D. E., Walden, H., & Shaw, G. S. (2014). RBR E3 ubiquitin ligases: new structures, new insights, new questions. *The Biochemical Journal*, 458(3), 421–437. <https://doi.org/10.1042/bj20140006>.
- Tanaka, K. (2020). The *PINK1-Parkin* axis: An Overview. *Neuroscience Research*, 159, 9–15. <https://doi.org/10.1016/j.neures.2020.01.006>.
- The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science (New York, N.Y.)*, 282(5396), 2012–2018. <https://doi.org/10.1126/science.282.5396.2012>.
- Thibaud-Nissen, F., Souvorov, A., Murphy, T., DiCuccio, M., & Kitts, P. (2013). Eukaryotic Genome Annotation Pipeline. <https://www.ncbi.nlm.nih.gov/books/nbk169439/>.
- Towler, D. A., Gordon, J. I., Adams, S. P., & Glaser, L. (2003). The biology and enzymology of eukaryotic protein acylation. *Annual review of biochemistry*, 57(1), 69–97. <https://doi.org/10.1146/annurev.bi.57.070188.000441>.
- Uesaka, K., Oka, H., Kato, R., Kanie, K., Kojima, T., Tsugawa, H., Toda, Y., & Horinouchi, T. (2022). Bioinformatics in bioscience and bioengineering: Recent advances, applications, and

- perspectives. *Journal of Bioscience and Bioengineering*, 134(5), 363–373. <https://doi.org/10.1016/j.jbiosc.2022.08.004>.
- Vihinen, M. (2023). Systematic errors in annotations of truncations, loss-of-function and synonymous variants. *Frontiers in Genetics*, 14. <https://doi.org/10.3389/fgene.2023.1015017>.
- Vizziello, M., Borellini, L., Franco, G., & Ardolino, G. (2021). Disruption of Mitochondrial Homeostasis: The Role of *PINK1* in Parkinson's Disease. *Cells*, 10(11). <https://doi.org/10.3390/cells10113022>.
- Von Coelln, R., Dawson, V. L., & Dawson, T. M. (2004). *Parkin*-associated Parkinson's disease. *Cell and Tissue Research*, 318(1), 175–184. <https://doi.org/10.1007/S00441-004-0924-4>.
- Walden, H., & Rittinger, K. (2018). RBR ligase-mediated ubiquitin transfer: a tale with many twists and turns. *Nature Structural & Molecular Biology*, 25(6), 440–445. <https://doi.org/10.1038/s41594-018-0063-3>.
- Waldmann, R., Nieberding, M., & Walter, U. (1987). Vasodilator-stimulated protein phosphorylation in platelets is mediated by cAMP- and cGMP-dependent protein kinases. *European Journal of Biochemistry*, 167(3), 441–448. <https://doi.org/10.1111/j.1432-1033.1987.TB13357.x>.
- Wang, X. Le, Feng, S. T., Wang, Z. Z., Yuan, Y. H., Chen, N. H., & Zhang, Y. (2021). *Parkin*, an E3 Ubiquitin Ligase, plays an essential role in mitochondrial quality control in Parkinson's Disease. *Cellular and Molecular Neurobiology*, 41(7), 1395–1411. <https://doi.org/10.1007/s10571-020-00914-2>.
- Wang, X. S., Cotton, T. R., Trevelyan, S. J., Richardson, L. W., Lee, W. T., Silke, J., & Lechtenberg, B. C. (2023). The unifying catalytic mechanism of the RING-between-RING

E3 ubiquitin ligase family. *Nature Communications*, 14(1), 168.
<https://doi.org/10.1038/s41467-023-35871-z>.

Yamamoto, M., Takai, Y., Inoue, M., Kishimoto, A., & Nishizuka, Y. (1978). Characterization of cyclic nucleotide-independent protein kinase produced enzymatically from its proenzyme by calcium-dependent neutral protease from rat liver. *Journal of Biochemistry*, 83(1), 207–212.
<https://doi.org/10.1093/oxfordjournals.jbchem.a131893>.