



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS  
DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

Unidad de Genómica Avanzada.  
Departamento de Ingeniería Genética

“Evolución de Redes de Regulación de microRNAs de  
Primates”

TESIS

Que presenta:

IBT Gilberto Alejandro Álvarez Canales

Para obtener el grado de:

MAESTRO EN CIENCIAS

en la especialidad de Biología Integrativa

Directores de Tesis.

Dr. Cei Abreu-Goodger

Dr. Luis José Delaye Arredondo

Irapuato, Guanajuato

Enero, 2018

El presente trabajo fue realizado bajo la asesoría del Dr. Cei Abreu-Goodger y el Dr. Luis José Delaye Arredondo en los laboratorios de Genómica Computacional del RNA y de Genómica Evolutiva en la Unidad de Genómica Avanzada y el Departamento de Ingeniería Genética de CINVESTAV Unidad Irapuato en Guanajuato, México.

Se agradece al Consejo Nacional de Ciencia y Tecnología, CONACYT por la beca otorgada durante el periodo de realización del presente trabajo.

## AGRADECIMIENTOS

Agradezco a los profesores Ceil Abreu y Luis Delaye por su valiosa guía que me aportaron en este trabajo. Además por permitirme trabajar en un tema un poco alejado de su área de más experiencia, creo que eso hizo a esta etapa una de las experiencias más enriquecedoras de mi vida. Su asesoría fue excelente junto con el apoyo que me brindaron los profesores Sean Rovito y Moisés Santillán tanto en la parte evolutiva como en la parte de redes de regulación, respectivamente. El discutir estos temas junto con ellos me ha hecho aprender mucho sobre la manera de hacer ciencia.

Agradezco al Posgrado de Biología Integrativa, a la Dra. Laila Partida y el Dr. Alfredo Herrero que fueron coordinadores del posgrado durante este periodo y a mis amigos de este posgrado Arely, los Danis, Diana, Fer, Gaby, Pavel y Víctor por todos esos momentos y buenas experiencias que pasamos juntos. Estoy muy feliz de conocerlos. También agradezco a todas las amistades que hice o mantuve aquí en CINVESTAV Irapuato, mis compañeros de casa, amigos del Fútbol, del Volley-ball y de reuniones para cenar y discutir sobre ciencia. A mis amigos de Clubes de ciencia Adrián, Manuel, Román, David, Diego, Andrés y Falcón que durante este período me volví a encontrar en distintos lugares, estoy muy feliz que sigamos creciendo como personas.

A los grupos de los laboratorios en los que estuve. A Gaby, Roberto, Beto, Dago, Javo, Pablo, Sergio, Manu, Dani, Mich, Rosy y Obed por las valiosas discusiones durante las reuniones de grupo y por la asesoría durante la preparación del proyecto. A Araceli por todo el apoyo en el uso de Mazorka y la instalación de software.

A mis amistades que hice en el MBL bajo la tutoría de mis grandes héroes durante la universidad: Rob Phillips y Hernán García. Gracias a todos ellos por esas buenas experiencias y estoy seguro que los volveré a encontrar en este camino de la ciencia. Muchas gracias a Alfred Simkin por su todo su apoyo y asesoría en la implementación y desarrollo de herramientas que utilicé.

A mi familia que siempre estuvo apoyándome en todos momentos. Claramente la influencia de mis padres y hermanos sobre la curiosidad y respecto por la naturaleza me ha hecho llegar hasta aquí. Estoy muy agradecido con ellos por su ayuda para crecer como persona.

A Tzitziki por este equipo que hemos formado, sigamos creciendo como lo hemos hecho. Gracias por todo tu gran apoyo en esta etapa de la vida, las buenas experiencias conociendo nuevos lugares y por los muchos retos que hemos superado.

Doy muchas gracias al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo de beca que me otorgaron durante la maestría durante el tiempo que realicé este proyecto.

# Índice

1. Resumen .....	11
2. Introducción. ....	12
2.1 La importancia de los elementos reguladores en la evolución.....	12
2.2 Clases de elementos reguladores de la transcripción. ....	22
2.3 ¿Cómo evolucionan las redes de regulación? Rasgos Generales.....	34
2.4 Evolución de los primates .....	45
2.5 Preguntas para abordar mediante el estudio de la evolución de las redes de regulación genética.....	49
3. Objetivos.....	50
3.1 Objetivo general.....	50
3.2 Objetivos específicos .....	50
4 Hipótesis.....	51
5 Justificación .....	51
6 Métodos.....	51
6.1 Definición de las métricas de recableado .....	51
6.2 Selección de las especies a analizar.....	56
6.4 Selección de familias de microRNAs.....	57
6.5 Análisis de recableado en 3'UTRs de Primates.....	58
6.6 Análisis de categorías funcionales. ....	69
6.7 Conjunto de 3'UTRs de un alineamiento de mamíferos .....	71
6.8 Complementación con datos de transcriptómica comparativa entre chimpancé y humano.....	72
6.9 Uso de familias no conservadas de microRNAs.....	73
6.10 Radiación adaptativa de los cíclidos de Malawi.....	73
6.11 Esquema Representativo de los métodos .....	76
6.12 Resumen general de la metodología.....	77
7. Resultados.....	78
7.1 Evaluación de las métricas de recableado de redes, mediante simulación .....	78
7.2 Análisis de las 3' UTR de la superfamilia de los Homínidos .....	82

7.3 Análisis de un conjunto alternativo de 3' UTRs de los primates obtenidas de un alineamiento de mamíferos.....	94
7.4 Familias de microRNAs con distintos niveles de conservación .....	105
7.5 Conjunto de secuencias de regiones 3'UTRs del grupo de los Cíclidos del Lago Malawi .....	107
8. Discusión .....	113
8.1 Efecto en los valores de recableado por las secuencias de regiones 3'UTR utilizadas .....	113
8.2 Implicaciones de los métodos de predicción de blancos .....	116
8.3 Implicaciones de los controles usados .....	117
8.4. Método con microRNA maduro en comparación con el método de k-meros .....	118
8.5. Métrica de recableado de Shou en comparación con la métrica de Simkin .....	118
8.6 Conservación en primates.....	120
8.7 Familias de microRNAs con tasas de recableado altas.....	120
8.8 Familias de microRNA con tasas de recableado bajas.....	123
8.9 Blancos con evidencia de selección.....	123
8.10 Discrepancia entre categorías funcionales obtenidas con el recableado por microRNAs y recableado por blanco .....	126
8.11 Efecto de otros elementos de regulación en la evolución fenotípica.....	127
8.12 Aplicación en cíclidos.....	127
9. Conclusiones .....	129
10 Perspectivas .....	130
10.1 Utilizar un grupo de controles basados en la aleatorización de secuencia .....	130
10.2 Compensación de pérdida de regulación de microRNA .....	131
10.3 Estimación de recableado en radiaciones adaptativas .....	132
10.4 Inclusión de interacciones Factor de Transcripción-microRNA-Blanco.....	133
11 Bibliografía .....	134
12. Anexos.....	143
12.1 Simulaciones de los valores de las métricas de Shou y de Simkin.....	143
12.2 Análisis con el conjunto de 3'UTRs de los Homínidos .....	146
12.3 Análisis de un conjunto alternativo de 3' UTRs de los primates obtenidas de un alineamiento de mamíferos.....	148
12.4 Análisis con familias de microRNAs más recientes .....	150

# Índice de figuras

Figure 1 Diagrama sobre el efecto de las moléculas inductoras en el desarrollo de las extremidades de un organismo (Modificado de Bénazet & Zeller 2009).....	12
Figura 2 Conjunto de genes relacionados con el desarrollo y sus patrones espaciales de expresión .....	14
Figura 3 Regulación por medio de moléculas cis y trans (www.khanacademy.org "NOTE: All Khan Academy content is available for free at www.khanacademy.org") .....	16
Figura 4 La evolución de extremidades en animales implica los mismos genes, sin embargo estos se regulan de diferente manera como podemos ver entre las extremidades de aves y humano (Modificado de Carroll et al. 2004) .....	18
Figura 5 Esquema de redes de regulación genética para distintos primates donde sus genes pueden estar regulados por 2 clases de elementos reguladores. ....	20
Figura 6 Ruta de síntesis de microRNAs en mamíferos (Modificacado de Winter et al. 2009) .....	24
Figura 7 Diagrama representativo de uno de los mecanismos de acción de los microRNAs en animales. Este mecanismo consiste en la inhibición del inicio de la traducción mediante la disociación de los elementos eIF4F( Modificado de da Sacco & Masotti 2013) .....	25
Figura 8 Modelo que representa el incremento del número de blancos conforme un microRNA es más antiguo (Nozawa et al. 2016).....	32
Figura 9 La evolución de familias de microRNAs y de complejidad morfológica a lo largo del tiempo (Modificado de Heimberg et al. 2008) .....	33
Figura 10 Recableado de red (Modificado de Shou et al. 2011).....	36
Figura 11 Diagrama representativo para evaluar recableado (Sarda & Hannehalli 2015) .....	37
Figura 12 Evolución de diferentes maneras de regulación relacionadas con la reproducción en levaduras, donde ganancia de la represión " $\alpha 2$ " permitió ganar sitios de regulación "Ste12" (Modificado de Sorrells et al. 2015).....	41
Figura 13 Radiación adaptativa de los cíclidos de África (Modificado de Brawand et al. 2014) .....	44
Figura 14 Filogenia de los primates, donde se señala al grupo de los hominoides que incluyen a los primates que sus genomas se analizaron en este proyecto (Levin 2011) .....	46
Figura 15 Cambio en la regulación de GDF6 relacionado con diferentes fenotipos.....	48
Figura 16 Esquema de las métricas de recableado entre genes y entre sitios, de Shou y de Simkin respectivamente. ....	54
Figura 17 Representación de los k-meros que pertenecen a la región semilla de la secuencia del microRNA let-7. ....	57
Figura 18 Diagrama de pre-procesamiento de las 3'UTR.....	58
Figura 19 Representación del sitio semilla conservado (Friedman et al. 2009) .....	61
Figura 20 Matriz de Incidencia que representa una red de regulación genética de microRNAs.....	62
Figura 21 Diagrama representativo de la comparación de las redes de regulación para estimar el recableado sufrido desde su ancestro .....	63

Figura 22 Diagrama representativo de la métrica de recableado por kilobase de gen blanco propuesta. ....	67
Figura 23 Obtención del valor de razón de recableado por blanco "p" .....	68
Figura 24 Obtención de los valores de recableado para cada blanco en los 3 posibles escenarios. ....	69
Figura 25 Filogenia de los primates donde se puede observar el grupo de los Catharrhini que incluye a los homínidos mencionados anteriormente y a los monos del viejo mundo (Naish 2016). ....	72
Figura 26 Efecto de la Proporción de Recableado entre genes (PRG ) sobre las dos métricas .....	78
Figura 27 Distribución de conectividades para cada especie de los primates analizados .....	79
Figura 28 Tasa de Simkin en función de la tasa de Shou para distintos valores del parámetro de la distribución de Poisson .....	80
Figura 29 Espacio de valores de recableado en el que se encuentran las redes a evaluar .....	81
Figura 30 Cantidad de secuencias disponibles después de cada uno de los procesamientos .....	82
Figura 31 Gráfico de dispersión de los valores de CS++ tomando en cuenta al $P_{CT}$ y no tomando en cuenta al $P_{CT}$ .....	83
Figura 32 Valor de recableado con métrica de Shou del conjunto de k-meros correspondientes a los microRNAs comparado con el modelo nulo de k-meros control.....	84
Figura 33 Valor de recableado con métrica de Simkin del conjunto de k-meros correspondientes a los microRNAs comparado con el modelo nulo de k-meros control.....	85
Figura 34 Tasa de recableado de Simkin para cada familia de microRNA utilizando los blancos más probables y los blancos menos probables.....	86
Figura 35 Tasa de recableado de Shou para cada familia de microRNA utilizando los blancos más probables y los blancos menos probables.....	87
Figura 36 Gráfico de dispersión para observar correlación de los valores de recableado para cada microRNA entre diferentes especies y sus respectivos controles. ....	88
Figura 37 Comparación de las tasas de recableado para cada microRNA nombrado en las filas para cada una de las especies mencionadas en las columnas.....	89
Figura 38 Pruebas de enriquecimientos categorías funcionales para las familias de microRNAs con más y menor recableado con diferentes cortes del CS++, en la barra de color se muestran los valores del logaritmo base 10 del FDR.....	91
Figura 39. Categorías funcionales de genes con valores extremos de recableado en humano.....	91
Figura 40 Comparación entre humano y chimpancé de las categorías funcionales de sus blancos con valores extremos de recableado .....	92
Figura 41 Resultado de ReviGO de las categorías funcionales enriquecidas para los blancos con mayor recableado .....	93
Figura 42 Resultado de ReviGO de las categorías funcionales enriquecidas para los blancos con menor recableado. ....	94
Figura 43 Diagrama de pasos del procesamiento de las 3'UTR con las cantidades de genes obtenidos en casa paso.....	94

Figura 44 Tasa de recableado de Shou para cada familia de microRNA utilizando los blancos más probables y los blancos menos probables.....	95
Figura 45 Comparación de las tasas de recableado para cada microRNA nombrado en las filas para cada una de las especies mencionadas en las columnas.....	97
Figura 46 Pruebas de enriquecimientos categorías funcionales para las familias de microRNAs con más y menor recableado con diferentes cortes de CS++, en la barra de color se muestran los valores del FDR.....	99
Figura 47 Pruebas de enriquecimientos de categorías funcionales para los blancos de microRNAs con valores extremos de recableado por kilobase, en la barra de color se muestran los valores del FDR.....	99
Figura 48 Categorías funcionales de los genes con valores extremos de recableado por gen blanco.....	100
Figura 49 Resultado de ReviGO de las categorías funcionales enriquecidas para los blancos con mayor recableado.....	101
Figura 50 Resultado de ReviGO de las categorías funcionales enriquecidas para los blancos con menor recableado.....	102
Figura 51 Gráfico de volcán para los niveles de expresión y valores de significancia estadística del contraste de los genes expresados en la corteza del cíngulo anterior entre humano y chimpancé. En amarillo se muestran los genes que perdieron sitios de regulación en humano y ganaron sitios de regulación en chimpancé. En azul se muestran los genes que perdieron sitios en chimpancé y ganaron sitios en humano.....	103
Figura 52 Gráfico de volcán para los niveles de expresión y valores de significancia estadística del contraste de los genes expresados en la corteza del núcleo caudado entre humano y chimpancé.....	104
Figura 53 Tasa de recableado de Shou para cada familia de microRNA utilizando los blancos más probables y los blancos menos probables.....	106
Figura 54 Tasas ganancia y de pérdidas para cada uno de los k-meros correspondiente a los sitios de pegado de los microRNAs. En rojo los correspondientes a las familias más conservadas y en verde los correspondientes a las familias más recientes en humano.....	107
Figura 55 Distribución de las tasas de recableado con métrica de Simkin para cada microRNA en diferentes linajes de cíclidos para las sitios blanco más probables en amarillo y los sitios blanco menos probables en azul.....	109
Figura 56 Comparación de las tasas de recableado con métrica de Simkin para cada microRNA nombrado en las filas para cada una de las especies mencionadas en las columnas.....	110
Figura 57 Distribución de las tasas de recableado con métrica de Shou para cada microRNA en diferentes linajes de cíclidos para las sitios blanco más probables en amarillo y los sitios blanco menos probables en azul.....	111
Figura 58 Comparación de las tasas de recableado con métrica de Shou para cada microRNA nombrado en las filas para cada una de las especies mencionadas en las columnas.....	112
Figura 59 Tasa de recableado de Shou para la intersección de los genes provenientes de diferentes alineamientos.....	114

Figura 60 Análisis funcional utilizando la intersección de secuencias de 3'UTR con el método de Simkin et al. del 2014 y Leclercq et al. del 2017.....	115
Figura 61 Diagramas de ejemplo de posibles situaciones conflictivas al utilizar la métrica de Simkin para análisis de categorías funcionales .....	119
Figura 62 Gráfico de volcán de expresión de genes en la corteza del cíngulo anterior. En rojo se nombran los genes que perdieron sitios de regulación en humano y ganaron sitios de regulación en chimpancé. La línea azul representa el punto de corte del FDR (0.1) .....	125
Figura 63 Esquema representativo de las implicaciones que tiene utilizar a los blancos de un microRNA con una tasa alta de recableado para un posterior análisis funcional.....	126
Figura 64 Gráfico de dispersión para el microRNA let-7.....	131
Figura 65 Número de pérdidas de regulación que no fueron compensadas por otro microRNA. ....	132

## Índice de tablas

Tabla 1 Adaptación conferida por mutaciones en regiones cis-reguladoras o codificantes a proteína.....	21
Tabla 2 Patrones observados en la evolución de familias y de blancos de microRNAs....	30
Tabla 3 Especies utilizadas y sus tiempos de divergencia .....	56
Tabla 4 Especies de cíclidos utilizadas para el análisis de recableado .....	74
Tabla 5 Tejidos con sus respectivos p-valores de las pruebas de Wilcoxon para los valores de expresión de los 2 conjuntos de genes .....	104
Tabla 6 Características de los genomas utilizados.....	116
Tabla 7 Familias de microRNAs con tasas de recableado altas, sus posibles funciones y linajes en los que están presentes .....	121
Tabla 8 Familias de microRNAs con tasas de recableado bajas, sus posibles funciones y linajes en los que están presentes .....	123

# 1. Resumen

La evolución afecta tanto a las proteínas como a los genes y sus elementos reguladores. Cada uno de ellos también tiene interacciones y los procesos evolutivos también afectan a las redes que resultan de dichas interacciones. Estas redes de regulación genética son elementos clave para la evolución de rasgos fenotípicos. Sin embargo, se conoce muy poco sobre la evolución de estas redes. Para entender mejor este proceso, en este trabajo se abordó la evolución de las redes de regulación de microRNAs en primates. Se estudió el cambio de la conectividad de estas redes, o recableado de redes, a lo largo de la evolución de diferentes primates. Se estimaron tasas de cambio de estas redes y se identificaron los microRNAs y blancos con valores extremos de recableado. Se hicieron análisis de enriquecimiento funcional para saber con qué funciones están asociados dichos elementos de regulación. Encontramos que los blancos que presentan una tasa alta de cambio están asociados con procesos del sistema nervioso y señalización entre células. Estos análisis se contrastaron con datos de expresión de tejidos de cerebro en humano y chimpancé. Sugerimos que existen cambios evolutivos de recableado que correlacionan con cambios en los niveles de expresión. Estos análisis ayudarán a explicar las diferencias de regulación génica entre los primates con respecto al humano.

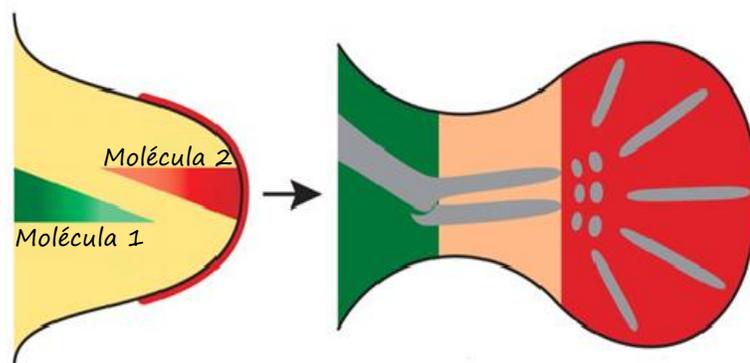
## Abstract

Evolution affects individual proteins, genes and regulatory elements. Each of these can also interact, and evolutionary processes also act on the resulting networks. Gene regulatory networks are key for the evolution of phenotypic traits, but the evolution of these networks is not well understood. To gain more information about this process, I focused on the evolution of microRNA regulatory networks in primates. I studied the changes in connectivity of these networks, or network rewiring, throughout the evolution of different primates. I calculated rates of change for the regulatory networks, and miRNAs and targets with the most extreme values for rewiring were identified. Functional enrichment analyses were done in order to predict the associated functions of the identified genes. I found that nervous system and cell signaling processes were enriched in targets with high rewiring. These analyses were compared to transcriptomic datasets from human and chimp brain tissues. I suggest that the rewiring events predicted here are correlated with the changes in the expression profiles. These analyses help explain the differences in genetic regulation between primates and humans.

## 2. Introducción.

### 2.1 La importancia de los elementos reguladores en la evolución.

Los mecanismos que dirigen la evolución de la gran diversidad morfológica de los seres vivos aún no se comprenden en su totalidad. Las primeras pruebas, de cómo la morfología de un organismo se puede alterar, se generaron mediante el estudio de los procesos de desarrollo. Los embriólogos del siglo XVII y XVIII observaron cómo es que se originan varias estructuras anatómicas a partir de un embrión. Mediante experimentos de trasplante se pudieron identificar las regiones discretas que generan determinadas partes del cuerpo. Al trasplantar estas regiones discretas en otras regiones del embrión notaron que seguían dando origen a las estructuras que originalmente generaban. A partir de estas observaciones se propuso que el efecto de moléculas inductoras producidas en estos sitios eran las causantes del desarrollo (Figura 1). Sin embargo, los mecanismos moleculares que producían este fenómeno no se conocieron hasta que se descubrieron algunos genes que se comportaban como las moléculas inductoras propuestas (Harrison 1918; Spemann & Mangold 1924).



*Figure 1 Diagrama sobre el efecto de las moléculas inductoras en el desarrollo de las extremidades de un organismo (Modificado de Bénazet & Zeller 2009)*

Con los primeros avances en el entendimiento de los genes, se empezaron a identificar cuáles son los grupos de genes que están implicados en el desarrollo. Estos análisis se hacían mediante la búsqueda de mutantes de genes que producen un fenotipo anormal.

Se identificó una pequeña fracción del total de genes que altera el desarrollo. Muchas de estas alteraciones ocurrían en partes discretas o segmentos específicos del organismo. Resultó entonces que estos genes que alteran el desarrollo son principalmente factores de transcripción o elementos involucrados en señalización. Después, mediante otro tipo de ensayos, se confirmó que las regiones espaciales afectadas durante el desarrollo correlacionan con el patrón espacial y temporal de expresión de los genes (Bridges & Morgan 1923; Lewis 1963).

Los genes implicados en el desarrollo se empezaron a clasificar de acuerdo a las funciones que tenían. Un ejemplo muy bien estudiado es el grupo de los genes homeóticos. Se les denominó homeóticos debido al fenotipo que mostraban sus mutantes. Este fenotipo consiste en la sustitución de una parte del cuerpo por otra, en una región que no le correspondía en un organismo. Uno de los ejemplos más famosos es cuando un insecto presenta patas en el lugar de las antenas. Al inicio los fenotipos se buscaron mediante mutaciones espontáneas que afectaban el desarrollo de la mosca. Estos análisis motivaron a seguir buscando sistemáticamente otros genes implicados en el desarrollo. Así fue como se llegó a descubrir el conjunto de los genes Hox. Este conjunto cuenta con 8 genes que afectan la especificación de segmentos particulares en el desarrollo. Los patrones de expresión de estos genes correlacionan con las partes del organismo que se afectan al mutarse estos genes (Figura 2). Lo que resultó aún más sorprendente fue que después de identificar este pequeño grupo de genes en *Drosophila melanogaster*, se descubrió que también están ampliamente conservados en diferentes linajes de animales (Lewis 1978; Nüsslein-Volhard & Wieschaus 1980).

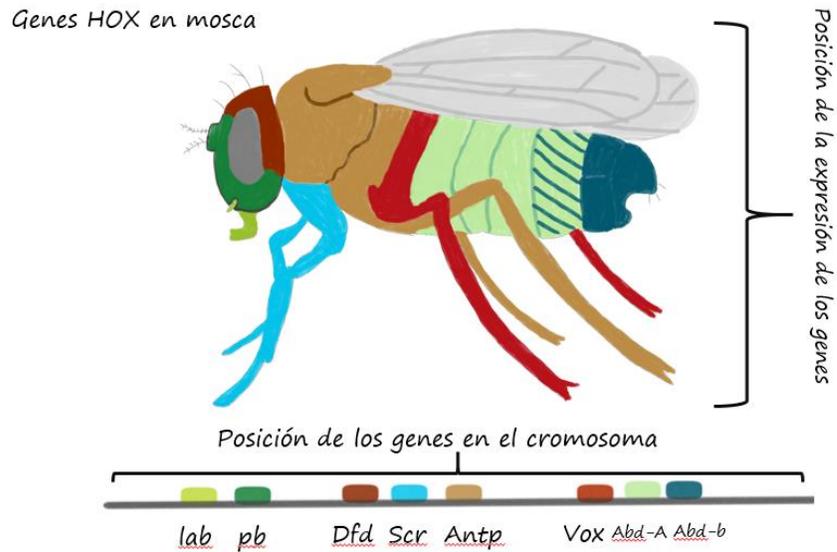


Figura 2 Conjunto de genes relacionados con el desarrollo y sus patrones espaciales de expresión

El hecho de que este grupo de genes estuvieran tan conservados y generaran aspectos morfológicos muy diferentes, intrigó más a los científicos. A estas alturas se conocían los elementos que generan la morfología pero lo que hace que los organismos fueran tan diversos en su forma, aún no se comprendía. Para poder entender como había evolucionado la morfología de estos organismos primero se debía conocer cómo funcionan los procesos evolutivos. Algunos autores consideran a la evolución como un proceso que consta de dos pasos que se van dando a través de generaciones sucesivas. Estos pasos son:

**Origen de la variación genética.** Existen diferentes fenómenos que generan variantes genéticas en un rango amplio de escalas. Estos cambios van desde el nivel de nucleótidos hasta cambios al nivel de cromosomas. Uno de ellos es la mutación puntual que consiste en la inserción, eliminación o sustitución de una sola base del DNA. Este tipo de mutación es importante tanto en bacterias como en eucariontes. Otro mecanismo de variación son los elementos transponibles que generan mutaciones por medio de recombinación, interrupción o inserción de

genes. Respecto a las mutaciones en escalas más grandes existen los rearrreglos, duplicaciones e inversiones cromosomales.

Existen mecanismos que generan nuevas variaciones a partir de los tipos de mutación mencionados anteriormente. En los organismos que pueden tener reproducción sexual existe una fuente adicional de variación por la combinación de DNA de individuos diferentes. Esta variación se debe al entrecruzamiento de los cromosomas homólogos y la segregación que ocurre en los procesos de meiosis. Estos procesos generan nuevos alelos y diferentes distribuciones de los mismos. Otra de las variaciones producidas por la recombinación es la duplicación en tándem de secuencias y ocurre cuando las secuencias homólogas no están perfectamente alineadas.

Otros procesos no generan nuevos alelos pero si afectan la distribución de los mismos. El flujo genético es uno de estos mecanismos y puede ser ocasionado por migraciones de poblaciones o individuos y el entrecruzamiento de los mismos. Otro de los mecanismos es el entrecruzamiento no aleatorio, donde eventos como la endogamia y exogamia afectan la composición genética de una población.

- **Destino de la variación genética.** Procesos como la deriva génica y la selección natural definirán lo que ocurrirá con la variación a causa de una reproducción diferencial. Estos procesos se llevan a cabo a nivel poblacional, por lo tanto los mecanismos de diversificación se llevan a cabo tanto entre poblaciones y en el interior de la misma (Futuyma 2005).

La diversidad morfológica primero se trató de explicar mediante cambios en el número de los genes del desarrollo, si presentaban duplicaciones o copias con variantes que estuvieran relacionadas con los cambios fenotípicos. Sin embargo, estas suposiciones resultaron ser incorrectas ya que se sugiere que algunos ancestros tuvieron más copias de este número de genes. Tan solo en el linaje de los mamíferos no se conoce duplicación alguna de este conjunto de genes. Por lo tanto, la duplicación de genes contribuye en la evolución, pero la frecuencia con la que ocurre no es suficiente para explicar la diversidad morfológica. Así se comenzó a sugerir y juntar evidencia de que la

variación morfológica se relaciona más con cambios en los patrones de expresión de estos genes (Meyer 1998; Carroll 2008).

Para entender el fenómeno de la evolución morfológica se tuvieron que comprender las bases de la regulación de genes. A escalas moleculares, la regulación genética puede ocurrir a dos niveles, cis y trans. El nivel cis de regulación genética consiste en los sitios de pegado de las moléculas reguladoras, estos sitios de pegado pueden encontrarse en el DNA y RNA. El nivel trans de regulación, son dichas moléculas reguladoras que van a interactuar con los sitios del nivel cis. En la Figura 3 se presenta un diagrama de estos elementos.

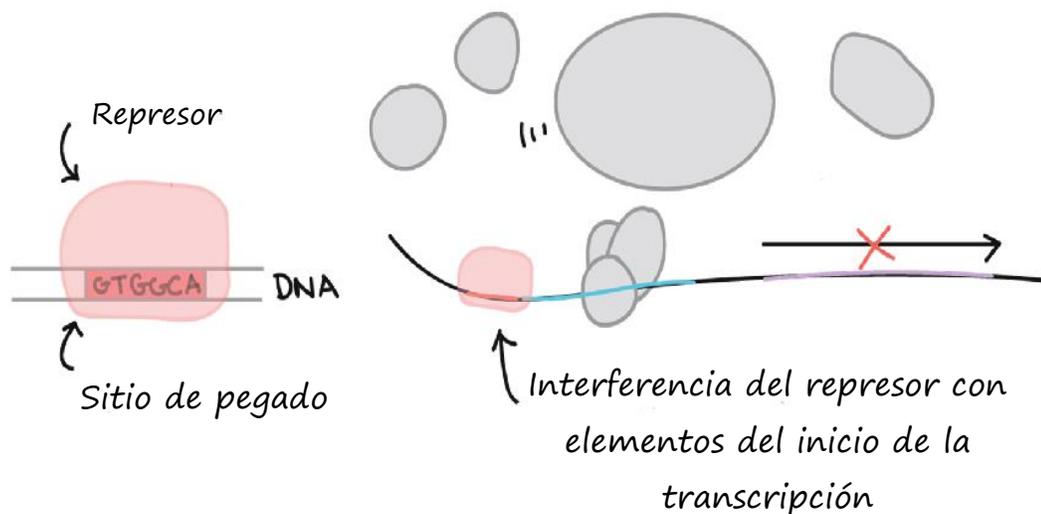
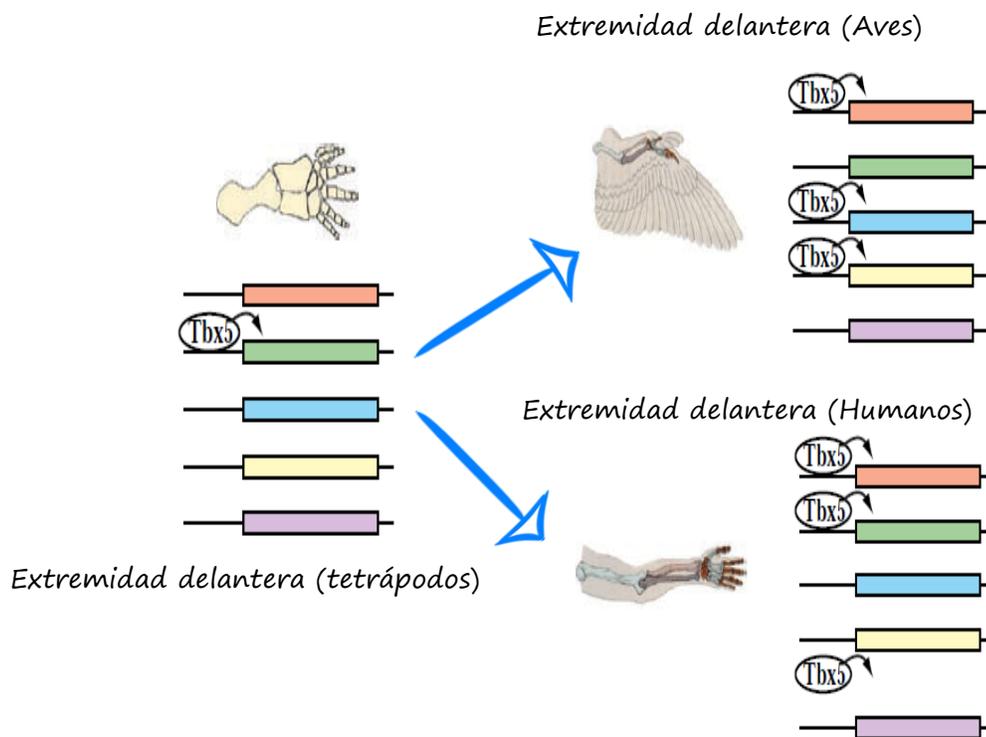


Figura 3 Regulación por medio de moléculas cis y trans ([www.khanacademy.org](http://www.khanacademy.org) "NOTE: All Khan Academy content is available for free at [www.khanacademy.org](http://www.khanacademy.org)")

En eucariontes complejos, solo una fracción de sus genomas contiene regiones codificantes para proteínas. Es aquí donde se encuentran codificados los factores de transcripción como los genes Hox. El resto del genoma se compone de regiones que codifican para moléculas reguladoras de RNA, los sitios de pegado de los elementos reguladores, intrones, elementos repetidos, y regiones llamadas intergénicas por no tener función conocida; todas estas regiones son llamadas no codificantes (Dunham et al. 2012).

Siguiendo otro enfoque, a mediados de la década de los noventa se comenzó a juntar evidencia que sustenta que los cambios en los patrones de expresión de los genes Hox en artrópodos, vertebrados y anélidos correlacionan con las transiciones fenotípicas que hubo entre estos linajes. También se observó que los sitios de pegado de los factores de transcripción son los que permiten principalmente la variación morfológica. Esto ocurre principalmente en situaciones donde los factores de transcripción están muy conservados. Así, es posible que los cambios morfológicos ocurran utilizando el mismo material genético y proteínas reguladoras. Un ejemplo es el gen Tbx5 que codifica para un factor de transcripción relacionado con el desarrollo de las extremidades en vertebrados. Como se muestra en la Figura 4 este factor de transcripción regula a diferentes genes en las extremidades de organismos distintos como mamíferos y aves. Se sugiere que los cambios en los sitios de pegado de este factor de transcripción ocasionan estas diferencias fenotípicas (Carroll 1995; Burke et al. 1995; Gellon & William McGinnis 1998).



*Figura 4 La evolución de extremidades en animales implica los mismos genes, sin embargo estos se regulan de diferente manera como podemos ver entre las extremidades de aves y humano (Modificado de Carroll et al. 2004)*

Si se analizan las propiedades intrínsecas de los elementos cis de regulación, se pueden distinguir 3 propiedades clave. Primero, su flexibilidad a los cambios es mucho más amplia que los cambios que pueden ocurrir en proteínas debido a que no existen restricciones tales como el marco de lectura. Además las distancias y posiciones en las que pueden regular la expresión pueden ser muy variadas. Otra propiedad es que estos elementos pueden cambiar de manera independiente sin afectarse permitiendo que haya cambios que se compensen. Por último, algunas configuraciones de regulación, pueden afectar la expresión a un nivel espacio-temporal generando nuevos fenotipos. Se llegó a concluir que las propiedades de estos elementos hacen que su dinámica evolutiva sea muy diferente a la de las proteínas (Carroll et al. 2004).

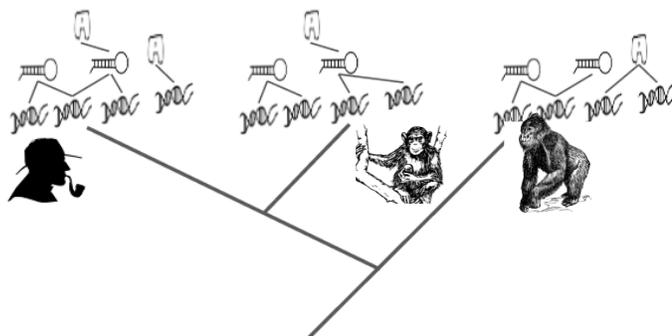
A mediados de los setenta el trabajo de King y Wilson comenzó la tendencia de asumir que la mayoría de cambios entre organismos con un genoma similar, pero con diferencias fenotípicas marcadas, ocurren en sitios de regulación. Este análisis se realizó comparando proteínas de chimpancé y de humano mediante diferentes tipos de ensayos, desde comparaciones indirectas mediante pruebas inmunológicas hasta comparaciones de las secuencias disponibles en aquella época (King & Wilson 1975).

Eric Davidson llegó a la conclusión de que el aspecto morfológico es el resultado de diferentes procesos del desarrollo que se llevan a cabo por la regulación de genes. Esto lo descubrió mientras estudiaba el desarrollo de los erizos de mar. Además, Davidson observó que la naturaleza de la regulación genética se puede abordar desde la perspectiva de redes, debido a que algunos elementos reguladores interactúan con sus respectivos genes blanco y algunos de estos elementos reguladores a su vez también pueden ser regulados por otros factores (Davidson 2006).

Toda esta evidencia llevó a la creación de una de las propuestas de evolución morfológica más ampliamente aceptadas. Sean B. Carroll en 2008, formuló la Teoría Genética de la Evolución Morfológica. En ocho puntos, él muestra los principios clave para el entendimiento de la evolución fenotípica:

1. Muchas de las proteínas que regulan el desarrollo son **pleiotrópicas**, es decir, tienen una amplia distribución e interacción en diferentes tejidos, células y etapas del desarrollo.
2. Los organismos que son diferentes morfológicamente tienen **conjuntos comunes** de genes participantes en el desarrollo.
3. Muchas de estas proteínas en común e implicadas en el desarrollo, por lo general tienen también la **misma función** en diferentes organismos.
4. Se piensa que muchas de las estructuras morfológicas evolucionaron de manera independiente mediante circuitos genéticos similares, siendo este un fenómeno llamado **homología profunda**.
5. La **duplicación** de estas familias de genes implicadas en el desarrollo es **escasa** y tal vez no es un elemento necesario para cambios en la morfología.
6. Existen cambios en la regulación espacial de estos genes esenciales del desarrollo y estos cambios afectan a la morfología; a este tipo de cambios se le llama **heterotopía**.
7. Regiones cis de regulación, que son **complejas**, contribuyen a la pleiotropía de este conjunto de genes esenciales para el desarrollo.
8. Un solo elemento regulador en trans puede controlar **muchos genes blanco**.

Este problema se puede abordar utilizando un enfoque de redes, permitiendo capturar la complejidad del fenómeno de evolución de la regulación. Las redes consideran cambios en los niveles de regulación cis y trans. Las redes de regulación genética controlan la expresión de miles de genes en patrones determinados de tiempo y espacio. Esta regulación es la que genera funciones específicas de las células (Alon 2006). Estas redes tienen un papel esencial para entender procesos biológicos del desarrollo y fenómenos evolutivos. Hasta el momento es poco claro cómo es que estas redes cambian en función del tiempo. En la Figura 5 a manera de diagrama se presentan varias redes de regulación para diferentes primates y se observan diferencias entre las conectividades de los diferentes elementos reguladores. Sin embargo nuestro conocimiento de las redes de regulación se limita a los organismos existentes en la actualidad y la información conocida de estos hasta este momento. Además el panorama es mucho más complejo ya que las redes son diferentes para cada tejido y para cada etapa del desarrollo.



*Figura 5 Esquema de redes de regulación genética para distintos primates donde sus genes pueden estar regulados por 2 clases de elementos reguladores.*

### 2.1.1 Importancia de la evolución en proteínas y elementos reguladores.

La evolución del genoma afecta tanto a las secuencias reguladoras como a las secuencias que codifican para las proteínas. Históricamente, la evolución de las secuencias codificantes ha sido más estudiada que la de regiones no codificantes ya que aún no se conocía la función de estas regiones no codificantes. Además, las primeras secuencias disponibles fueron las de los aminoácidos de las proteínas que estaban relacionadas con rasgos funcionales. A pesar de que ya se habían hecho esfuerzos de comparar moléculas biológicas entre especies mediante ensayos indirectos de hibridación e inmunología, fue hasta la década de los cincuenta que se empezaron a secuenciar péptidos como la insulina. Estas secuencias proteicas permitieron realizar comparaciones entre especies. Sin embargo, pasaron más de 15 años para obtener las primeras secuencias de ácidos nucleicos. Así fue como las bases de la evolución molecular empezaron a cambiar para entender la complejidad de los genomas (Hagen 1999).

En la actualidad, con el conocimiento de la regulación de genes y con las tecnologías de secuenciación, se pueden estudiar los sitios de pegado de los elementos reguladores y cómo se afectan los genes regulados al modificar algunas de las moléculas reguladoras. Esta información permite estudiar la evolución molecular de las redes de regulación y cómo es afectada por diferentes procesos.

Como se discutió en la sección anterior, tanto las secuencias que codifican a las proteínas como las secuencias reguladoras pueden tener un papel en la evolución de la morfología (Hoekstra & Coyne 2007; Carroll 2008). En la Tabla 1 se presentan varios ejemplos de mutaciones en regiones reguladoras y en regiones codificantes que tienen efectos en la adaptación y en el fenotipo.

Tabla 1 Adaptación conferida por mutaciones en regiones cis-reguladoras o codificantes a proteína.

Regiones cis-reguladoras		Genes codificantes de proteínas	
Rasgo fenotípico	Mecanismo molecular	Rasgo fenotípico	Mecanismo molecular
Armadura esquelética diferente en peces espinosos marinos y los bénticos	Diferentes patrones de expresión en el gen <i>Pitx1</i> (Shapiro <i>et al.</i> 2004, 2006)	Capacidad de las aves para migrar a grandes altitudes	Cambios en la estructura de la hemoglobina
Pigmentación en las alas de <i>Drosophila</i>	Expresión diferencial de proteína amarilla, por pérdida y ganancia de sitios de pegado en el gen <i>yellow</i> (Gompel <i>et al.</i> 2005)	Capacidad de algunos peces de vivir en aguas frías	Proteínas con propiedad anticongelante
Densidad de marcas en larvas de <i>Drosophila</i>	Expresión diferencial de <i>ovo Shavenbaby (svb)</i> en <i>D. melanogaster</i> y <i>D. sechelia</i>	Incremento en la herbivoría en algunos monos	RNAasa pancreática

El trabajo de Hoekstra y Coyne llega para recordar la importancia que juegan también las variaciones en la región codificante y no centrar toda la atención en los elementos reguladores. Este trabajo se enfoca principalmente en aquellas variaciones genéticas que afectan la adaptación e incluso se hace mención de genes involucrados en eventos de

especiación. Estos genes contribuyen al aislamiento reproductivo, el cual es un paso relacionado con el origen de nuevas especies. Un ejemplo son los genes *OdsH*, *Hmr*, *Lhr* y *Nup96* relacionados con esterilidad de los híbridos, los cuales tienen señales de cambios evolutivos rápidos. Hoekstra y Coyne argumentan que, por sí solos, los cambios en regiones cis-reguladoras no pueden causar aislamiento reproductivo. (Presgraves et al. 2003; Hoekstra & Coyne 2007). Sin embargo esta última parte aún no es concluyente debido a la poca información de los elementos reguladores que existe hoy en día.

Los eventos de especiación y adaptación probablemente proceden por mutaciones en ambas regiones: codificantes y no codificantes. Anteriormente se explicó la importancia de los elementos reguladores en la evolución fenotípica y la existencia de una gran cantidad de evidencia que lo soporta, proveniente de los trabajos hechos por científicos como Sean B. Carroll y Eric Davidson. El escenario ideal en el futuro será evaluar en conjunto tanto a las proteínas como a los elementos reguladores.

## 2.2 Clases de elementos reguladores de la transcripción.

Existen varios mecanismos que regulan la expresión de genes en animales. Esto puede ocurrir en distintos momentos, desde que se genera la accesibilidad del DNA, al inicio de la transcripción de genes, durante procesos pos-transcripcionales, durante la traducción y también procesos pos-traduccionales (Farrell & Bassett 2007).

El DNA se encuentra empaquetado en el núcleo formando una estructura tridimensional llamada cromatina. Algunas conformaciones de este arreglo estructural impide que la RNA polimerasa y otras moléculas reguladoras tengan acceso para iniciar el proceso de la transcripción. Varias interacciones tienen que ocurrir antes para remodelar la cromatina y permitir el acceso de la maquinaria que da inicio a la transcripción. Entre los componentes de esta maquinaria se encuentran algunas proteínas con la capacidad de unirse a ciertas regiones del DNA y que pueden afectar la transcripción ya sea positiva o negativamente. Estas moléculas son los factores de transcripción y regulan la expresión de genes a nivel transcripcional (Farrell & Bassett 2007; Li et al. 2007).

Otro proceso que afecta a la expresión de genes es el splicing alternativo. Este proceso permite que un solo gen codifique para más de una proteína en diferentes proporciones. Esto ocurre al procesar a los transcritos de diferentes maneras. A nivel pos-

transcripcional también tenemos a los procesos de poli-adenilación y la regulación mediada por RNAs no codificantes. Entre las moléculas de RNAs no codificantes con capacidad reguladora están los RNAs largos no codificantes y los RNAs pequeños. Los RNAs largos no codificantes desempeñan diversas funciones, entre ellas, guían moléculas para generar modificaciones epigenéticas (Halley et al. 2014). Entre los RNAs pequeños están los que interactúan con proteínas Piwi. A estos RNAs se les denomina piRNAs y están relacionados con la regulación del ciclo celular y de la actividad de los transposones (Siomi et al. 2011). También están los RNAs pequeños de interferencia que tienen efecto de represión de genes a nivel pos-transcripcional. Existen otras moléculas de especial interés debido a su capacidad reguladora de un gran número de RNAs mensajeros, estas moléculas son los microRNAs. A continuación se presentarán más detalles de los microRNAs, su síntesis, mecanismos de acción, su capacidad de regular distintos fenotipos y su evolución (Kaikkonen et al. 2011).

### 2.2.1 microRNAs

Los microRNAs son pequeñas moléculas de aproximadamente 22 nucleótidos que regulan la expresión de genes a nivel pos-transcripcional. Estas moléculas interactúan con una proteína llamada Argonauta para formar un complejo de silenciamiento. Los microRNAs actúan como guías de dicho complejo hacia los transcritos blanco. La característica principal que ayuda a la represión de genes es el apareamiento Watson-Crick entre la región semilla del microRNA con sitios complementarios en las regiones 3' UTR de los RNA blanco. Esta región semilla comprende los nucleótidos 2-7 del microRNA maduro (Lewis et al. 2003).

#### 2.2.1.1 Biosíntesis y mecanismo de acción de los microRNAs

La producción de los microRNAs inicia con la síntesis de transcritos primarios (pri-miRNA) principalmente por la RNA polimerasa II (aunque en algunos casos puede ser por RNA Pol III). Después ocurre un proceso de corte por DROSHA-DGCR8 (Pasha), liberando una estructura de tallo y asa, llamada pre-miRNA como se puede ver en la Figura 6. Este pre-miRNA es exportado al citoplasma por la Exportina-5-Ran-GTP. En el citoplasma DICER, una RNAasa, con una proteína de unión al RNA (TRBP) corta al pre-miRNA para generar un microRNA maduro. Este microRNA maduro es cargado junto con Argonauta (AGO2), DICER y TRBP al complejo de silenciamiento inducido por RNA (RISC, por sus

siglas en inglés). El microRNA maduro guía a este complejo hacia RNAs mensajero blanco para silenciar la expresión mediante represión traduccional o de-adenilación (Winter et al. 2009).

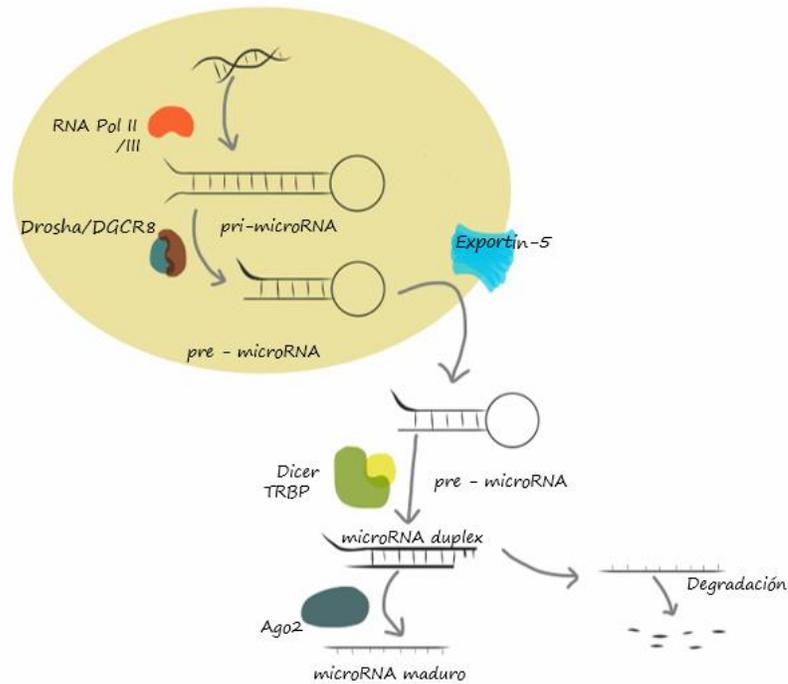


Figura 6 Ruta de síntesis de microRNAs en mamíferos (Modificado de Winter et al. 2009)

El silenciamiento de los microRNAs en mamíferos puede deberse a diferentes mecanismos. Uno de ellos consiste en la inhibición del inicio de la traducción. Esta inhibición de la traducción se realiza mediante la disociación de los elementos eIF4F a través de inhibidores traduccionales (Figura 7). Los elementos eIF4F son factores que promueven el inicio de la traducción mediante la formación del complejo de los ribosomas. El otro mecanismo es la desestabilización del RNA mensajero mediante el acortamiento de las colas poliA, reclutando deadenilasas y factores para remoción de estas moléculas (Iwakawa & Tomari 2015).

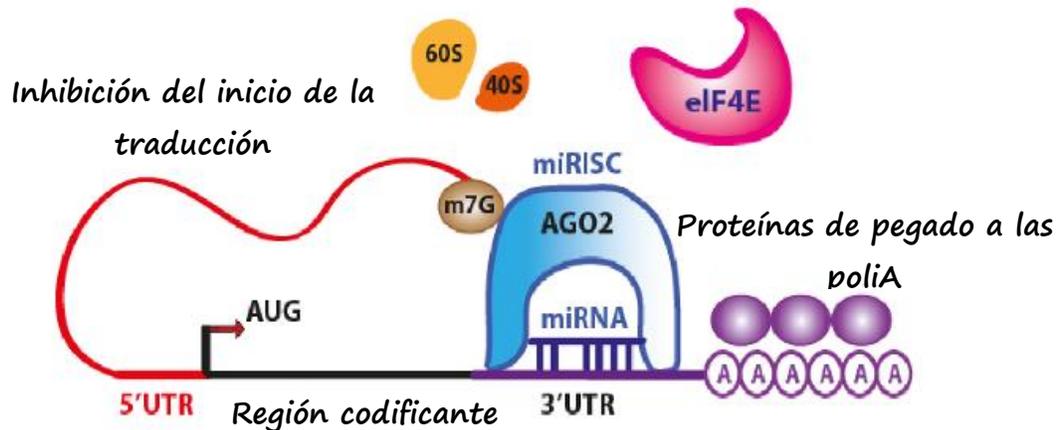


Figura 7 Diagrama representativo de uno de los mecanismos de acción de los microRNAs en animales. Este mecanismo consiste en la inhibición del inicio de la traducción mediante la disociación de los elementos eIF4E (Modificado de da Sacco & Masotti 2013) .

Aún quedan por aclarar los mecanismos en su totalidad, pero la información que se conoce en la actualidad es suficiente para permitir identificar y predecir posibles sitios de interacción en RNAs mensajeros.

#### 2.2.1.2 Identificación de blancos de microRNA

##### 2.2.1.2.1 Métodos experimentales

Existen varios métodos de identificación de blancos de microRNA. Los métodos experimentales de identificación de blancos se pueden clasificar en 3 categorías (Hausser & Zavolan 2014; Broughton & Pasquinelli 2016) .

- **Mediante screening genético.** Estos ensayos se realizaron originalmente para conocer la función de los genes blanco de microRNAs. Sin embargo, como se ha visto en experimentos de knockout y silenciamiento de los microRNAs, no todos los microRNAs son esenciales o presentan un fenotipo visible.
- **Mediante estimaciones de cambios de expresión génica.** Estas predicciones se pueden hacer mediante 2 clases de experimentos. En la primer clase de experimentos se adicionan moléculas anti-sentido a los microRNAs para silenciar a un microRNA específico. En la segunda clase de experimentos se sobreexpresan algunos microRNAs de manera específica. Después mediante

medidas transcriptómicas se observan los genes cuya expresión se ve alterada. Se consideran como blancos el grupo de genes que se alteró su expresión y que también tienen capacidad de ser blanco mediante un análisis de predicción de sitios de pegado de microRNAs.

- **Mediante entrecruzamiento e inmunoprecipitación de AGO.** Los experimentos CLIP (Cross-Linking Immunoprecipitation) y CLASH (Cross-linking, ligation and sequencing of hybrids) son ejemplos de métodos de alto rendimiento para encontrar RNAs mensajeros que interactúen con las proteínas Argonauta o con otros RNAs. Estos métodos se basan en un entrecruzamiento mediado por radiación UV. Después se realiza una inmunoprecipitación y se identifican interacciones entre la proteína Argonauta y los RNAs. Cada uno de estos métodos cuenta con diversas variantes y se complementan con datos de secuenciación para identificar los posibles blancos de microRNAs.

#### *2.2.1.2.2 Métodos computacionales*

Existen varios programas computacionales que permiten la predicción de blancos de microRNAs. Estos programas utilizan distintas fuentes de información para hacer dichas predicciones. Utilizan modelos basados en diferentes criterios como estructura secundaria, complementariedad de sitios, información de interacción con proteínas Argonauta o conservación del sitio blanco. Entre los métodos más usados se encuentran miRanda, TargetScan, DIANA micro-T, entre otros (Broughton & Pasquinelli 2016; Leclercq et al. 2017; Riffo-Campos et al. 2016).

**miRanda.** Se basa en la búsqueda de interacciones que sean estables termodinámicamente. Las versiones más recientes de miRanda incluyen algoritmos como mirSVR que utilizan información del contexto y de perfiles de expresión para mejorar las predicciones (Betel et al. 2010).

**Diana-microT.** Identifica elementos en las 3'UTR que puedan ser reconocidos por el microRNA de acuerdo a ciertas reglas de interacción, algunas basadas en la energía de pegado (Paraskevopoulou et al. 2013; Riffo-Campos et al. 2016).

**TargetScan.** Toma en cuenta diferentes criterios de información que utiliza para la predicción de blancos y que coinciden con los modelos propuestos mediante reportes experimentales. El modelo que utiliza TargetScan resume gran parte de la complejidad del fenómeno en principios básicos de la interacción del microRNA y sus blancos, además de que permite que algunos de estos criterios puedan ser evaluados por separado.

Las versiones recientes de TargetScan además de tomar en cuenta a la región semilla, el tipo de sitio de pegado y número de sitios, también toman en cuenta el contexto. Para ello TargetScan evalúa la accesibilidad estructural del sitio de pegado de los microRNAs. Esto lo hace mediante el grado de apareamiento que tienen las regiones alrededor del sitio de pegado en la 3'UTR. La accesibilidad también se ve afectada por la composición de AU y también se toma en cuenta este factor. Otros de los criterios que se toman en cuenta son la posición de los sitios en la 3'UTR, la cooperatividad que pueda haber entre sitios distintos, la longitud de las 3'UTRs y la estabilidad del apareamiento del blanco con el microRNA (Agarwal, Bell, Nam, & Bartel 2015).

Al compararse los distintos métodos antes mencionados se observa que TargetScan7 tiene mejor precisión. También TargetScan7 genera más predicciones que otros métodos utilizando niveles de precisión similares. miRanda tiene un número más grande de predicciones que TargetScan7 pero con el costo de que estas predicciones incluyen más falsos positivos (Riffo-Campos et al. 2016; Oliveira et al. 2017).

#### 2.2.1.3 Mecanismos de evolución de microRNAs

Un microRNA puede originarse en una región del genoma simplemente si tiene una adecuada estructura secundaria de horquilla de RNA que sea reconocible por Drosha y Dicer. También esta horquilla debe tener la capacidad de formar un complejo de silenciamiento. Estas propiedades hacen que la generación de genes nuevos de microRNA sea mucho más probable que la emergencia de genes que codifiquen para proteínas. Sin embargo, el problema no reside en que puedan generarse estas horquillas sino en que se transcriban adecuadamente (Chen & Rajewsky 2007).

Los genes codificantes para microRNAs pueden aparecer por duplicación génica y también pueden aparecer en intrones. Las regiones intrónicas son propensas a producir estructuras de horquilla y no requieren de un promotor propio. Se considera que una gran proporción de los microRNAs en mamíferos están dentro de regiones intrónicas. Otras regiones que proveen de estas características son los pseudogenes, snoRNAs, tRNAs y las secuencias antisentido de transcritos de microRNAs. Otra fuente de innovación de los microRNAs son los transposones. Existen reportes que establecen que muchos de los microRNAs vienen de los elementos transponibles (Rodríguez et al. 2004; Berezikov 2011).

Los genes de microRNAs también pueden generarse *de novo*. Existen cientos de miles de horquillas similares a las de los microRNAs predichas en el genoma del humano. Probablemente muchas de estas horquillas puedan ser transcritas por procesos de transcripción generalizada del genoma. La transcripción generalizada del genoma consiste en que una gran parte del genoma se transcribe. Sin embargo aún no se conoce la relevancia biológica de este proceso. Se ha propuesto que esta generación *de novo* es de manera gradual para que las horquillas puedan ser procesadas correctamente por Drosha y Dicer. Existen formas transicionales que soportan este gradualismo en la evolución de familias de microRNAs (Liu et al. 2009; Clark et al. 2011).

Las secuencias de microRNAs pueden diversificarse de distintas maneras. Las mutaciones en las secuencias del microRNA maduro, en las estructuras precursoras y en la maquinaria de edición pueden generar nuevas familias. Los microRNAs se pueden producir de manera diferente aun viniendo de la misma horquilla. Debido a la precisión con que Drosha y Dicer los procesan se pueden generar variantes con pequeñas diferencias al inicio o al final del microRNA maduro. A estas variantes de microRNAs se les llama isomiRs y son una fuente más de diversidad de los microRNAs. Otra fuente de variación viene dada por los microRNAs\*. El microRNA\* proviene de la hebra del dúplex que por lo general se degrada. Estas secuencias proveen material para diversificación ya que pueden dar origen a nuevas familias de microRNAs funcionales. Además, las modificaciones en la horquilla y la maquinaria de edición como las desaminasas de adenosina pueden afectar también la evolución del microRNA (Okamura et al. 2008).

#### 2.2.1.4 Historia evolutiva de las familias de microRNAs

Los microRNAs como clase de reguladores han estado presentes desde etapas tempranas en la evolución de los animales. Por ejemplo, mir-100 está compartido entre los eumetazoos. Alrededor de 34 microRNAs están conservados entre protostomados y deuterostomados. Esto sugiere una explosión en la innovación en la base del linaje de los bilaterales. Un gran incremento en las familias de los microRNAs también parece haber ocurrido durante la emergencia de los vertebrados y de los mamíferos placentarios (Berezikov 2011; Zhu et al. 2012).

Se ha estimado la tasa de adquisición de microRNAs durante la emergencia de los vertebrados. Esta tasa es de 0.9 a 2.7 familias por millón de años. Muchas de estas familias muestran expresión específica de tejido. Este patrón podría estar relacionado con la complejidad de los organismos. Se ha reportado que microRNAs más antiguos se expresan más altamente y extendidamente que los microRNAs más jóvenes. Además, un knockout de un microRNA más antiguo es más probable que resulte en un cambio fenotípico notable. Los microRNAs más recientes en ocasiones tienen efectos deletéreos pero son sustrato para la selección natural si generan un beneficio a la adecuación del organismo (Zhu et al. 2012).

Algunos trabajos indican que después de que un gen de microRNA emerge en un linaje particular, rara vez este microRNA se perderá. Se han reportado que 12 pérdidas de familias de microRNA han ocurrido a lo largo de la evolución de los deuterostomados. Este patrón parece indicar una tasa baja de pérdidas secundarias y una adquisición continua de nuevas familias de microRNAs (Berezikov 2011). Sin embargo, se ha propuesto que hay más pérdidas de familias de lo que se pensaba. Mientras se probaba si los microRNAs son buenos marcadores filogenéticos se encontró que los análisis anteriores no estaban hechos correctamente. Estos análisis típicamente utilizan métodos de parsimonia para estimar relaciones filogenéticas. Sin embargo, los métodos de parsimonia no pueden estimar correctamente las pérdidas de familias y llevan a un patrón en el que el número de familias de microRNAs incrementa con el tiempo. Se encontró que las pérdidas de familias de microRNAs en ocasiones no son detectadas por los métodos de observación estándar. También se encontró que las pérdidas de familias de microRNAs no son raras como se pensó anteriormente (Thomson et al. 2014).

### 2.2.1.5 Patrones en la evolución de blancos de microRNAs

Se han reportado diferentes tendencias para los blancos conservados y los no conservados:

- a) Los RNAs mensajero con **sitios conservados** tienden a estar expresados en el mismo tejido en el que se expresa su microRNA regulador. El RNA mensajero y el microRNA se co-expresan a niveles muy bajos a comparación de tejidos donde no se expresa el microRNA. Los tejidos que tienen alta concentración del microRNA y altos niveles del RNA mensajero blanco tienden a estar adyacentes unos de otros (Stark et al. 2005).
- b) Los RNAs mensajero con **sitios no conservados**, en ocasiones se expresan en tejidos con su microRNA asociado no expresado. De hecho, los RNAs mensajero co-expresados altamente con su microRNA tienen menor número de sitios no conservados. Por lo tanto, se sugiere que existen conjuntos de RNA mensajeros que evitan la regulación de microRNAs implicando cambios frecuentes en sus UTRs(Farh 2005).

Xu y colaboradores en el 2013 observaron que la proporción de sitios conservados cambia con el tiempo. Existe mucha conservación cuando se estudian los primates y muy poca cuando se comparan especies distantes. Otros trabajos como el de Simkin *et al.* en 2014 sugieren esta alta conservación en primates en el caso de familias de microRNA conservadas. Otros reportes también mencionan que los microRNAs de humano y sus secuencias blanco en las 3' UTRs evolucionan bajo selección purificadora (Loh *et al.* 2010). Las observaciones mencionadas anteriormente se condensan en la Tabla 2.

*Tabla 2 Patrones observados en la evolución de familias y de blancos de microRNAs*

<b>Molécula</b>	<b>Patrón Global</b>	<b>Posible causa</b>	<b>Referencia</b>
<b>Familias / Sitios blanco</b>	Familias conservadas con blancos conservados	Selección purificadora	Simkin <i>et al.</i> 2014
<b>Familias / Sitios blanco</b>	Familias conservadas con blancos no	Potencial de selección positiva	Chen & Rajewsky 2007; Gallego <i>et al.</i> 2016

	conservados		
<b>Familias de microRNA</b>	Exceso de ganancia	Selección positiva, evolución neutral	Loh <i>et al.</i> 2011
<b>Familias de microRNA</b>	Exceso de conservación	Selección purificadora, menos variación en la expresión	Simkin <i>et al.</i> 2014
<b>Sitios blanco</b>	Exceso de ganancia	1.- Incremento en la canalización, mayor robustez y menos variación 2.- Evolución neutral debido a poca presión de selección	Peterson <i>et al.</i> 2009
<b>Sitios blanco</b>	Exceso de conservación	Selección purificadora	Simkin <i>et al.</i> 2014

En el 2007, Chen y Rajewsky hicieron un modelo de control transcripcional de nuevos microRNAs. Ellos proponen que los microRNAs recientes se expresan a bajos niveles en tejidos específicos para limitar los efectos deletéreos y gradualmente los blancos deletéreos se van eliminando del transcriptoma mediante selección natural. Después de esto, el rango espacial y el nivel de expresión de los microRNAs podría aumentar. Esto se basa en experimentos que muestran que nuevos microRNAs si se sobreexpresan podrían ser deletéreos en el caso de *Drosophila*. Sin embargo en 2016, Nozawa y colaboradores, haciendo un estudio más detallado considerando la coexpresión de regulador y blanco en tejidos específicos, encontraron el patrón contrario sugiriendo un modelo donde el número de blancos incrementa a lo largo de la evolución (Figura 8).



individuales sean difíciles de ver, a nivel de red, los microRNAs son muy importantes. La no adecuada expresión de los microRNAs o mutaciones en los blancos podrían producir enfermedades, anomalías o cambios morfológicos visibles (Loh *et al.* 2010).

Existe una correlación entre expansiones de familias de microRNAs con un incremento en la complejidad morfológica (Figura 9). Esto sugiere que probablemente la adquisición de nuevas familias de microRNAs está relacionada con la emergencia de organismos complejos (Heimberg *et al.* 2008; Zhu *et al.* 2012).

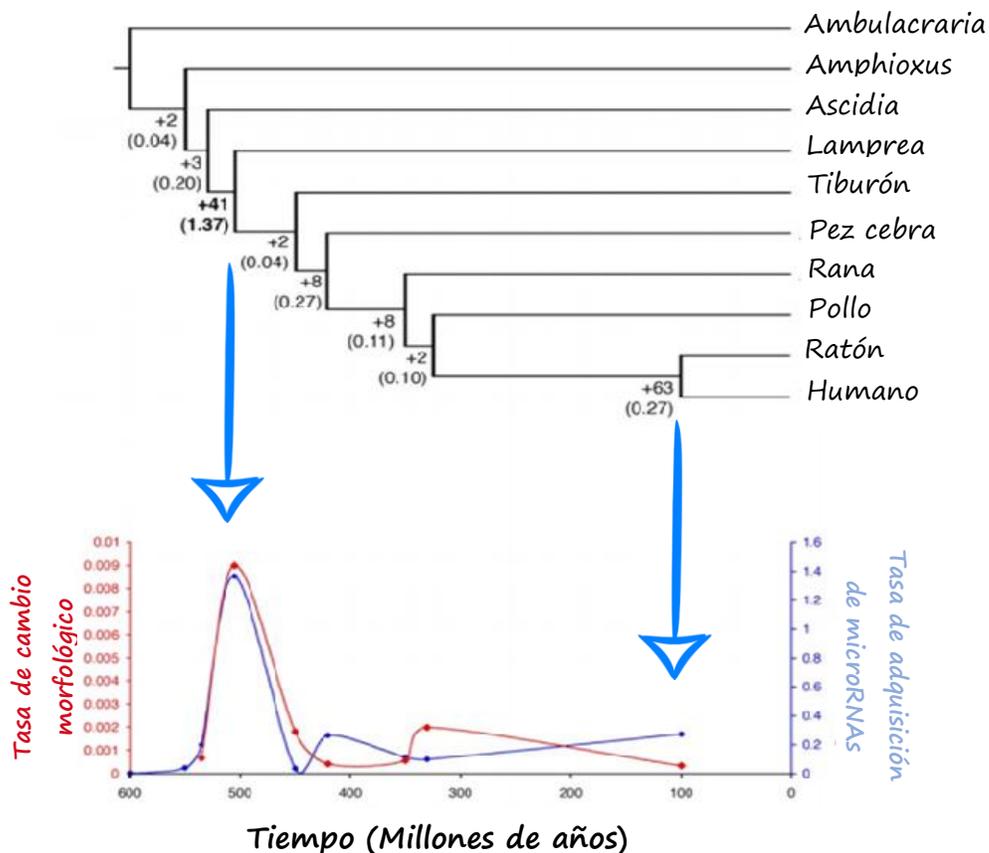


Figura 9 La evolución de familias de microRNAs y de complejidad morfológica a lo largo del tiempo (Modificado de Heimberg *et al.* 2008)

Los patrones evolutivos sugieren que: i) la variación fenotípica intraspecífica disminuye a lo largo del tiempo geológico para diferentes géneros y familias de trilobites, ii) los microRNAs disminuyen la variabilidad en expresión génica, y iii) el número de microRNAs

dentro de cada linaje aumenta con el tiempo. Al analizar distintos niveles taxonómicos como organismos triploblásticos, vertebrados y primates se observaron grandes incrementos de complejidad morfológica y de adición de familias de microRNAs. Esto lleva a pensar que los microRNAs pueden canalizar el desarrollo de una manera que la variación fenotípica disminuye con el tiempo por el incremento de precisión de la expresión. Este incremento en la precisión de la expresión hace que aumente la heredabilidad de un rasgo fenotípico. Al aumentar la heredabilidad de un rasgo fenotípico se genera material para que actúe la selección natural y se puedan generar estructuras más complejas morfológicamente. En resumen, los microRNAs estabilizan la expresión haciendo que los rasgos relacionados sean más capaces de ser seleccionados (Peterson et al. 2009).

### 2.3 ¿Cómo evolucionan las redes de regulación? Rasgos Generales

Las redes de regulación genética no son del todo flexibles. Algunos sub-circuitos se encuentran más conservados que otras regiones de la red, debido a las repercusiones que puedan tener mutaciones en ellos. Esta conservación ocurre en los circuitos genéticos relacionados con el desarrollo (Peter y Davidson 2011).

Desde una perspectiva de Evolución del Desarrollo se establece que los nuevos fenotipos emergen debido a variaciones ancestrales. Estas variaciones surgen mediante diferentes combinaciones y la reutilización de módulos implicados en el desarrollo. Esto ocurre bajo conservación y homología profunda de caracteres, causando complejidad fenotípica. La evolución del desarrollo, ha confirmado que los sistemas de desarrollo de los animales tienen una variación genética sesgada que restringe los efectos de la selección natural y por lo tanto a la evolución fenotípica (Moczek et al. 2015). Evidencia de esto, son los paralelismos evolutivos, es decir especies diferentes que comparten un mismo rasgo fenotípico debido a que tienen variaciones genéticas similares. Por ejemplo, las mutaciones en el gen MC1R ocasionan cambios en la pigmentación y coloración de varios grupos taxonómicos como peces y mamíferos. Otro caso son los sistemas donde el desarrollo de una parte del cuerpo está íntimamente relacionado con el desarrollo de otra parte del cuerpo. Por lo tanto, las variaciones que afecten a una sección también afectarán a la otra, imponiendo mayor cantidad de restricciones para que ambos cambios puedan ocurrir (Brakefield 2006; Peter & Davidson 2011).

### 2.3.1 Recableado de redes de regulación

La evolución de las redes de regulación se puede estudiar enfocándose en la cantidad y la tasa de eventos de recableado que ocurren en diferentes linajes. Estos eventos de recableado se han definido como cambios en la conectividad de las redes a lo largo de la evolución de los organismos. El recableado se puede evaluar de distintas maneras. Se puede abordar de manera global comparando las redes de regulación completas o también estudiando el comportamiento de grupos de genes y de reguladores específicos que han estado bajo diferentes regímenes de cambios en su regulación.

#### 2.3.1.1 Estimadores de la tasa de recableado de las redes de regulación

Algunos autores han propuesto medidas para la evolución de redes en Biología. El tipo de medidas de tasas de recableado de las redes está limitado por la distancia evolutiva, tiempo y el número de cambios en las interacciones (Shou et al. 2011). Durante este proyecto se documentaron las distintas medidas propuestas en la literatura para estimar el cambio y evolución en redes de regulación. Algunas de estas medidas de recableado se presentan a continuación.

##### 2.3.1.1.1 Métrica de Shou

La métrica propuesta por Shou y colaboradores en 2011 representa el número total de conexiones que se recablearon. Esta cantidad es normalizada por el número de conexiones posibles y el tiempo de divergencia. Para calcular la tasa de recableado se comparan las redes de las especies usando una como referencia. La estimación se realiza identificando los nodos comunes y conexiones comunes que tengan las 2 redes comparadas y los nodos que se pierdan y se ganen.

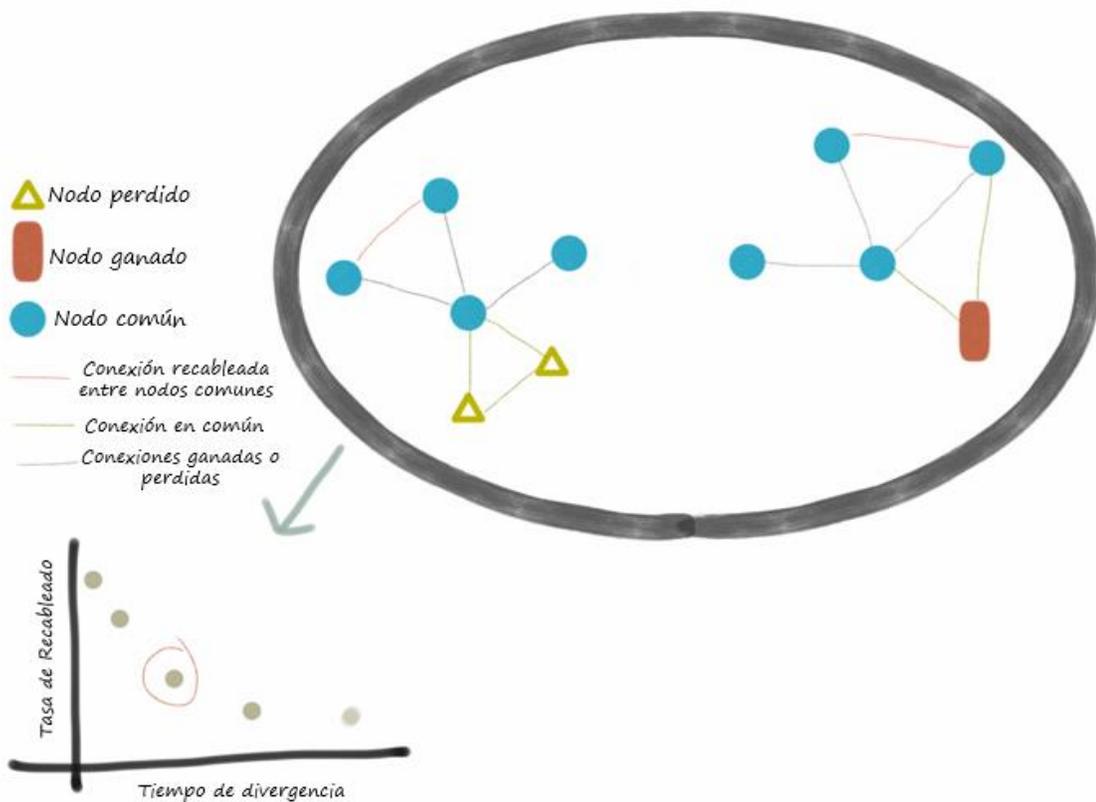


Figura 10 Recableado de red (Modificado de Shou et al. 2011)

### 2.3.1.1.2 Métrica de Simkin

Esta métrica consiste en estimar la tasa de ganancia y pérdida de sitios de pegado de familias de microRNAs. Las estimaciones se hacen para cada familia de microRNA y se contrastan con un control de secuencias con composición de nucleótidos similar. Estos valores de ganancia y pérdida de sitios para cada microRNA, se normalizan por la frecuencia de sitios de cada uno de los microRNAs que estén presentes en el ancestro. Es decir, aquellos que son más abundantes es más probable que tengan un evento de pérdida o de ganancia.

Las comparaciones, a diferencia de la métrica de Shou, se hacen con respecto a la red del ancestro. Esta red del ancestro se genera mediante un método de inferencia ancestral

de las secuencias de las 3'UTR de los organismos a estudiar. Otra diferencia es que la métrica de Shou cuenta eventos de recableado considerando una conexión máxima por gen con su regulador, y la métrica de Simkin considera todos los sitios de pegado posibles pudiendo ser más de uno entre un blanco y un determinado regulador (Simkin *et al.* 2014).

### 2.3.1.1.3 Métrica para medidas probabilísticas de elementos recableados, para valores continuos

En casos donde tenemos valores de recableado que no sean cantidades discretas existe una métrica que se ha aplicado para medir el recableado en levaduras (Sarda & Hannenhalli 2015). En este trabajo se utiliza una medida llamada Puntaje de Recableado, que es una función que indica qué tan probable es que un gen haya cambiado a su regulador por otro.  $P(\text{TF}, \text{gen}, \text{Especie})$  representa a la probabilidad con que un factor de transcripción se une a un promotor en una especie. Este factor de transcripción se representa en la Figura 11 por X ó Y. En la Figura 11 se observa que las especies S1, S2, S3 y S4 pudieron tener recableado donde otro factor de transcripción regula a otro gen en especies diferentes. Este método se utiliza para comparar 2 grupos de organismos que posean características diferentes para conocer los eventos de recableado relacionados con estos cambios fenotípicos.

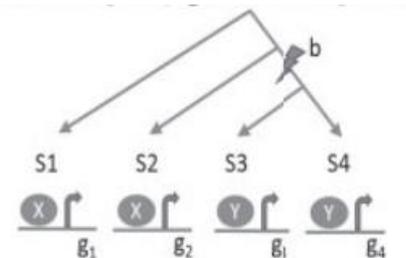


Figura 11 Diagrama representativo para evaluar recableado (Sarda & Hannenhalli 2015)

### 2.3.1.2 Pruebas estadísticas para el recableado

La información acerca de las redes se genera mediante experimentos o predicciones computacionales. Las estimaciones de las redes pueden estar sujetas a muchos falsos positivos y falsos negativos. Existe un método general para verificar la influencia de falsos positivos y negativos en la estimación del recableado de redes (Beltrao & Serrano 2007). Para ello se realiza un análisis de sensibilidad basado en muestreo. Con este análisis se evalúa la robustez de las métricas de recableado de redes de interacción de proteínas. El efecto se observa mediante muestreo al azar, añadiendo y removiendo fracciones de interacciones aleatoriamente entre las 2 redes. Así se generan redes simuladas y se estima la tasa de recableado. Se usan además fracciones permitidas de falsos positivos y negativos en diferentes proporciones para evaluar su efecto. Este método también fue usado por Shou *et al.* en 2011 para evaluar sus estimaciones de recableado de distintas redes biológicas.

Otra manera de generar soporte estadístico a las estimaciones de recableado es la generación de grupos de secuencias control. Este grupo control permitirá contrastar el comportamiento de secuencias correspondientes a elementos de regulación funcionales con secuencias que probablemente estén evolucionando de manera neutral (Friedman *et al.* 2009; Simkin *et al.* 2014).

### 3.3.2 Patrones evolutivos de redes de regulación

#### 2.3.2.1 Saturación de cambios

El proceso de saturación inicialmente fue observado en secuencias de proteínas. Mientras Emile Zuckerkandl y Linus Pauling comparaban secuencias proteicas, observaron una tasa lineal de acumulación de cambios de aminoácidos (Zuckerkandl & Pauling 1962; Zuckerkandl & Pauling 1965). Después de esto, Kimura y Ohta propusieron de que una gran proporción de cambios de aminoácidos podrían ser neutrales siendo estos los que reflejen esta tasa lineal (Kimura 1968; Kimura & Ohta 1971). En escalas más grandes de tiempo es cuando surge el problema ya que se llega a un punto donde no se ven más cambios o diferencias. A este fenómeno se le llamó efecto de saturación de sustituciones (Gojobori 1983).

De manera similar a las mutaciones en las secuencias de DNA, las redes están bajo un proceso de saturación que, a cierta distancia, no permite ver el número real de cambios. Esto genera dificultades para saber cómo las redes evolucionan a lo largo de eventos de

especiación. En el trabajo de Shou *et al.* del 2011 para las redes de regulación de microRNA con sus blancos realizaron comparaciones entre *H. sapiens*, *M. musculus*, *D. rerio*, *C. elegans* y *D. melanogaster*. Sin embargo, en ese momento eran muy pocas las redes disponibles para especies cercanas y el método de detección de blancos que ellos tenían era muy propenso a tener errores. En este mismo trabajo especies que tenían un tiempo de divergencia similar tendían a tener tasas de recableado similares indicando que la métrica de la tasa de recableado era dependiente del tiempo de divergencia pero no del par de especies comparadas. A distancias evolutivas grandes, la tasa de recableado se encuentra bajo saturación y por lo tanto es difícil de comparar (Shou et al. 2011).

#### 2.3.2.2 Recableado bajo presiones selectivas

A lo largo de la historia de la vida existen patrones causados por extinciones en masa seguidas de rápidas radiaciones evolutivas. Muchos paleontólogos sugieren que la discrepancia entre el reloj molecular y los fechados paleontológicos son debido a relojes moleculares imprecisos. Por ejemplo el proceso de radiaciones adaptativas podría acelerar el reloj molecular. La capacidad de interacción de un regulador con múltiples sitios blanco, permite que el recableado de las redes de regulación sea un mecanismo importante de adaptación visto desde un nivel más allá de proporción de cambios en las secuencias (Bromham 2003).

#### 2.3.3 Ejemplos de recableado en la naturaleza

Algunos análisis sugieren al recableado como una explicación de la divergencia en los patrones de expresión, sin embargo, pocos esfuerzos se han enfocado en las propiedades del recableado de la red y la manera causal de cómo evolucionan las redes en la naturaleza.

##### 2.3.3.1 microRNAs

Se han hecho estimaciones de recableado a nivel general en diferentes animales. A continuación se mencionan algunos de los trabajos que evalúan la evolución de redes de regulación de microRNAs en distintos linajes de organismos. En la siguiente sección también se hablará de otros ejemplos de organismos cuyos microRNAs se han estudiado con el fin de entender su influencia en la evolución fenotípica.

#### 2.3.3.1.1 Primates

En primates se ha observado que para el conjunto de familias de microRNA más conservadas, sus sitios de pegado parecen tener una tasa baja de cambio. Lo que implica que los sitios blanco también están conservados comparados con otras secuencias del genoma (Simkin *et al.* 2014). También en otro trabajo se ha observado que dentro de los primates hay menos cambios comparados con otros grupos (Xu *et al.* 2013).

Se han reportado varias expansiones de familias de microRNAs en el grupo de los primates. Específicamente en el grupo Catharrini, donde se encuentran los grandes simios, ocurrió una expansión de familias de microRNAs (Hertel & Stadler 2015). Además de que se han encontrado expansiones específicas de familias ya existentes como lo son miR-301, miR-130 entre otras (Guerra-assunção & Enright 2012).

#### 2.3.3.2 Factores de Transcripción

La literatura para el recableado en redes de factores de transcripción es mucho más extensa. A continuación se presentarán algunos ejemplos de recableado de factores de transcripción en varios grupos de organismos.

##### 2.3.3.2.1 Levaduras

Un caso específico de recableado es en la reproducción de levaduras. La reproducción de *Saccharomyces cerevisiae* está regulada mediante la respuesta a una feromona. A su vez, el factor de transcripción Ste12 regula esta respuesta. Lo interesante es que este mecanismo de respuesta por parte de Ste12 está conservado en levaduras mucho más lejanas como *Candida albicans*. El tiempo de divergencia es de casi 200 millones de años de divergencia entre estas especies. Sin embargo, la red que controla la respuesta de la feromona, evolucionó de dos maneras diferentes para unirse a los genes blanco. Mediante la reconstrucción de varios intermediarios evolutivos se demostró que uno de los sitios de pegado no era accesible para el ancestro, como se puede ver en la Figura 12. En este caso se sugiere que elementos de regulación de estos organismos estuvieron bajo eventos de recableado que solo se pueden entender a nivel de red (Sorrells *et al.* 2015).

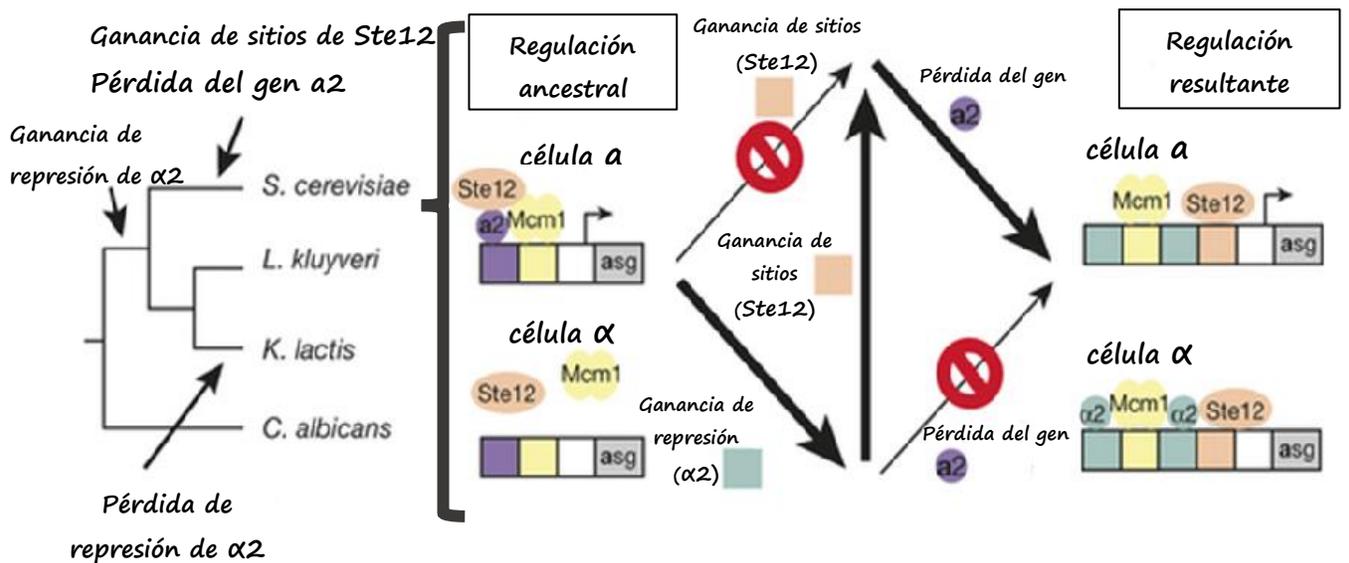


Figura 12 Evolución de diferentes maneras de regulación relacionadas con la reproducción en levaduras, donde ganancia de la represión “ $\alpha 2$ ” permitió ganar sitios de regulación “Ste12” (Modificado de Sorrells et al. 2015)

En otro de los estudios de recableado realizado en levaduras, se reporta la cantidad de recableado que ha ocurrido en diferentes grupos de la filogenia de 23 levaduras (Sarda & Hannenhalli 2015). Esta cantidad de recableado se estimó mediante un modelo que toma en cuenta probabilidades de sitios de pegado. Se seleccionaron grupos de diferentes levaduras para comparar los eventos de recableado. Esta selección se realizó con base al estilo de vida, ya sea i) si estuvieron bajo un proceso de duplicación genómica, o ii) si son domesticadas o no. En este trabajo se encontró que algunos linajes presentan más recableado que otros, especialmente el grupo de levaduras relacionadas con *Saccharomyces cerevisiae* comparado con el grupo de *Candida albicans*. Parece ser que

parte de este alto recableado se debe a cambios importantes en los sitios de pegado para la regulación de genes ribosomales y del metabolismo de azúcares.

#### 2.3.3.2.2 *Erizos de mar*

En un trabajo del grupo de Eric Davidson se enfocaron en la divergencia de 2 erizos de mar. Se estudió específicamente el proceso de esqueletogénesis donde se hizo un análisis funcional dirigido hacia una especie de erizo (*Eucidaris tribuloides*). Este análisis se hizo para demostrar que había diferencias de regulación genética con respecto al erizo que es modelo de laboratorio (*Strongylocentrotus purpuratus*). Se sugiere que hubo una reorganización de los circuitos de regulación implicados en la esqueletogénesis. También se sugiere que los elementos importantes para este fenómeno fueron el surgimiento de un nuevo factor de transcripción y cambios en los blancos regulados (Erkenbrack & Davidson 2015).

#### 2.3.4 Contribución de la evolución de redes de regulación de microRNAs en la evolución fenotípica

Algunos patrones sugieren que los microRNAs pueden jugar un papel importante en la evolución morfológica. Uno de los principales ejemplos para evaluar la contribución de algún elemento de regulación genética en la evolución fenotípica, es el fenómeno de las radiaciones adaptativas. Lo que permite describir la evolución y comportamiento de las redes de regulación bajo diferentes presiones de selección y etapas de cambio morfológico.

##### 2.3.4.1 Radiaciones adaptativas

Las radiaciones adaptativas ocurren cuando la selección natural propicia la formación de nuevas especies las cuales tienen la habilidad de explotar diferentes nichos ecológicos. Las definiciones modernas tienen 3 características principales: 1) diversificación de especies provenientes de un ancestro común, 2) generación de rasgos adaptativos por selección natural, y 3) diversificación extraordinaria por el tiempo ocurrido o una diversidad no común relativa a clados cercanos. Sin embargo esta última característica sigue en discusión y no se considera estrictamente necesaria para una radiación adaptativa (Glor 2010).

#### 2.3.4.1.1 Peces espinosos

Los peces espinosos (*Gasterosteus aculeatus*) son un modelo importante de especiación ecológica. Sus recientes eventos de cambio morfológico y especiación los convierte en candidatos ideales para estudiar la variación fenotípica relacionada con la interacción del ambiente y el material genético (Jones et al. 2012; Leder et al. 2015).

Los peces espinosos son un buen modelo para responder preguntas sobre cómo los elementos de regulación cambian a un nivel global. Respecto a los microRNAs existe un trabajo que se enfoca en la predicción de familias de microRNA y sus posibles blancos (Chaturvedi et al. 2014). Se estudió cómo los SNPs identificados en estas especies podrían estar presentes en algunos de los blancos predichos. Estos SNPs presentes en los blancos de los microRNAs podrían estar relacionados con la rápida evolución de estos peces. Sin embargo respecto a la contribución de los microRNAs en estas especies solo se encontró un polimorfismo asociado a sus sitios de regulación.

#### 2.3.4.1.2 Cíclidos

La evolución de los cíclidos es otro ejemplo de radiación adaptativa (Figura 13). Estos peces han sido estudiados en aspectos ecológicos, filogeográficos y de la genética clásica. Para los cíclidos del Lago Malawi en África se ha reportado que este evento de radiación adaptativa estuvo relacionado con cambios de regulación de la expresión. Loh y colaboradores en 2011 sugirieron que estos cambios muy probablemente tuvieron que ver con la evolución de los microRNAs y selección positiva en los sitios blanco de los microRNAs. Estudios más recientes que incluyen los ensamblajes completos de los genomas de los cíclidos pero con grupos más lejanos filogenéticamente han reportado evidencia de selección purificadora en este grupo de peces. Quedaría por confirmar si esta selección purificadora también se ve con los mismos cíclidos estudiados por Loh y colaboradores utilizando ensamblajes completos de sus genomas (Loh et al. 2011; Brawand et al. 2014).

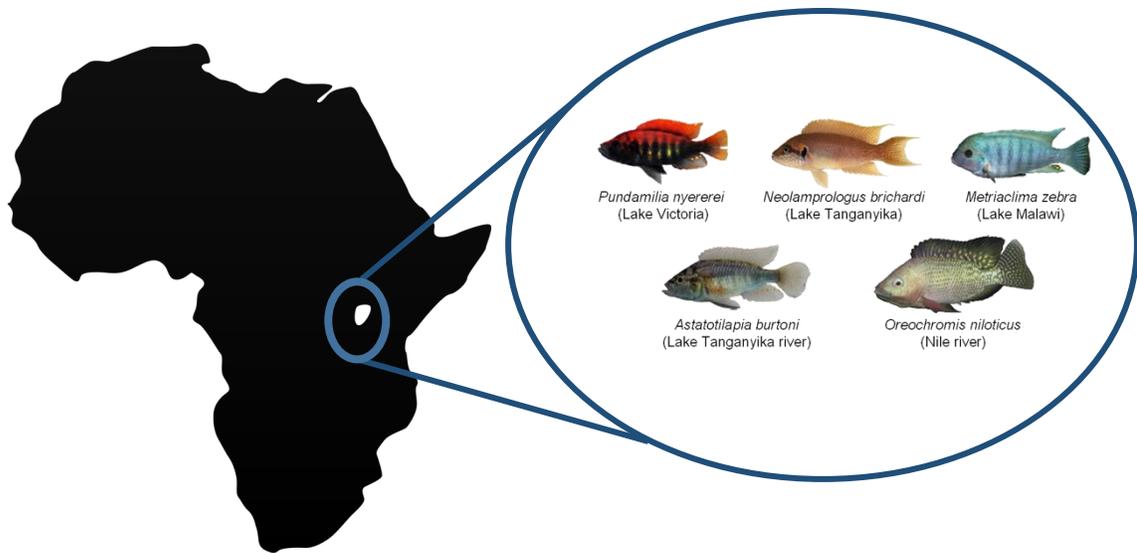


Figura 13 Radiación adaptativa de los cíclidos de África (Modificado de Brawand et al. 2014)

Existe otro grupo de cíclidos también muy estudiado que corresponde a los cíclidos Rey Midas de Nicaragua. Respecto a los microRNAs y sus blancos se han estudiado 6 especies de cíclidos Rey Midas. Se encontró que la mayoría de cíclidos presentan también selección purificadora en los sitios de pegado de microRNAs. Sin embargo para una especie se encontró que la selección está relajada. Esto puede deberse a que este cíclido habita en el lago del cráter donde la especiación simpátrica se lleva a cabo. Los eventos de especiación simpátrica están relacionados con una alta tasa de cambio morfológico. Una posible explicación es que al relajarse la presión de selección purificadora se permite a una población tener más variación en la regulación. Al haber mayor variación en la regulación genética se propiciaría mayor variación morfológica. La selección natural actuaría sobre esta variación propiciando los eventos de especiación (Franchini et al. 2016).

#### 2.3.4.2 Otros grupos de organismos con información genómica disponible

En mamíferos también se han observado expansiones de familias de microRNAs. Estas expansiones se han visto sobre todo en la radiación de los mamíferos placentarios. También dentro de los mamíferos placentarios, en el grupo de los Catharrini que incluye a los grandes simios, hubo otra expansión de familias de microRNAs. Se ha reportado que para las familias más jóvenes de microRNAs en mamíferos los genes blanco también tienen menos conservación en las secuencias de sus sitios de pegado. En cambio, para

los microRNAs más antiguos se observa que los sitios de pegado están más conservados. Con la disponibilidad actual de genomas y sus anotaciones es factible realizar estimaciones de evolución de las redes de regulación de varios grupos de mamíferos (Meunier et al. 2013; Hertel & Stadler 2015).

## 2.4 Evolución de los primates

El orden de los primates consiste en el grupo de mamíferos que se originó aproximadamente hace 65 millones de años partiendo de un ancestro similar a una musaraña. Los primates tienen rasgos anatómicos que los distinguen de otros mamíferos. Estos rasgos incluyen modificaciones en estructuras del cráneo debido al incremento del tamaño del cerebro, un hocico reducido debido a una menor dependencia del olfato, disminución del tamaño de las mandíbulas debido a cambios en la dieta y mayor agudeza visual que modificó las cavidades oculares. Existen otros rasgos característicos como la presencia en ocasiones de un pulgar oponible y uñas en lugar de garras (Willoughby & Willoughby 2005).

Dentro del orden de los primates se encuentra la superfamilia de los Homínidos. En este grupo se encuentran los grandes simios, que incluyen al humano, y los simios menores como el gibbon (Figura 14). Este grupo se caracteriza principalmente por el desarrollo de rasgos fenotípicos relacionados con la evolución del hombre. La evolución del tamaño del cerebro en primates es uno de los fenotipos que más ha despertado interés. La capacidad volumétrica del craneo humano es mucho mayor que la de los otros homínidos. El humano cuenta con aproximadamente 1200 cm<sup>3</sup>, y le siguen el gorila y el chimpancé con 400 cm<sup>3</sup> y 300 cm<sup>3</sup> respectivamente. También la evolución del bipedismo, la capacidad de hablar, la respuesta del sistema inmunológico y la capacidad de fabricar herramientas son fenotipos que han sido estudiados para entender la evolución humana (Willoughby & Willoughby 2005; Singleton 2013).

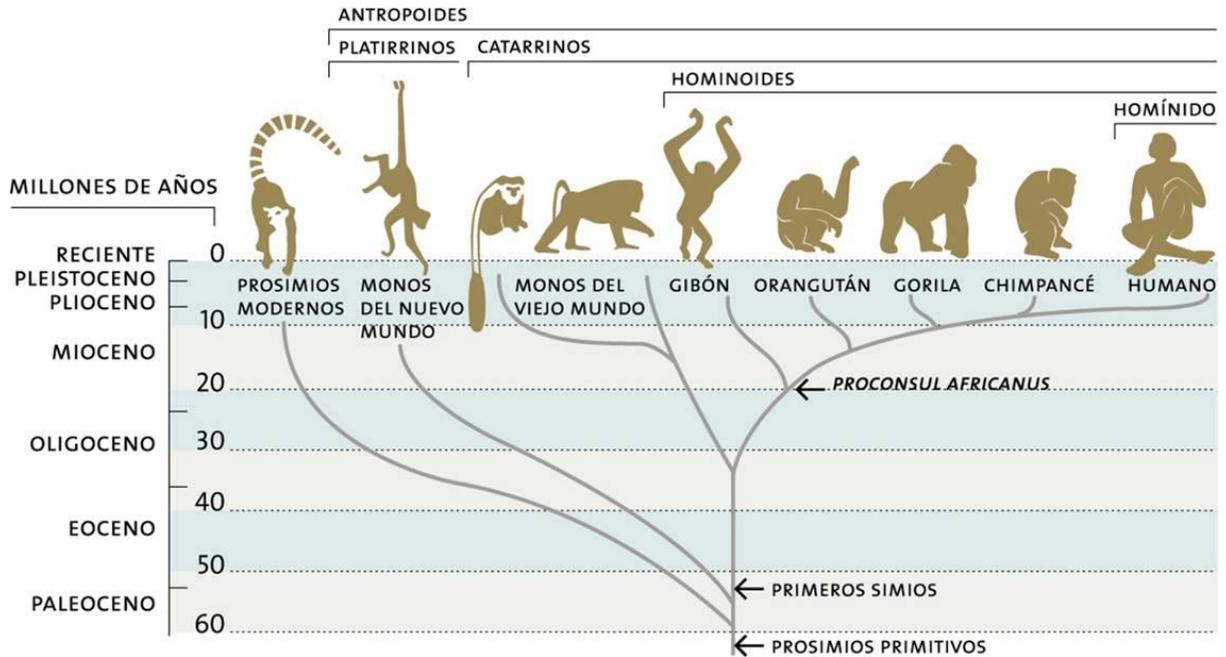


Figura 14 Filogenia de los primates, donde se señala al grupo de los homínidos que incluyen a los primates que sus genomas se analizaron en este proyecto (Levin 2011)

Otro tipo de evidencia que nos ayuda a entender la evolución humana es el registro fósil. Sobre todo se han encontrado fósiles del grupo de los homínidos. Los homínidos son el grupo hermano de los chimpancés de los cuales el humano es el único sobreviviente. Con base al registro fósil se dice que hubo muchos eventos de cambio morfológico en este grupo. Estos eventos implicaron etapas de gran cambio donde varios rasgos anatómicos surgieron de manera independiente en los diferentes linajes que derivaron de esta radiación, como se puede ver en la Figura 16. Los principales rasgos anatómicos que cambiaron en este grupo son el tamaño del cerebro y el bipedismo (Carroll 2003).

La superfamilia de los Homínidos se encuentra bien documentada a nivel genómico, con la mayoría de los genomas con un nivel de ensamblaje y anotación bueno. Además, entre los genomas de los grandes simios hay una relación de homología aproximadamente de 1:1. También entre las regiones alineables existe una gran similitud; por ejemplo, la divergencia de secuencias genómicas alineables entre humano y chimpancé es aproximadamente 1.4%. Estas características lo hacen un grupo adecuado para resolver preguntas acerca de la evolución de sus secuencias genómicas y rasgos fenotípicos (Khaitovich et al. 2006; Lewis et al. 2006).

La información de los primates a nivel molecular en la actualidad es uno de los recursos más utilizados para entender la evolución del hombre. Existe una gran cantidad de información acerca de los genes y elementos reguladores relacionados con los cambios fenotípicos que ocurrieron a lo largo del tiempo durante la evolución humana. Mediante la comparación de diferentes genomas de primates se han encontrado familias de genes con distintos niveles de selección natural. A nivel de las regiones codificantes de proteínas se encuentran como positivamente seleccionados algunos genes codificantes para factores de transcripción. Por otro lado, se encuentran negativamente seleccionados genes codificantes para proteínas del citoesqueleto (Bustamante et al. 2005).

En las regiones no codificantes también se han identificado sitios con diferentes niveles de selección. Las regiones promotoras de los genes involucrados en procesos del sistema nervioso y metabolismo muestran evidencia de selección positiva. Otra de las características que más varía entre los genomas de los primates es la inserción de elementos retrotransponibles. La tasa de acumulación de estos elementos difiere mucho en humano, chimpancé y orangután con un estimado de 5,000, 2,300 y 500 inserciones respectivamente (Haygood et al. 2007; Locke et al. 2011).

También se han comparado los niveles de expresión de genes y se han encontrado categorías funcionales de genes que están diferencialmente expresados entre primates en determinados tejidos. Por ejemplo, al comparar tejidos de corteza frontal entre chimpancé y humano se encontraron diferencialmente expresados genes relacionados con transmisión sináptica, transporte, fosforilación oxidativa y metabolismo de lípidos. Además, estos análisis se han complementado con la evidencia de selección positiva de estudios como el de Haygood et al. del 2007. Los genes con evidencia de selección positiva y expresión diferencial están enriquecidos en funciones como transporte en general y cadena de transporte de electrones (Babbitt et al. 2010).

Se han hecho análisis de expresión para otros tejidos entre diferentes especies de primates. De manera general se ha encontrado expresión diferencial en genes relacionados con procesos del sistema nervioso, señalización, metabolismo y transporte iónico. Estos patrones se han reportado en diferentes estudios donde comparan la expresión de tejidos principalmente del cerebro. Sin embargo también estos resultados se

presentan cuando se incluyen tejidos de hígado, tejido adiposo, tejido musculoesquelético y de cerebelo en las comparaciones (Bauernfeind et al. 2015; Babbitt et al. 2017).

La evolución de rasgos morfológicos en el humano está relacionada con múltiples tipos de cambios en la regulación de genes que afectan al desarrollo. Un ejemplo es la evolución del bipedismo, de la cual se han propuesto varios mecanismos por los que pudo haber surgido. Uno de ellos implica el cambio de regulación en el gen GDF6. Anteriormente se había reportado que este gen tiene efectos en el tamaño de las placas óseas de los peces espinosos. Ahora se sabe que existe una regulación única de este gen en humano a diferencia del chimpancé y otros mamíferos (Figura 15) . Se cree que este cambio en regulación está relacionado con la evolución del bipedismo ya que el gen se expresa en regiones anatómicas que se modificaron de manera particular durante esta transición (Indjeian et al. 2016).

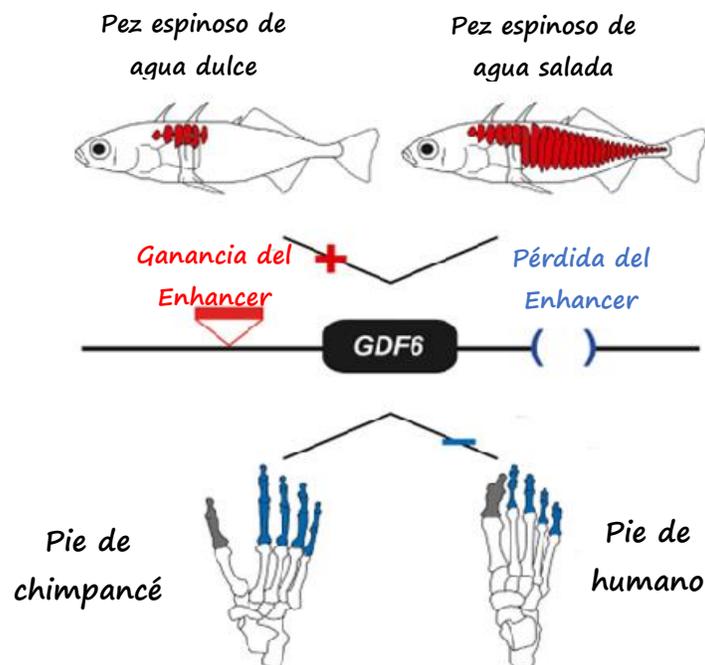


Figura 15 Cambio en la regulación de GDF6 relacionado con diferentes fenotipos

Un ejemplo de un gen que se sugiere que evolucionó bajo selección positiva es FOXP2. Este gen codifica para un factor de transcripción relacionado con la evolución del habla. Algunas personas con mutaciones en este gen presentan problemas del habla. Se ha observado que los blancos de FOXP2 tienen señales de selección positiva. Otro ejemplo es el caso de EDAR. Este gen también tiene señales de selección positiva. Mediante mutantes en ratón se encontró que EDAR está relacionado con la densidad de glándulas sudoríparas. Este último fenotipo se comparte con las variantes de este gen en diferentes poblaciones de humanos en China (Lai et al. 2001; Ayub et al. 2013; Kamberov et al. 2013)

## 2.5 Preguntas para abordar mediante el estudio de la evolución de las redes de regulación genética

En el siguiente cuadro se presenta el tipo de preguntas que se pueden responder mediante el estudio de la evolución de redes de regulación genética.

### **¿Cuáles son las bases moleculares de la evolución fenotípica?**

Con la información que se dispone en la actualidad se puede contestar parcialmente la pregunta anterior. Sin embargo se pueden ir contestando preguntas específicas para tener un mejor entendimiento de los mecanismos de evolución de las redes de regulación.

La teoría de la dinámica evolutiva de redes de regulación genética tiene todavía varias preguntas que deben responderse. A continuación se plantean algunas preguntas que faltan por responder con base en la revisión de literatura científica.

- **1.- ¿Las redes de regulación evolucionan más rápido que otro tipo de redes biológicas (Redes de interacción proteína-proteína, Redes metabólicas entre otras)?**
- **2.- Si se toma en cuenta la gran cantidad de ruido, etapas de desarrollo, expresión tejido específico y diferencias entre especies, ¿cómo se pueden explicar las diferencias en los perfiles de expresión entre especies?**
- **3.- ¿Existen secciones de las redes de regulación genética que son más**

**variables que otras?**

- **4.- ¿Qué factores afectan la dinámica evolutiva de las redes de regulación genética?**

En este presente proyecto se estudió la evolución de redes de regulación de microRNAs en primates y su relación con la evolución del fenotipo. Sin embargo, también se abordaron algunas de las preguntas aquí propuestas.

## 3. Objetivos

### 3.1 Objetivo general

Determinar las tasas de recableado de las redes de regulación de microRNAs en primates y las funciones de los genes asociadas a estos cambios.

### 3.2 Objetivos específicos

- Comparar las métricas de recableado propuestas por Simkin y Shou.
- Determinar las implicaciones de las consideraciones de cada métrica para evaluar la tasa de recableado.
- Establecer controles para la estimación de cambios en las redes de regulación de microRNAs.
- Construir las redes de regulación de microRNAs en primates.
- Comparar las redes de regulación de microRNAs en primates.
- Estimar las tasas de cambio para las redes de regulación de microRNAs.
- Identificar familias de microRNAs con valores extremos de recableado.
- Identificar blancos de microRNAs con valores extremos de recableado.
- Identificar categorías funcionales enriquecidas de dichos blancos con valores extremos.
- Comparar categorías funcionales enriquecidas entre especies.
- Aplicar el método desarrollado en otras situaciones biológicas relevantes.

## 4 Hipótesis

Los cambios ocurridos en el sistema nervioso, estructura del cráneo y regulación diferencial de genes asociados con estos fenotipos dentro del linaje de los Homínidos fueron ocasionados por el recableado de las redes de regulación de los microRNAs que ocurrió en este grupo.

## 5 Justificación

El poder predictivo de blancos de microRNAs mediante herramientas como TargetScan junto con la gran cantidad de datos genómicos existentes permite construir las redes de regulación de microRNAs para diferentes especies. Mediante comparaciones entre las redes de regulación de diferentes grupos de organismos se pueden responder preguntas relacionadas con la manera en que evolucionan las redes de regulación de microRNAs y cómo contribuyen a la evolución del fenotipo, un fenómeno que hasta la fecha se entiende muy poco.

## 6 Métodos

### 6.1 Definición de las métricas de recableado

Primero se hizo una revisión en la literatura de las métricas de recableado que se podrían utilizar para entender la evolución de las redes de regulación de microRNAs. A continuación se explican las dos métricas que se utilizarán para resolver las preguntas planteadas en este proyecto.

#### 6.1.1 Métrica de Shou

En las estimaciones propuestas en el trabajo de Shou *et al.* en 2011, se calcula una tasa de recableado basándose en la ortología de los nodos de la red. La red de una especie se toma como referencia y las de las otras especies se utilizan para compararlas con la referencia. Esta tasa de recableado consiste en el número total de conexiones que se recablearon divididas por el tamaño de la red. El tamaño de la red se representa por el número total de posibles conexiones entre las dos redes completas. También se divide

por el tiempo de divergencia. Además para esta métrica cabe resaltar que solamente se permite máximo una conexión entre regulador y blanco.

Para estimar esta tasa se consideran 3 conjuntos de nodos. Los nodos comunes son los que están presentes en ambas redes, la de referencia y la que se compara. Los nodos perdidos solo están presentes en la red de referencia y los nodos ganados solo están en la red que se compara. Además se miden también 4 tipos de eventos de recableado que son: 1) ganancia de nodos comunes, 2) pérdida de nodos comunes, 3) pérdidas de conexiones por pérdida de nodos y 4) ganancia de conexiones por ganancia de nodos.

El valor de la tasa de recableado se define por la siguiente ecuación:

$$Tasa\ de\ recableado = \frac{R}{C * Tiempo\ de\ divergencia}$$

Donde R es la cantidad de conexiones recableadas y C es el total de número de eventos de recableado posibles. Se divide también por el tiempo de divergencia entre los pares de especies a comparar. El valor de C viene de la combinatoria:

*C para redes de microRNAs*

$$C = RegCNs \times TarCNs + RegGNs \times TarGNs + RegLNs \times TarLNs \\ + RegCNs \times (TarGNs + TarLNs) + TarCNs \times (RegGNs + TarLNs)$$

Donde CN son los nodos comunes, LN son los nodos perdidos, GN nodos ganados, Reg son los nodos que corresponden a reguladores y Tar son los nodos que corresponden a sus Blancos.

En este trabajo de tesis se tiene como referencia a una red ancestral y por ello no es necesario dividir por el tiempo de divergencia, haciendo una modificación a la manera en que se estima esta métrica.

### 6.1.2 Métrica de Simkin

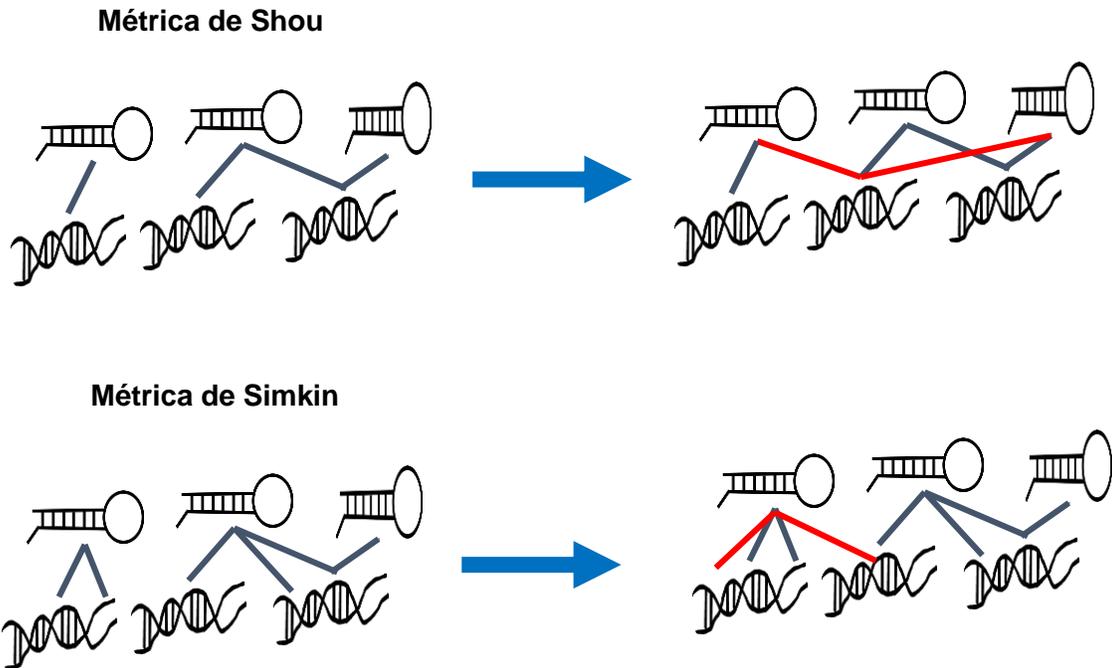
Para entender cómo funciona esta métrica primero se debe definir la palabra k-mero. Un k-mero hace referencia a una de todas las posibles palabras de una longitud *k* que se

pueden formar con  $n$  posibles letras. El número de palabras posibles se puede estimar con la fórmula  $n^k$  donde  $k$  es la longitud de la palabra y  $n$  son las posibles letras. Por ejemplo para el código genético se tienen 4 posibles letras, y si se quiere saber cuántos  $k$ -meros se pueden formar de longitud 8, también llamados octámeros, se calcula que son  $n^k = 4^8$  o 65,536 octámeros posibles.

La métrica de Simkin consiste en estimar la tasa de ganancia o pérdida de posibles sitios de pegado de familias de microRNAs. También se estima la ganancia y pérdida del resto de  $k$ -meros de la misma longitud que la región semilla. En este método se normaliza por la frecuencia de cada uno de los  $k$ -meros debido a que aquellos que son más abundantes tienen más probabilidad de tener al menos un evento de pérdida o ganancia. También la normalización ayuda a corregir por otro tipo de sesgos que afecten la frecuencia de  $k$ -meros como los sesgos composicionales. En este trabajo de tesis, como las tasas de cambio son para cada microRNA, se realizó la suma de todas las tasas de cambio para tener la tasa de cambio total de la red. Esta métrica originalmente fue aplicado con  $k$ -meros pero en este caso también se utilizó el microRNA maduro para la predicción de sus blancos.

### 6.1.3 Comparación entre métricas

La métrica de Shou y la de Simkin estiman diferentes tipos de eventos que se muestran en la Figura 16. La métrica de Simkin permite múltiples conexiones de un mismo microRNA a un mismo blanco. Por otro lado, la métrica de Shou permite máximo una conexión de un microRNA a un blanco.



*Figura 16 Esquema de las métricas de recableado entre genes y entre sitios, de Shou y de Simkin respectivamente.*

Otra de las diferencias entre las dos métricas es la manera de normalizar de cada una. En la métrica de Shou la normalización está basada en el tamaño de la red asumiendo que la conectividad es de 1 entre el regulador y sus blancos. En la métrica de Simkin la normalización se hace con base a la frecuencia de sitios de pegado que posee cada una de las familias de microRNAs. Esta normalización se hace debido a que las familias con más sitios de pegado tienen mayor probabilidad de presentar pérdidas o ganancias. Estas diferencias de las métricas podrían llevar a estimar diferencias de recableado que no necesariamente sean por el recableado evolutivo sino por causas como saturación de conexiones en la red. Para demostrar este tipo de situaciones, y determinar si podrían afectar las estimaciones hechas con las redes de regulación genética, se evaluó bajo qué condiciones estas métricas podrían ser afectadas por un incremento de tamaño y saturación de las redes.

Debido a que son diferentes variables en juego y la complejidad de resolver analíticamente esta situación, se realizaron simulaciones para evaluar si las métricas

estaban correlacionadas entre sí, y después para ver si la saturación y el tamaño de las redes podrían llevar a interpretaciones incorrectas.

Un gen puede tener múltiples sitios de pegado para un mismo microRNA. También se ha aproximado que los sitios de pegado potencialmente verdaderos aparecen raramente. Considerando estas 2 características se puede aproximar el número de los sitios de pegado de los microRNAs mediante una distribución de Poisson (Rehmsmeier et al. 2004; Ying 2006).

#### 6.1.3.1 Simulaciones cambiando la proporción de recableado entre genes contra recableado de sitios de pegado.

Se realizó una simulación para comprobar si las métricas pueden arrojar resultados contrastantes. A continuación se explican los pasos realizados para este análisis.

- 1.- Se utilizaron redes de 4 reguladores y 4 posibles blancos como modelo de juguete.
- 2.- Se fijaron las proporciones de recableado entre genes con respecto a una red de referencia. Se utilizaron 3 proporciones de recableado entre genes con respecto al número de eventos totales de recableado (Proporción de Recableado entre Genes, PRG). Las PRGs se fijaron a 0.5, 0.75 y 0.875.
- 3.- Se construyeron redes de manera arbitraria que al compararse cumplieran estas proporciones.
- 4.- Se estimaron los valores de las métricas de Shou y de Simkin para las redes construidas anteriormente con sus respectivas PRGs.
- 5.- Estos valores se contrastaron con respecto a un modelo nulo que sigue una distribución de Poisson en sus conectividades. Este último paso se hizo para saber si por azar era posible observar valores similares a los que se obtuvieron con las redes construidas arbitrariamente.

#### 6.1.3.2 Efectos de la distribución de conectividades en las métricas.

Se utilizó la distribución de Poisson para representar las distribuciones de conectividad de los microRNAs. Con la finalidad de comparar con las redes biológicas se estimó el parámetro de la distribución de Poisson que corresponde a los valores de conectividad de las redes reales de microRNAs de primates. Después se varió el parámetro de la distribución de Poisson para ver el efecto de diferentes distribuciones de conectividades sobre la estimación de las tasas de recableado.

## 6.2 Selección de las especies a analizar

Los primates, especialmente los grandes simios, son un grupo adecuado para construir redes de regulación genética debido a la calidad de anotación de sus genomas. Además no presentan grandes expansiones o pérdidas de genes y por lo tanto para la mayoría de sus genes existe una relación 1:1 de homología. Estas características permiten obtener una mejor perspectiva de cómo las redes de regulación de microRNAs evolucionan. Además estas características también permiten entender los mecanismos que subyacen la evolución del fenotipo dentro del grupo de los primates debido a los microRNAs.

Las especies de primates utilizadas se muestran en la Tabla 3.

Tabla 3 Especies utilizadas y sus tiempos de divergencia

<b>Especie</b>	<b>Tiempo de divergencia (en millones de años)</b>	
<b>Humano (<i>Homo sapiens</i>)</b>	Referencia	
<b>Chimpancé (<i>Pan troglodytes</i>)</b>	6.2	
<b>Gorila (<i>Gorilla gorilla gorilla</i>)</b>	8.3	
<b>Orangután (<i>Pongo pygmaeus abelii</i>)</b>	15.1	

<b>Gibón (<i>Nomascus leucogenys</i>)</b>	19.3	
---	------	---

#### 6.4 Selección de familias de microRNAs

En este proyecto se utilizó la definición de familia de microRNA usada por TargetScan. Una familia consiste en el grupo microRNAs que comparte la misma secuencia de la región semilla, es decir las posiciones 2-8 del microRNA maduro. Para el conjunto de microRNAs más ampliamente conservados en vertebrados se utilizó el conjunto definido por TargetScan. Este conjunto proviene de grupos de microRNAs de humano que cuentan con evidencia de secuenciación de RNAs pequeños. Después fueron comparados con los alineamientos de 100 especies (100-way Multiz), para cada microRNA maduro, si los nucleótidos 2-8 están conservados en más de 60 especies se consideraron ampliamente conservados en vertebrados. En la Figura 17 se muestra a let-7 que es un microRNA ampliamente conservado y que todos los vertebrados lo tienen (Agarwal *et al.* 2015).

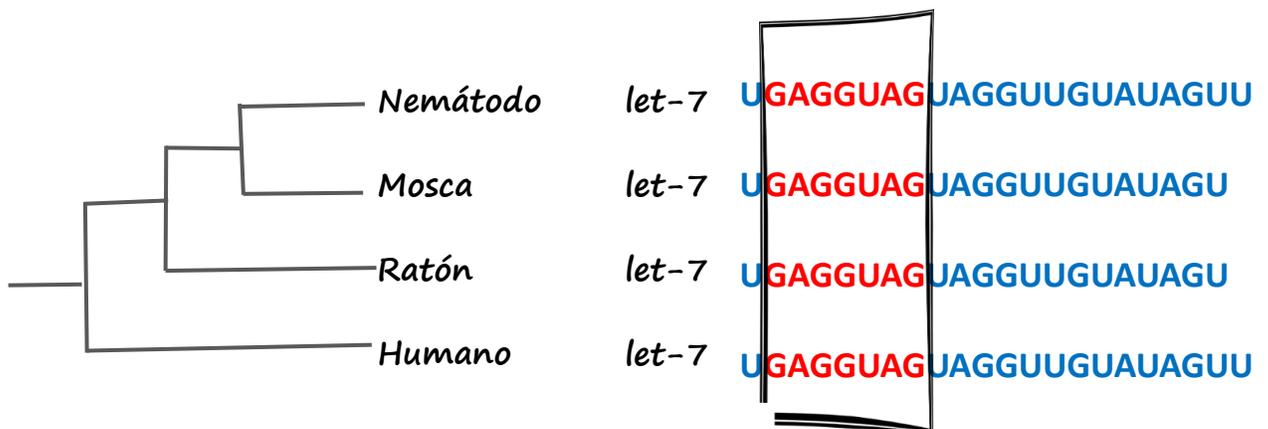


Figura 17 Representación de los k-meros que pertenecen a la región semilla de la secuencia del microRNA let-7.

El conjunto seleccionado contiene 109 familias de microRNAs que están profundamente conservadas en vertebrados. Estas familias están principalmente conservadas en

alineamientos que abarcan vertebrados hasta el pez cebra. En este proyecto se trabajó con las familias conservadas debido a que están presentes en todos los grupos de organismos que se van a estudiar. Además, la mayoría de estos microRNAs conservados participan en procesos importantes del desarrollo y es mucho más probable que estén asociados a un fenotipo.

Para el caso del análisis utilizando TargetScan se utilizó la secuencia del microRNA maduro. En el análisis de k-meros se utilizaron los octámeros preferentemente sobre otros k-meros más pequeños de la región semilla. Esto es debido a que los octámeros presentan mayor señal sobre ruido (Simkin *et al.* 2014)

## 6.5 Análisis de recableado en 3'UTRs de Primates

### 6.5.1 Obtención de los conjuntos de 3'UTRs

Se realizó el pre-procesamiento de las regiones 3'UTR sugerido por Simkin *et al.* en 2014 para realizar una inferencia ancestral de estas secuencias con menor incertidumbre. Esta metodología se describe en la Figura 18

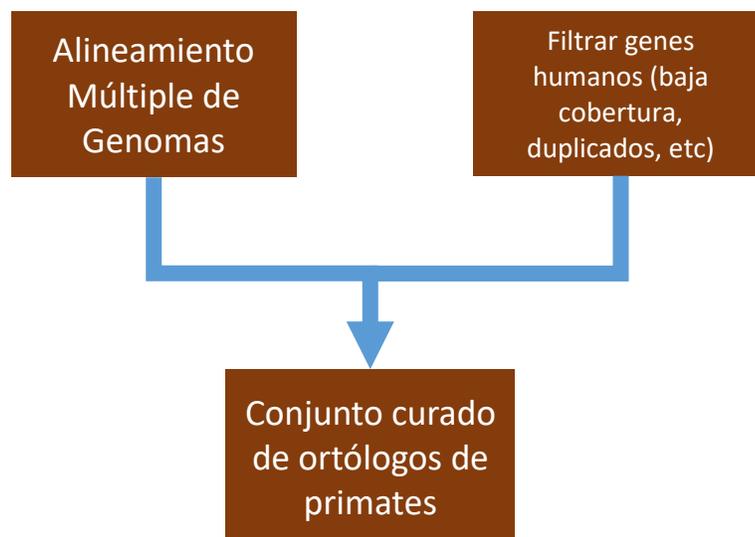


Figura 18 Diagrama de pre-procesamiento de las 3'UTR

Primero se descargaron los genomas de cada uno de los primates a analizar, junto con la base de datos de secuencias de RNAs mensajeros y sus isoformas del humano. Primero, estas secuencias de humano se filtraron para excluir genes sin 3'UTR, genes que ocurren varias veces en el genoma y genes con regiones 3' UTR multiexónicas.

Después se realizaron alineamientos mediante el programa LASTZ de cada genoma de primate con respecto a las secuencias filtradas de humano. Este procedimiento se hizo para quitar aquellas secuencias que pudieran estar repetidas y que tuvieran una baja cobertura. Esto permitió obtener un conjunto de ortólogos curados. Los parámetros que se utilizaron para LASTZ son match 1, mismatch: -3 y un tamaño de paso de 200. Los genes con menos de 95% de identidad al de humano se descartaron (Harris 2007).

De los resultados de los alineamientos con LASTZ se tomaron las secuencias que alinearon solamente con un gen de humano con un porcentaje de identidad mayor al 95%. Se obtuvieron algunos genes con múltiples alineamientos asociados a ellos. Si estos múltiples alineamientos por gen estaban acomodados de manera lineal, se consideraban como transcritos del gen. Estos transcritos se ordenaron con respecto a su porcentaje de cobertura. Si el transcrito con mayor cobertura tenía un valor menor de 80% también se descartaba. De los aceptados, aquellos con un segundo transcrito con un porcentaje de cobertura mayor al 20% con el mejor transcrito también fueron descartados debido a que podría pertenecer a un gen duplicado. De los transcritos restantes los que tienen un porcentaje de cobertura de 20% con un gen de manera redundante también se descartaron debido a posibles duplicaciones internas. Aquellos transcritos que se alinearan con más de una 3'UTR también se descartaron.

A partir de este conjunto curado de ortólogos, se obtuvieron las regiones que corresponden a las 3' UTR de los alineamientos múltiples de primates con respecto al gorila. Se eligió este alineamiento debido a que contiene las versiones más recientes de genomas de los grandes simios. Los alineamientos usados son los de Multiz de la base de datos de UCSC (Blanchette et al. 2004).

### 6.5.2 Obtención del conjunto de secuencias 3'UTR ancestrales

Se utilizó DNAML para la inferencia ancestral de las secuencias 3'UTR (Felsenstein & Churchill 1996). Para realizar la inferencia ancestral de secuencias de DNA este programa necesita la filogenia de los organismos a los que se les quiere predecir sus secuencias ancestrales. También este programa utiliza otras fuentes de información como un modelo de sustitución y las secuencias actuales de los organismos a analizar. DNAML realiza la inferencia ancestral de secuencias mediante el método de Máxima verosimilitud considerando una tasa constante de sustitución.

### 6.5.3 Predicción de Blancos

#### 6.5.3.1 Método de región semilla basado en k-meros

Para este método se definió como blanco de un microRNA, cualquier secuencia de 3'UTR que tuviera al menos un 8-mero perfectamente complementario a la región semilla de dicho microRNA. Respecto a las predicciones con los k-meros, se buscaron los sitios complementarios a los octámeros correspondientes a los microRNA y sus controles.

#### 6.5.3.2 Método usando la secuencia del microRNA maduro

La conservación de un sitio de pegado es una característica importante para la predicción de blancos de microRNA (Figura 19). Esta conservación se mide mediante el criterio del valor de Probabilidad de conservación del blanco ( $P_{CT}$ ). Este valor es usado para mejorar las predicciones de sitios blanco. Sin embargo en este caso dicho valor podría limitar nuestras estimaciones del recableado observado en las redes. Esto es debido a que solo toma en cuenta aquellos sitios conservados y no a los que se pierden y se ganan. Por lo tanto se utilizó el Context Score ++ (CS++). El CS++ es otro criterio de TargetScan para predecir blancos que se basa en 14 factores. Estos factores son el tipo de región semilla (ya sea hexámero, heptámero u octámero), frecuencia de sitios de pegado en las 3' UTR, distancia entre los sitios de pegado, estructura secundaria que pueda tomar la 3' UTR, entre otros. Además, en la última versión de TS7 se incluye el criterio de conservación antes mencionado.



Figura 19 Representación del sitio semilla conservado (Friedman et al. 2009)

Se realizó una modificación en el pipeline de TS7 para evitar la contribución del criterio de conservación a los resultados de CS++. Este análisis se realizó con las 11 especies establecidas en TS7, ya que solo para estos organismos se conocen los parámetros del modelo lineal de TS7. Estos parámetros sirven para hacer las estimaciones de las probabilidades de conservación de blancos.

Además, para cada microRNA se hizo la suma de los valores de CS++ de sus sitios de pegado en cada RNA mensajero. Al final se obtuvo un solo valor que conecta a cada microRNA con cada uno de sus genes blanco. Se seleccionaron como posibles blancos aquellos que pertenecieran a una determinada proporción con los valores de CS++ menores (entre más negativo, mejor es la predicción de acuerdo al CS++). Se probaron distintos porcentajes para estudiar cómo los resultados eran afectados al variar el corte de CS++.

#### 6.5.4 Construcción de las redes

Los datos de los microRNAs y sus blancos predichos se pueden representar con las llamadas redes de 2 modos. Estas redes tienen 2 clases de nodos y solamente puede haber conexiones entre los diferentes tipos de nodos. En este caso las 2 clases de nodos son los microRNAs y sus blancos predichos (Opsahl 2013).

Para estimar la tasa de recableado se generó una matriz de incidencia que incluye las 2 clases de nodos. Una matriz de incidencia es aquella matriz donde las columnas pertenecen a una clase de nodos, por ejemplo los genes blanco, y las filas a la otra clase de nodos, por ejemplo los reguladores. En la Figura 20, la celda que corresponde al regulador 1 (R1) y al blanco 1 (T1), tiene el número 2, ya que estos nodos están conectados 2 veces entre ellos.

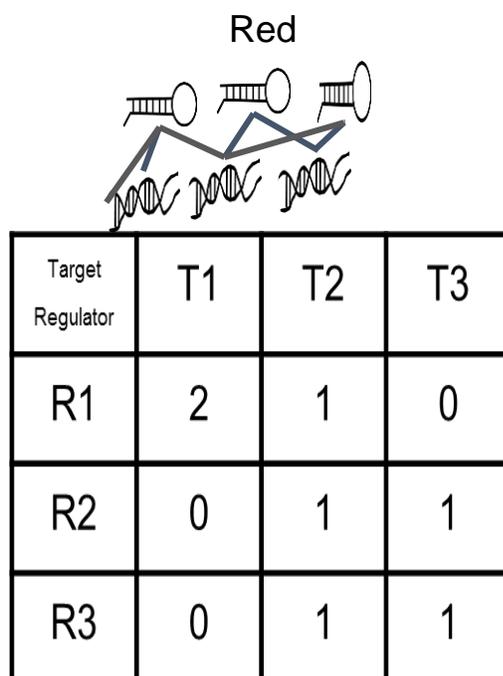


Figura 20 Matriz de Incidencia que representa una red de regulación genética de microRNAs

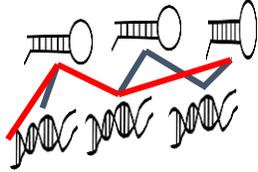
Para generar las matrices de incidencia utilizando la métrica de Simkin se tomaron en cuenta todos los sitios existentes entre microRNA y blanco. En el caso de la métrica de Shou se tomó en cuenta un sitio por microRNA como máximo. Se generó esta matriz para cada organismo analizado.

### 6.5.5 Estimación del recableado entre genes en primates utilizando k-meros

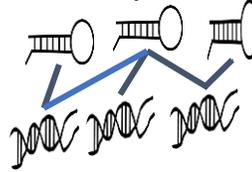
#### 6.5.5.1 Estimación de las Tasas de recableado de Shou y Simkin

Partiendo de las redes construidas anteriormente se midieron las tasas de recableado con las métricas de Shou y de Simkin. Al final se obtiene únicamente un valor de recableado para cada red comparada con la red ancestral de referencia. A continuación se explica con un modelo de juguete. En la Figura 21, en color rojo se observan 3 conexiones perdidas y en color azul se observa una conexión ganada.

Red Ancestral



Red de Especie



Target Regulator	T1	T2	T3
R1	2	1	0
R2	0	1	1
R3	0	1	1

Target Regulator	T1	T2	T3
R1	1	0	0
R2	1	1	1
R3	0	0	1

$Abs([Especie1] - [Ancestral]) =$

Target Regulator	T1	T2	T3
R1	1	1	0
R2	1	0	0
R3	0	1	0

$= \text{Recableado}$

Figura 21 Diagrama representativo de la comparación de las redes de regulación para estimar el recableado sufrido desde su ancestro

Utilizando la métrica de Shou se observa que el número de eventos de recableado entre genes es 3. El valor de la métrica de Shou es 3/9, ya que son 3 eventos de recableado del total de 9 posibles conexiones.

$$\text{Métrica de Shou} = \frac{\sum_i^m \sum_j^n \text{Recableado}}{m \cdot n} = \frac{3}{9}$$

Donde “j” es el índice de las columnas y que representa a los blancos. Y por otro lado, “i” es el índice de los renglones que representa a los reguladores.

Con la métrica de Simkin se observa que el número total de eventos es 4. El valor de la métrica de Simkin se calcula para cada regulador. Para el regulador 1 la tasa de recableado es 2/3 ya que sufrió 2 cambios de 3 conexiones que tenía su ancestro. Para los otros reguladores es 1/2 para cada uno y la suma total es de 10/6.

$$\text{Métrica de Simkin} = \frac{\sum_i^m \frac{\sum_j^n \text{Recableado}}{\sum_j^n \text{Ancestral}}}{n} = \frac{\sum_i^m \left[ \begin{array}{c} 2 \\ 3 \\ 1 \\ 2 \\ 1 \\ 2 \end{array} \right]}{3} = \frac{10}{18}$$

6.5.5.2 Generación de soporte estadístico con el método de k-meros de la región semilla. Para generar el soporte estadístico de este análisis basado en k-meros se utilizaron controles composicionales. Se hicieron secuencias similares en composición y que no corresponden a secuencias reales de semillas de microRNAs. Estos controles consistieron de secuencias de 8 nucleótidos que estuvieran a 1 nucleótido de diferencia en la región correspondiente al hexámero central de la región semilla de los microRNAs. El hexámero de la región semilla corresponde a los nucleótidos de las posiciones 2 a la 7

del microRNA maduro. De estas secuencias, solo se tomaron en cuenta las que tengan un Adenina en la posición complementaria a la posición 1 del microRNA. Esta Adenina tiene la capacidad de estabilizar la interacción entre los microRNAs y sus blancos, no necesariamente mediante un apareamiento tipo Watson y Crick (Lewis 2005).

Se obtuvo un conjunto de 800 k-meros control. Para hacer el contraste se hicieron muestreos independientes de este conjunto es decir se hizo un modelo nulo del valor de recableado. Para los muestreos se usó un tamaño de muestra similar al del conjunto de k-meros pertenecientes a las regiones semilla de los microRNAs de vertebrados. Se hicieron 1,000 repeticiones y en cada una de ellas se estimó el valor ya sea de la métrica de Shou ó la de Simkin para ese conjunto de k-meros control. Se compararon los valores de los controles con el valor de recableado para los microRNAs de los vertebrados, para obtener un p-valor empírico. Es decir la fracción de los valores del modelo nulo que fueron al menos tan extremos como el valor real observado.

#### 6.5.6 Estimación del recableado entre genes en primates utilizando la secuencia del microRNA maduro

##### 6.5.6.1 Estimación con métrica de Shou y Simkin

Se utilizaron las secuencias completas de los microRNAs maduros para mejorar la predicción de blancos. El uso del microRNA maduro permite tomar en cuenta otros factores adicionales al pegado de la región semilla. Estos factores están relacionados con otras interacciones de los microRNAs y sus blancos. Para esta predicción de los blancos a partir de los microRNAs maduros se utilizó el algoritmo de TargetScan7 CS++, el cual regresa una calificación numérica relacionada con la probabilidad de que un mRNA puede ser reprimido por el microRNA analizado. Debido a que no hay un valor de corte ampliamente recomendado para definir lo que es un verdadero blanco, se evaluaron 3 diferentes umbrales, basados en los percentiles de los valores del CS++. Los diferentes umbrales se usaron, para conocer el efecto de ser muy estrictos, o no, con los valores del CS++. Habiendo definido entonces tres conjuntos de blancos predichos para cada microRNA, se realizaron las estimaciones de las tasas de recableado de Shou y de Simkin. Estos resulta dos permitieron saber si las estimaciones de recableado son robustas a los valores de predicción del CS++. En este análisis se estimaron los eventos

de recableado de manera individual para cada microRNA y sus respectivos controles (descritos a continuación).

#### 6.5.6.2 Generación de soporte estadístico para el método del microRNA maduro

En este caso, en lugar de generar secuencias aleatorias como controles, éstas se obtuvieron a partir de los microRNAs. Esto se hizo mediante el criterio de evaluación del CS++ que dice qué tan probable es que cada mRNA predicho sea un verdadero blanco. Los blancos del conjunto control se seleccionaron mediante el puntaje del CS++. Los blancos del conjunto control son aquellos que tienen malos puntajes del CS++. En otras palabras, son aquellos sitios que muy probablemente no son blancos a pesar de tener cierta complementariedad a la región semilla del microRNA. Al final, para cada conjunto de blancos verdadero, se tomó un número equivalente de genes, pero del lado opuesto de la distribución de puntajes del CS++.

#### 6.5.6.3 Métrica de recableado por blanco

Se midió la cantidad de pérdidas y ganancias de sitios de pegado de los microRNAs que sufre cada blanco. Se construyeron vectores para cada blanco contando cada uno de los sitios de regulación que posee cada uno. Como se mencionó anteriormente si un microRNA tiene varios sitios dentro de un gen blanco se puede contar como una sola predicción, haciendo la suma del CS++ de estos sitios (Agarwal et al. 2015). Después se seleccionaron como blancos aquellos que cumplieran el criterio de un determinado porcentaje de CS++. Se evaluaron varios porcentajes de corte del CS++. Para análisis posteriores se delimitaron las predicciones al 20% más alto y más bajo de los valores de CS++. A partir del conjunto de familias de microRNAs y blancos predichos se cuentan las diferencias entre los blancos de los descendientes con el ancestro común para estimar la proporción de sitios que se pierden o ganan por blanco. Se debe normalizar por longitud para obtener un recableado por kilobase del gen blanco debido a que una UTR más larga puede tener más sitios de pegado y por ello aparentar tener más recableado (Figura 22).

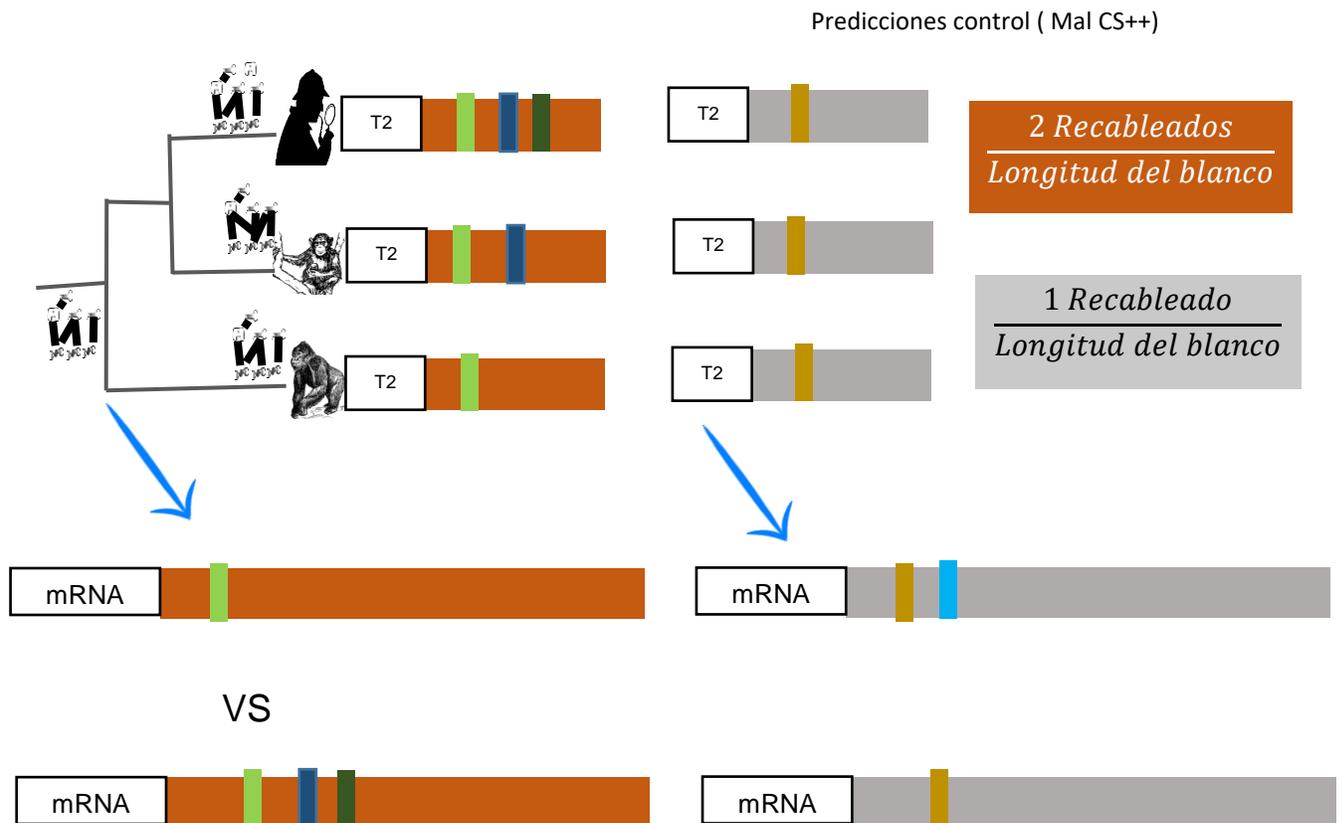


Figura 22 Diagrama representativo de la métrica de recableado por kilobase de gen blanco propuesta.

A este valor obtenido de recableado por kilobase del gen blanco se le hace una división sobre el mismo valor pero utilizando predicciones de sitios que no son microRNAs. Es decir, se dividió por la tasa de recableado que se esperaba de manera neutral. A esta razón se le llamó Razón de Recableado y se representa por “ $\rho$ ” (Figura 23).

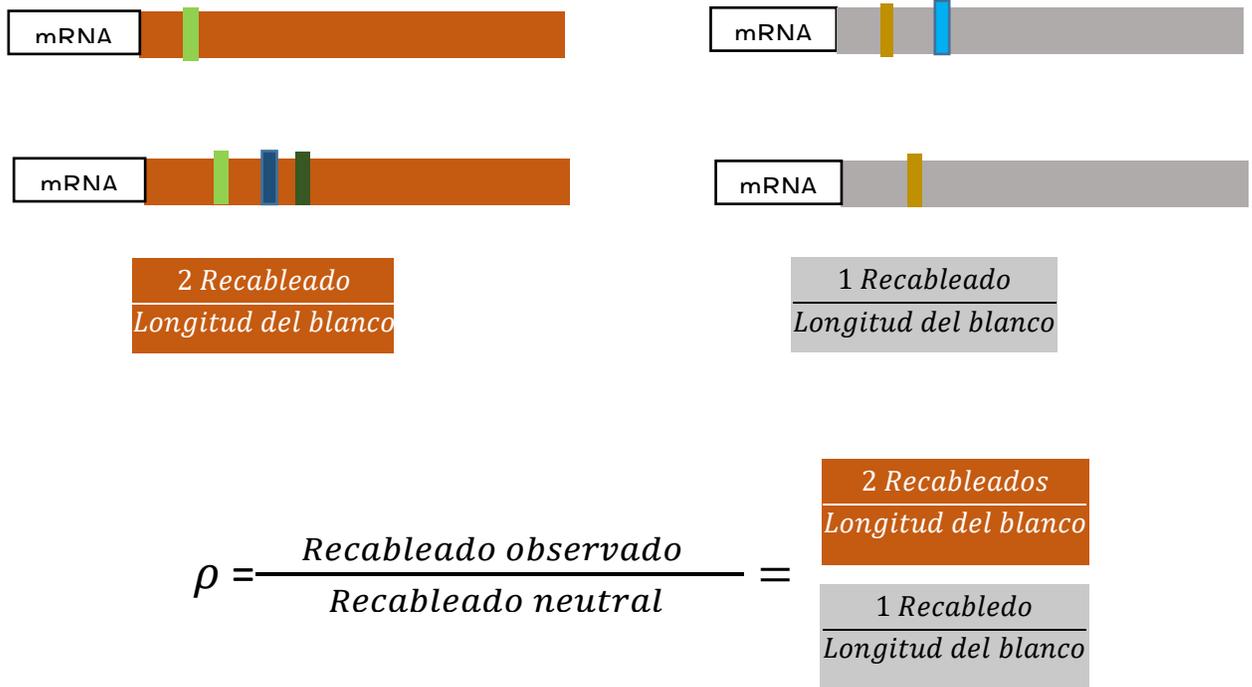


Figura 23 Obtención del valor de razón de recableado por blanco "ρ"

Esta razón de recableado por blanco se obtiene para cada uno de los blancos y se calcula para 3 escenarios posibles. Estos escenarios son la razón de recableado para la ganancia de sitios de pegado, razón de recableado para la pérdida de sitios de pegado y la razón de recableado total, es decir la razón de recableado para el número total de eventos, tanto de pérdida como de ganancia de sitios de pegado (Figura 24).

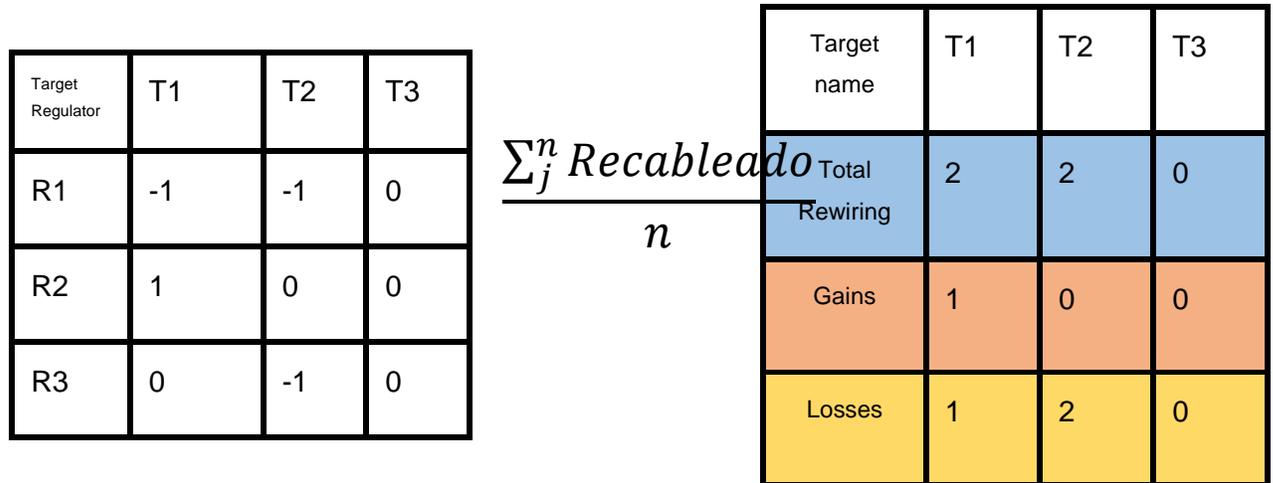


Figura 24 Obtención de los valores de recableado para cada blanco en los 3 posibles escenarios.

## 6.6 Análisis de categorías funcionales.

Para el análisis funcional solamente se evaluaron las tasas de recableado de Shou, ya que lo que se quiere medir es el cambio a nivel de genes y no a nivel de sitios. Esto se hace asumiendo que los cambios en categorías funcionales se miden mejor a nivel de genes.

De los análisis anteriores se obtuvieron distribuciones tanto de microRNAs con sus respectivas tasas de recableado así como de blancos con sus respectivas medidas de recableado por kilobase. Para conocer las funciones de los blancos que probablemente estén bajo un tipo de presión selectiva se realizaron pruebas de enriquecimiento de funciones de estos grupos de genes blancos. Se utilizaron los términos de GO (Gene Ontology) para obtener las posibles funciones de los conjuntos de genes. Los términos de GO vienen ordenados de manera jerárquica. Particularmente se utilizó el conjunto de términos GOSlim, el cuál es un subconjunto de términos de GO, conteniendo solamente los términos de las categorías más generales o que estén más arriba en la jerarquía. Por

lo tanto, reducir los términos de GO a este subconjunto sirve para obtener un panorama general de las funciones en las que participan estos grupos de genes blanco (Ashburner et al. 2000).

Otra manera de reducir los términos GO para tener una idea general de las funciones es utilizando ReviGO. ReviGO elimina la redundancia buscando los términos que pertenecen a una misma categoría semántica. Es decir, agrupa los términos GO con base a los términos parentales en común que puedan tener y la distancia a la que se encuentran estos términos. Esto permite tener mayor resolución que con GOSlim, ya que los términos de GO que provee GOSlim son 68 (Schlicker et al. 2006; Supek et al. 2011).

#### 6.6.1 Categorías funcionales de blancos de los microRNAs con valores extremos de tasa recableado

Se seleccionaron el 5% de familias de microRNAs con más alto recableado, así como el 5% de las familias con más bajo recableado. De estas familias se seleccionaron sus posibles blancos regulados y se realizaron las pruebas de enriquecimiento. En esta situación se tiene el problema de que los tamaños de los conjuntos a comparar son muy diferentes, es decir el número de blancos de microRNAs que han estado bajo tasas altas de recableado es muy pequeño comparado con el número de blancos de microRNAs que tienen tasas bajas de recableado. Para ver los efectos de este problema, se realizaron distintos cortes de CS++ para obtener distintos tamaños de conjuntos de blancos y los efectos que puedan tener sobre las categorías funcionales enriquecidas.

#### 6.6.2 Categorías funcionales de blancos con valores extremos en tasa de recableado por kilobase de gen blanco

Del análisis de recableado por kilobase de los genes blanco, se tomaron el 30% de genes que tuvieran mayor y el 30% con el menor recableado, y se realizaron las pruebas de enriquecimiento. En este análisis también se tiene el problema del tamaño de conjuntos de genes con alto recableado y bajo recableado y por ello se decidió tomar los conjuntos de genes en función de los percentiles.

### 6.6.3 Soporte estadístico de los análisis de categorías funcionales

Se hicieron pruebas de enriquecimiento para cada función asociada a los genes blanco. Se utilizó la distribución hipergeométrica para el cálculo de la probabilidad de ver un cierto número de genes o más con una determinada categoría funcional. Como se estimó la probabilidad para cada GO se hizo una corrección de pruebas múltiples mediante la tasa de falsos descubrimientos (Benjamini & Hochberg 1995). Los gráficos de enriquecimiento se representaron mediante heatmaps y mediante treemaps.

### 6.7 Conjunto de 3'UTRs de un alineamiento de mamíferos

Este conjunto de 3'UTRs fueron obtenidos del trabajo de Leclerq *et al.* de 2017. Se trata de un conjunto de regiones 3'UTRs ya procesadas incluyendo la inferencia de las secuencias ancestrales. Sin embargo, no todas las especies de primates comparten los mismos ortólogos incluyendo algunos de sus ancestros, entonces se tomó la intersección de secuencias ortólogos que estuvieran presentes en todos los primates analizados y sus ancestros. En este caso se seleccionó el grupo de los primates Catharrhini (Figura 25) presente en el alineamiento Multiz 28. Es decir las secuencias de Humano, Chimpancé, Gorila, Orangután, Babuino y Macaco. Este grupo se seleccionó ya que incluye al ancestro de los hominoides.

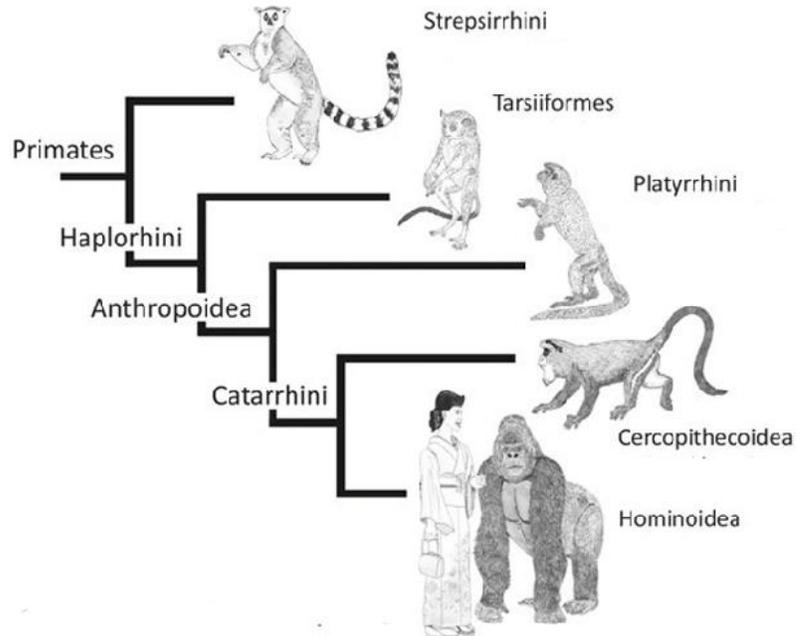


Figura 25 Filogenia de los primates donde se puede observar el grupo de los Catharrhini que incluye a los hominoides mencionados anteriormente y a los monos del viejo mundo (Naish 2016).

Se siguió la metodología planteada anteriormente, donde se estimó la tasa de Shou de recableado de las redes de cada uno de estos primates. Después se hizo el análisis de categorías funcionales. Se revisó si existía consistencia con los resultados del conjunto de 3'UTRs anteriores y también se revisaron las diferencias ya que este conjunto de 3'UTRs es más del doble que las obtenidas con el método anterior.

#### 6.8 Complementación con datos de transcriptómica comparativa entre chimpancé y humano.

Se utilizaron los datos obtenidos a partir de muestras de tejidos de corteza del cíngulo anterior y núcleo caudado de chimpancé y humano. Estos datos provienen de un análisis de expresión diferencial entre especies hecho por Bauernfeind *et al.* del 2015. Cuando se complementan las predicciones a nivel genómico con datos de expresión diferencial se puede obtener información de genes que hayan sufrido cambios evolutivos importantes y que a su vez estos cambios sean funcionales en un tejido determinado.

Para este análisis se identificaron grupos de genes con marcas de selección positiva en sus sitios de pegado, es decir, que sufrieron más cambios en sus sitios de pegado de lo

que se esperaría por azar. Para facilitar la búsqueda de genes con alto recableado y expresión diferencial entre especies se identificaron genes que en una especie perdieron sitios de pegado y que en la otra ganaron sitios, en ambos casos los cambios fueron más extremos de lo que se esperaría por azar. Se obtuvieron diferentes grupos de genes. Un grupo consistió de aquellos genes que ganaron más sitios de pegado en humano y que perdieron más sitios en chimpancé, con respecto al azar. El otro grupo consistió en aquellos genes que ganaron sitios en chimpancé y que perdieron en humano. Se comparó el nivel de expresión de ambos grupos de genes para cada tejido.

## 6.9 Uso de familias no conservadas de microRNAs

TS7 provee de una clasificación de familias de acuerdo a 3 niveles de conservación. Familias de microRNAs conservadas hasta vertebrados, familias de microRNAs conservadas principalmente hasta mamíferos y familias no conservadas.

### 6.9.1 Inclusión de un conjunto de familias conservadas hasta mamíferos

Se incluyó un conjunto de familias de 109 microRNAs que están principalmente conservados hasta mamíferos. Se realizó un análisis incluyendo tanto a los microRNAs conservados principalmente en vertebrados como a los microRNAs conservados principalmente en mamíferos y se estimó la tasa de recableado de todos estos microRNAs.

### 6.9.2 Familias con variantes únicas en humano

Para el caso de las secuencias de microRNAs específicas de humano se utilizaron las reportadas en el trabajo de Gallego *et al.* en 2016. Alrededor de 320 secuencias de microRNAs maduros se consideraron para este análisis ya que contenían mutaciones específicas en humano con respecto a otros primates y se analizó a nivel de k-meros sus tasas de cambio, tanto pérdidas como ganancias.

## 6.10 Radiación adaptativa de los cíclidos de Malawi.

En este trabajo se analizaron algunos eventos evolutivos que pueden brindar información de la influencia de los elementos reguladores en la evolución de los rasgos fenotípicos. Situaciones donde los organismos estuvieron bajo presiones evolutivas fuertes para

cambios en su morfología son buenos modelos de estudio para entender los mecanismos que dirigen la evolución de las redes de regulación genética.

Después de establecer el método se sentaron las bases para trabajar con otras especies que sean ejemplos adecuados de mucho cambio morfológico en poco tiempo. Estos tipos de especies son buenos candidatos para sugerir que los elementos de regulación juegan un papel importante en estos cambios. Se realizó un pequeño análisis con 130 3'UTRs provenientes del trabajo de Loh *et al.* del 2011. En la Tabla 4 se muestra la información de las especies de cíclidos utilizadas para el análisis de recableado.

Tabla 4 Especies de cíclidos utilizadas para el análisis de recableado

Especie	
<i>Tyrannochromis maculiceps</i>	
<i>Docimodus evelynae</i>	
<i>Nimbochromis polystigma</i>	
<i>Mchenga conophorus</i>	
<i>Labeotropheus fuelleborni</i>	
<i>Maylandia zebra</i>	
<i>Rhamphochromis esox</i>	

(Imágenes por Oliver 1997)

#### 6.10.1 Obtención de las secuencias de microRNAs

Para las familias de microRNA, se seleccionó el conjunto de 115 familias de microRNAs conservadas en los peces que fueron predichas mediante búsqueda de precursores de microRNAs de peces teleósteos en los genomas de los cíclidos (Loh *et al.* 2011).

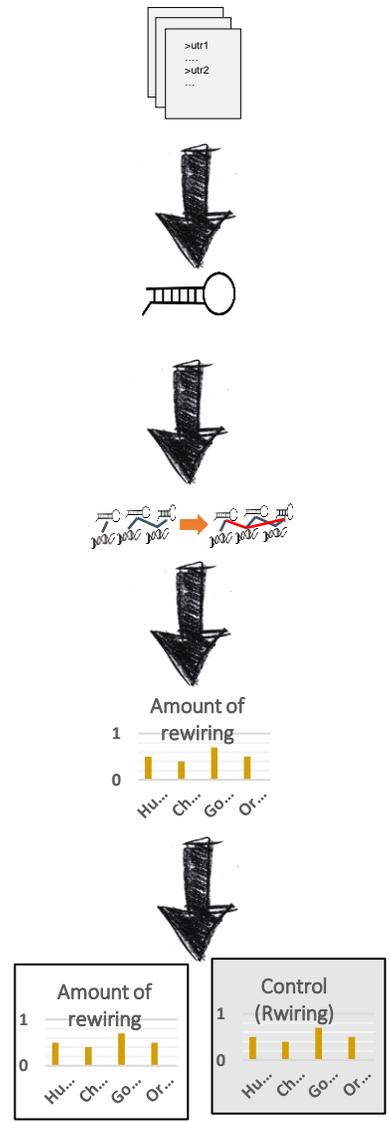
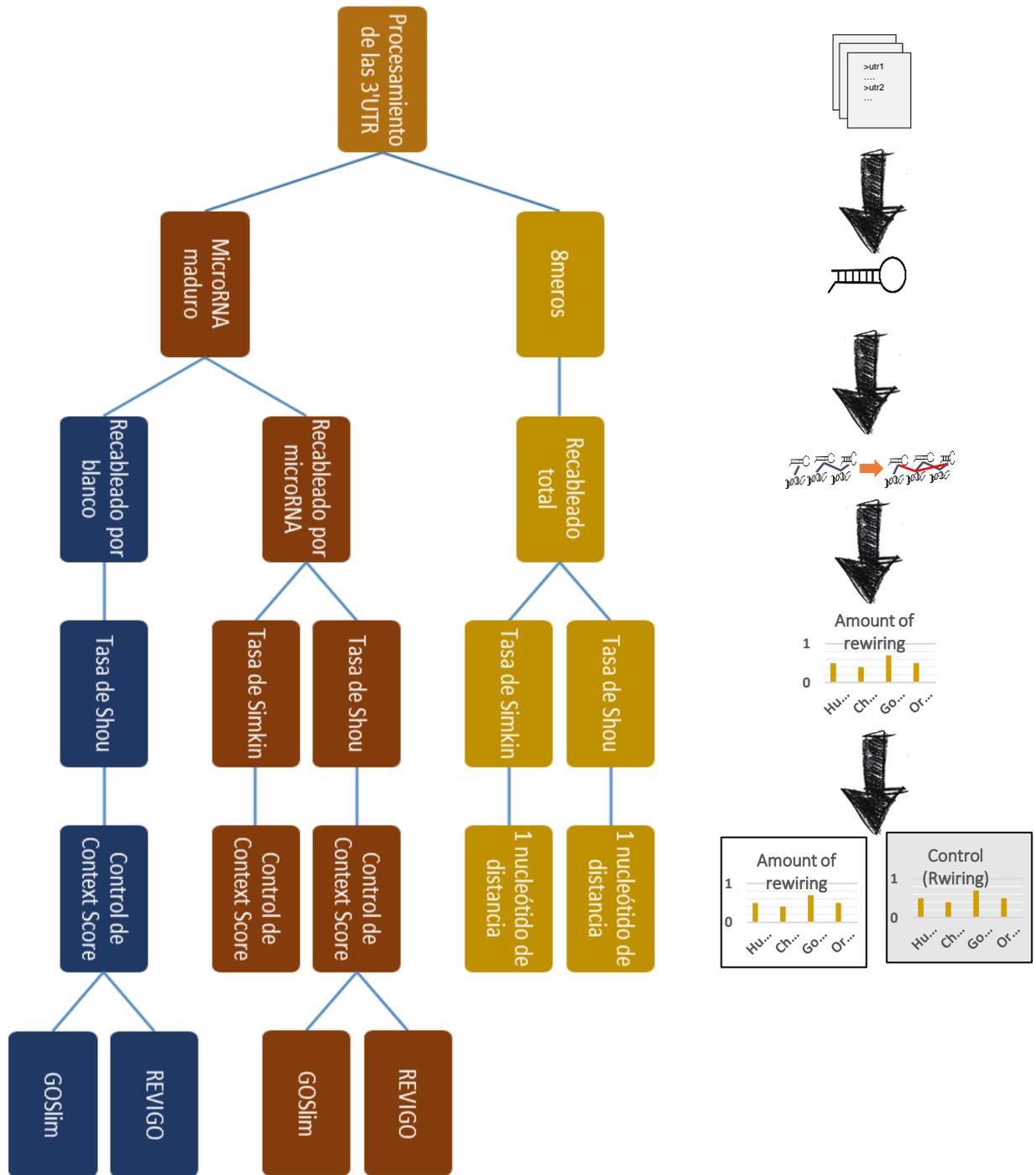
#### 6.10.2 Conjunto de las regiones 3' UTR

Se tomaron las secuencias de las UTRs validadas mediante resecuenciación en el trabajo de Loh *et al.* en 2011. Estas secuencias consisten de alineamientos múltiples de 130 3'UTRs que fueron secuenciadas en los 8 linajes de cíclidos. Se curaron estas secuencias ya que algunas de ellas estaban incompletas.

#### 6.10.3 Análisis de recableado en cíclidos

Se utilizó el método establecido en primates donde primero se estimaron las tasas de recableado para las redes de regulación de cada especie. En este caso se estudiaron los cíclidos del Lago Malawi. Después se generó el soporte estadístico utilizando un control de TargetScan7 CS++. La única diferencia fue que en este caso se utilizó a la especie de cíclido *Ramphochromis esox* como referencia debido a que no se posee un método concreto para la inferencia ancestral en este grupo.

## 6.11 Esquema Representativo de los métodos



## 6.12 Resumen general de la metodología

Los métodos para la estimación de las tasas de recableado para los microRNAs, de manera general requieren primero de las secuencias 3'UTRs. Para obtener estas secuencias se debe hacer un procesamiento de los datos genómicos que se disponen. Para la predicción de blancos durante este proyecto se llevó a cabo por medio de 2 métodos. Uno de los métodos está basado en los octámeros de la región semilla y el otro está basado en predicciones a partir de la secuencia de microRNA maduro. A partir de cada una de estas predicciones se construyen las redes de regulación para cada organismo y sus ancestro. Se mide el recableado de las redes y después se realizan pruebas para generar soporte estadístico a los resultados encontrados. Por último, para aquellos grupos de genes o familias de microRNAs con valores extremos de tasas de cambio se asocian a sus respectivas funciones con Gene Ontology.

# 7. Resultados

## 7.1 Evaluación de las métricas de recableado de redes, mediante simulación

Se evaluaron las 2 métricas de manera diferente. El primer análisis es para contrastar los comportamientos que tienen ambas tasas de recableado dada una Proporción de Recableado entre Genes (PRG). Se utiliza un modelo nulo con una distribución de Poisson de los posibles valores que pueden tomar las tasas. En la Figura 26A, los valores de la tasa de recableado de la métrica de Shou aumentan conforme la PRG aumenta. Sin embargo, el hecho de que aumente la proporción de recableado entre genes no implica que aumente el valor de recableado con la métrica de Simkin como se puede ver en la Figura 26B.

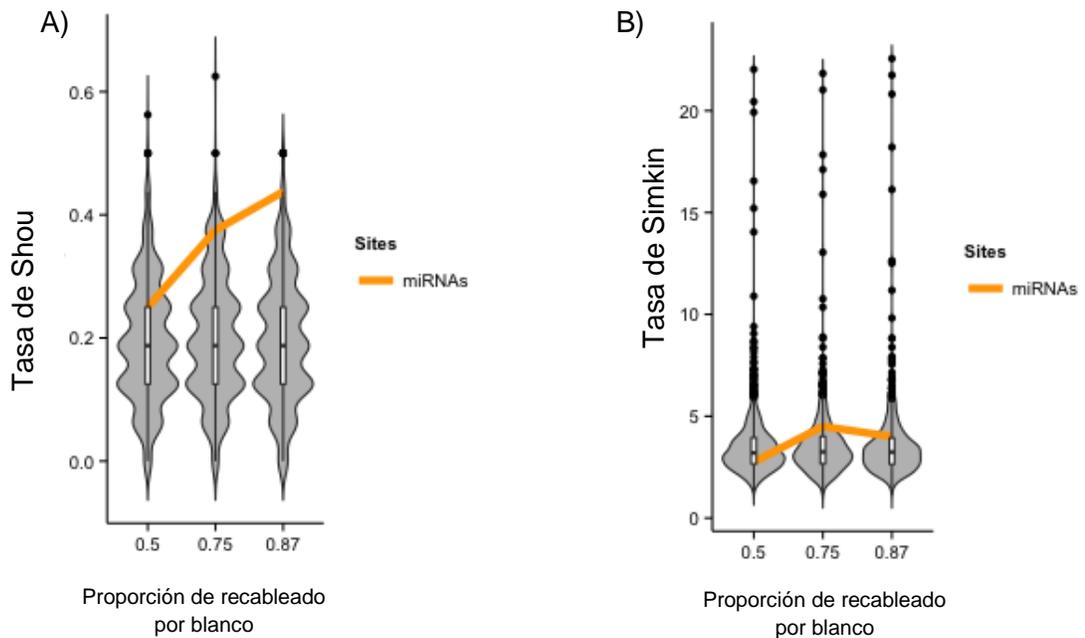


Figura 26 Efecto de la Proporción de Recableado entre genes (PRG ) sobre las dos métricas

El segundo análisis se hizo para ver los efectos de saturación y cambio de tamaño de la red en las tasas de recableado. Se evaluó el espacio de valores de las tasas de

recableado de las dos métricas para ver bajo qué parámetros de las distribuciones de conectividades y tamaño de la red se pueden ver situaciones contrastantes como el mostrado en el análisis anterior.

Se obtuvo la distribución de la frecuencia de todos los k-meros que pueden aparecer en cada blanco de cada especie de los primates analizados. En la Figura 27 se muestran estas distribuciones de conectividades en escala logarítmica para cada especie.

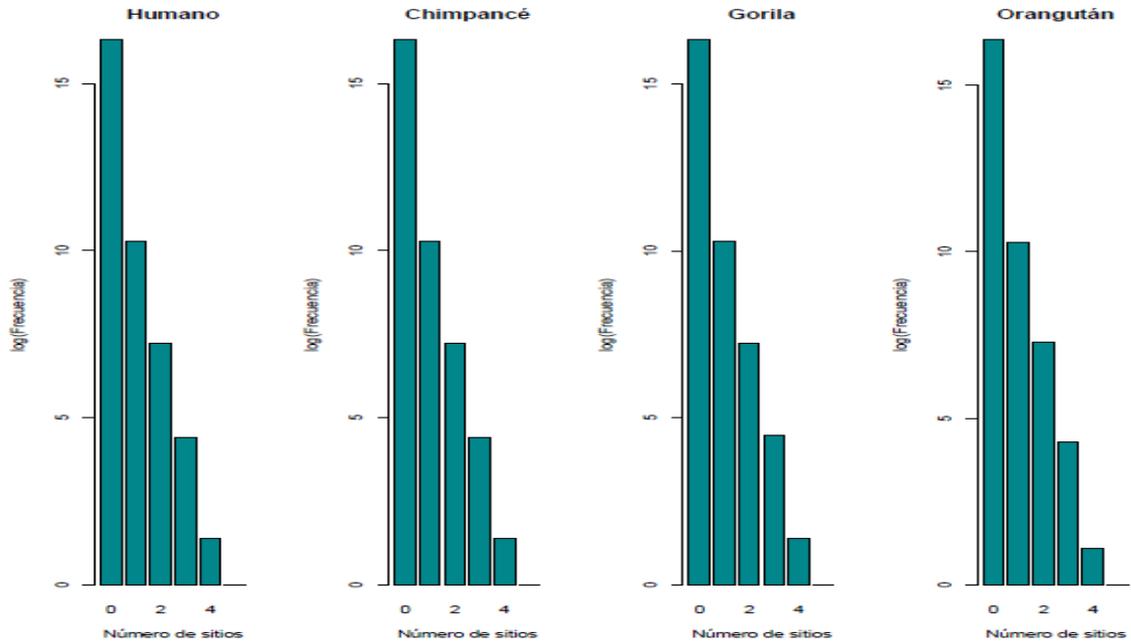


Figura 27 Distribución de conectividades para cada especie de los primates analizados

Las simulaciones se realizaron utilizando la distribución Poisson debido a los valores observados de las conectividades en las redes de regulación de microRNAs para estas especies de primates. Se hicieron las simulaciones variando el parámetro de la distribución entre valores extremos de saturación. Los resultados de las tasas de recableado se graficaron y se muestran en la Figura 28. Como se puede ver, en el eje Y, la tasa de Simkin disminuye conforme el parámetro Lambda de la distribución de Poisson aumenta. La tasa de Shou (eje X) alcanza su máximo en un punto determinado y comienza a decrecer.

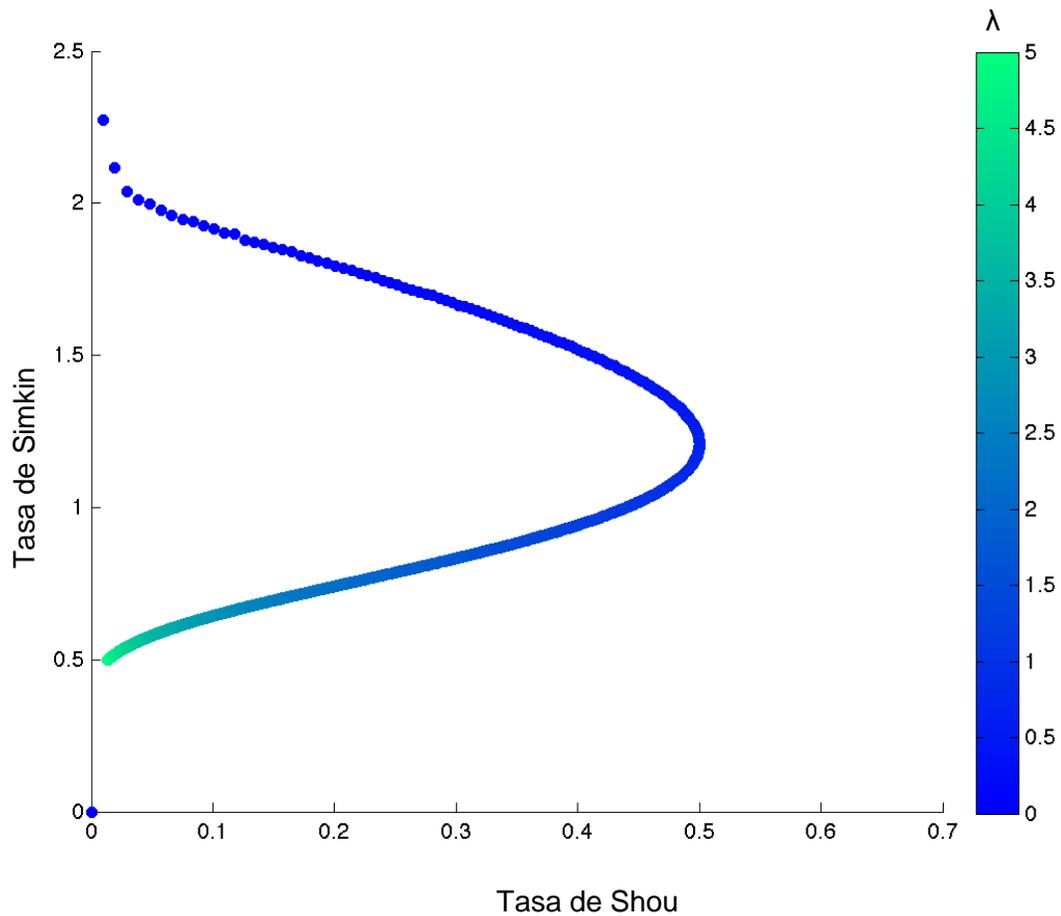


Figura 28 Tasa de Simkin en función de la tasa de Shou para distintos valores del parámetro de la distribución de Poisson

Se probaron distintos tamaños de red y se observó un cambio en los patrones de las tasas de recableado en función del número de nodos utilizado en la simulación. Para redes pequeñas vemos que la tasa de Simkin aumenta al inicio hasta que alcanza su punto de saturación (Figura S1). Para redes más grandes con una distribución similar a las redes de regulación de primates se observó que la tasa de Simkin se encuentra saturada. Conforme más conectividades se agreguen cambiando el parámetro de la distribución de Poisson la tasa de Simkin irá disminuyendo. Por otro lado, para redes de

distintos tamaños, la tasa de Shou aumenta hasta llegar a un punto de saturación y después comienza a disminuir (Figuras S1,S2,S3).

### 7.1.1 Efectos de saturación sobre las tasas de recableado

Para encontrar la región que corresponde al parámetro que está presente en nuestras redes reales de distribución de conectividades en primates se utilizó la función *fitdistr* de R para calcular el parámetro de la distribución de Poisson observado con las redes biológicas. Como se puede ver en la Figura 29 esta región no se ve afectada por los efectos de saturación. Sin embargo tampoco existe correlación positiva entre las métricas.

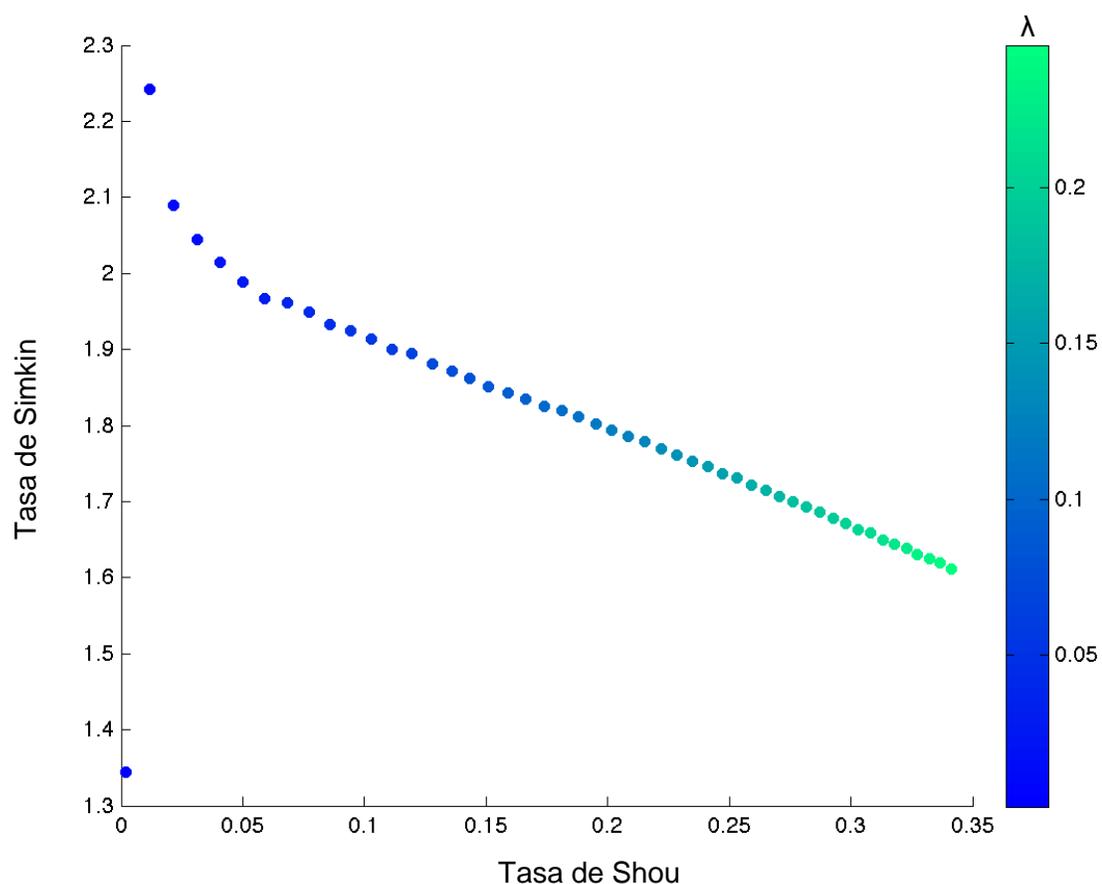


Figura 29 Espacio de valores de recableado en el que se encuentran las redes a evaluar

Mediante el ajuste que se realizó del parámetro de la distribución con respecto a las distribuciones reales de conectividades por k-meros en primates, se identificó la región en

la que las métricas de recableado se podrían ver afectadas para evitar efectos ajenos al recableado.

## 7.2 Análisis de las 3' UTR de la superfamilia de los Homínidos

### 7.2.1 Procesamiento de las 3'UTRs

Las secuencias de la región 3'UTR, en este proyecto se procesaron originalmente para cada una de las metodologías de manera diferente. Utilizando la metodología propuesta por Simkin *et al.* en 2014, al final se obtuvieron 8,260 secuencias de 3'UTRs debido a los procesos de inferencia ancestral. En la Figura 30 se muestran los pasos para obtener estas secuencias y la cantidad de secuencias obtenidas después de cada paso del procesamiento.



Figura 30 Cantidad de secuencias disponibles después de cada uno de los procesamientos

### 7.2.2 Efecto del criterio de conservación en la predicción de blancos

No es conveniente utilizar el criterio de conservación de TS7 para medir cambios en las redes de regulación. Al modificar parte de la metodología de TS7 para no considerar este criterio se midió cómo se afectan los valores del CS++ cuando se elimina el criterio de  $P_{CT}$ . Se realizó un gráfico de dispersión para observar la correlación entre los valores de CS++ con o sin el  $P_{CT}$ , (Figura 31). La correlación de Pearson resultó mayor de 0.99 con un p-valor menor de 0.01. También es importante señalar que a valores altos y negativos de CS++ (buena probabilidad de ser un verdadero blanco), el efecto de eliminar el  $P_{CT}$  es aún más pequeño.

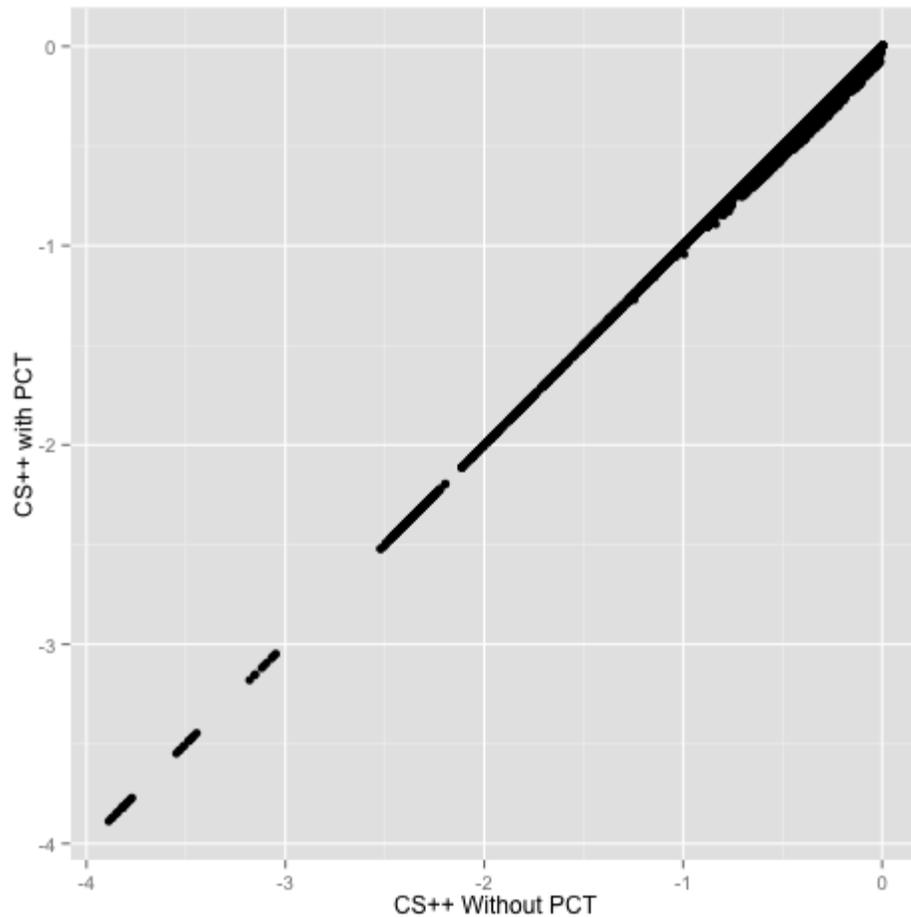


Figura 31 Gráfico de dispersión de los valores de CS++ tomando en cuenta al  $P_{CT}$  y no tomando en cuenta al  $P_{CT}$

### 7.2.3 Estimación del recableado entre genes de primates utilizando k-meros

Se utilizaron los k-meros correspondientes a la región semilla y como control los k-meros que estén a 1 nucleótido de distancia de semillas de microRNA (ver Métodos). Para cada métrica se contrastó el comportamiento de las 109 familias conservadas de microRNA con respecto al comportamiento de un modelo nulo generado a partir de los k-meros control.

#### 7.2.3.1 Estimación con métrica de Shou

Con la métrica de Shou se observa que el valor de recableado correspondiente a este grupo de microRNAs muestra una tasa de recableado que se comporta similar a la del

modelo nulo, se obtuvieron p-valores empíricos de 0.92, 0.89, 0.78 y 0.76 para humano, chimpancé, gorila y orangután respectivamente (Figura 32).

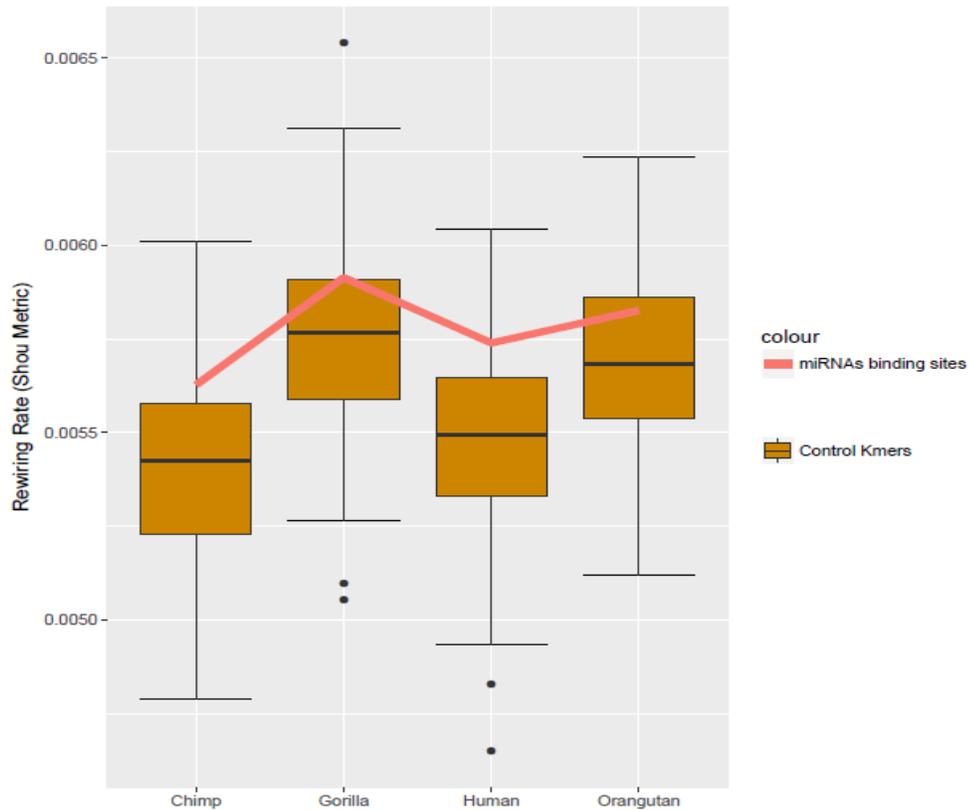


Figura 32 Valor de recableado con métrica de Shou del conjunto de k-meros correspondientes a los microRNAs comparado con el modelo nulo de k-meros control

### 7.2.3.2 Estimación con métrica de Simkin

Utilizando la métrica de Simkin, se estimó el recableado total de los sitios de pegado. Se encontró que para este conjunto de k-meros, correspondientes a los microRNAs más conservados, sus valores de recableado son bajos y significativamente menores con respecto a los k-meros control. Estos resultados se obtuvieron mediante p-valores empíricos menores de 0.01 para todas las especies (Figura 33) .

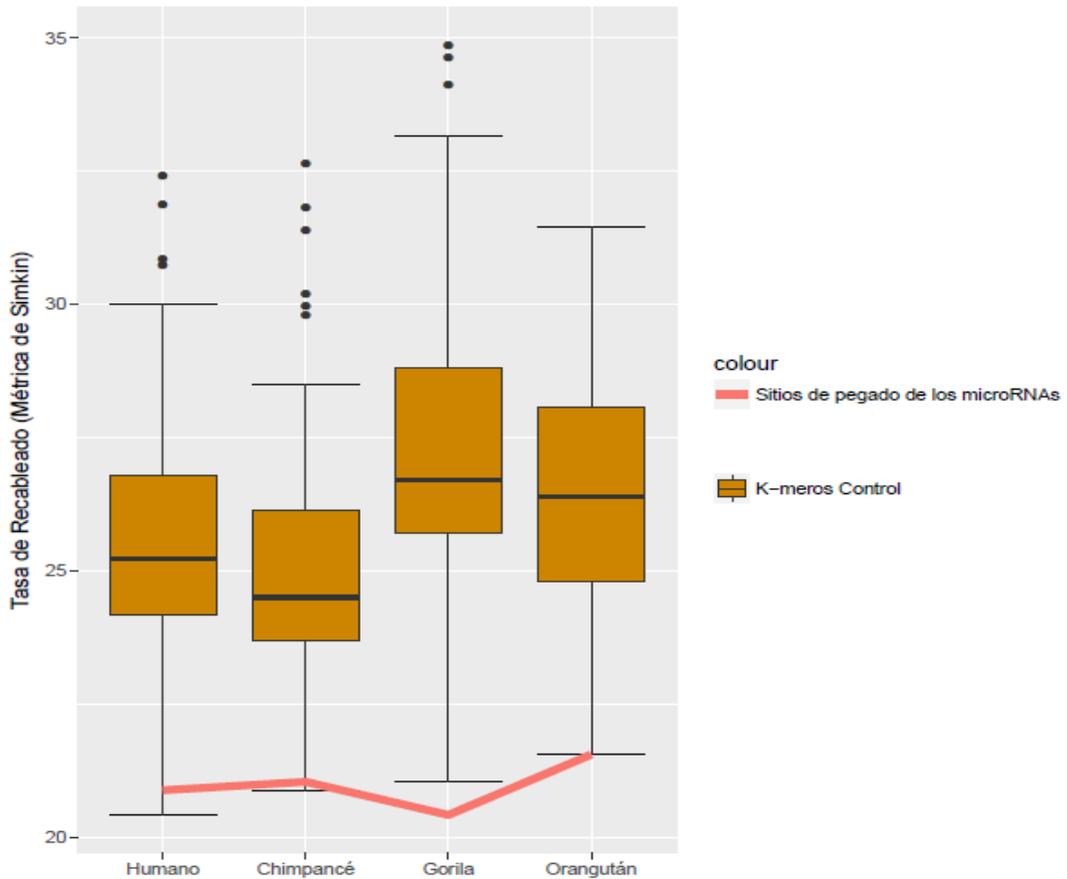


Figura 33 Valor de recableado con métrica de Simkin del conjunto de  $k$ -meros correspondientes a los microRNAs comparado con el modelo nulo de  $k$ -meros control

#### 7.2.4 Estimación del recableado entre genes de primates utilizando la secuencia completa del microRNA maduro

A partir de las predicciones realizadas con TS7 se construyeron las redes correspondientes a cada primate. Se estimó el recableado de las 109 familias de microRNA conservadas con sus respectivos blancos contrastando con el grupo control basado en los valores del CS ++.

#### 7.2.4.1 Estimación con métrica de Simkin

Se estimó el recableado total de sitios de pegado de los microRNA utilizando la métrica de Simkin. Al igual que con la métrica de Shou, se puede ver en la Figura 34, que el recableado de las redes construidas con el conjunto de las mejores predicciones de blancos (anaranjado), es menor comparado con el recableado del control hecho con el conjunto de las peores predicciones de blancos (rojo). Las diferencias son estadísticamente significativas de acuerdo a una prueba de Wilcoxon ( $p < 0.01$ )

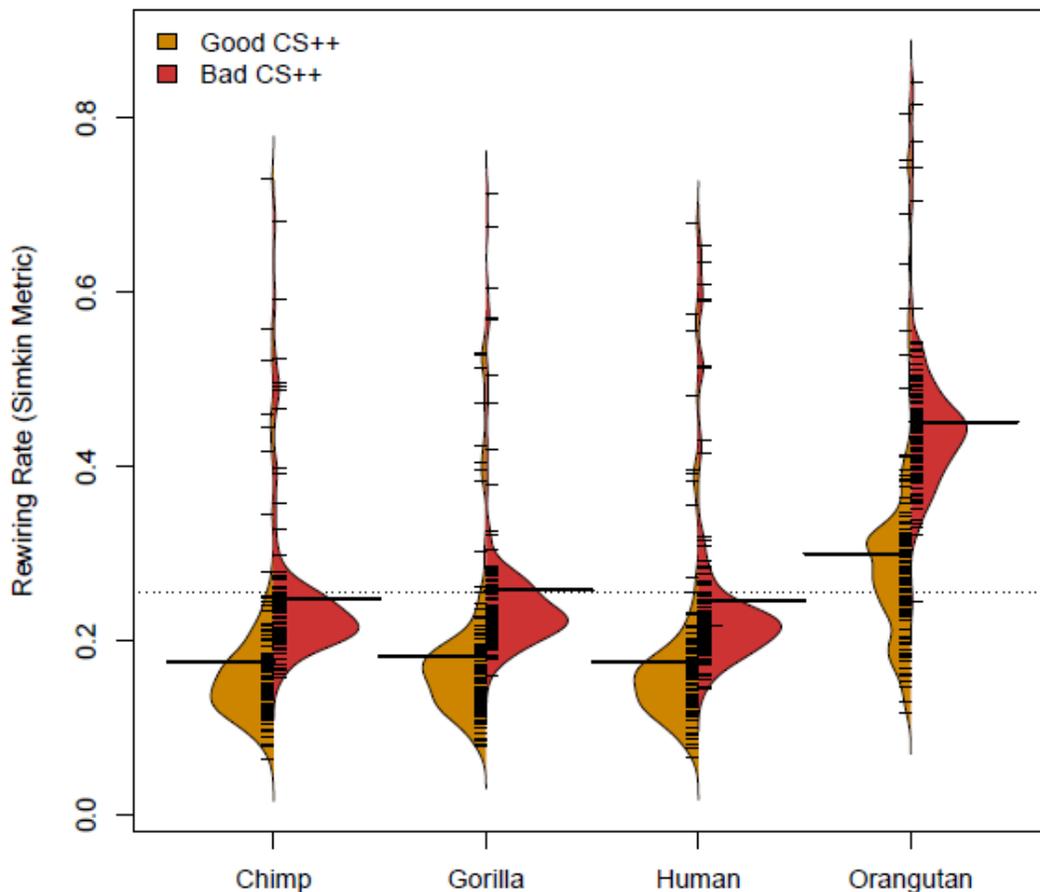


Figura 34 Tasa de recableado de Simkin para cada familia de microRNA utilizando los blancos más probables y los blancos menos probables

#### 7.2.4.2 Estimación con métrica de Shou

Se estimó el recableado entre genes, utilizando la métrica de Shou. En general, como se puede ver en la Figura 35, el recableado de las redes construidas con el conjunto de las

mejores predicciones de blancos, con su distribución representada en anaranjado, es menor comparado con el recableado del control. La tasa de recableado control está hecha con el conjunto de las peores predicciones de blancos cuya distribución se representa en rojo (Prueba de Wilcox,  $p < 0.01$ )

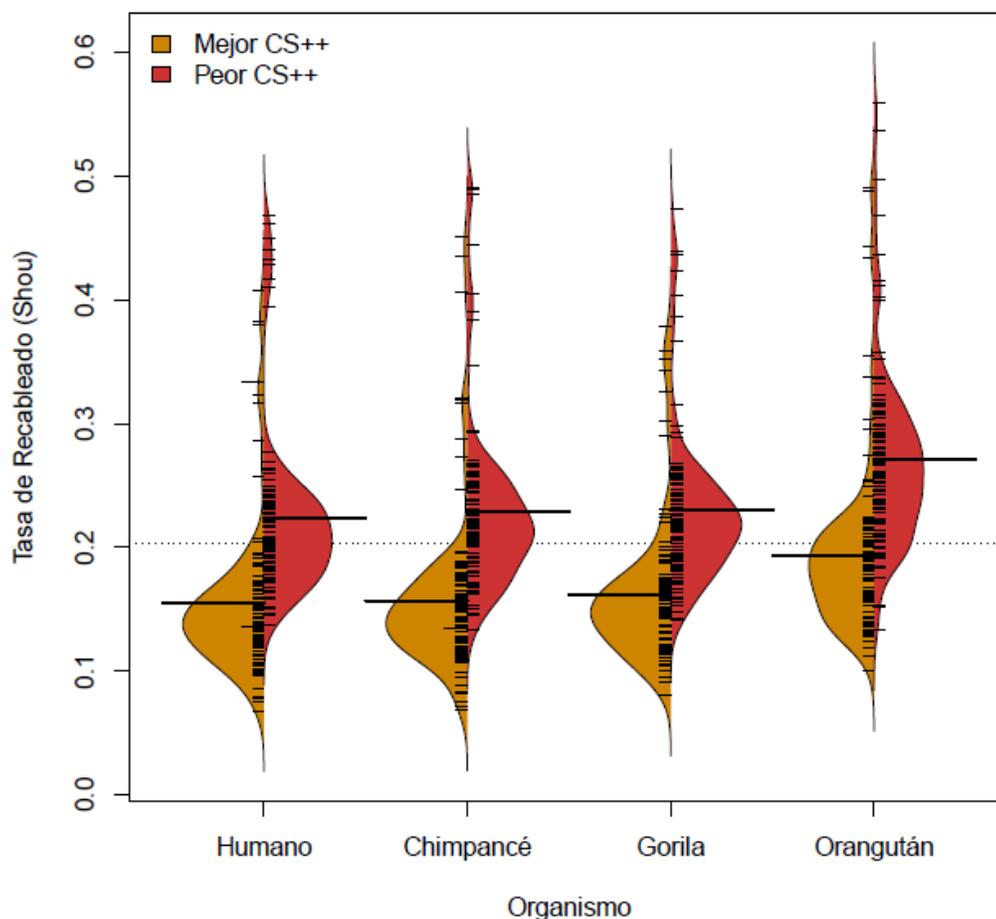


Figura 35 Tasa de recableado de Shou para cada familia de microRNA utilizando los blancos más probables y los blancos menos probables

#### 7.2.4.2.1 Comparación de tasas de recableado entre especies para cada microRNA

El siguiente análisis se hizo para ver si existen diferencias entre las tasas de recableado de los mismos microRNAs entre diferentes primates. Primero se realizaron comparaciones pareadas entre especies para saber si había correlación con los valores de tasas de recableado (Figura 36). Por otro lado, para saber si estas diferencias se observan de igual manera en los controles, pudiendo no ser algo único de los blancos reales, se compararon las tasas de recableado reales con las tasas de recableado del grupo control de blancos mediante el criterio de CS++.

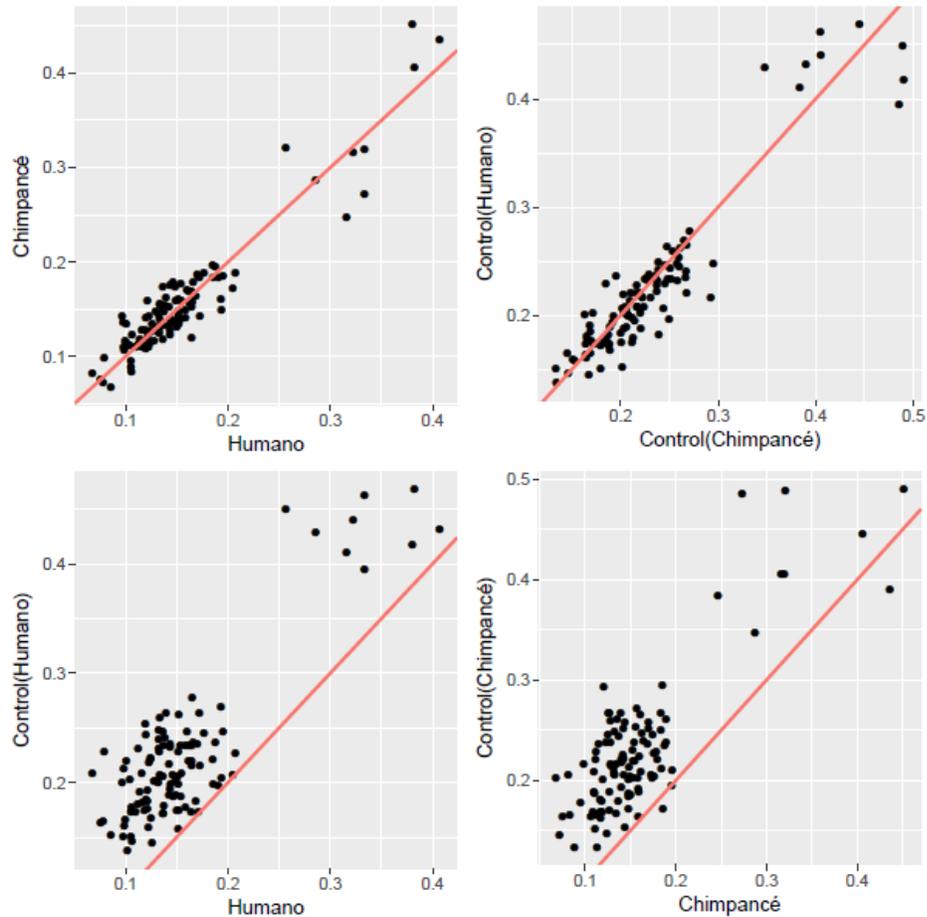


Figura 36 Gráfico de dispersión para observar correlación de los valores de recableado para cada microRNA entre diferentes especies y sus respectivos controles.

En los gráficos de dispersión de la Figura 36 y mediante correlaciones de Pearson, se puede observar que la correlación es más alta cuando se trata de comparar entre especies ( Humano vs Chimpancé,  $R=0.945$ ) que con el control de la misma especie, tanto en el caso de humano (Humano vs Humano Control ,  $R=0.859$ ) como el del chimpancé (Chimpancé vs Chimpancé Control,  $R=0.827$ ). Sin embargo cuando se selecciona el control de humano y chimpancé, y se compara entre ellos, se observa que también es alta la correlación (Humano Control vs Chimpancé Control,  $R=0.934$ ).

Para tener una visión más generalizada entre el total de especies se decidió representar estas comparaciones mediante un heatmap (Figura 37). Cada una de las filas corresponde a una familia de microRNA, y las columnas a cada primate. Las familias que tienen menor recableado están coloreadas de rojo mientras que las familias que tienen tasas más altas de recableado se representan con color azul.

Tasa de Shou Normalizada

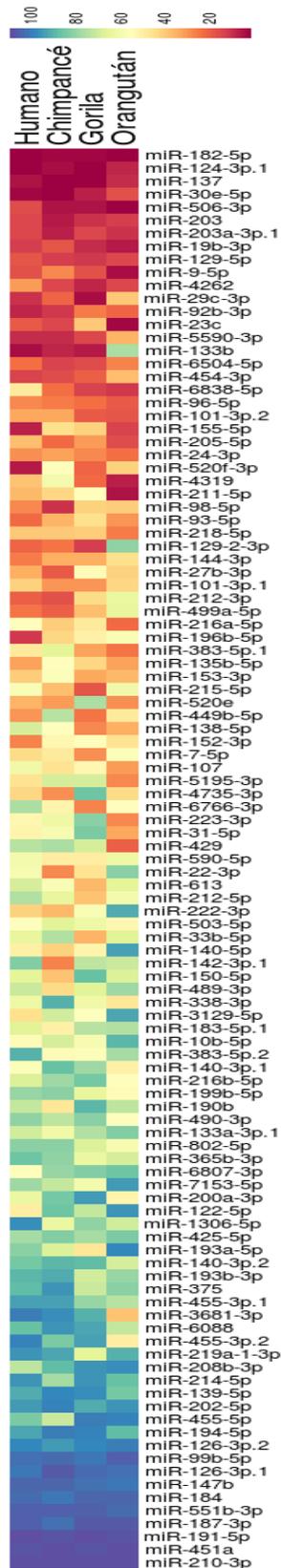


Figura 37 Comparación de las tasas de recableado para cada microRNA nombrado en las filas para cada una de las especies mencionadas en las columnas

En este análisis se puede observar que los microRNAs con más recableado y los que tienen menos recableado se comparten entre especies. Esto se observa en las zonas con tonos más rojos y más azules. Los 5 microRNAs con mayor tasa de recableado corresponden a las familias miR-210, miR-451, miR-191, miR-187 y miR-551. Mientras que las 5 familias con menor tasa de recableado son miR-182, miR-124, miR-137, miR-30e y miR-506.

### 7.2.5 Análisis de categorías funcionales

Se realizaron análisis de enriquecimiento de categorías funcionales de los grupos de blancos de familias con valores extremos de recableado y de los blancos que presentaron mayor recableado por gen.

#### 7.2.5.1 Análisis de categorías funcionales provenientes del recableado por microRNA

Debido a las diferencias del tamaño del conjunto de los blancos de microRNAs con alto recableado y los blancos con bajo recableado, se evaluaron 3 puntos de corte del valor de CS++. Esto se hizo para saber si las categorías obtenidas con distintos tamaños de conjuntos eran robustas a este criterio. A continuación se muestra en un heatmap a las filas como las categorías funcionales de GOSlim (Figura 38). Las columnas representan las funciones de los distintos conjuntos de genes que se obtuvieron variando el corte de CS++ para los blancos de microRNAs con más bajo recableado y más alto recableado.

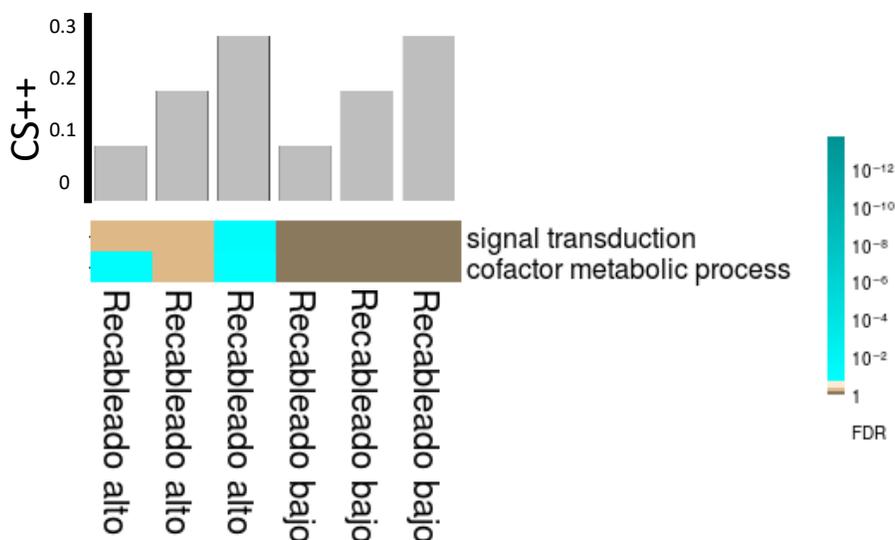


Figura 38 Pruebas de enriquecimientos categorías funcionales para las familias de microRNAs con más y menor recableado con diferentes cortes del CS++, en la barra de color se muestran los valores del logaritmo base 10 del FDR

Para este análisis de blancos de microRNAs con valores extremos en recableado, se puede observar que ninguna categoría funcional fue significativa de acuerdo a los valores de FDR para el caso de bajo recableado. Mientras que para el grupo de blancos de microRNAs con alto recableado las categorías con valores de FDR más bajo están implicadas en procesos metabólicos de cofactores y de catálisis, además en categorías como transducción de señales.

### 7.2.5.2 Categorías funcionales de blancos con valores extremos de recableado

#### 7.2.5.2.1 Análisis con los términos de GOSlim

##### 7.2.5.2.1.1 Recableado por blanco en humano

Como el punto de corte de CS++ no afectó los resultados de enriquecimientos funcionales de recableado por gen (Figura S3), se fijó el punto de corte de CS++ a 0.2 para todos estos análisis de recableado por gen blanco. Como se puede observar en la Figura 39 solamente los genes de procesos de transporte transmembranal tuvieron una tasa alta de recableado y específicamente se puede ver que fue por pérdida de sitios de pegado de los microRNAs.

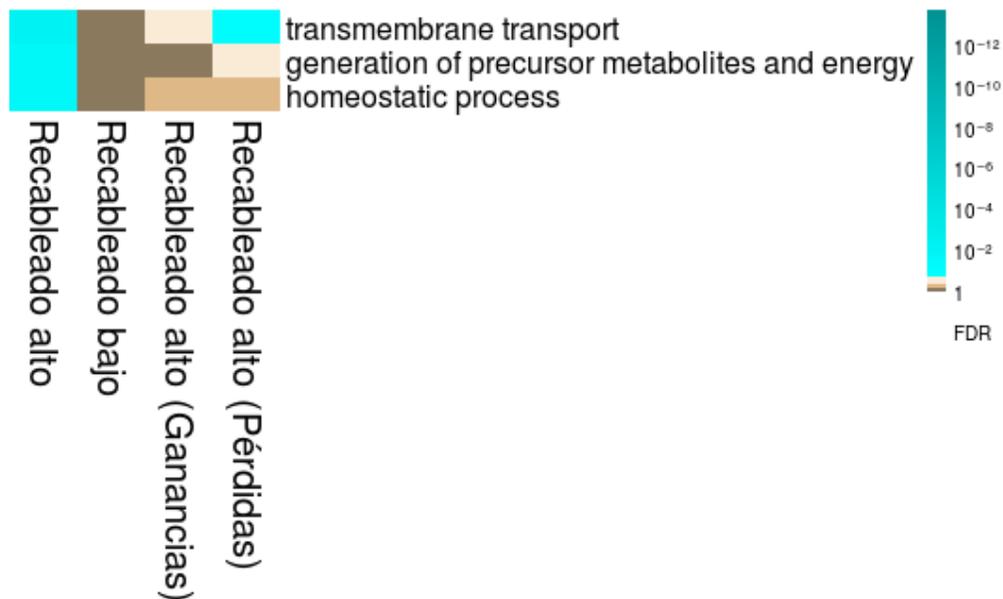


Figura 39. Categorías funcionales de genes con valores extremos de recableado en humano

#### 7.2.5.2.1.2 Comparación de categorías funcionales de blancos con valores extremos de recableado entre humano y chimpancé

Se compararon las categorías enriquecidas en chimpancé con las del humano (Figura 40). A diferencia del humano, en chimpancé se ve un conjunto más grande de categorías funcionales enriquecidas cuyos genes tienen valores extremos de recableado. Para el caso de genes con bajo recableado se observa que están involucrados en procesos de movilidad celular, locomoción y organización de membrana. Por otro lado se observa que aquellos que tienen alto recableado están relacionados también con procesos de transporte transmembranal y además con pigmentación.

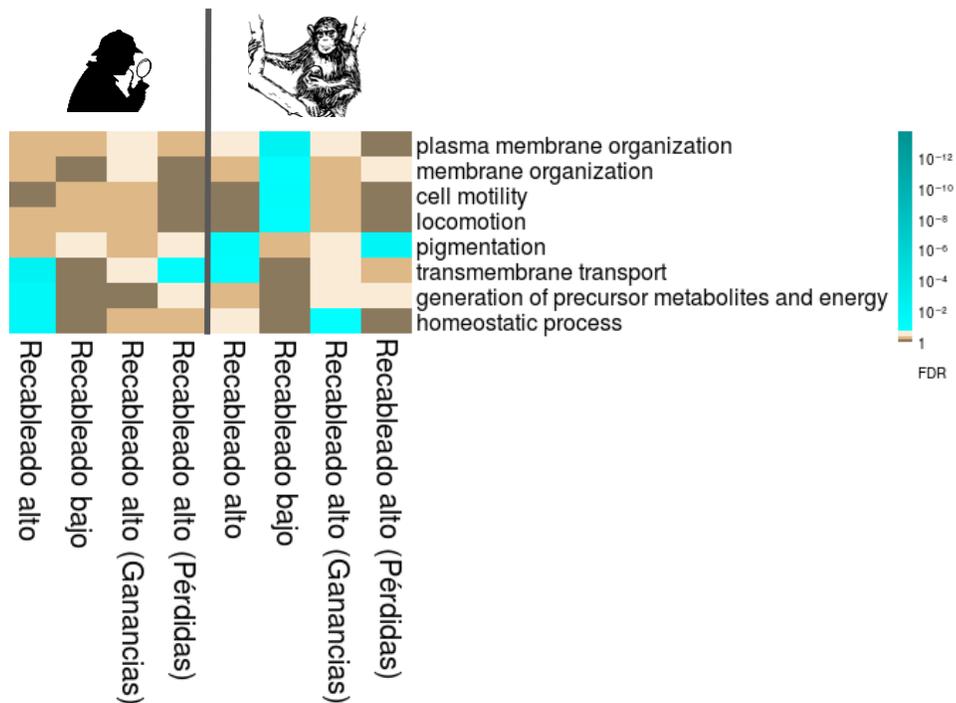


Figura 40 Comparación entre humano y chimpancé de las categorías funcionales de sus blancos con valores extremos de recableado

#### 7.2.5.2.2 Análisis con ReVIGO

Los resultados de ReVIGO para las categorías funcionales enriquecidas en humano se presentan a manera de TreeMap. Este tipo de gráficos muestran de manera general los términos GO que estén enriquecidos. A estos términos los agrupa de acuerdo a su

relación en la jerarquía para disminuir el número de términos redundantes. Los grupos se representan con diferentes colores y se mencionan las categorías de ese grupo que tengan mayor enriquecimiento. Las categorías funcionales enriquecidas de manera general están relacionadas con señalización, algunos procesos neuronales, procesos de transporte, respuesta celular entre otras (Figura 41).

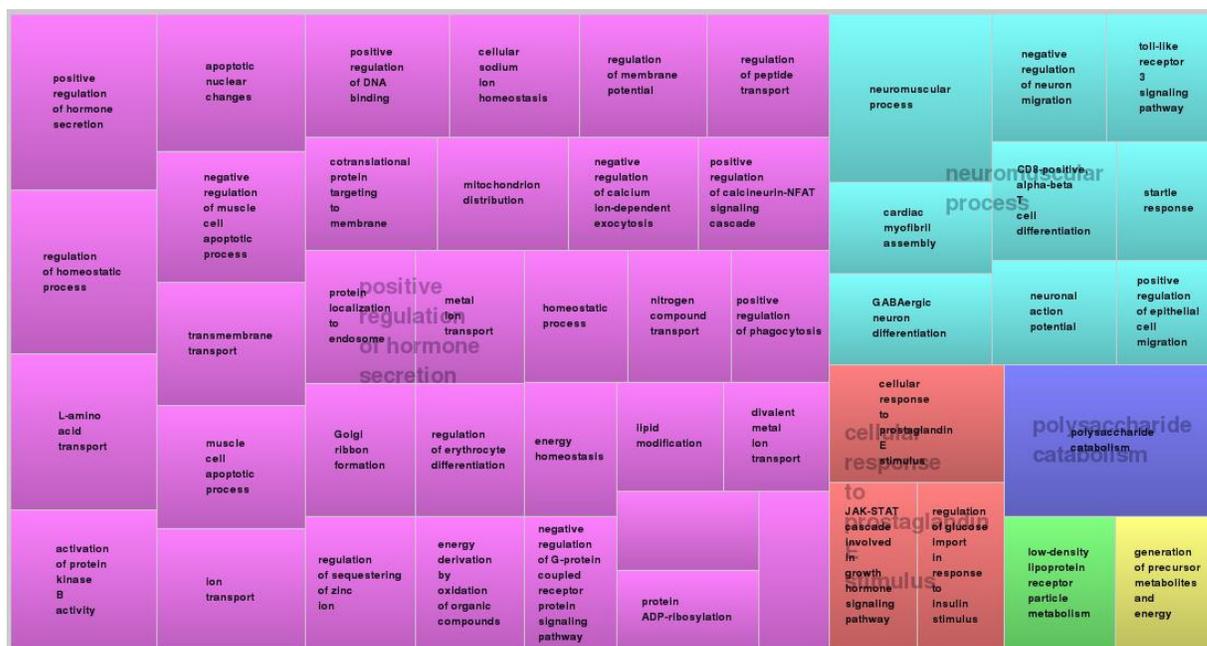


Figura 41 Resultado de ReviGO de las categorías funcionales enriquecidas para los blancos con mayor recableado

Para el caso de las categorías funcionales de blancos con valores bajos de recableado se tienen funciones relacionadas con procesamiento de RNA, organización de membrana, otras relacionadas con la respuesta a daño y otros tipos de estrés (Figura 42).



Figura 42 Resultado de ReviGO de las categorías funcionales enriquecidas para los blancos con menor recableado.

## 7.3 Análisis de un conjunto alternativo de 3' UTRs de los primates obtenidas de un alineamiento de mamíferos

### 7.3.1 Procesamiento de las 3'UTRs

Utilizando las secuencias de 3'UTRs del conjunto de datos de Leclerq *et al.* de 2017, se obtienen las siguientes cantidades de genes en cada paso del proceso que se les hizo (Figura 43).



Figura 43 Diagrama de pasos del procesamiento de las 3'UTR con las cantidades de genes obtenidos en casa paso

### 7.3.2 Estimación de recableado entre genes mediante métrica de Shou

En este análisis se obtuvieron tasas de recableado para cada familia de microRNA y se compararon con sus respectivos controles hechos a partir de los blancos menos probables. Similar a lo observado anteriormente, se puede ver en la Figura 44 que los microRNAs tienen tasas de recableado más bajas con respecto a sus controles.

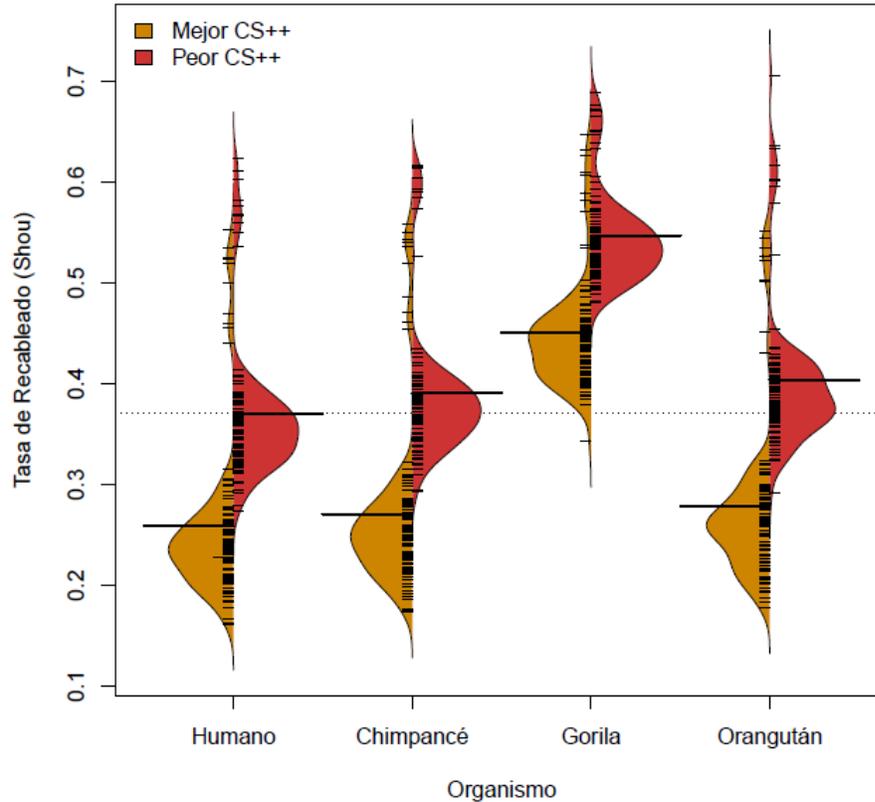
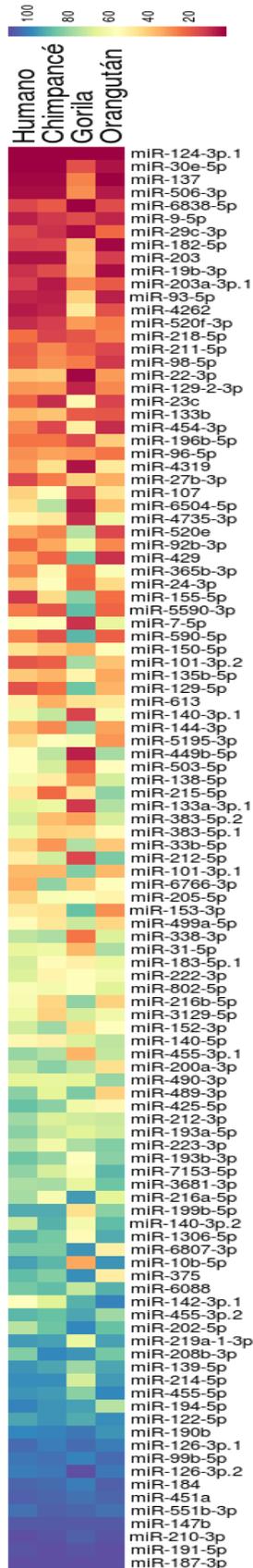


Figura 44 Tasa de recableado de Shou para cada familia de microRNA utilizando los blancos más probables y los blancos menos probables

Para saber si también había consistencia con las familias de microRNAs con alto recableado y las familias de microRNAs con bajo recableado se hicieron heatmaps donde se representa la tasa de recableado para cada especie. Como se puede observar en la Figura 45, son las mismas familias las que tienen más recableado y las que tienen menos recableado en las diferentes especies.

Tasa de Shou Normalizada



*Figura 45 Comparación de las tasas de recableado para cada microRNA nombrado en las filas para cada una de las especies mencionadas en las columnas*

### 7.3.3 Análisis de categorías funcionales

Se realizaron análisis de enriquecimiento funcional con este conjunto de secuencias 3'UTR. Además se puede ver en la Figura 46 que el nivel de significancia aumentó considerablemente con este conjunto más grande de secuencias.

#### 7.3.3.1 Blancos de familias de microRNAs con valores extremos de recableado

Para el análisis de categorías funcionales de este conjunto se utilizaron los genes blanco de los microRNAs con mayor y menor tasa de recableado. Se hicieron pruebas de enriquecimiento y se obtuvo el valor de FDR de estas pruebas. Como se puede ver en el heatmap de la Figura 46, los valores azules corresponden a las categorías más significativas que corresponden a procesos de desarrollo de estructuras anatómicas y desarrollo pero estas categorías salieron enriquecidas tanto en microRNAs con alto como bajo recableado. En el caso de los blancos de microRNAs con menor tasa de recableado solamente tenemos a la pigmentación y algunos tipos de transporte. Por otro lado, para los blancos de microRNAs con tasas altas de recableado, se tienen categorías funcionales relacionadas con maduración del desarrollo y procesos neurológicos. Sin embargo, se observa una gran cantidad de categorías enriquecidas para ambas situaciones, microRNAs con alto y bajo recableado.

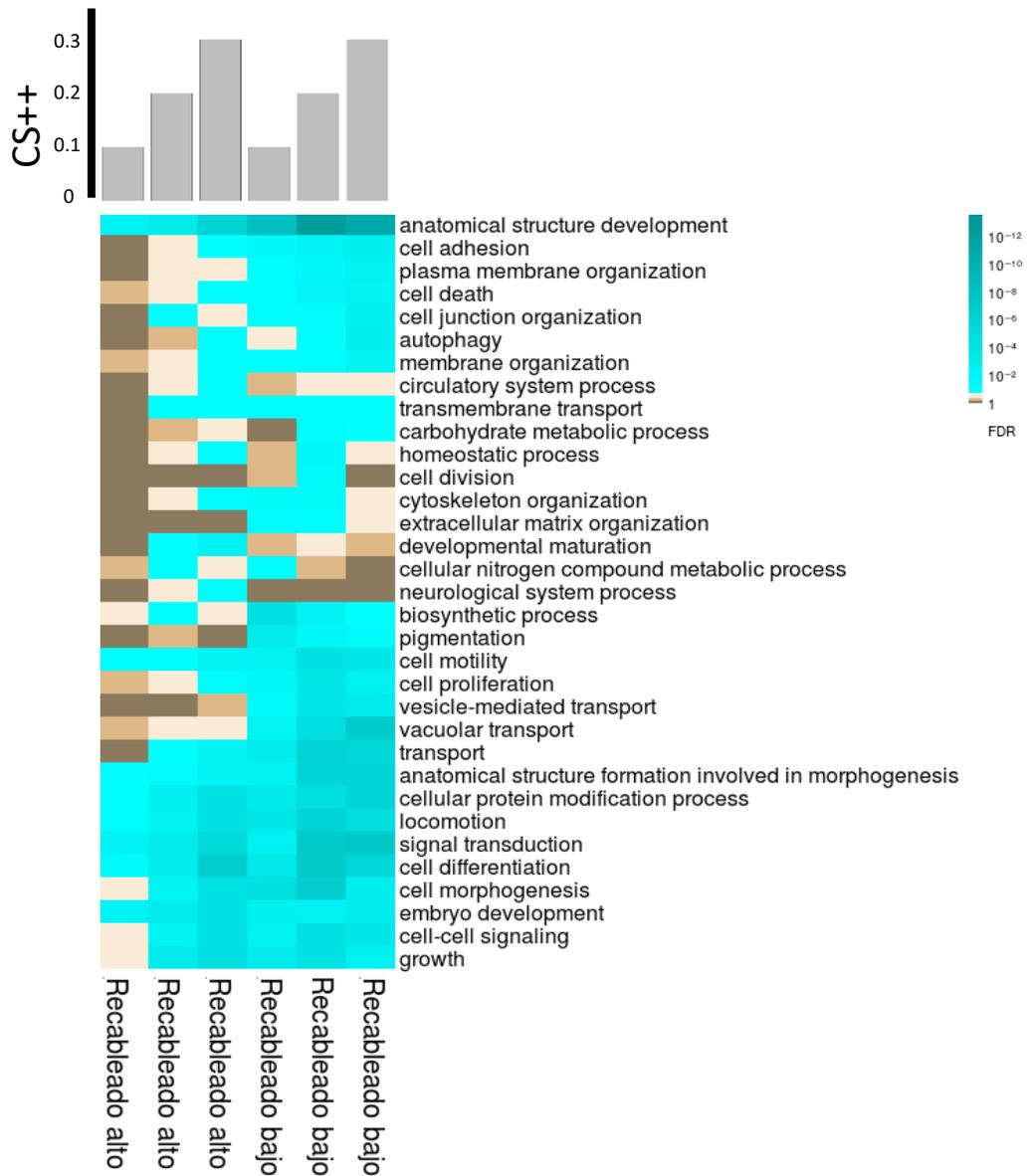


Figura 46 Pruebas de enriquecimientos categorías funcionales para las familias de microRNAs con más y menor recableado con diferentes cortes de CS++, en la barra de color se muestran los valores del FDR

### 7.3.3.2 Categorías funcionales de blancos con valores extremos de recableado

#### 7.3.3.2.1 Análisis con los términos de GOSlim

Se analizaron los enriquecimientos de categorías funcionales para los blancos que tuvieran valores de recableado muy altos y muy bajos, para identificar si algunas de estas funciones estaban bajo alguna presión de selección. Como se puede ver en el heatmap de la Figura 47, los valores azules corresponden a las categorías más significativas, por ejemplo para los blancos con mayor recableado se observan procesos como los del sistema nervioso, desarrollo de estructuras anatómicas, algunos procesos de señalización y locomoción. Para el caso de los blancos que tienen menor recableado se observan al transporte nucleocitoplasmático, locomoción, proliferación y muerte celular.

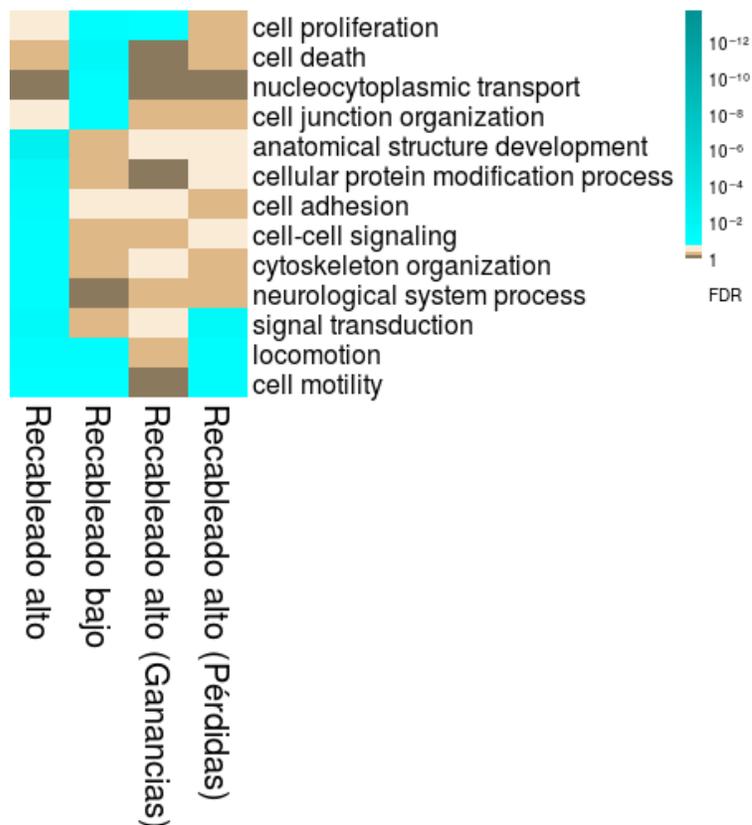


Figura 47 Pruebas de enriquecimientos de categorías funcionales para los blancos de microRNAs con valores extremos de recableado por kilobase, en la barra de color se muestran los valores del FDR



### 7.3.3.2.2 Análisis con ReVIGO

Para los blancos con alto recableado las categorías funcionales que resultaron significativas son algunas relacionadas de manera general con desarrollo del sistema nervioso, sinapsis y señalización. A continuación en la Figura 49 se muestran las categorías enriquecidas en un diagrama TreeMap.

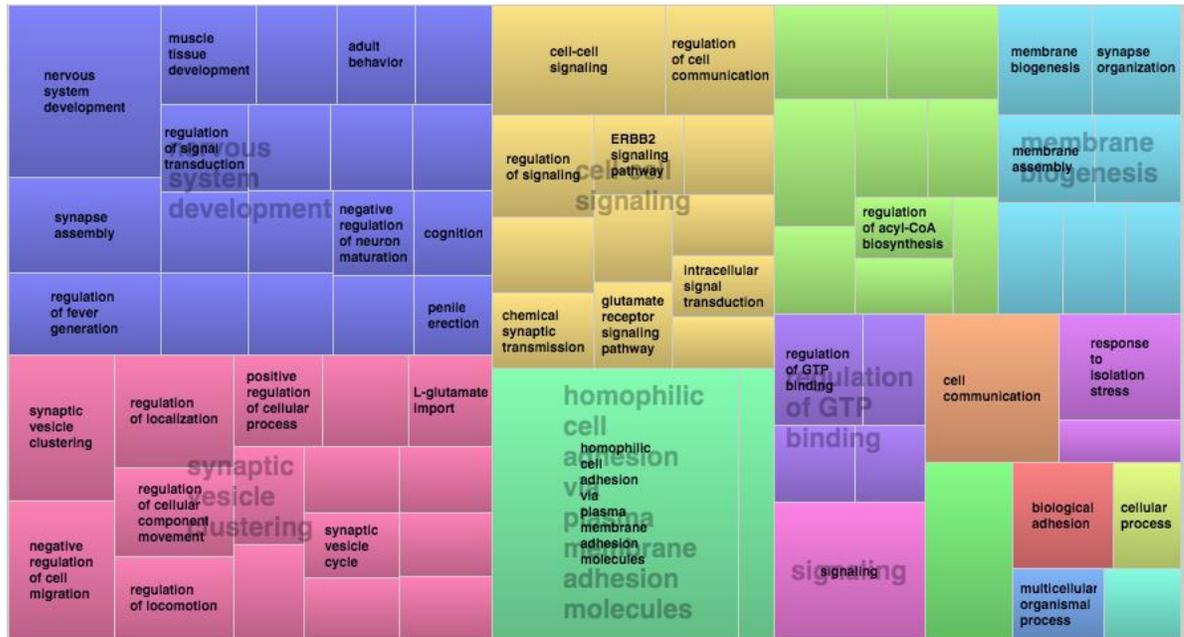


Figura 49 Resultado de ReviGO de las categorías funcionales enriquecidas para los blancos con mayor recableado.

Por otro lado, para los blancos con bajo recableado se pueden ver algunas categorías relacionadas con el procesamiento del RNA, y con la regulación de algunos procesos de señalización como el pegado de receptores (Figura 50). También se observa una categoría relacionada con la cornificación de la piel y que a su vez involucra procesos centrales como muerte celular.

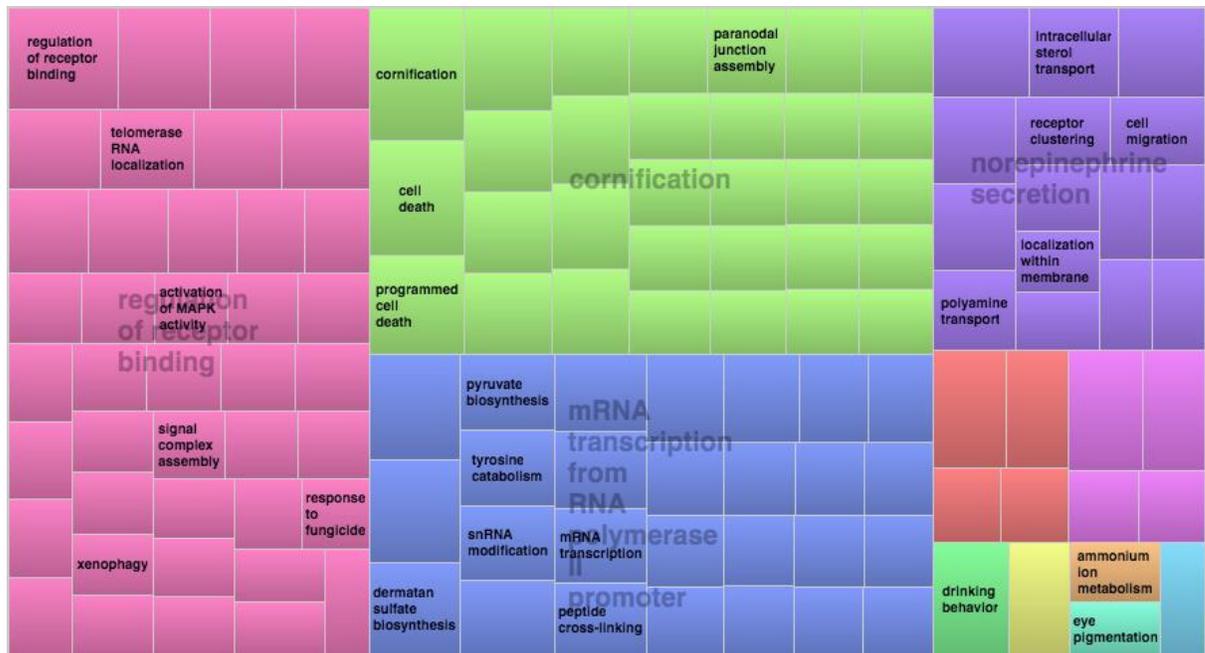


Figura 50 Resultado de ReviGO de las categorías funcionales enriquecidas para los blancos con menor recableado.

### 7.3.4 Uso de datos de expresión diferencial a partir de transcriptomas de humano y chimpancé

Se obtuvieron los valores de expresión de genes reportados en el trabajo de Bauernfeind *et al.* del 2015 para los tejidos de corteza del cíngulo anterior y del núcleo caudado en humano y chimpancé. Se identificaron los valores para los genes que tienen tasas grandes de recableado y se observó el nivel de expresión de estos genes. Esto se realizó para saber si las ganancias de sitios de pegado en un gen de una especie y pérdidas de sitios de pegado en el mismo gen en la otra especie se ven reflejadas en sus niveles de expresión para alguno de los tejidos.

A continuación, en las Figura 51 y 52 se muestran los gráficos de volcán que representan el nivel de expresión de genes entre humano y chimpancé. En el eje de las X se tienen los valores de expresión representados por el Fold Change en escala logarítmica. En el eje de las Y se representa el valor de significancia estadística para ese gen.

Estos gráficos corresponden al contraste chimpancé con humano. Es decir del lado izquierdo están los genes con menos expresión en chimpancé con respecto al humano.

Por otro lado la parte derecha de este gráfico, donde el logaritmo del Fold Change es mayor a cero, representa aquellos genes que tienen mayor expresión en chimpancé con respecto a humano. En la Figura 51 se muestra en color amarillo aquellos genes que sufrieron pérdida de sitios de pegado en humano y ganancia de sitios de pegado en chimpancé esperando que estos genes estuvieran principalmente orientados a la izquierda. Para el mismo tejido de corteza del cíngulo anterior se identificaron los genes que sufrieron ganancia de sitios de pegado en humano y pérdida de sitios de pegado en chimpancé esperando que estos genes se encontraran del lado derecho de la gráfica de volcán. Estos genes se muestran en color azul en la Figura 51.

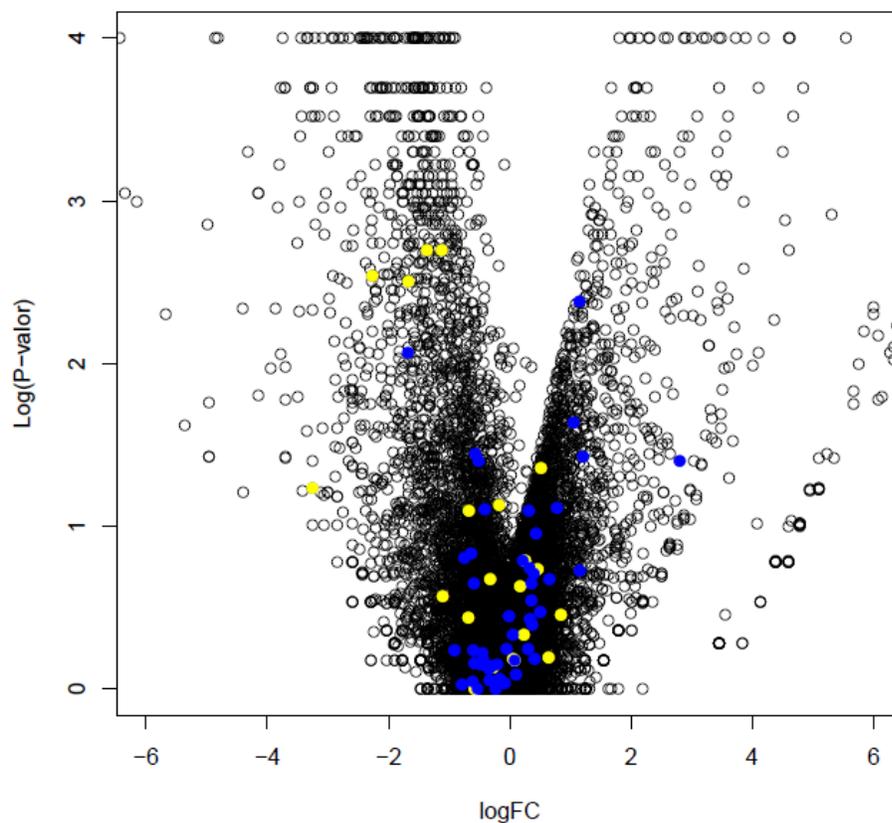


Figura 51 Gráfico de volcán para los niveles de expresión y valores de significancia estadística del contraste de los genes expresados en la corteza del cíngulo anterior entre humano y chimpancé. En amarillo se muestran los genes que perdieron sitios de regulación en humano y ganaron sitios de regulación en chimpancé. En azul se muestran los genes que perdieron sitios en chimpancé y ganaron sitios en humano.

Para el núcleo caudado se realizó el mismo análisis identificando genes que en una especie tuvieran ganancias de sitios mientras que en la otra pérdidas. Se comparó también el logaritmo del cambio en el nivel de expresión entre los 2 conjuntos de genes y

se observa también una tendencia para orientarse al lado correspondiente con los niveles de expresión esperados. (Figura 52).

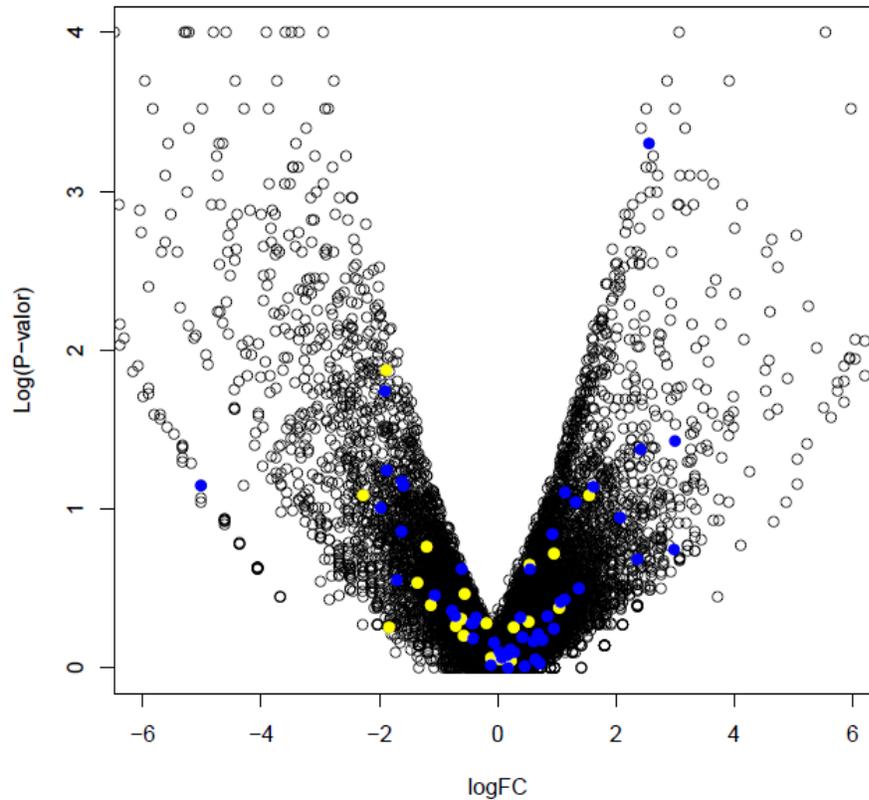


Figura 52 Gráfico de volcán para los niveles de expresión y valores de significancia estadística del contraste de los genes expresados en la corteza del núcleo caudado entre humano y chimpancé.

Se realizaron pruebas de Wilcoxon para realizar los contrastes de los valores de expresión de los 2 conjuntos de genes los que ganaron sitios en humano y perdieron sitios en chimpancé y viceversa (Tabla 5).

Tabla 5 Tejidos con sus respectivos p-valores de las pruebas de Wilcoxon para los valores de expresión de los 2 conjuntos de genes

Tejido	Wilcoxon Test
	0.067
	0.032

## 7.4 Familias de microRNAs con distintos niveles de conservación

### 7.4.1 Análisis de recableado con un conjunto de familias de microRNAs conservadas en mamíferos.

Se realizó el mismo tipo de análisis utilizando ahora el conjunto de las familias de microRNAs conservadas solamente dentro del linaje de los mamíferos. En la Figura 53 se observa que los valores de recableado entre los blancos más probables son menores que cuando se usa el grupo control con un conjunto de malas predicciones de blancos. Este comportamiento es similar al caso de las familias de microRNAs conservadas en todos los vertebrados (Figura 35 y 44).

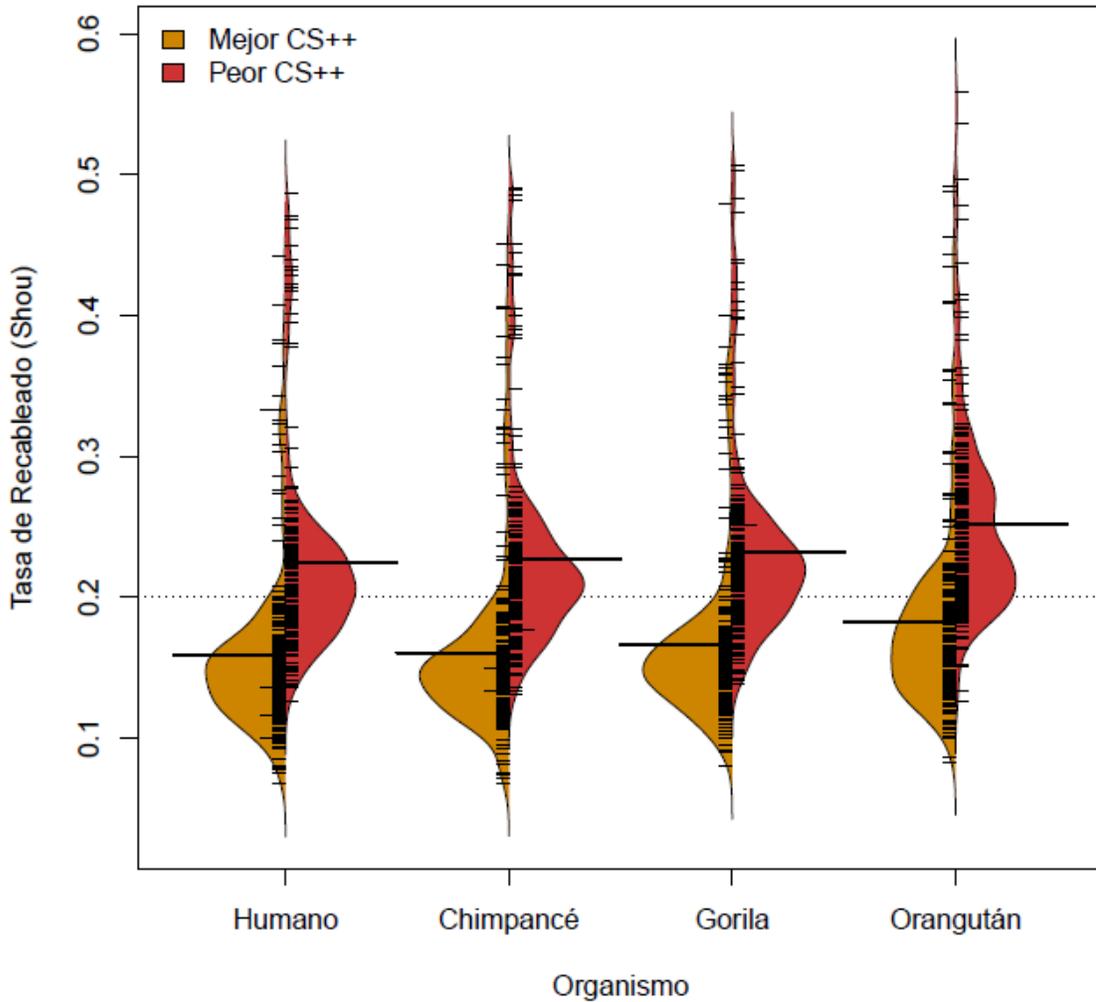
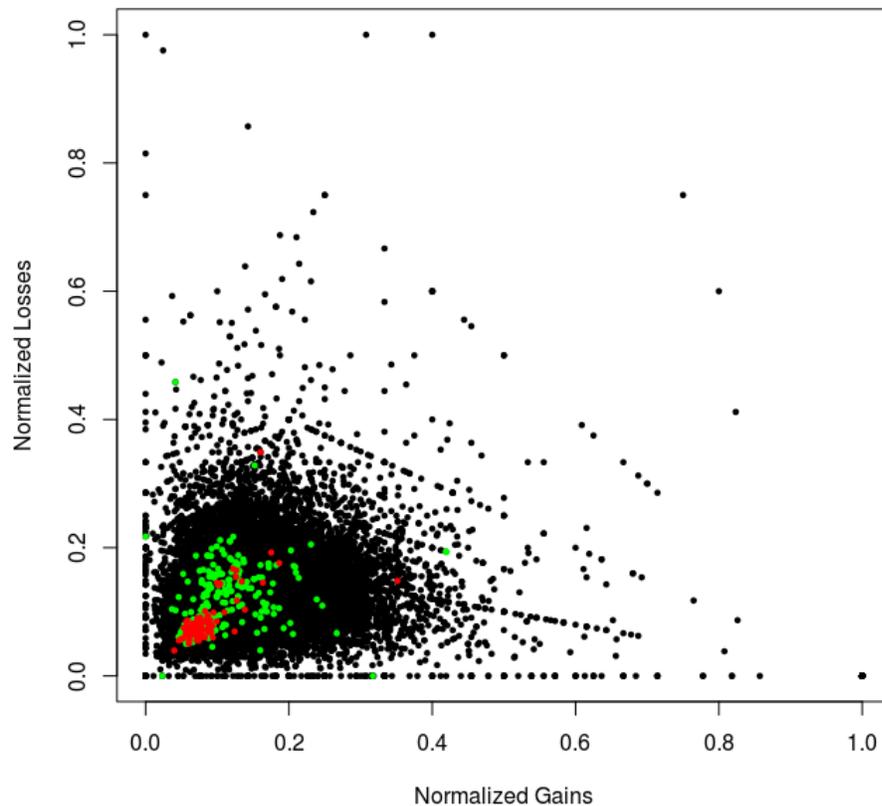


Figura 53 Tasa de recableado de Shou para cada familia de microRNA utilizando los blancos más probables y los blancos menos probables

7.4.2 Familias de microRNAs que han sufrido cambios exclusivamente en humano  
 Se realizaron estimaciones con las aproximadamente 320 familias únicas a humano para saber si también sus blancos se encontraban bajo la misma presión de selección. Este análisis se hizo a nivel de los 8-meros correspondientes a la región semilla de estas familias y se colorearon en verde en la Figura 54. Esta figura muestra el universo de las 65,536 8-meros posibles. Se resaltaron aquellos que corresponden a la semilla de cierto tipo de microRNA. Los ejes X y Y corresponden a la tasa de ganancias y la tasa de pérdidas respectivamente.

En rojo se observa que los k-meros correspondientes a sitios blanco de las semillas de los microRNAs más conservados, tienen tasas bajas de pérdida y ganancia comparado con los k-meros correspondientes a los sitios blanco de las semillas de los microRNAs más recientes. Se realizaron pruebas de enriquecimiento para saber si el cuadrante con menores tasas de recableado se encuentra enriquecido en los blancos de microRNAs recientes (Prueba hipergeométrica;  $p < 0.01$ ).



*Figura 54 Tasas ganancia y de pérdidas para cada uno de los k-meros correspondiente a los sitios de pegado de los microRNAs. En rojo los correspondientes a las familias más conservadas y en verde los correspondientes a las familias más recientes en humano.*

## 7.5 Conjunto de secuencias de regiones 3'UTRs del grupo de los Cíclidos del Lago Malawi

La siguiente etapa consistió en aplicar la metodología desarrollada al conjunto de las 3'UTRs que se han estudiado en el grupo de los cíclidos del lago Malawi publicadas por Loh *et al.* en 2011. Las 3'UTR de los cíclidos son de interés para entender los efectos de la regulación de los microRNAs en la evolución fenotípica de los mismos.

### 7.5.1 Recableado de microRNA maduro

#### 7.5.1.1 Análisis de recableado de Simkin

Se utilizaron alrededor de 115 familias de microRNAs que estuvieran presentes en peces. Se estimó la tasa de recableado de Simkin para cada familia de microRNA de cíclidos y se graficaron las distribuciones de las tasas siguiendo la metodología planteada. Como se puede observar en la Figura 55 a diferencia de los primates el recableado de las redes de regulación solamente en algunos cíclidos es más alto para el conjunto de los microRNAs con buenas predicciones de blancos en comparación con el control generado con las predicciones menos probables de blancos.

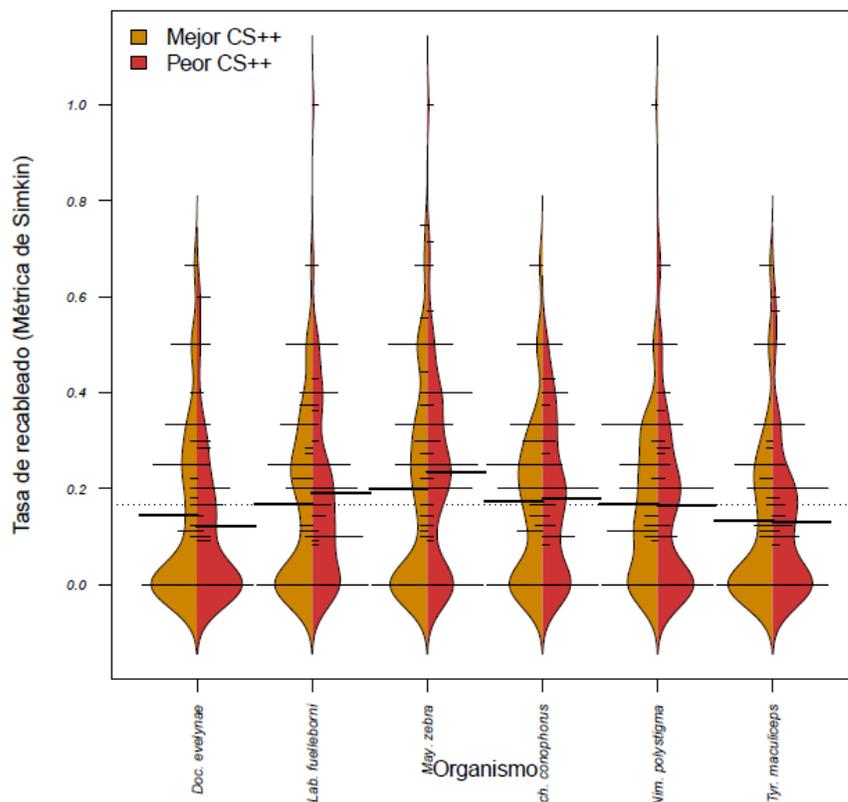


Figura 55 Distribución de las tasas de recableado con métrica de Simkin para cada microRNA en diferentes linajes de cíclidos para los sitios blanco más probables en amarillo y los sitios blanco menos probables en azul.

Se realizó la comparación entre las diferentes especies de cíclidos de sus métricas de recableado para cada microRNA, para identificar aquellos que tuvieran patrones de alta tasa de recableado y baja tasa de recableado (Figura 56). La mayoría de los comportamientos de recableado no se comparten entre todas las especies, a excepción de miR-135a que tiene alta tasa de recableado casi en todas las especies. Por otro lado, miR-210 y miR-26 tienen valores bajos de recableado consistentes en varias especies. Sin embargo no hay patrones que sean en su totalidad claros.

Tasa de Shou Normalizada

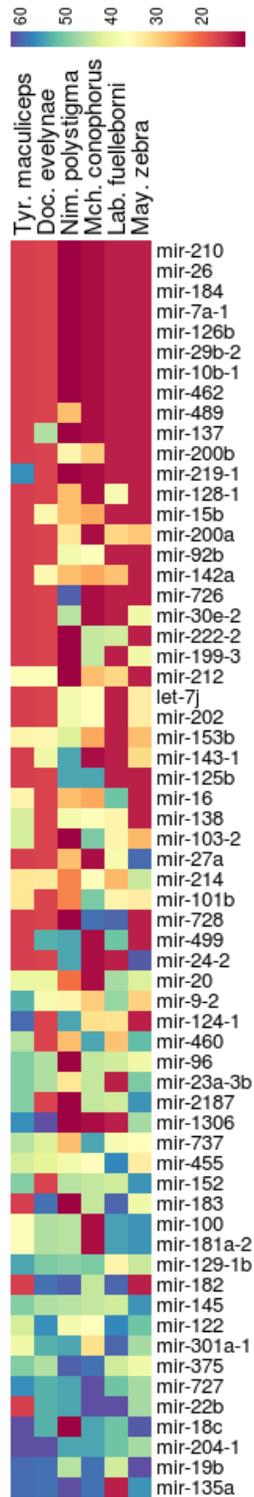


Figura 56 Comparación de las tasas de recableado con métrica de Simkin para cada microRNA nombrado en las filas para cada una de las especies mencionadas en las columnas

### 7.5.1.2 Análisis de recableado de Shou

También se graficaron las tasas de recableado utilizando la métrica de Shou (Figura 57). En esta situación no hay un patrón claro y las diferencias no son significativas entre los controles de CS++ y las medidas con buen criterio de CS++.

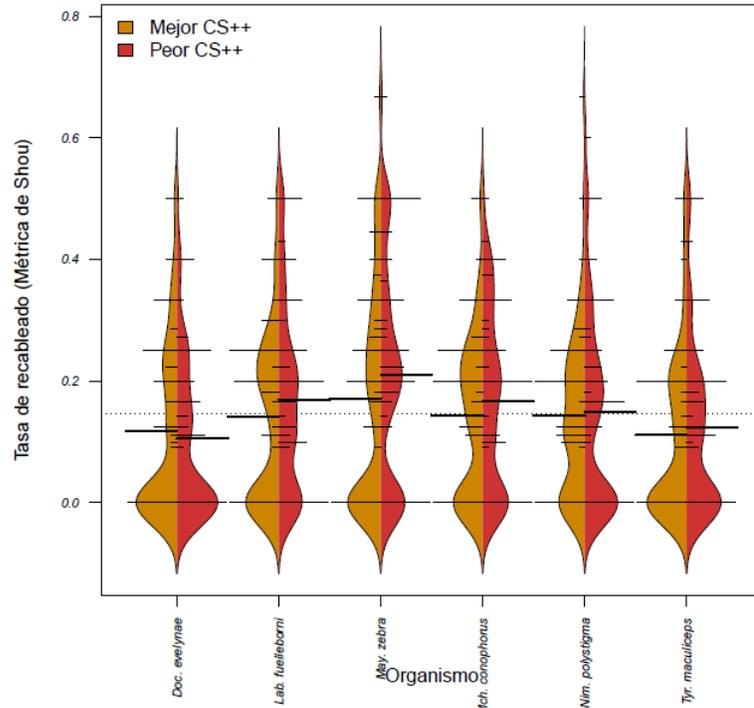


Figura 57 Distribución de las tasas de recableado con métrica de Shou para cada microRNA en diferentes linajes de cíclidos para los sitios blanco más probables en amarillo y los sitios blanco menos probables en azul

A continuación se presentan los microRNAs y sus respectivas tasas de recableado entre diferentes especies (Figura 58). Como se puede ver los microRNAs con tasas más bajas de recableado en todas las especies son miR-210, miR-26 y miR-184 y los que tienen tasa más alta de cambio en todas las especies son miR-135a, miR-204 y miR-19b.

Tasa de Shou Normalizada

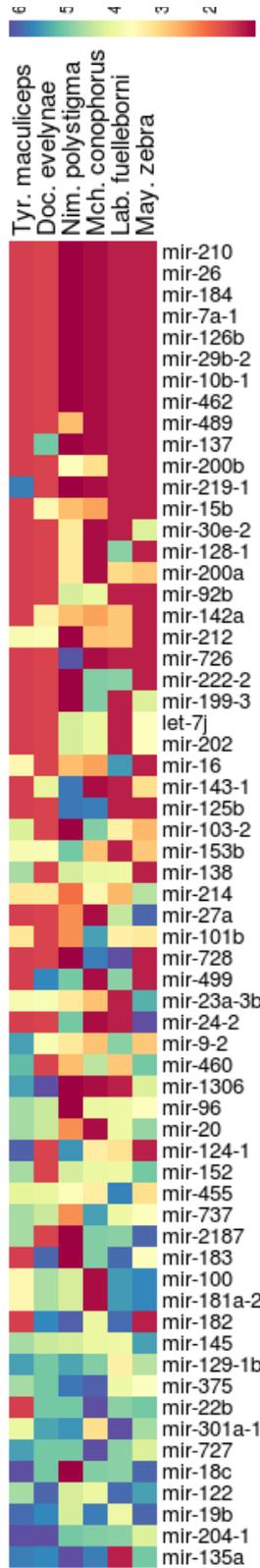


Figura 58 Comparación de las tasas de recableado con métrica de Shou para cada microRNA nombrado en las filas para cada una de las especies mencionadas en las columnas

## 8. Discusión

### 8.1 Efecto en los valores de recableado por las secuencias de regiones 3'UTR utilizadas

Se procesaron de 2 maneras diferentes las 3'UTRs en primates. En dichos análisis se pueden observar patrones diferentes en las tasas de recableado. Con el procesamiento de Simkin *et al.* propuesto en el 2014 se observa al orangután con tasas de recableado mayores que el resto de los primates comparados en este estudio. Por otro lado con el método de procesamiento de Leclercq *et al.* de 2017 el gorila es el que tiene valores de tasa de recableado más altos con respecto a los otros primates comparados.

Primero se pensó que esto podría ser debido a la gran diferencia que existe entre el número de 3'UTRs obtenidas con uno u otro método, 8,260 en el caso de Simkin *et al.* y 14,800 en el caso de Leclercq *et al.* Para saber si esta era la causa, se utilizó la intersección de las secuencias de ambos métodos, es decir si las secuencias que tenían diferentes entre estos métodos hacía que el gorila tuviera recableado más alto en un análisis y en el otro el orangután. Utilizando la intersección de las secuencias, se pudieron seleccionar alrededor de 5,000 secuencias de las regiones 3'UTRs presentes en ambos conjuntos de datos. El resultado esperado era que la tasa de recableado fuera similar al resultado con las secuencias de Simkin *et al.* donde el orangután tenía recableado más alto. Como se puede observar en la Figura 59 el resultado no fue el esperado y el gorila seguía teniendo recableado más alto.

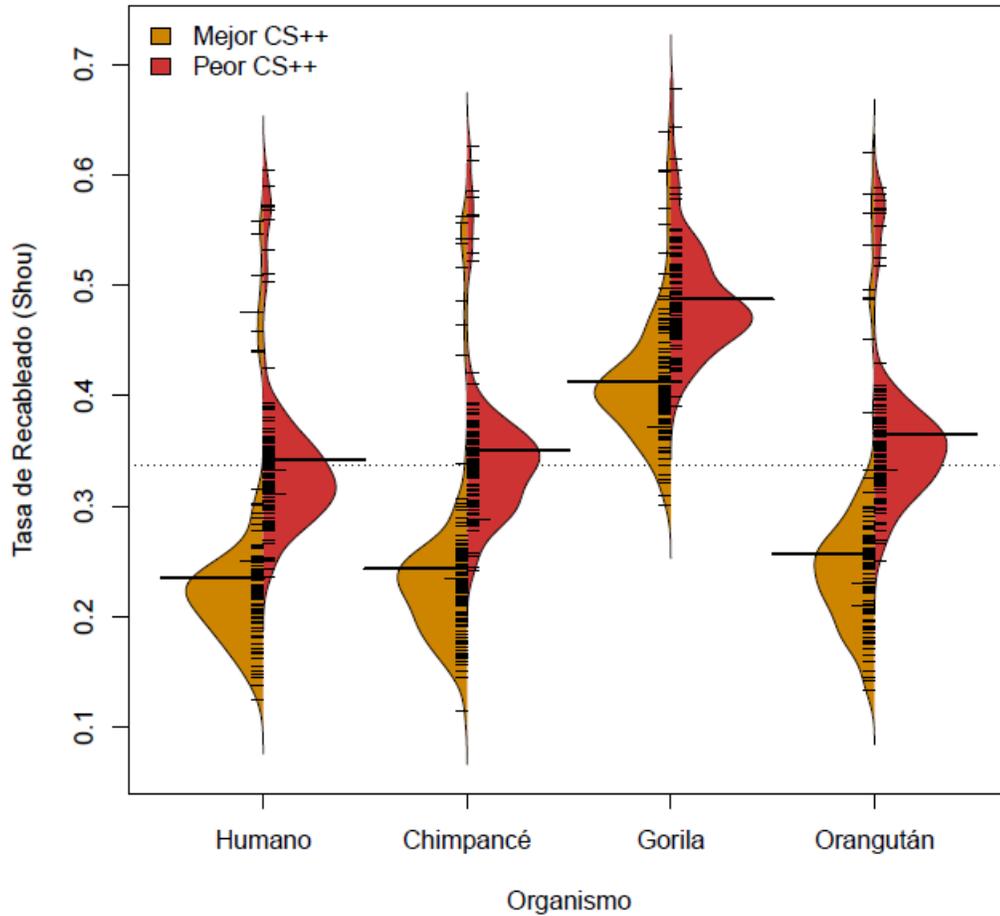


Figura 59 Tasa de recableado de Shou para la intersección de los genes provenientes de diferentes alineamientos

También se observaron algunas diferencias en los enriquecimientos funcionales usando los diferentes métodos. A continuación se muestran los resultados de realizar la intersección de blancos provenientes del alineamiento de mamíferos Leclercq et al. del 2017 con los provenientes de Simkin et al. del 2014. Esto se hizo para ver si estas diferencias se deben a que son conjuntos de genes diferentes. Como se puede ver en la Figura 60, las categorías enriquecidas en el método de Leclercq con la intersección solo se conservan las categorías de locomoción y desarrollo de estructuras anatómicas en común.

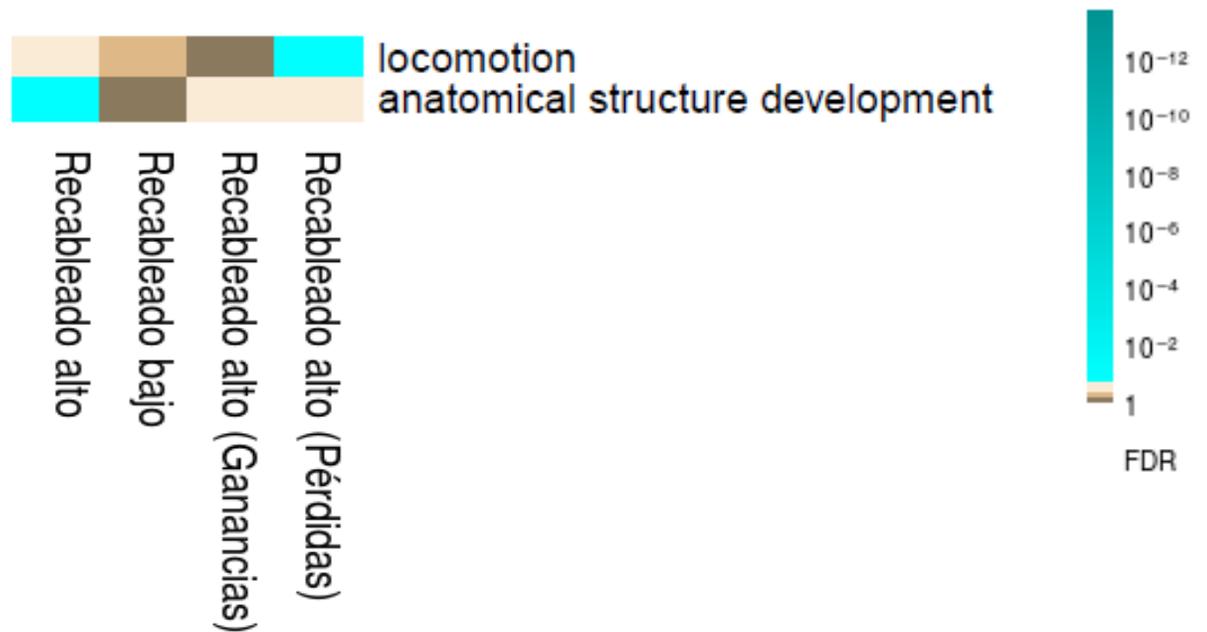


Figura 60 Análisis funcional utilizando la intersección de secuencias de 3'UTR con el método de Simkin et al. del 2014 y Leclerq et al. del 2017

Otro factor que podría estar generando diferentes resultados con ambos métodos podría ser que están utilizando versiones diferentes de los genomas de primates. El método de Leclerq utiliza versiones más antiguas de los genomas y están alineadas con el genoma del humano como referencia. El método de Simkin utiliza versiones más recientes de los genomas de los primates alineadas con el gorila de referencia (Tabla 6). En el análisis anterior en la figura 60 se muestra que las categorías funcionales enriquecidas también son similares con las del método de Leclerq. Estas diferencias podrían deberse a diferentes versiones de genomas y criterios de procesamiento de las 3'UTR y por lo tanto este análisis se debería volver a retomar cuando se tengan mejores genomas ensamblados para los hominoides.

Tabla 6 Características de los genomas utilizados

<b>Organismo</b>	<b>Simkin: Año (Versión)</b>	<b>Leclerq: Año (Versión)</b>
<b>Humano</b>	2009 (hg19)	2009 (hg19)
<b>Chimpancé</b>	2010 (panTro3)	2006 (panTro2)
<b>Gorila</b>	2011 (gorGor3)	2008 (gorGor1)
<b>Orangutan</b>	2007 (ponAbe2)	2007 (ponAbe2)
<b>Gibón</b>	2010 (NomLeu1)	X

## 8.2 Implicaciones de los métodos de predicción de blancos

Se debe tomar en cuenta que los resultados dependen directamente del método de predicción de blancos. Este tipo de análisis se deben de realizar nuevamente, cuando existan mejoras en dichos métodos, para disminuir la incertidumbre de las predicciones.

### 8.2.1 Mamíferos

Los métodos computacionales y los métodos experimentales siguen perfeccionándose ya que hasta la fecha resulta complicado predecir a los blancos de los microRNAs en animales. El simple pegado de la maquinaria de silenciamiento no garantiza un efecto en la regulación como se ha observado con los experimentos de CLIP. El caso ideal sería tener las validaciones de los blancos de familias de microRNAs para todo un genoma y así estimar la dinámica evolutiva de sus redes de regulación. Como no existe tal fuente de información, una buena aproximación son los métodos predictivos a nivel computacional. Debido a que se está analizando al grupo de los primates, no existe problema al usar el método de TS7, ya que este método ha sido entrenado con conjuntos de datos de líneas celulares de humano y experimentos en ratón, permitiendo hacer predicciones adecuadas en mamíferos. Como se puede observar, las categorías funcionales se pueden ver afectadas variando el valor de corte del CS++. El ser muy

estrictos con el valor de corte de CS++ hace que se pierdan muchos verdaderos blancos pero también el ser muy permisivo incluiría muchos falsos positivos. Lo ideal será en un futuro rehacer el análisis complementando con listas de verdaderos blancos validados experimentalmente (Hausser & Zavolan 2014; Agarwal et al. 2015).

### 8.2.2 Cíclidos

En el caso de los cíclidos el uso de TS7 con todos sus parámetros podría ser conflictivo. Sin embargo, como no se requirió la estimación de los valores del  $P_{CT}$  ni parámetros relacionados con las proporciones de las isoformas de los transcritos en humano, se decidió continuar usando el programa de TS7. Por otro lado ya se han realizado predicciones de blancos con herramientas computacionales en el grupo de los cíclidos. Se han utilizado distintos programas como miRanda, TargetScan y PACMIT donde algunas de estas predicciones se han evaluado de manera experimental. No se pudieron realizar variantes de los análisis utilizando valores de CS++ estrictos ya que son pocas las 3'UTR utilizadas (Loh et al. 2011; Brawand et al. 2014; Franchini et al. 2016).

## 8.3 Implicaciones de los controles usados

Se requiere de una correcta generación de controles para poder contrastar la dinámica de los sitios de pegado de microRNAs con un sistema que capture la naturaleza de las secuencias de los microRNAs pero que evolucione de manera neutral. La obtención de un buen control para una familia de microRNA sigue siendo un tema de investigación hoy en día y existen varios criterios para la selección de controles de microRNAs como energía libre del plegamiento, composición de nucleótidos, probabilidad de apareamiento de pares de bases entre otros (NG Kwang Loong & Mishra 2006).

### 8.3.1 Controles composicionales.

Se usaron controles producidos con secuencias a 1 nucleótido de distancia de los octámeros correspondientes a las regiones semilla. Estos controles pueden estar incluyendo sitios de pegado que puedan ser considerados de microRNAs reales. Para evitar dicho problema se filtró al nivel de hexámeros que se considera la secuencia base para la región semilla, sin embargo el espacio de posibles de secuencias se reduce

bastante. Por otro lado se desconoce qué efectos composicionales se puedan estar generando con estas secuencias y tal vez algunas de estas secuencias control poseen rasgos que alteran su dinámica evolutiva. Por ejemplo, si se tiene algún contenido composicional atípico cuando se hace la aleatorización de dinucleótidos de las secuencias por métodos heurísticos, se ha reportado que las propiedades intrínsecas de los microRNAs nativos se afectan. Esto ocurre ya que no preservan exactamente la frecuencia de dinucleótidos y mononucleótidos afectando a la energía del plegamiento (Workman & Krogh 1999; CLOTE 2005).

### 8.3.2 Controles basados en CS++.

El uso de este tipo de controles fue propuesto en este trabajo bajo la premisa de que los blancos que tienen un mal puntaje de CS++ probablemente contienen una mayor frecuencia de secuencias neutrales. Esto es debido a que la mayoría de estos sitios no son funcionales y cambios en su secuencia no afectarían a la adecuación de un organismo.

## 8. 4. Método con microRNA maduro en comparación con el método de k-meros

Se decidió continuar trabajando con el método basado en predicciones partiendo del microRNA maduro debido a las implicaciones que tienen los controles composicionales. Por ejemplo, el hecho de que una región semilla esté presente no significa que realmente sea un blanco e incrementa la incertidumbre del análisis comparado con predicciones más precisas que incluyen otros criterios además de la región semilla.

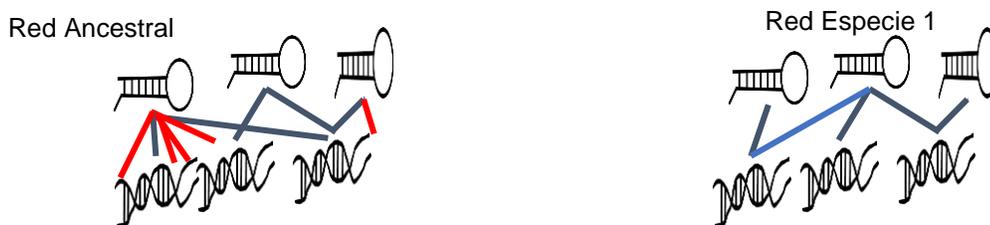
## 8.5. Métrica de recableado de Shou en comparación con la métrica de Simkin

Las 2 métricas de recableado estudiadas miden eventos diferentes de cambio en las conectividades. Sin embargo el utilizar la métrica de Simkin para estudios de categorías funcionales podría traer consigo algunas complicaciones. Por ejemplo, valores altos de recableado de un microRNA mediante la métrica de Simkin se pueden deber solamente a grandes cambios de sitios, en una fracción pequeña de todos sus blancos, y los demás

blancos realmente no sufran cambios (Figura 51a). Si se considera a esta familia de microRNA como interesante debido a su tasa de recableado se incluirían genes blanco de estas familias que no tienen recableado. Otro problema con la métrica de Simkin, en cuanto al análisis de categorías funcionales de blancos, podría ser la normalización, ya que un microRNA con 2 sitios de pegado, si pierde un sitio, tendría una tasa de  $\frac{1}{2}$  y otro microRNA con 100 sitios de pegado si pierde 50, también tendría el mismo valor de  $\frac{1}{2}$ . Por lo tanto, existen situaciones en que los microRNAs con muy pocos sitios puedan tener valores altos de recableado (Figura 51b). Como vimos además en la Figura 61, en la variación de la métrica de Simkin con respecto al parámetro de la distribución de conectividades, para redes más saturadas de conexiones la tasa de Simkin resultante es menor.

Situaciones conflictivas de recableado

a) Recableado principalmente en solo un gen blanco



b) Efecto de la normalización

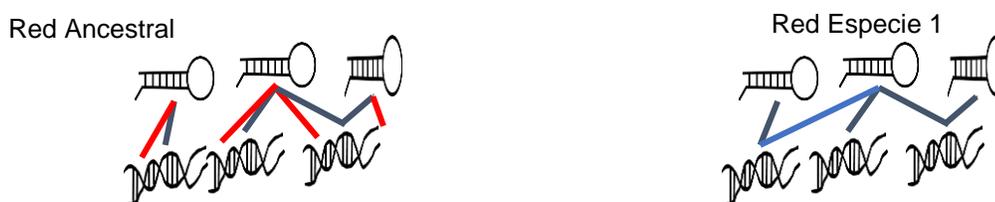


Figura 61 Diagramas de ejemplo de posibles situaciones conflictivas al utilizar la métrica de Simkin para análisis de categorías funcionales

## 8.6 Conservación en primates.

En todos los análisis hechos con el método de las familias de microRNAs maduros más conservadas se puede ver que las tasas de recableado son bajas. Esto se observa para dichas familias utilizando las distintas métricas en cada una de las especies. Este análisis se confirmó cuando se aplicó la misma metodología en el otro conjunto de 3'UTRs de primates provenientes del trabajo de Leclercq *et al.* de 2017, donde contamos con un conjunto más amplio de 3'UTRs y sus respectivas secuencias ancestrales, haciendo esta observación más robusta. Además se observa que las familias más conservadas y con más cambios son consistentes con ambos métodos.

Las observaciones sobre conservación de los sitios de pegado de los microRNAs coinciden con lo que Xu *et al.* en 2013 y Simkin *et al.* en 2014 habían reportado. Además se observó que no solo la ganancia y pérdida de sitios son bajas sino que el recableado entre genes es bajo, con respecto al control, en cada una de las especies. Estos análisis se hicieron con las familias más conservadas de microRNAs. Este mismo patrón se observó con las familias de microRNAs correspondientes a mamíferos. Las redes para los microRNAs conservados en mamíferos también tienen menor recableado que las redes construidas con los blancos control. Para saber si este comportamiento es intrínseco de los microRNAs funcionales el caso ideal será utilizar a las familias de reciente origen y comparar las tasas de recableado contra un control de CS++ similar al usado durante este trabajo.

## 8.7 Familias de microRNAs con tasas de recableado altas.

Se podría pensar como hipótesis nula que los microRNAs con tasas de recableado altas del conjunto de las familias más conservadas en vertebrados, no tendrían algún enriquecimiento funcional. Esto es porque a diferencia de lo visto con las familias con valores más bajos de recableado, muchos de los cambios podrían ser neutrales como se menciona a continuación.

Es posible que para el caso de los microRNAs más conservados, aquellos que tienen una tasa de recableado más alta sea debido a cambios neutrales y no estemos obteniendo información sobre selección positiva, esto es debido a que existe una correlación de conservación de blancos dada la antigüedad del microRNA. Para ello sería necesario enfocarse en familias más recientes de microRNAs que no estén bajo un proceso fuerte de selección negativa, sin embargo de todos modos no hay que pasar por alto los enriquecimientos funcionales que se vean en los microRNAs con recableado alto ya que lo que estamos evaluando es el cambio de sus blancos y no la conservación de la familia como tal.

Otra manera de darle la vuelta a este problema es incluir microRNAs más recientes y estimar la tasa de recableado de todos los microRNAs. Bajo este análisis no se tendría el problema de que los cambios solo sean una relajación de la presión de selección ya que estamos analizando microRNAs conservados y no conservados.

Se realizó una búsqueda en la literatura sobre los microRNAs con mayores tasas de recableado. Se buscaron sus posibles funciones asociadas y el linaje donde surgieron. Dicha información se puede ver en la Tabla 7.

*Tabla 7 Familias de microRNAs con tasas de recableado altas, sus posibles funciones y linajes en los que están presentes*

microRNA	Función asociada	Grupo taxonómico de origen
<b>miR-451</b>	Eritropoyesis y estrés	Vertebrata
<b>miR-191</b>	Desarrollo cortical y proliferación celular	Tetrapoda
<b>miR-187</b>	Crecimiento celular y apoptosis	Euteleostomi
<b>miR-210</b>	Respuesta a hipoxia	Bilateria
<b>miR-214</b>	Formación de huesos	Euteleostomi

De manera interesante se encontró que la tasa de recableado de miR-210 es alta en el caso de los primates pero es muy baja en el caso de los cíclidos. Este contraste podría ser informativo sobre la evolución de la respuesta a hipoxia en peces y mamíferos, sin embargo falta información para poder validarlo. En peces se ha reportado que miR-210 aumenta su expresión en condiciones de hipoxia en *Pelteobagrus vachelli*. Debido a su efecto pleiotrópico reportado podría ser que por eso se encuentra con una tasa muy baja

de recableado. En humano miR-210 está relacionado con los procesos de hipoxia dentro de tumores cancerígenos (Zhang et al. 2016; Wan et al. 2016).

Como se puede observar las funciones de muchas de estas familias y las familias de la Tabla 8 no están muy bien establecidas. Probablemente muchas de las que están reportadas solo son una pequeña fracción de las verdaderas funciones que poseen ya que principalmente se han asociado con enfermedades debido a los fines biomédicos con que se estudian muchas veces los microRNAs.

#### 8.7.1 microRNAs analizados con reportes de evidencia de selección positiva.

Las familias miR-130 y miR-301 son familias altamente conservadas analizadas durante este proyecto. Anteriormente se reportó que estas familias sufrieron una rápida expansión en el linaje de los primates y se han expandido mediante duplicaciones en tándem. Parece que estas expansiones son específicas de primates y además varias de estas familias son expresadas en células troncales embrionarias. Algunas hipótesis apuntan a que este tipo de expansiones de familias de expresión en desarrollo temprano de los embriones están relacionadas con los incrementos de complejidad morfológica y longevidad en este linaje (Guerra-assunção & Enright 2012).

Sin embargo, en nuestro análisis se observa que las tasas de recableado de estas familias, miR-130 y miR-301 (miR-454 para 130-3p) están en un valor intermedio. En otro reporte se ha visto que miR-454 regula algunos genes de manera diferencial solamente en humano y tiene que ver con la plasticidad del desarrollo del sistema nervioso (Somel et al. 2011). Esto no necesariamente contradice los resultados que se obtuvieron en este proyecto ya que puede haber situaciones donde pocos eventos de recableado puedan tener un cambio fenotípico notorio. Se podría pensar que familias que han tenido expansiones recientes podrían haber cambiado su repertorio de blancos, con base a los reportes del 2007 hechos por Chen y Rajewsky. Sin embargo como se puede observar la expansión de familias de los microRNAs no siempre lleva a un re-arreglo de la red.

## 8.8 Familias de microRNA con tasas de recableado bajas.

En la Tabla 8 se muestran los microRNAs con menor tasa de recableado junto con sus posibles funciones asociadas y el linaje donde surgieron. Estas familias de microRNAs están asociadas principalmente a procesos neuronales y procesos de desarrollo del sistema nervioso.

*Tabla 8 Familias de microRNAs con tasas de recableado bajas, sus posibles funciones y linajes en los que están presentes*

microRNA	Función asociada	Grupo taxonómico de origen
<b>miR-124</b>	Diferenciación neuronal	Bilateria
<b>miR-30e</b>	Proliferación celular	Vertebrata
<b>miR-182</b>	Desarrollo y apoptosis	Bilateria
<b>miR-137</b>	Desarrollo neuronal	Bilateria
<b>miR-506</b>	Supresor tumoral	Eutheria

### 8.8.1 microRNAs analizados con evidencia de selección negativa.

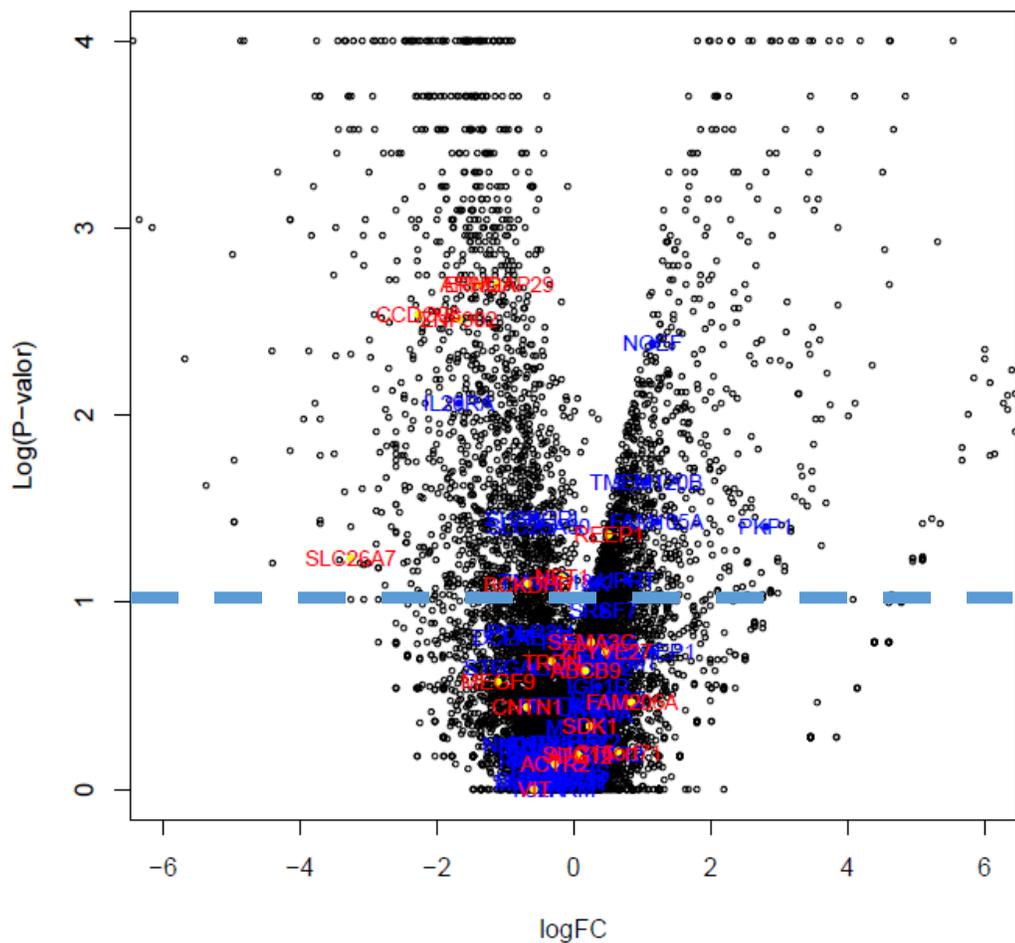
En este trabajo de tesis se puede observar que las bajas tasas de recableado podrían ser reflejo de la gran presión de selección para conservar sus interacciones con sus genes blanco. Es posible que estas familias de microRNAs estén participando en procesos esenciales o se expresen en varios tejidos y por lo tanto estén más restringidos sus eventos de recableado. (Chen & Rajewsky 2006).

## 8.9 Blancos con evidencia de selección.

Una vez identificados los microRNAs con mayor recableado también se identificaron los genes con mayor recableado. Para ello se hicieron estimaciones de recableado total por blanco. Las categorías funcionales obtenidas con ambos conjuntos de 3'UTRs para los valores extremos de recableado son consistentes con categorías funcionales reportadas en literatura. Los promotores con marcas de selección positiva, por ejemplo, están relacionados con procesos neuronales y del metabolismo (Haygood et al. 2007). Además, las funciones de genes con alto recableado coinciden con las funciones de genes que han cambiado su expresión entre humano y chimpancé (Babbitt et al. 2010; Bauernfeind et al. 2015; Babbitt et al. 2017).

Existen muchos tipos de cambios en las secuencias, algunos pueden ser a nivel de regulación y otros a nivel de secuencia codificante para proteína. Por ejemplo, de manera interesante se encontró que entre las funciones de los genes blanco con menor recableado está la cornificación de la piel. Sin embargo, los genes de esta categoría funcional a nivel de aminoácidos tienen una tasa alta de cambio. Lo inverso ocurre con las categorías del citoesqueleto y señalización: a nivel de regulación tienen muchos cambios, pero a nivel de la región codificante están muy conservadas. Por otro lado procesos del sistema nervioso y metabolismo tienen una tasa alta de cambio tanto de regulación como en las secuencias codificantes (Bustamante et al. 2005; Haygood et al. 2007; Cagan et al. 2016; Babbitt et al. 2017).

Se buscaron aquellos genes que hayan ganado más sitios de lo esperado en chimpancé y que hayan perdido más sitios de lo esperado en humano. También se buscaron aquellos que tuvieran el comportamiento inverso. De la corteza del cíngulo anterior a continuación se muestran aquellos que aumentaron su expresión en humano con respecto a chimpancé debido a posibles eventos de recableado de los microRNAs. Se seleccionaron aquellos que su valor de expresión diferencial era estadísticamente significativo.



*Figura 62 Gráfico de volcán de expresión de genes en la corteza del cíngulo anterior. En rojo se nombran los genes que perdieron sitios de regulación en humano y ganaron sitios de regulación en chimpancé. La línea azul representa el punto de corte del FDR (0.1)*

Entre estos genes se encuentra ARHGAP29, EPM2A, FAM107A, ZNF302 y CCD93. De manera interesante la mayoría de estos genes están involucrados en enfermedades del sistema nervioso y del desarrollo. ARHGAP29 es un gen involucrado en el desarrollo craneofacial y alteraciones en su nivel de expresión se han sugerido como posibles causantes en las malformaciones del paladar. Sin embargo, no se han podido explicar las causas de sus alteraciones en los niveles de expresión. El gen FAM107A codifica para una proteína que está implicada en la regulación de la transcripción de genes. Se sabe que este gen participa en el desarrollo neuronal y cambios en su expresión se ha asociado con desórdenes psiquiátricos y respuesta a estrés (Mangold et al. 2011; Attanasio et al. 2014; Nakajima & Koizumi 2014).

El gen EPM2A también está involucrado en el sistema nervioso, específicamente en la supervivencia de células nerviosas y mutaciones en este gen están relacionadas con un tipo de epilepsia. Los otros genes, ZNF302 y CCDC93, participan en señalización (Ganesh et al. 2001; Gayarre et al. 2014; Phillips-Krawczak et al. 2015). Estos resultados son consistentes con los análisis de categorías funcionales donde las funciones enriquecidas están asociadas con sistema nervioso y señalización.

Se hizo también el análisis de los genes que ganaron sitios en humano pero que perdieron sitios en chimpancé (Figura S9). Algunos de estos genes están más expresados en chimpancé con respecto al humano. El gen SRGAP3 además de cumplir este criterio está estadísticamente significativo representado del lado de los genes con mayor expresión en chimpancé. Este gen participa en la morfogénesis neuronal y se ha asociado que alteraciones en este gen están relacionadas con retraso mental (Bacon et al. 2013).



## 8.11 Efecto de otros elementos de regulación en la evolución fenotípica

Aún queda mucho por explorar acerca de los microRNAs y la evolución fenotípica, ya que la mayoría de los trabajos se han enfocado a los efectos de los factores de transcripción. Además como dicen Hoekstra y Coyne en 2007, no hay que olvidarse de las regiones codificantes para proteínas. Un trabajo que soporta esta perspectiva es el trabajo de Gojobori *et al.* en 2007, donde realizan un análisis genómico de polimorfismos en humanos e hicieron una estimación de que aproximadamente 10% de las sustituciones en aminoácidos en humanos y chimpancé son adaptativas.

## 8.12 Aplicación en cíclidos

El análisis realizado con las 130 secuencias de las 3'UTRs de cíclidos, solo provee un pequeño acercamiento de lo que en realidad podría estar pasando. En el trabajo de Loh *et al.* de 2011, se reporta para este mismo conjunto de 3'UTRs y para otro conjunto de 700 secuencias con datos de polimorfismos, que existe una tasa rápida de cambio en estas regiones en comparación con las secuencias codificante. Análisis más recientes utilizando el grupo de los cíclidos Midas observan este patrón de tasas de cambio rápidas en las 3'UTRs con conjuntos más completos de secuencias pero también observan que otros grupos de estos mismos cíclidos tienen tasas de cambio bajas. Por otro lado, en otras especies de cíclidos del Este de África, más cercanos a los que se analizaron en este trabajo, se ha reportado el comportamiento contrario poniendo en duda las conclusiones de Loh *et al.* en 2011. Ellos observan menor tasa de cambio en los blancos de microRNAs que la tasa de cambio de regiones que no son blancos de microRNAs, utilizando genomas completos. Sin embargo, este análisis lo realizaron con 5 especies de cíclidos de grupos más lejanos lo cual aún no deja claro si el comportamiento de tasas de cambio altas reportado anteriormente es válido, ya que como se mencionó anteriormente, entre más tiempo de divergencia se podrían tener problemas de saturación de cambios (Shou *et al.* 2011; Loh *et al.* 2011; Brawand *et al.* 2014; Franchini *et al.* 2016).

Estos trabajos comparativos proveen información valiosa de los genes que puedan estar implicados en morfogénesis. Para este trabajo, dado que la evolución es un proceso complejo donde muchos factores influyen, no se puede asumir correctamente todas las

consideraciones y conclusiones acerca de las tasas de recableado que en un futuro con datos más precisos, mayor disponibilidad de genomas y conjuntos de datos de microRNAs, se resolverán. Sin embargo es posible conseguir información importante a partir de los datos de los genomas disponibles, y aumentar la resolución de trabajos previos mediante la evaluación de diferentes maneras en que las redes pueden evolucionar.

## 9. Conclusiones

Los análisis de las métricas de recableado de redes nos sugieren que hace falta la generación de una métrica que sea adecuada a las redes de regulación de los microRNAs. Los microRNAs conservados en vertebrados y mamíferos tienen tasas bajas de recableado con respecto a un modelo neutral. Esto se confirmó mediante las predicciones utilizando tanto k-meros como la secuencia completa del microRNA maduro. Las funciones asociadas a los genes que han presentado mayor número de cambios están relacionadas con procesos de desarrollo del sistema nervioso y señalización, consistente con la literatura basada en datos de expresión de diferentes tejidos de primates. Utilizando datos de expresión de cerebro se encontró que los posibles eventos evolutivos de recableado de microRNAs podrían verse reflejados en los niveles de expresión de sus blancos cuando se comparan tejidos de chimpancé y humano. Estas observaciones sugieren que los microRNAs pudieron jugar un papel clave en la evolución fenotípica de los homínidos.

# 10 Perspectivas

## 10.1 Utilizar un grupo de controles basados en la aleatorización de secuencia

Durante el proyecto se generó un grupo de secuencias similares a los microRNA mediante un método de aleatorización controlado por la composición y abundancia de estas secuencias. Las secuencias fueron aleatorizadas con un generador de secuencias que realiza permutaciones de dinucleótidos. En este caso se controló por la frecuencia de las secuencias de los microRNA. Se generaron alrededor de 800 k-meros diferentes con este método. Para la realización de los controles, además de que se iban a generar regiones semilla aleatorizadas, también se requiere del resto de microRNA maduro.

A la fecha existen varios simuladores de secuencias de microRNA maduro que se basan en criterios de composición y plegamiento. Se podrían utilizar estos simuladores para generar controles composicionales de microRNAs maduros y contrastar con los resultados que obtuvimos con los controles de 8-meros y de CS++. Por el momento nos limitamos a hacer una prueba para saber los efectos de aleatorizar la parte que no es la semilla del microRNA maduro. En futuros análisis se podrán utilizar estas secuencias para cuantificar los eventos de recableado en lugar de hacerlo solamente al nivel de la región semilla.

El análisis de TargetScan requiere utilizar la secuencia de un microRNA maduro. Utilizando los microRNAs let-7 y miR-124 se hicieron pruebas como un primer paso para saber el efecto de utilizar regiones aleatorizadas de la secuencia madura sobre el puntaje CS++ para la predicción de blancos. Se encontró que los valores no se ven influenciados casi por el uso de regiones aleatorizadas, ya que mantienen una relación muy alta cuando se utiliza una región aleatorizada o la original. Esto se midió mediante una correlación lineal y que además resultó ser significativa con un p-valor  $< 0.01$ . El gráfico de dispersión se muestra en la Figura 64.

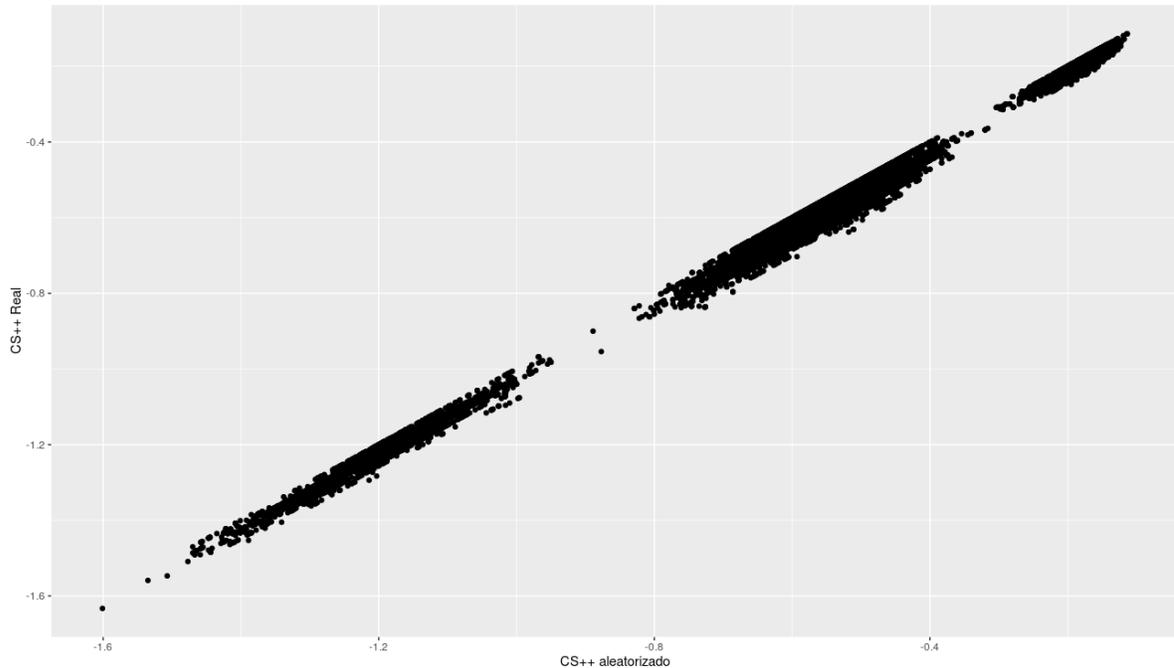


Figura 64 Gráfico de dispersión para el microRNA let-7.

## 10.2 Compensación de pérdida de regulación de microRNA

En el caso de las ganancias, se podría evaluar la situación de aquellos blancos que estaban siendo regulados por algún microRNA y ganaron regulación por otro microRNA manteniendo la que ya tenían. Se realizó un análisis bajo esta premisa pero este método debe ser mejorado ya que no se conoce si todos esos microRNAs que están reemplazando un sitio de pegado se expresan en ese mismo tejido. Una perspectiva sería tomar en cuenta el tejido y etapa de desarrollo donde se expresan determinados microRNAs y saber si las ganancias de sitios de pegado son realmente compensatorias.

A continuación se presenta el análisis de si las ganancias de sitios son compensatorias sin tomar en cuenta tejido ni etapa de desarrollo. De las 109 familias conservadas de microRNA se evaluó si cuando un sitio de pegado de una determinada familia de microRNA se pierde, es otra familia de microRNA la que llega a sustituirlo para compensar esta pérdida, vimos que esto es muy poco común al menos para este

conjunto de familias estudiadas, cuando comparamos con los k-meros que están a un nucleótido de distancia (Figura 65).

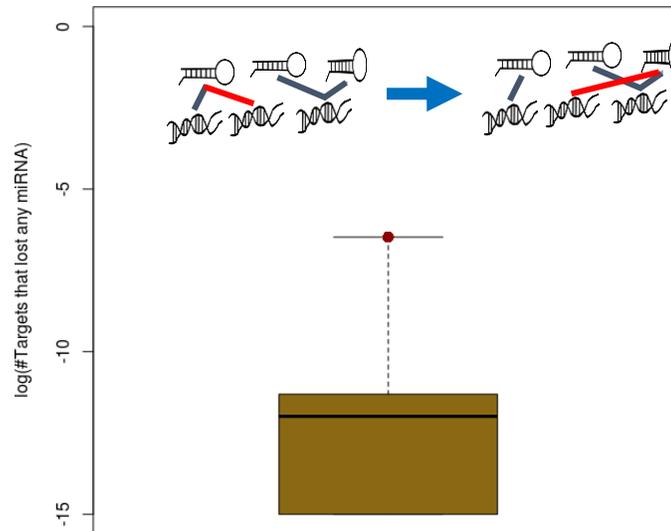


Figura 65 Número de pérdidas de regulación que no fueron compensadas por otro microRNA.

### 10.3 Estimación de recableado en radiaciones adaptativas

En la actualidad se tiene información a nivel de genomas suficiente para estimar tasas de recableado en Cíclidos y Peces espinosos. Sin embargo para el caso de los cíclidos se necesita tener un conjunto confiable de secuencias de las 3'UTRs para predecir los posibles sitios de pegado de los microRNAs. Este paso es limitante en tiempo ya que no se encuentran disponibles estas secuencias. Los trabajos publicados anteriormente han utilizado secuencias de las 3'UTRs procesadas de manera independiente por cada grupo de investigación que las ha trabajado. Una vez procesadas estas secuencias se podrían estimar las tasas de recableado de microRNAs para este grupo de peces.

Otro ejemplo muy bien estudiado, también de radiaciones adaptativas, es el de los pinzones de Darwin de las Islas Galápagos. Recientemente se han publicado análisis con datos de secuenciación. Estos datos podrían utilizarse para entender cómo los elementos

de regulación han afectado su evolución y sus cambios fenotípicos. Sin embargo, las 3'UTRs no se encuentran bien anotadas para estas especies y es un proceso que se debe hacer para saber el grado de divergencia de los sitios blanco de microRNAs (Lamichhane et al. 2015).

#### 10.4 Inclusión de interacciones Factor de Transcripción-microRNA-Blanco

Se pueden estudiar interacciones más complejas como entre factores de transcripción y microRNAs. Se utilizan datos validados experimentalmente de interacciones reportadas en diferentes bases de datos. Cada tipo de circuito obtenido con este método resulta de la combinación de interacciones de inducción y represión. Se pueden identificar motivos de red que estén conservados a lo largo de la historia evolutiva y de eventos de especiación. De estos motivos, se podría probar si se han seleccionado debido a sus efectos fenotípicos.

# 11 Bibliografia

- Agarwal, V. et al., 2015. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4(AUGUST2015), pp.1–38.
- Alon, U., 2006. *An Introduction to Systems Biology. Design principles of biological circuits.*, Chapman and Hall/CRC. Taylor and Francis Group.
- Ashburner, M. et al., 2000. Gene ontologie: Tool for the unification of biology. *Nature Genetics*, 25(1), pp.25–29.
- Attanasio, C. et al., 2014. Fine Tuning of Craniofacial Morphology by Distant-Acting Developmental Enhancers. *Science*, 342(6157), pp.1–20. Available at: <http://dx.doi.org/10.1126/science.1241006>.
- Ayub, Q. et al., 2013. FOXP2 targets show evidence of positive selection in European populations. *American Journal of Human Genetics*, 92(5), pp.696–706. Available at: <http://dx.doi.org/10.1016/j.ajhg.2013.03.019>.
- Babbitt, C.C. et al., 2010. Both noncoding and protein-coding rnas contribute to gene expression evolution in the primate brain. *Genome Biology and Evolution*, 2(1), pp.67–79.
- Babbitt, C.C. et al., 2017. Gene expression and adaptive noncoding changes during human evolution. *BMC Genomics*, 18(1), p.435. Available at: <http://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-017-3831-2>.
- Bacon, C., Endris, V. & Rappold, G.A., 2013. The cellular function of srGAP3 and its role in neuronal morphogenesis. *Mechanisms of Development*, 130(6–8), pp.391–395. Available at: <http://dx.doi.org/10.1016/j.mod.2012.10.005>.
- Bauernfeind, A.L. et al., 2015. Evolutionary divergence of gene and protein expression in the brains of humans and chimpanzees. *Genome Biology and Evolution*, 7(8), pp.2276–2288.
- Beltrao, P. & Serrano, L., 2007. Specificity and evolvability in eukaryotic protein interaction networks. *PLoS Computational Biology*, 3(2), pp.0258–0267.
- Bénazet, J.D. & Zeller, R., 2009. Vertebrate limb development: moving from classical morphogen gradients to an integrated 4-dimensional patterning system. *Cold Spring Harbor perspectives in biology*, 1(4), pp.1–15.
- Benjamini, Y. & Hochberg, Y., 1995. Controlling the False Discovery Rate: A practical and Powerful approach to multiple testing. *JSTOR*, 57(1), pp.289–300.

- Berezikov, E., 2011. Evolution of microRNA diversity and regulation in animals. *Nature Reviews Genetics*, 12(12), pp.846–860. Available at: <http://www.nature.com/doi/10.1038/nrg3079>.
- Betel, D. et al., 2010. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biology*, 11(8), p.R90. Available at: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-8-r90>.
- Blanchette, M. et al., 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Research*, 14(4), pp.708–715.
- Brakefield, P.M., 2006. Evo-devo and constraints on selection. *Trends in Ecology and Evolution*, 21(7), pp.362–368.
- Brawand, D. et al., 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513(7518), pp.375–381. Available at: <http://www.nature.com/doi/10.1038/nature13726>.
- Bridges, C. & Morgan, T., 1923. The third-chromosome group of mutant characters of *Drosophila melanogaster*. *The Carnegie Institution, Washington, DC. 1923*.
- Broughton, J.P. & Pasquinelli, A.E., 2016. A tale of two sequences: microRNA-target chimeric reads. *Genetics Selection Evolution*, 48(1), p.31. Available at: <http://gsejournal.biomedcentral.com/articles/10.1186/s12711-016-0209-x>.
- Burke, A.C. et al., 1995. Hox genes and the evolution of vertebrate axial morphology. *Development*, 121(2), pp.333–346. Available at: [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=7768176](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7768176).
- Bustamante, C.D. et al., 2005. Natural selection on protein-coding genes in the human genome. *Nature*, 437(7062), pp.1153–1157. Available at: <http://www.nature.com/doi/10.1038/nature04240>.
- Cagan, A. et al., 2016. Natural selection in the great apes. *Molecular Biology and Evolution*, 33(12), pp.3268–3283.
- Carroll, S.B., 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell*, 134(1), pp.25–36.
- Carroll, S.B., 2003. Genetics and the making of *Homo sapiens*. *Nature*, 422(6934), pp.849–857. Available at: <http://www.nature.com/doi/10.1038/nature01495>.
- Carroll, S.B., 1995. Homeotic genes and the evolution of arthropods and chordates.
- Carroll, S.B., Grenier, J.K. & Weatherbee, S.D., 2004. *From DNA to Diversity*, Available at: [papers3://publication/uuid/35335DAF-2D4E-4EBD-AEF9-34FBFA264E34](http://papers3://publication/uuid/35335DAF-2D4E-4EBD-AEF9-34FBFA264E34).

- Chaturvedi, A., Raeymaekers, J.A.M. & Volckaert, F.A.M., 2014. Computational identification of miRNAs, their targets and functions in three-spined stickleback (*Gasterosteus aculeatus*). *Molecular Ecology Resources*, 14(4), pp.768–777.
- Chen, K. & Rajewsky, N., 2006. Natural selection on human microRNA binding sites inferred from SNP data. *Nature Genetics*, 38(12), pp.1452–1456. Available at: <http://www.nature.com/doi/10.1038/ng1910>.
- Chen, K. & Rajewsky, N., 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews Genetics*, 8(2), pp.93–103. Available at: <http://www.nature.com/doi/10.1038/nrg1990>.
- Clark, M.B. et al., 2011. The reality of pervasive transcription. *PLoS Biology*, 9(7), pp.5–10.
- CLOTE, P., 2005. Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *Rna*, 11(5), pp.578–591. Available at: <http://www.rnajournal.org/cgi/doi/10.1261/rna.7220505>.
- Davidson, E.H., 2006. *The Regulatory Genome. Gene Regulatory networks*, Elsevier.
- Dunham, I. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74. Available at: <http://www.nature.com/doi/10.1038/nature11247>.
- Erkenbrack, E.M. & Davidson, E.H., 2015. Evolutionary rewiring of gene regulatory network linkages at divergence of the echinoid subclasses. *Proceedings of the National Academy of Sciences*, 112(30), pp.E4075–E4084. Available at: <http://www.pnas.org/lookup/doi/10.1073/pnas.1509845112>.
- Farh, K.K.-H., 2005. The Widespread Impact of Mammalian MicroRNAs on mRNA Repression and Evolution. *Science*, 310(5755), pp.1817–1821. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.1121158>.
- Farrell, R. & Bassett, C., 2007. *Regulation of Gene Expression in Plants*, Springer.
- Felsenstein, J. & Churchill, G.A., 1996. A Hidden Markov Model Approach Evolution to Variation Among Sites in Rate of. , 13(1), pp.93–104.
- Franchini, P. et al., 2016. The Role of microRNAs in the Repeated Parallel Diversification of Lineages of Midas Cichlid Fish from Nicaragua. *Genome biology and evolution*, 8(5), pp.1543–1555.
- Friedman, R.C. et al., 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, 19(1), pp.92–105.
- Gallego, A. et al., 2016. Functional Implications of Human-Specific Changes in Great Ape

- microRNAs. *PLoS ONE*, 11(4), pp.1–20.
- Ganesh, S. et al., 2001. Regional and developmental expression of Epm2a gene and its evolutionary conservation. *Biochemical and Biophysical Research Communications*, 283, pp.1046–1053.
- Gayarre, J. et al., 2014. The phosphatase activity of laforin is dispensable to rescue Epm2a<sup>-/-</sup> mice from lafora disease. *Brain*, 137(3), pp.806–818.
- Gellon, G. & William McGinnis, 1998. Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *BioEssays*, 20, pp.116–125. Available at: [http://dx.doi.org/10.1002/\(SICI\)1521-1878\(199802\)20:2%3C116::AID-BIES4%3E3.0.CO;2-R](http://dx.doi.org/10.1002/(SICI)1521-1878(199802)20:2%3C116::AID-BIES4%3E3.0.CO;2-R).
- Glor, R.E., 2010. Phylogenetic Insights on Adaptive Radiation. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), pp.251–270.
- Gojobori, T., 1983. Codon substitution in evolution and the “saturation” of synonymous changes. *Genetics*, 105(4), pp.1011–1027.
- Guerra-assunção, J.A. & Enright, A.J., 2012. Large-scale analysis of microRNA evolution.
- Hagen, J.B., 1999. Naturalists, Molecular Biologists, and the Challenges of Molecular Evolution Author (s): Joel B. Hagen Source: *Journal of the History of Biology*, Vol. 32, No. 2 (Autumn, 1999), pp. 321–341 Published by: Springer Stable URL: <http://www.jsto.org>, 32(2), pp.321–341.
- Halley, P. et al., 2014. Epigenetic regulation of the apolipoprotein gene cluster by a long non protein coding RNA. *Cell Reports*, 50(4), pp.381–389.
- Harris, R., 2007. Improved pairwise alignment of genomic DNA. *PhD Thesis. The Pennsylvania State University*.
- Harrison, R., 1918. Experiments on the development of the forelimb of *Amblystoma*, a self-differentiating equipotential system. *JEZ-A Ecological and integrative physiology*, 25(2), pp.413–461.
- Hausser, J. & Zavolan, M., 2014. Identification and consequences of miRNA–target interactions — beyond repression of gene expression. *Nature Reviews Genetics*, 15(9), pp.599–612. Available at: <http://www.nature.com/doi/10.1038/nrg3765>.
- Haygood, R. et al., 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nature Genetics*, 39(9), pp.1140–1144. Available at: <http://www.nature.com/doi/10.1038/ng2104>.
- Heimberg, A.M. et al., 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proceedings of the National Academy of Sciences of the United States of*

- America*, 105(8), pp.2946–50. Available at:  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2268565&tool=pmcentrez&rendertype=abstract>.
- Hertel, J. & Stadler, P., 2015. The Expansion of Animal MicroRNA Families Revisited. *Life*, 5(1), pp.905–920. Available at: <http://www.mdpi.com/2075-1729/5/1/905/>.
- Hoekstra, H.E. & Coyne, J.A., 2007. The locus of evolution: Evo devo and the genetics of adaptation. *Evolution*, 61(5), pp.995–1016.
- Indjeian, V.B. et al., 2016. Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell*, 164(1–2), pp.45–56. Available at:  
<http://dx.doi.org/10.1016/j.cell.2015.12.007>.
- Iwakawa, H. & Tomari, Y., 2015. The Functions of MicroRNAs : mRNA Decay and Translational Repression. *Trends in Cell Biology*, 25(11), pp.651–665. Available at:  
<http://dx.doi.org/10.1016/j.tcb.2015.07.011>.
- Jones, F.C. et al., 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), pp.55–61. Available at:  
<http://www.nature.com/doi/10.1038/nature10944>.
- Kaikkonen, M.U., Lam, M.T.Y. & Glass, C.K., 2011. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research*, 90(3), pp.430–440.
- Kamberov, Y.G. et al., 2013. Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell*, 152(4), pp.691–702. Available at:  
<http://dx.doi.org/10.1016/j.cell.2013.01.016>.
- Khaitovich, P. et al., 2006. Evolution of primate gene expression. *Nature Reviews Genetics*, 7(9), pp.693–702. Available at:  
<http://www.nature.com/doi/10.1038/nrg1940>.
- Kimura, M., 1968. Evolutionary Rate at the molecular level. *Nature*, 217, pp.671–672. Available at: <http://linkinghub.elsevier.com/retrieve/pii/B0122270800004328>.
- Kimura & Ohta, 1971. Protein polymorphism as a phase of molecular evolution. *Nature*, 232(5313), pp.655–657.
- King, M. & Wilson, a C., 1975. Humans and Chimpanze es. *Science*, 188(4184), pp.107–116.
- Lai, C.S.L. et al., 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413(6855), pp.519–523. Available at:  
<http://www.nature.com/doi/10.1038/35097076>.
- Lamichhaney, S. et al., 2015. Evolution of Darwin’s finches and their beaks revealed by

- genome sequencing. *Nature*, 518(7539), pp.371–375. Available at:  
<http://www.nature.com/doi/10.1038/nature14181>.
- Leclercq, M., Diallo, A.B. & Blanchette, M., 2017. Prediction of human miRNA target genes using computationally reconstructed ancestral mammalian sequences. *Nucleic Acids Research*, 45(2), pp.556–566.
- Leder, E.H. et al., 2015. The evolution and adaptive potential of transcriptional variation in sticklebacks - Signatures of selection and widespread heritability. *Molecular Biology and Evolution*, 32(3), pp.674–689.
- Levin, L., 2011. Coevolución. Available at:  
<https://www.educ.ar/recursos/14292/coevolucion>.
- Lewis, B., Robert, J. & Kilgore, L., 2006. *Understanding humans: An introduction to physical anthropology and archaeology*, WADSWORTH CENGAGE Learning.
- Lewis, B.P. et al., 2003. 33-Prediction of Mammalian MicroRNA Targets. *Cell*, 115(7), pp.787–798. Available at:  
<http://linkinghub.elsevier.com/retrieve/pii/S0092867403010183>.
- Lewis, E.B., 1978. A gene complex controlling segmentation in *Drosophila*. *Nature*, 276(5688), pp.565–570. Available at:  
<http://www.nature.com/doi/10.1038/276565a0>.
- Lewis, E.B., 1963. Genes and developmental pathways. *American Zoologist*, 3(February), pp.33–56.
- Li, B., Carey, M. & Workman, J.L., 2007. The Role of Chromatin during Transcription. *Cell*, 128(4), pp.707–719.
- Liu, N. et al., 2009. The evolution and functional diversification of animal microRNA genes. *October*, 18(10), pp.985–996.
- Locke, D.P. et al., 2011. Comparative and demographic analysis of orang-utan genomes. *Nature*, 469(7331), pp.529–533. Available at:  
<http://www.nature.com/doi/10.1038/nature09687>.
- Loh, Y.H.E., Yi, S. V. & Strelman, J.T., 2011. Evolution of MicroRNAs and the diversification of species. *Genome Biology and Evolution*, 3(1), pp.55–65.
- Mangold, E., Ludwig, K.U. & Nöthen, M.M., 2011. Breakthroughs in the genetics of orofacial clefting. *Trends in Molecular Medicine*, 17(12), pp.725–733.
- Meunier, J. et al., 2013. Birth and expression evolution of mammalian microRNA genes. *Genome research*, pp.1–12.
- Meyer, A., 1998. Hox gene variation and evolution. *Nature*, 391(6664), p.225,227-228.

- Available at: [papers3://publication/doi/10.1038/34530](https://doi.org/10.1038/34530).
- Moczek, A.P. et al., 2015. The significance and scope of evolutionary developmental biology: A vision for the 21st century. *Evolution and Development*, 17(3), pp.198–219.
- Naish, D., 2016. Tetrapod zoology. *Scientific american*. Available at: <https://blogs.scientificamerican.com/tetrapod-zoology/>.
- Nakajima, H. & Koizumi, K., 2014. Family with sequence similarity 107: A family of stress responsive small proteins with diverse functions in cancer and the nervous system (Review). *Biomedical reports*, 2(3), pp.321–325. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3990222&tool=pmcentrez&rendertype=abstract>.
- NG Kwang Loong, S. & Mishra, S.K., 2006. Unique folding of precursor microRNAs: Quantitative evidence and implications for de novo identification. *Rna*, 13(2), pp.170–187. Available at: <http://www.rnajournal.org/cgi/doi/10.1261/rna.223807>.
- Nüsslein-Volhard, C. & Wieschaus, E., 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287(5785), pp.795–801.
- Okamura, K. et al., 2008. The regulatory activity of microRNA\* species has substantial influence on microRNA and 3' UTR evolution. *Nature Structural & Molecular Biology*, 15(4), pp.354–363. Available at: <http://www.nature.com/doi/10.1038/nsmb.1409>.
- Oliveira, A.C. et al., 2017. Combining Results from Distinct MicroRNA Target Prediction Tools Enhances the Performance of Analyses. *Frontiers in Genetics*, 8(May), pp.1–10. Available at: <http://journal.frontiersin.org/article/10.3389/fgene.2017.00059/full>.
- Oliver, M.K., 1997. *The Cichlid Fishes of Lake Malawi, Africa*. Available at: <https://malawicichlids.com/index/.htm>.
- Opsahl, T., 2013. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, 35(2), pp.159–167. Available at: <http://dx.doi.org/10.1016/j.socnet.2011.07.001>.
- Paraskevopoulou, M.D. et al., 2013. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic acids research*, 41(Web Server issue).
- Peter, I.S. & Davidson, E.H., 2011. Evolution of gene regulatory networks controlling body plan development. *Cell*, 144(6), pp.970–985. Available at: <http://dx.doi.org/10.1016/j.cell.2011.02.017>.

- Peterson, K.J., Dietrich, M.R. & McPeck, M.A., 2009. MicroRNAs and metazoan macroevolution: Insights into canalization, complexity, and the Cambrian explosion. *BioEssays*, 31(7), pp.736–747.
- Phillips-Krawczak, C.A. et al., 2015. COMMD1 is linked to the WASH complex and regulates endosomal trafficking of the copper transporter ATP7A. *Molecular Biology of the Cell*, 26(1), pp.91–103. Available at: <http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E14-06-1073>.
- Presgraves, D.C. et al., 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature*, 423(6941), pp.715–719. Available at: <http://www.nature.com/doi/10.1038/nature01679>.
- Rehmsmeier, M. et al., 2004. Fast and effective prediction of microRNA / target duplexes. *Spring*, (2003), pp.1507–1517.
- Riffo-Campos, Á.L., Riquelme, I. & Brebi-Mieville, P., 2016. Tools for sequence-based miRNA target prediction: What to choose? *International Journal of Molecular Sciences*, 17(12).
- Rodriguez, A. et al., 2004. Identification of Mammalian microRNA Host Genes and Transcription Units Identification of Mammalian microRNA Host Genes and Transcription Units. , pp.1902–1910.
- da Sacco, L. & Masotti, A., 2013. Recent insights and novel bioinformatics tools to understand the role of MicroRNAs binding to 5' untranslated region. *International Journal of Molecular Sciences*, 14(1), pp.480–495.
- Sarda, S. & Hannenhalli, S., 2015. High-throughput identification of cis-regulatory rewiring events in yeast. *Molecular Biology and Evolution*, 32(12), pp.3047–3063.
- Schlicker, A. et al., 2006. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*, 7(1), p.302. Available at: <http://dx.doi.org/10.1186/1471-2105-7-302>  
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1559652&tool=pmcentrez&rendertype=abstract>.
- Shou, C. et al., 2011. Measuring the evolutionary rewiring of biological networks. *PLoS Computational Biology*, 7(1).
- Simkin, A.T. et al., 2014. Inferring the evolutionary history of primate microRNA binding sites: Overcoming motif counting biases. *Molecular Biology and Evolution*, 31(7), pp.1894–1901.
- Singleton, M., 2013. Primate Cranial Diversity. *Nature Education Knowledge*, 4(12).

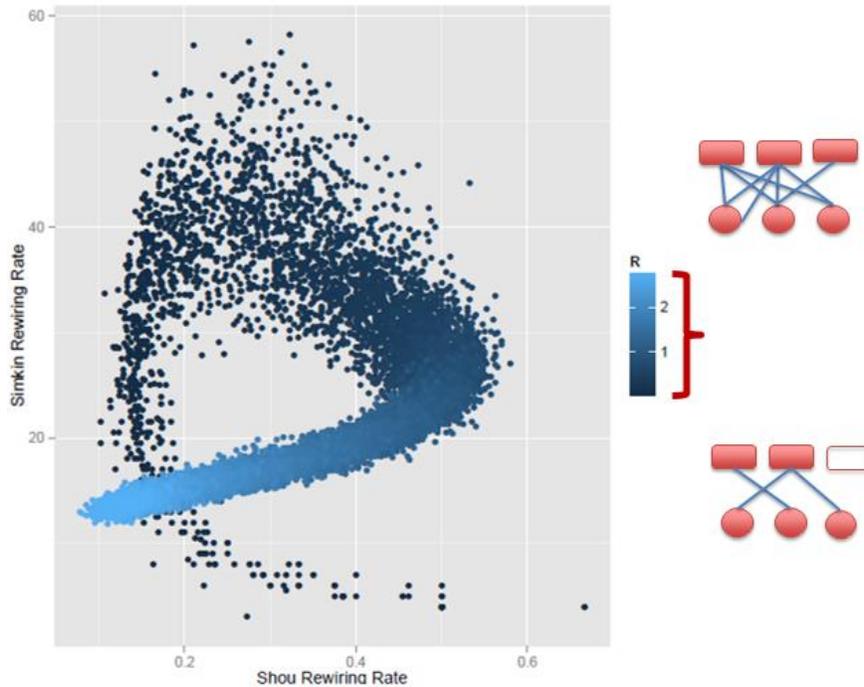
- Siomi, M.C. et al., 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology*, 12(4), pp.246–258. Available at: <http://www.nature.com/doi/10.1038/nrm3089>.
- Somel, M. et al., 2011. MicroRNA-Driven Developmental Remodeling in the Brain Distinguishes Humans from Other Primates. , 9(12).
- Sorrells, T.R. et al., 2015. Intersecting transcription networks constrain gene regulatory evolution. *Nature*, 523(7560), pp.361–365. Available at: <http://www.nature.com/doi/10.1038/nature14613>.
- Spemann, H. & Mangold, H., 1924. Induction of embryonic primordia by implantation of organizers from a different species. *International Journal of Developmental Biology*, 45(1 SPEC. ISS. 1), pp.13–38.
- Stark, A. et al., 2005. Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*, 123(6), pp.1133–1146.
- Supek, F. et al., 2011. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, 6(7).
- Thomson, R.C. et al., 2014. A critical appraisal of the use of microRNA data in phylogenetics. *Proceedings of the National Academy of Sciences*, 111(35), pp.E3659–E3668. Available at: <http://www.pnas.org/cgi/doi/10.1073/pnas.1407207111>.
- Wan, S.-M. et al., 2016. Dynamic mRNA and miRNA expression analysis in response to intermuscular bone development of blunt snout bream (*Megalobrama amblycephala*). *Scientific Reports*, 6(July), p.31050. Available at: <http://www.nature.com/articles/srep31050>.
- Willoughby, P. & Willoughby, P.R., 2005. Palaeoanthropology and the Evolutionary Place of Humans in Nature. *International Journal of Comparative Psychology*, 18(1), pp.60–91.
- Winter, J. et al., 2009. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature cell biology*, 11(3), pp.228–234.
- Workman, C. & Krogh, A., 1999. No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. *Nucleic acids research*, 27(24), pp.4816–4822.
- Xu, J. et al., 2013. The evolution of evolvability in microRNA target sites in vertebrates. , pp.1810–1816.
- Ying, S., 2006. *MicroRNA Protocols*, Humana Press.

- Zhang, G. et al., 2016. Integrated analysis of mRNA-seq and miRNA-seq in the liver of *Pelteobagrus vachelli* in response to hypoxia. *Scientific Reports*, 6(1), p.22907. Available at: <http://www.nature.com/articles/srep22907>.
- Zhu, Y. et al., 2012. Evolutionary relationships between miRNA genes and their activity. *BMC Genomics*, 13(1), p.718. Available at: <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-718>.
- Zuckerandl, E. & Pauling, L., 1965. Evolutionary divergence and convergence in proteins. *Evolving Genes and Proteins*, pp.97–166. Available at: [http://www.yanaiweb.com/genome/Clocks/Zuckerandl\\_1965.pdf](http://www.yanaiweb.com/genome/Clocks/Zuckerandl_1965.pdf).
- Zuckerandl, E. & Pauling, L., 1962. Molecular Disease, Evolution, and Genic Heterogeneity. *Horizons in Biochemistry*, pp.189–222.

## 12. Anexos

### 12.1 Simulaciones de los valores de las métricas de Shou y de Simkin Diferentes tamaños de red para las simulaciones.

20x20 Nodos



*Figura S 1 Comportamiento de las tasas de recableado en una red de 20x20 utilizando la distribución de Poisson*

Para esta red simulada de 20 nodos reguladores y 20 nodos blancos, se observaron distintos patrones de saturación. A lo largo del eje X que corresponde a la tasa de recableado de Shou, podemos observar que en 2 ocasiones ocurren saturaciones. La primera saturación es debido a que se llega al tamaño máximo de red. La segunda saturación se debe a que se alcanza un punto de saturación de conectividades, es decir, cada microRNA está conectado al menos una vez con cada uno de los blancos. En el eje Y, correspondiente a la tasa de recableado de Simkin, se puede observar que la tasa va aumentando hasta que llega el punto donde se satura y la normalización por el número de sitios hace que disminuya.

100 Nodos

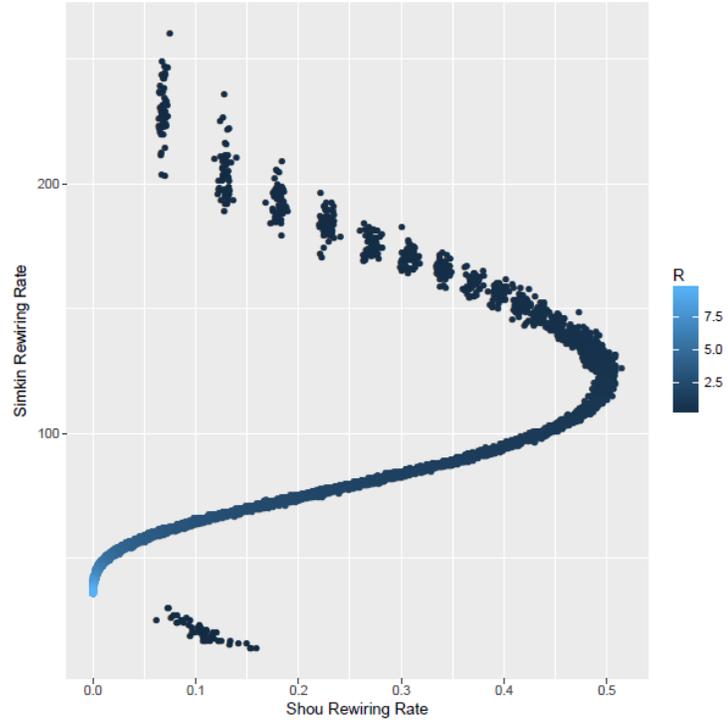


Figura S 2 Comportamiento de las tasas de recableado en una red de 100x100 utilizando la distribución de Poisson

150 Nodos.

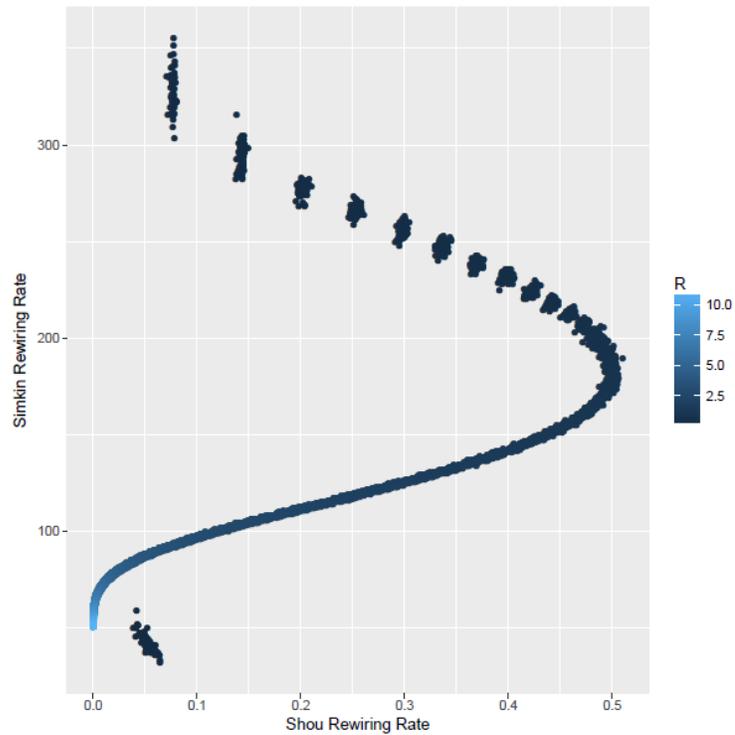


Figura S 3 Comportamiento de las tasas de recableado en una red de 150x150 utilizando la distribución de Poisson

## 12.2 Análisis con el conjunto de 3'UTRs de los Homínidos

### 12.2.1 Efecto del uso de diferentes puntos de corte de CS++ para la métrica de Shou

Valor de corte de CS++: 10%

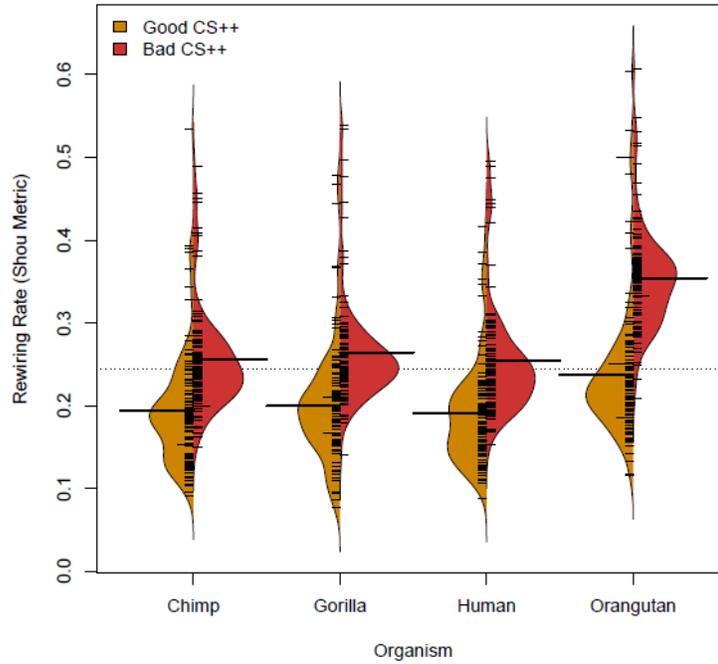


Figura S 4 Tasa de recableado de Shou para los microRNAs conservados principalmente en vertebrados con blancos predichos utilizando un corte de Cs++ del 10% más extremo

Valor de corte de CS++: 20%

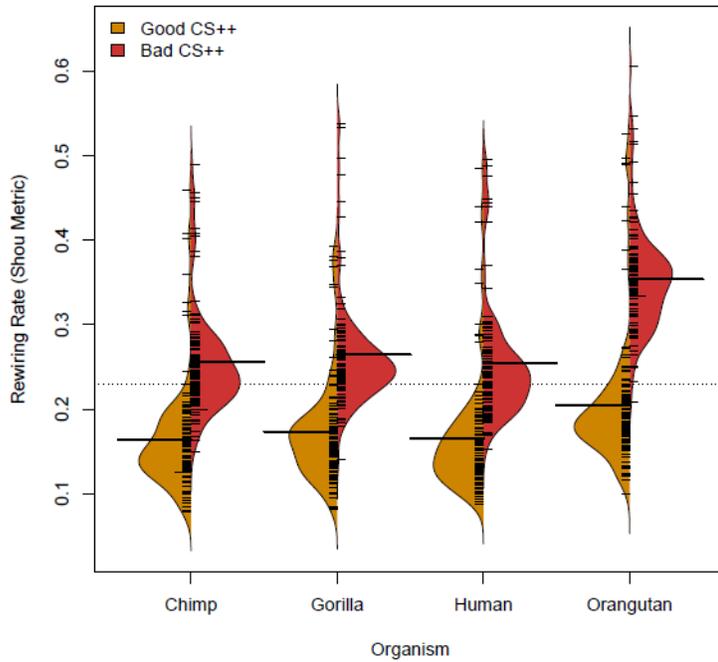


Figura S 5 Tasa de recableado de Shou para los microRNAs conservados principalmente en vertebrados con blancos predichos utilizando un corte de Cs++ del 20% más extremo

### 12.2.2 Recableado por kilobase de blanco utilizando diferentes cortes de CS++ (utilizando las 3'UTRs de Simkin)

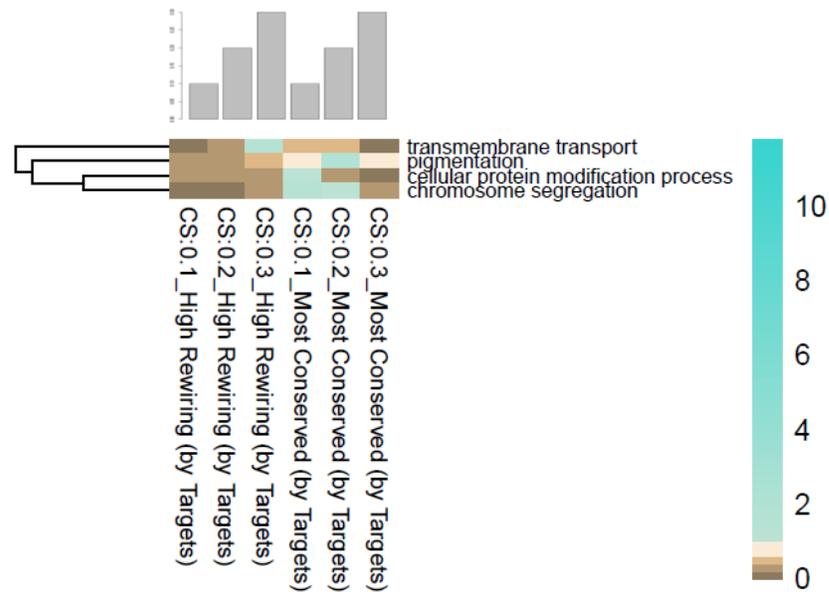


Figura S 6

### 12.3 Análisis de un conjunto alternativo de 3' UTRs de los primates obtenidas de un alineamiento de mamíferos

#### 12.3.1 Expresión de genes con alto recableado por blanco en núcleo caudado.

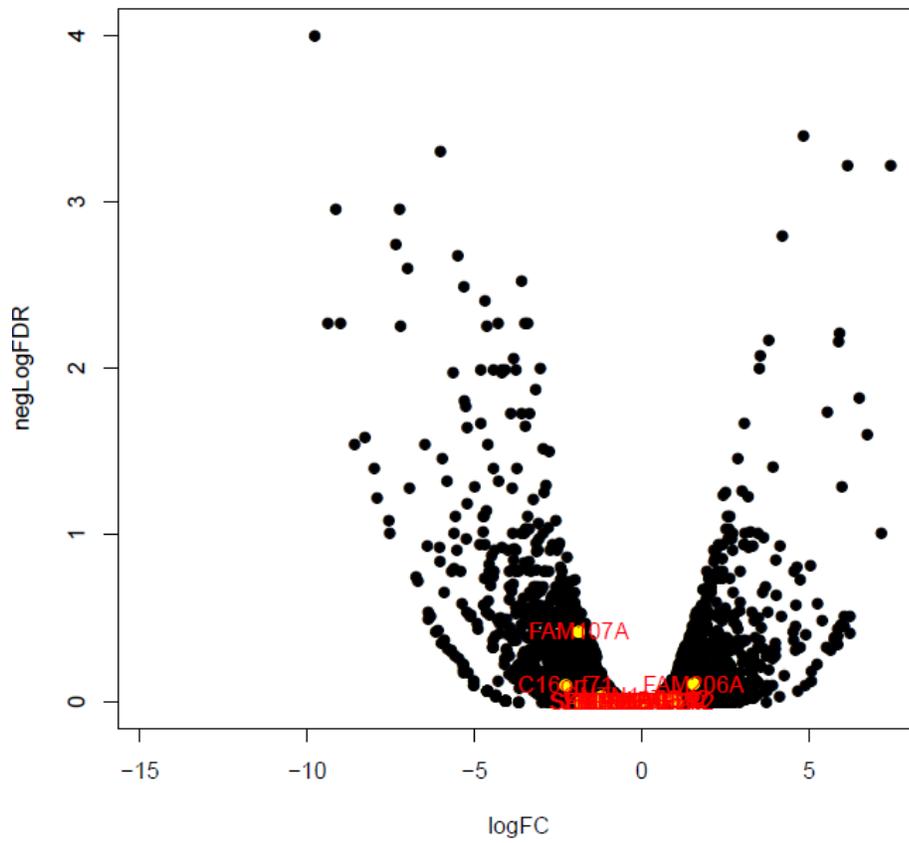


Figura S 7 Gráfico de Volcán para la comparación de niveles de expresión de genes en núcleo caudado entre humano y chimpancé. En rojo se representan los nombres de los genes que ganaron sitios de pegado en chimpancé y perdieron sitios de pegado en humano.

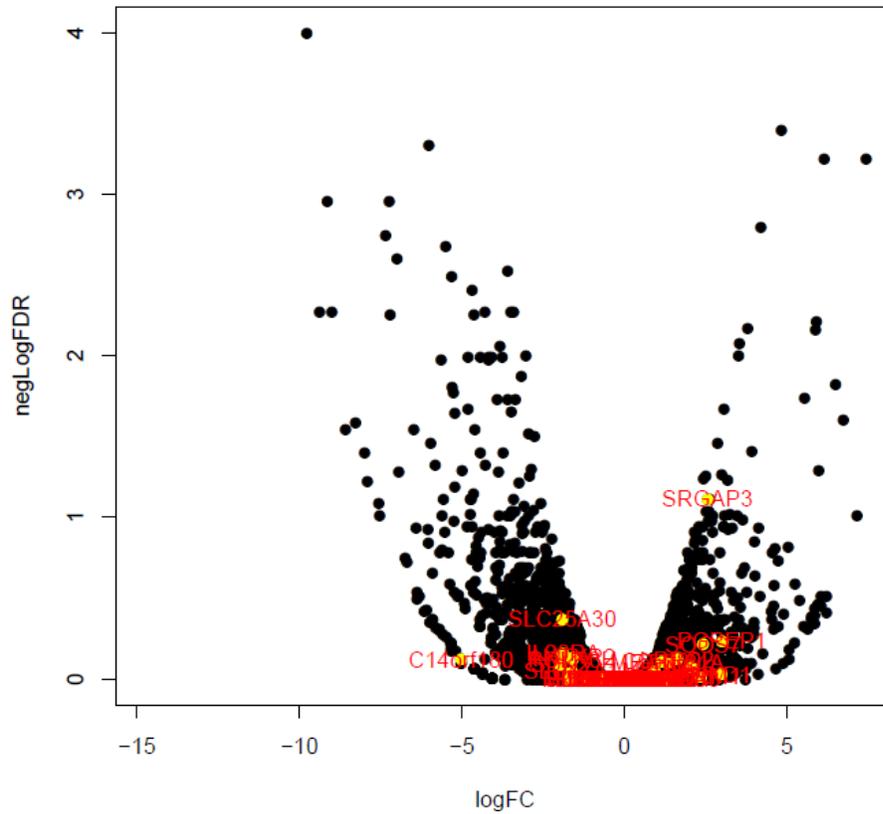


Figura S 8 Gráfico de Volcán para la comparación de niveles de expresión de genes en núcleo caudado entre humano y chimpancé. En rojo se representan los nombres de los genes que perdieron sitios de pegado en chimpancé y ganaron sitios de pegado en humano.

## 12.4 Análisis con familias de microRNAs más recientes

### 12.4.1 Familias conservadas en mamíferos

#### 12.4.1.1 Tasas de recableado con diferentes cortes de CS++ para miRNAs de mamíferos

Valor de corte de CS++: 10%

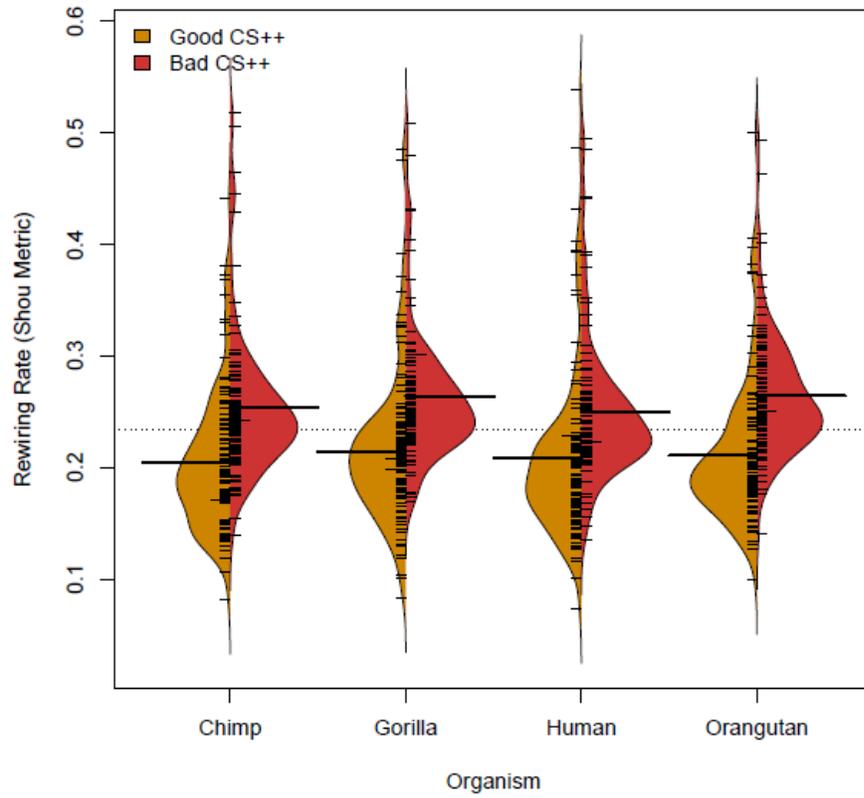
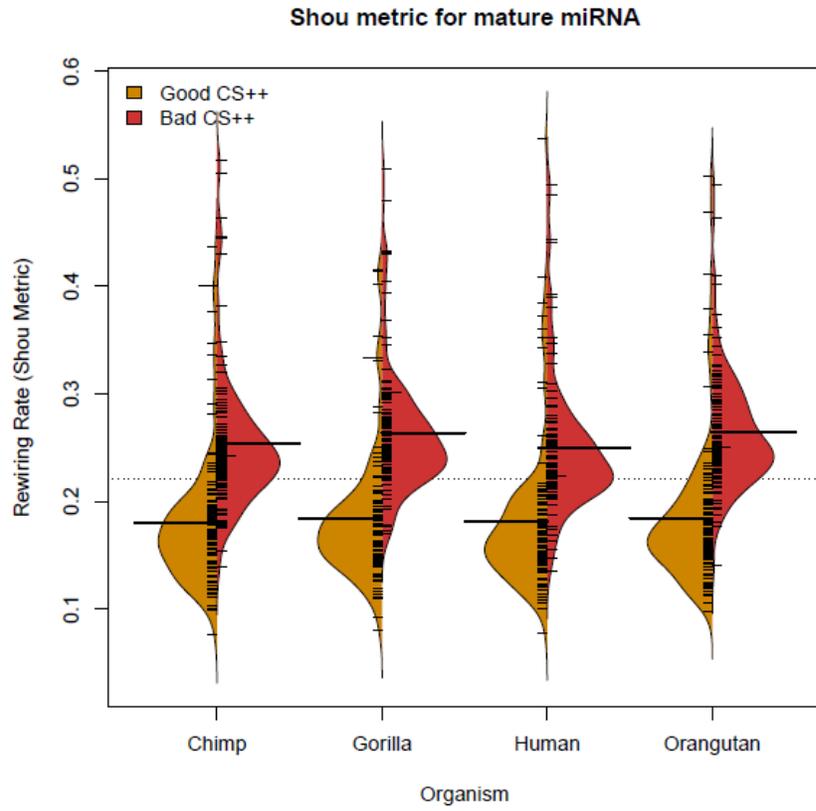


Figura S 9 Tasa de recableado de Shou para los microRNAs conservados principalmente en mamíferos con blancos predichos utilizando un corte de Cs++ del 10% más extremo

Valor de corte de CS++: 20%



*Figura S 10 Tasa de recableado de Shou para los microRNAs conservados principalmente en mamíferos con blancos predichos utilizando un corte de Cs++ del 20% más extremo*

### 12.4.1.2 Análisis de categorías funcionales

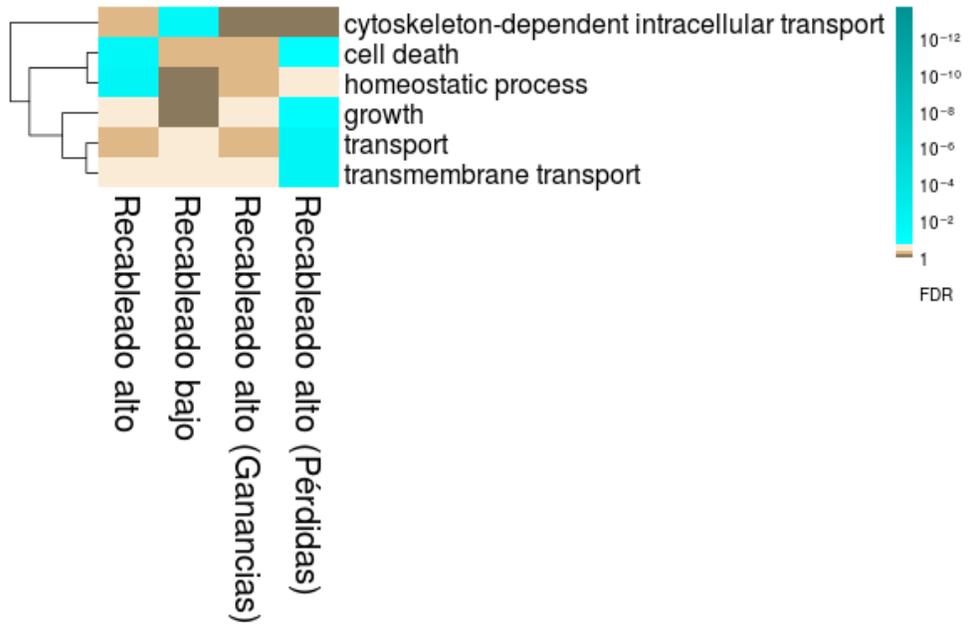


Figura S 11 Categorías funcionales enriquecidas para genes que tuvieron valores extremos de recableado en humano utilizando tanto las familias de microRNAs conservadas en vertebrados como las conservadas hasta mamíferos.