PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE INGENIERÍA

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

Expresión génica diferencial de oocitos claros y oscuros de *Espadarana prosoblepon* (Centrolenidae).

Disertación previa a la obtención del título de Magister en Biología Computacional

Henry Fabián Macías Bazante

Quito, 2023

Certifico que la Disertación de Maestría en Biología Computacional del Sr. Henry Fabián Macías Bazante ha sido concluida de conformidad con las normas establecidas; por lo tanto, puede ser presentada para la calificación correspondiente.



Dr. Andrés Romero Carvajal

Director de la Disertación

Quito, 05 de septiembre del 2023

En memoria de mi padre

AGRADECIMIENTOS

Agradezco especialmente a mi padre Narciso Antonio Macías Molina al que siempre recuerdo con amor y gratitud.

A mis padres por el apoyo que me han brindado toda la vida, por ser mi ejemplo a seguir y por su cariño incondicional. A mis hermanos por las aventuras y su compañía en todo momento.

A mi abuelita por siempre haberme dado todo su cariño y lo mejor de sí. A Violetita y Miguelito que han sido como mis segundos padres.

Agradezco al Doctor Andrés Romero-Carvajal por haberme permitido formar parte de su laboratorio, por su guía durante el desarrollo de este trabajo; además, por su confianza y por incentivarnos a aprender cosas nuevas cada día.

A Santy, Ricardo, Andrés, Christopher, Diego, Christian y Ángel por su amistad y todas las experiencias compartidas.

A Julissa por su cariño, compañía y apoyo incondicional.

Agradezco a la Pontificia Universidad Católica del Ecuador por el financiamiento para la realización de este proyecto. A la Balsa de los Sapos por la colaboración con los especímenes.

TABLA DE CONTENIDOS

1. RESUMEN 1
2. ABSTRACT 2
3. INTRODUCCIÓN 3
3.1 OBJETIVOS 5
3.1.1 OBJETIVO GENERAL 5
3.1.2 OBJETIVOS ESPECÍFICOS 5
4. MATERIALES Y MÉTODOS 6
4.1 EXTRACCIÓN DE ARN Y SECUENCIACIÓN 6
4.2 CONTROL DE CALIDAD Y FILTRADO DE LOS TRANSCRITOS PREVIO AL ENSAMBLAJE 7
4.3 ENSAMBLAJE DE TRANSCRITOS 7
4.4 ANOTACIÓN FUNCIONAL 7
4.5 ANÁLISIS DE EXPRESIÓN GÉNICA DIFERENCIAL 8
5. RESULTADOS 9
5.1 SECUENCIACIÓN DE ARN Y ENSAMBLAJE DEL TRANSCRIPTOMA 9
5.2 EXPRESIÓN GÉNICA DIFERENCIAL 9
5.3 ANOTACIÓN FUNCIONAL 10
5.3.1 DOMINIOS PROTEICOS 10
5.3.2 ONTOLOGÍA GÉNICA (GO) 10
5.3.3 VÍAS METABÓLICAS 11
6. DISCUSIÓN 12
6.1 ENSAMBLAJE 12
6.2 EXPRESIÓN GÉNICA DIFERENCIAL ENTRE OOCITOS CLAROS Y OSCUROS Y ANOTACIÓN FUNCIONAL 12

6.2.1 DOMINIOS PROTEICOS 13
6.2.2 ONTOLOGÍA GÉNICA (GO) 13
6.2.3 VÍAS METABÓLICAS 15
7. CONCLUSIONES 16
8. RECOMENDACIONES 17
9. REFERENCIAS BIBLIOGRÁFICAS 18
10. FIGURAS 23
11. TABLAS 30
12. ANEXOS 33

LISTA DE FIGURAS

Figura 1. Flujo de trabajo del ensamblaje, anotación funcional y análisis de expresión génica diferencial. 23
Figura 2. Gráfico de volcán de la expresión génica diferencial con genes. 24
Figura 3. Mapa de calor del agrupamiento jerárquico de las muestras de oocitos de *Espadarana prosoblepon* de acuerdo a sus niveles de expresión. 25
Figura 4. Clasificación de Ontología Génica (GO). 26
Figura 5. Gráfico circular de Ontología Génica (GO) con los términos relacionados a la melanogénesis. 27
Figura 6. Mapa de la vía metabólica de la maduración de los oocitos mediada por la progesterona. 28

LISTA DE TABLAS

Tabla 1. Estadísticas del ensamblaje global de dos muestras de huevos claros y dos de huevos oscuros de *Espadarana prosoblepon.* 30
Tabla 2. Datos de completitud del ensamblaje de los transcritos. 30
Tabla 3. Datos del porcentaje de mapeo de secuencias por cada muestra 31
Tabla 4. Datos de expresión diferencial de la comparación entre los transcritos de oocitos claros y oscuros $p < 0.05$ 31
Tabla 5. Lista de los diez dominios proteicos más comunes 32

LISTA DE ANEXOS

Anexo 1. Individuos de *Espadarana prosoblepon* utilizados para la extracción de oocitos. 33
Anexo 2. Resumen de las estadísticas de las secuencias limpias. 34
Anexo 3. Isoformas diferencialmente expresadas en los términos GO relacionados a la melanogénesis y pigmento..... 38

1. RESUMEN

El pigmento es una de las adaptaciones más comunes en el reino animal y ha sido ampliamente estudiado en anuros; a pesar de esto, se conoce muy poco sobre su función y control molecular durante el desarrollo embrionario. Las hembras de la rana de cristal *Espadarana prosoblepon* pueden poner oocitos con pigmentación reducida (claros) y otros de color café oscuro, los cuales llegan a formar renacuajos y adultos con pigmento normal. El fenotipo claro está determinado por el alelo *pel* (pale eggs like) que es recesivo y su herencia está limitada al sexo, ya que los machos no producen oocitos; mientras que el fenotipo oscuro se expresa por el alelo salvaje que es el dominante. Debido a la presencia de los dos fenotipos, esta especie ha sido propuesta como un modelo para estudiar el papel adaptativo de la pigmentación del huevo. El objetivo principal de este estudio es determinar los procesos moleculares involucrados en la pigmentación melánica de los oocitos y así tener más información sobre el rol del pigmento durante el desarrollo embrionario. En este estudio se realizó un análisis de expresión génica diferencial entre los oocitos claros y oscuros. Se identificaron 4.672 genes expresados diferencialmente entre los oocitos claros y oscuros. Entre estos, se identificaron 25 isoformas que pertenecen a 11 genes que están relacionados con la pigmentación y melanogénesis. Estos genes diferencialmente expresados están presentes en procesos de biosíntesis y metabolismo del pigmento. Los resultados obtenidos en este estudio indican que en los oocitos se están llevando a cabo procesos moleculares involucrados en la formación, transporte y establecimiento del pigmento. Además, se puede concluir que en la pigmentación melánica normal y reducida existen controles y procesos moleculares distintos.

Palabras clave: *Espadarana prosoblepon,* pigmentación melánica, oocitos oscuros, oocitos con pigmentación reducida (claros), expresión génica diferencial, melanogénesis.

2. ABSTRACT

The melanic pigment is one of the most common adaptations in the animal kingdom and has been extensively studied in anurans; Despite this, very little is known about its function and molecular control during embryonic development. The females of the glass frog *Espadarana prosoblepon* can lay oocytes with reduced pigmentation (clear) and others of dark brown color, which become tadpoles, and adults with normal pigment. The light phenotype is determined by the *pel* allele (pale eggs like) which is recessive and its inheritance is limited to sex, since males do not produce oocytes; while the dark phenotype is expressed by the wild allele which is dominant. Due to the presence of the two phenotypes, this species has been proposed as a model to study the adaptive role of egg pigmentation. The main objective of this study is to determine the molecular processes involved in the melanic pigmentation of oocytes and thus have more information about the role of pigment during embryonic development. In this study, an analysis of differential gene expression between light and dark oocytes was performed. They identified 4,672 genes differentially expressed between light and dark oocytes. Among these, 25 isoforms belonging to 11 genes that are related to pigmentation and melanogenesis were identified. These differentially expressed genes are present in processes of biosynthesis and pigment metabolism. The results obtained in this study indicate that molecular processes involved in the formation, transport, and establishment of pigment take place in oocytes. In addition, it can be concluded that in normal and reduced melanic pigmentation there are different controls and molecular processes.

Keywords: *Espadarana prosoblepon*, melanic pigmentation, dark oocytes, oocytes with reduced pigmentation (light), differential gene expression, melanogenesis.

3. INTRODUCCIÓN

El pigmento es una de las adaptaciones más comunes en el reino animal y ha sido ampliamente estudiada en anuros; a pesar de esto, se conoce muy poco sobre su función y control molecular durante el desarrollo embrionario (Rudh y Qvarnström, 2013; Salazar-Nicholls *et al*., 2021).

El desarrollo embrionario temprano es el proceso mediante el cual un huevo fertilizado pasa a ser un individuo completamente formado. Durante esta etapa los embriones de anuros poseen varias adaptaciones que les permiten sobrevivir y formarse con normalidad en su medio ambiente (Elinson y del Pino, 2012). Una de estas adaptaciones es el pigmento, que en adultos y renacuajos sirve para enviar señales, camuflaje, termorregulación y protección de rayos UV (Duellman y Trueb, 1994; Rudh y Qvarnström, 2013). Durante las etapas de desarrollo embrionario se conoce muy poco sobre su función.

En la especie *Espadarana prosoblepon* las hembras pueden poner huevos de dos fenotipos distintos, uno tiene huevos claros debido a la pigmentación reducida y el otro tiene pigmentación normal y es de color café oscuro. El fenotipo de huevos claros está determinado por el alelo *pel* (pale egg like) que es recesivo, y su herencia está limitada al sexo, ya que los machos no producen oocitos. Se asume que el fenotipo *pel* es causado por un alelo que sigue un patrón de herencia mendeliana (Salazar-Nicholls *et al*., 2021). El fenotipo oscuro se expresa por el alelo salvaje que es el dominante. Ambos fenotipos de oocitos llegan a formar renacuajos y adultos con pigmento normal. Debido a la presencia de los dos fenotipos esta especie ha sido propuesta como un modelo para estudiar el papel adaptativo de la pigmentación del huevo (Salazar-Nicholls *et al*., 2021).

En el desarrollo de anuros existen cambios en la expresión génica, que están relacionados con las adaptaciones que les permiten sobrevivir en distintos hábitats (Elinson y del Pino, 2012; Gilbert y Barresi, 2017; Piprek *et al*., 2018). El análisis de transcriptomas permite determinar la expresión de los genes en tejidos y estadios de vida específicos (Gall, 2014; Smirnov *et al*., 2022).

En *Espadarana prosoblepon* se han secuenciado los transcritos de tres tejidos relacionados al sistema inmune de los cuales se obtuvo la anotación de 8881 genes (Savage *et al*., 2014). En otras especies se han realizado estudios sobre la actividad transcripcional en oocitos, pero no se han centrado en la función del pigmento (Gall, 2014; Simeoni *et al*., 2012). Además, no existe un genoma de *Espadarana prosoblepon* que pueda servir de referencia para realizar la reconstrucción del transcriptoma. Por lo tanto, es necesario realizar un ensamblaje de novo de los transcritos de los oocitos.

Los transcriptomas disponibles para *Espadarana prosoblepon* no permiten resolver la pregunta sobre la función del pigmento en desarrollo porque corresponden a otros estadios de desarrollo y no se conoce si provienen de individuos que pertenecen al polimorfismo pel (Salazar-Nicholls *et al*., 2021; Savage *et al*., 2014). Debido a esto, en este estudio se analizará el transcriptoma de los oocitos de los dos fenotipos para determinar la expresión diferencial de genes durante esta etapa de desarrollo y poder identificar genes relacionados con la melanogénesis y el pigmento.

El análisis del transcriptoma de los dos fenotipos de los oocitos de *Espadarana prosoblepon* nos permitirá conocer las diferencias de expresión génica y nos brindará más información sobre la función del pigmento melánico durante el desarrollo embrionario y sus mecanismos moleculares (Salazar-Nicholls *et al*., 2021).

En general se conoce muy poco sobre los mecanismos biológicos durante el desarrollo embrionario de especies no modelo, por lo que realizar el análisis del transcriptoma completo de esta especie ayudará a posicionarla como un modelo emergente para el estudio de la pigmentación durante el desarrollo. Además, este estudio tiene el potencial de contribuir a los esfuerzos para la conservación de esta y otras especies de anfibios y facilitar el planteamiento de nuevos proyectos de investigación en el mismo campo.

3.1 OBJETIVOS

3.1.1 OBJETIVO GENERAL

Describir la expresión diferencial de genes de oocitos claros y oscuros de *Espadarana prosoblepon*

3.1.2 OBJETIVOS ESPECÍFICOS

- Identificar los genes que están relacionados con la melanogénesis
- Determinar las diferencias de la expresión de los genes relacionados con la melanogénesis y pigmentación entre los dos fenotipos de oocitos
- Determina la función del pigmento melánico durante el desarrollo embrionario

4. MATERIALES Y MÉTODOS

Los individuos utilizados en este estudio han sido criados en cautiverio por la Iniciativa de Conservación “Balsa de los Sapos” de la PUCE.

Los análisis computacionales se realizaron en el supercomputador de CEDIA, siguiendo el flujo de trabajo presentado en la Figura 1.

4.1 EXTRACCIÓN DE ARN Y SECUENCIACIÓN

Los oocitos claros y oscuros en estadio de vitelogénesis fueron extraídos de los ovarios de ranas adultas de la especie *Espadarana prosoblepon*. Las hembras se pusieron en una caja petri con la solución Steinberg´s 1X estéril, cubriendo únicamente la zona abdominal. Se colocaron gradualmente gotas de metanosulfonato de tricaína 0.5% (MS-222) hasta que las hembras estaban completamente sedadas, para comprobar esto se las colocaba boca arriba y si no tenían la capacidad de regresar a posición ventral se procedía con la cirugía. Se realizó una incisión de 5 mm en la posición ventro-lateral del abdomen, con ayuda de pinzas estériles se extrajeron los ovarios y se suturó la herida con hilos biodegradables. Después del procedimiento las ranas fueron lavadas con Steinberg´s 1X y se recuperaron en un contenedor con suficiente humedad y ventilación (Salazar-Nicholls *et al*., 2021).

Los oocitos vitelogénicos extraídos se almacenaron a -80ºC en tubos cónicos de 2 ml con la solución RNAlater y se enviaron para la secuenciación del ARN. La creación de librerías y secuenciación se hizo con la tecnología de Illumina Novaseq, se realizó la secuenciación de lecturas pareadas con una longitud de 151 pares de bases.

A los oocitos claros se les asignaron los códigos B-4033 y B-4161. Mientras que a los oocitos oscuros los códigos N-3568 y N-3713. La letra B corresponde al fenotipo “blanco”, la letra N al fenotipo “negro” y los números corresponden al terrario al que pertenecen las hembras utilizadas en las cirugías (Anexo 1).

4.2 CONTROL DE CALIDAD Y FILTRADO DE LOS TRANSCRITOS PREVIO AL ENSAMBLAJE

Después de la secuenciación se obtuvieron las lecturas de secuencias crudas de los oocitos claros y oscuros, se evaluó la calidad de las secuencias con FastQC v0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Se recortaron adaptadores y filtraron las secuencias cortas y de baja calidad, utilizando los parámetros preestablecidos en el preprocesador fastp v0.23.4 (Chen *et al*., 2018). Después de la limpieza, nuevamente se evaluó la calidad de las secuencias y se procedió con el ensamblaje de los transcritos.

4.3 ENSAMBLAJE DE TRANSCRITOS

El ensamblaje de novo de todos los transcritos se realizó con Trinity v2.15.1, el cual realiza la reconstrucción con el enfoque de k-mers (Grabherr *et al*., 2011). Se utilizó el parámetro --SS_lib_type FR, que sirve para identificar correctamente la cadena conductora (FORWARD) y la retardada (REVERSE). Se redujo la redundancia de las secuencias con CD-HIT v4.8.1, utilizando los valores predeterminados (Fu *et al*., 2012). Las estadísticas del ensamblaje se obtuvieron con el script 'TrinityStats.pl' que viene integrado en Trinity v2.15.1 (Grabherr *et al*., 2011).

Finalmente, se evaluó la completitud de los transcritos ensamblados con los linajes de eucariotas, metazoos, vertebrados y tetrápodos utilizando BUSCO v5.4.7 (Manni *et al*., 2021).

4.4 ANOTACIÓN FUNCIONAL

Después del ensamblaje se realizó la predicción de regiones codificantes con TransDecoder v5.7.0 (Haas *et al*., 2013). Primero se extrajeron los marcos de lectura abiertos largos y después se realizó la predicción de las regiones codificantes. Se obtuvieron tres archivos, entre estos un archivo con las secuencias de aminoácidos de las regiones codificantes.

La anotación de las proteínas se realizó con Trinotate v4.0.0. Para esto se utilizaron los archivos del ensamblaje de los transcritos en formato fasta, el archivo gene trans map y las secuencias de péptidos obtenidos en los pasos previos. Se realizó la anotación utilizando las bases de datos de Swiss-Prot (Bairoch y Apweiler, 1999; UniProt Consortium, 2021) y Pfam (Mistry *et al*., 2021). La anotación de las rutas metabólicas se hizo con las secuencias de péptidos en el servidor KAAS (Moriya *et al*., 2007) y posteriormente se hizo la reconstrucción de los mapas en el servidor de KEGG Mapper (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa *et al*., 2022; Kanehisa y Sato, 2020)

Los gráficos de ontología génica se realizaron con el paquete GOplot v1.0.2 (Walter *et al*., 2015) en el programa R v4.3.1 (R Core Team, 2022).

4.5 ANÁLISIS DE EXPRESIÓN GÉNICA DIFERENCIAL

El alineamiento y la estimación de la abundancia de los transcritos se realizó con el script 'align_and_estimate_abundance.pl' de Trinity (Grabherr *et al*., 2011); para estimar la abundancia se usó Salmon v1.10.2 (Patro *et al*., 2017) y el alineamiento de secuencias se hizo con bowtie2 v2.5.1 (Langmead y Salzberg, 2012). Con los datos de las abundancias estimadas se construyeron las matrices de abundancia de genes e isoformas con el script 'abundance_estimates_to_matrix.pl', que también viene incluido en Trinity (Grabherr *et al*., 2011).

Los análisis de expresión génica diferencial se realizaron con el script "run_DE_analysis.pl" de Trinity (Grabherr *et al*., 2011). En las opciones, se especificó el uso del paquete edgeR de Bioconductor para la identificación de genes e isoformas diferencialmente expresadas entre tratamientos. Se realizó la normalización utilizando el método TMM (Trimmed Mean of M-values), el cual que ajusta las diferencias en la profundidad de la secuenciación en las muestras (Robinson *et al*., 2010). Las figuras de la expresión diferencial de genes como el gráfico de volcán y el mapa de calor se diseñaron con los paquetes tidyverse v2.0.0, ggrepel v0.9.3 y ggplot2 v3.4.2 (Slowikowski, 2023; Wickham, 2009; Wickham *et al*., 2019) en R v4.3.1 (R Core Team, 2022)

5. RESULTADOS

5.1 SECUENCIACIÓN DE ARN Y ENSAMBLAJE DEL TRANSCRIPTOMA

La secuenciación del ARN de los oocitos claros y oscuros de *Espadarana prosoblepon* se realizó con illumina. Se obtuvieron secuencias crudas de 151 pares de bases que se limpiaron y filtraron. Las secuencias limpias tuvieron una longitud de 136 pb, debido a la remoción de los adaptadores, y una puntuación promedio en la calidad de Phred de 36 en todas las muestras (Anexo 2).

Se realizó el ensamblaje con las secuencias limpias de las cuatro muestras en una sola corrida en Trinity y se obtuvo un total de 479,370 transcritos ensamblados, 239,967 genes, un contenido de GC de 46.10%, un promedio de contigs de 653.91, y un N50 de 1003 bases (Tabla 1). Además, en el ensamblaje se identificaron las isoformas génicas, ya que el método utilizado por Trinity agrupa las secuencias de transcritos relacionadas y permite su identificación.

La evaluación de la calidad de los ensamblajes se realizó mediante la identificación de genes ortólogos de cuatro linajes con BUSCO. Se encontró que el ensamblaje final tuvo una completitud de 83.8% para tetrápodos, 92.5% para vertebrados, 99.4% para metazoos y 99.6 % para eucariotas (Tabla 2).

5.2 EXPRESIÓN GÉNICA DIFERENCIAL

Se realizó el alineamiento y la estimación de la abundancia de los transcritos ensamblados y se obtuvo un promedio de secuencias alineadas de 96.5% para las cuatro muestras (Tabla 3). Después se hizo la construcción de las matrices de expresión de genes e isoformas, que posteriormente se utilizaron para los análisis de expresión génica diferencial utilizando el paquete edgeR.

Para medir la expresión génica se utilizó como medida de abundancia relativa los conteos por millón (CPM) y se consideraron como transcritos expresados los que tenían un valor de CPM > 1 en al menos una de las muestras. La expresión génica diferencial en

oocitos claros y oscuros se estimó mediante el valor de FDR < 0.05 (tasa de descubrimiento falso o valor de p-ajustado). En el análisis de expresión génica diferencial se identificaron un total de 66.047 genes expresados entre las cuatro muestras de oocitos, de los cuáles 61.375 no presentaron cambios significativos en la expresión génica entre ambos fenotipos. Por otro lado, se encontró que 4.672 genes se expresan diferencialmente entre los oocitos con fenotipo claro y oscuro, de estos 1.773 genes se encuentran subexpresados y 2.899 sobreexpresados (Figura 2 y 3) (Tabla 4).

5.3 ANOTACIÓN FUNCIONAL

Se identificaron los posibles marcos abiertos de lectura (ORF) con TransDecoder, y a partir de estos se realizó la predicción de las regiones codificantes. En base a estas regiones, se hizo la anotación de los transcritos con Trinotate usando las bases de datos de Swiss-Prot (Bairoch y Apweiler, 1999; UniProt Consortium, 2021) y Pfam (Mistry *et al*., 2021) . En total se anotaron 54.494 secuencias de los transcritos.

5.3.1 DOMINIOS PROTEICOS

En la anotación se identificaron los dominios proteicos de acuerdo a la base de datos de Pfam (Mistry *et al*., 2021). Se identificaron 2402 dominios proteicos en los transcritos, siendo el más común el de C2H2 Zinc Finger con 788 transcritos anotados, seguido de KRAB box y WD domain, G-beta repeat, cada uno con 251 y 236 transcritos anotados respectivamente (Tabla 5).

5.3.2 ONTOLOGÍA GÉNICA (GO)

En la anotación funcional con GO se identificaron 24.893 transcritos que corresponden a 691 genes, los cuales se clasificaron en las tres categorías de ontología génica que son: procesos biológicos (BP), componentes celulares (CC) y funciones moleculares (FM), las siglas corresponden a sus nombres en inglés. Los tres términos más enriquecidos son los de procesos celulares (BP), agrupamiento o binding (MF) y organelas (CC) (Figura 4).

Entre los términos de GO se identificaron 19 que están relacionados con la melanogénesis y pigmentación, de los cuales 18 están enriquecidos. Siendo el término de pigmentación (GO:0043473) el más enriquecido, seguido del término de "desarrollo del pigmento" (GO:0048066) y la "localización del pigmento granular" (GO:0051875). El único término que no tiene ningún valor de enriquecimiento es el de "proceso biosintético de melanina a partir de tirosina" (GO:0006583). Los 19 términos identificados pertenecen a la categoría de procesos biológicos (BP) y están involucradas 95 isoformas génicas (Figura 4 y 5).

De los transcritos identificados en los términos de la pigmentación, 25 isoformas se encuentran diferencialmente expresadas entre los oocitos claros y oscuros, 18 están sobreexpresadas y 7 subexpresadas. Las isoformas sobreexpresadas tienen rangos de valores de log2FC que van desde 12.46 a 15.48, mientras que las isoformas subexpresadas tienen valores de log2FC desde -12.44 a -13.63. Todas las isoformas diferencialmente expresadas pertenecen a los términos de "procesos de biosíntesis de pigmentos" (GO:0046148) y "procesos de metabolismo de pigmentos" (GO:0042440) (ANEXO 3).

5.3.3 VÍAS METABÓLICAS

En la anotación funcional realizada en el servidor de KAAS (Moriya *et al*., 2007) se identificaron 33.756 términos funcionales, los cuales se utilizaron para hacer la reconstrucción de mapas en el servidor KEGG mapper (Kanehisa *et al*., 2022; Kanehisa y Sato, 2020).Después de realizar los mapas se identificaron un total 428 vías metabólicas en los transcritos de los oocitos.

Entre las vías metabólicas se identificó la ruta de la maduración de oocitos mediada por la progesterona, que se encuentra completa y en la que se expresan 62 isoformas de los genes identificados en los transcritos (Figura 6). Además, se identificó la vía metabólica de la melanogénesis, que está parcialmente completa e involucra la expresión de 61 isoformas de los genes de oocitos. En esta ruta metabólica no está presente la proteína alfa-melanocito estimulante (αMSH) y la proteína de señalización agouti (ASP) (Figura 7).

6. DISCUSIÓN

6.1 ENSAMBLAJE

Los resultados obtenidos demuestran que el ensamblaje de novo de los transcritos es de alta calidad. Antes del ensamblaje se realizó la limpieza y filtración de las lecturas obtenidas de la secuenciación del ARN de los oocitos de *Espadarana prosoblepon*. Las secuencias limpias tuvieron un puntaje de Phred de 36 en todas las muestras, lo que significa que los ácidos nucleicos fueron identificados con una tasa de error de 1 en 4000 (Anexo 2) (Ewing *et al*., 1998). Este resultado en el puntaje de Phred nos indica que la secuenciación fue de alta calidad en todas las muestras, por lo que se procedió con el ensamblaje.

Después de realizar el ensamblaje se obtuvo un valor N50 de 1003 bases y se evaluó la calidad a través del análisis de completitud. La completitud del ensamblaje tuvo un valor 83.8% para el linaje de los tetrápodos (Tabla 2). El porcentaje de completitud mínimo recomendado para ensamblajes de alta calidad es de 80%, por lo tanto, nuestro ensamblaje se puede considerar como tal (Raghavan *et al*., 2022).

Durante el proceso de limpieza y ensamblaje de los transcritos, existieron varias limitaciones. Inicialmente, para la limpieza y filtrado de secuencias se utilizó Trimmomatic v0.39 (Bolger *et al*., 2014); sin embargo, las estadísticas de la limpieza y la calidad del ensamblaje no fueron tan buenas como las que se obtuvieron con fastp (Chen *et al*., 2018). Después del ensamblaje, en el análisis de completitud, se identificaron más grupos duplicados que simples. Debido a esto, se decidió utilizar la herramienta CD-HIT para disminuir la redundancia (Fu *et al*., 2012).

6.2 EXPRESIÓN GÉNICA DIFERENCIAL ENTRE OOCITOS CLAROS Y OSCUROS Y ANOTACIÓN FUNCIONAL

Se analizó la expresión de genes de los oocitos con fenotipo claro y oscuro. En el análisis se identificaron un total de 66.047 genes expresados entre las cuatro muestras. En la comparación entre muestras del mismo fenotipo se observa que hay una expresión génica

similar (Figura 3). Mientras que entre los oocitos claros y oscuros se encontraron 4.672 genes expresados diferencialmente de los cuales 2899 están sobreexpresados y 1773 genes subexpresado (Figura 2 y 3). Estos datos indican que en las muestras pertenecientes a un mismo fenotipo los procesos moleculares son similares, mientras que entre muestras de diferentes fenotipos existen diferencias en algunos procesos.

6.2.1 DOMINIOS PROTEICOS

En la anotación funcional del transcriptoma ensamblado se identificaron 54.494 transcritos conocidos que se agrupan en 17.556 genes. Entre estos se encontró que los dominios proteicos más comunes fueron los dedos de zinc C2H2, KRAB box (o caja asociada a Kruppel) y WD, G-beta repeat (Tabla 5).

Los dedos de zinc C2H2 forman parte de la familia más grande de factores de transcripción en eucariotas superiores (Fedotova *et al*., 2017). El dominio KRAB box forma parte de aproximadamente un tercio de las proteínas con dedos de zinc C2H2 y está presente solo en tetrápodos. Las proteínas que tienen el dominio KRAB box participan en el mantenimiento del nucléolo, la diferenciación celular, la proliferación celular, la apoptosis y la transformación neoplásica (Urrutia, 2003). Las proteínas WD40 son una familia grande que se encuentra en todos los eucariotas y están implicadas en el crecimiento, desarrollo, transducción de señales, control de ciclo celular, transporte intracelular, remodelación de la cromatina, organización citoesquelética, regulación transcripcional, entre otras (Jain y Pandey, 2018).

En general, los dominios proteicos más comunes del ensamblaje de novo están involucrados en funciones biológicas esenciales (Albagli *et al*., 1995; Isalan, 2013; Kobe y Kemp, 2003; Li *et al*., 2006; Nalefski y Falke, 1996; SenGupta, 2013).

6.2.2 ONTOLOGÍA GÉNICA (GO)

En la anotación funcional realizada con ontología génica, se determinó que la mayoría de los términos enriquecidos están relacionados con funciones metabólicas y celulares (Figura 4). Esto coincide con la información obtenida en los dominios proteicos,

lo cual valida la anotación de nuestros transcritos y nos da información sobre los procesos biológicos que suceden en los oocitos.

Entre los términos más enriquecidos se encontró el de “unión de ácidos nucleicos” (GO: 0003676) (Figura 5). Este término representa la unión de proteínas a los ácidos nucleicos, dichas proteínas se suelen caracterizar por tener dominios de dedos de zinc. Esto concuerda con la identificación de los dominios proteicos en los cuales el más común fue el de dedos de zinc C2H2 (Tabla 5). En general, los términos enriquecidos de GO cumplen funciones biológicas esenciales.

En cuanto a los términos relacionados a la melanogénesis y pigmentación los más enriquecidos son los de “pigmentación” (GO:0043473), “desarrollo de pigmento” (GO:0048066) y la “localización del pigmento granular” (GO:0051875). De acuerdo a estos resultados en los oocitos están activos genes involucrados en la producción, distribución y localización de pigmentos (Ashburner *et al*., 2000; Consortium *et al*., 2023).

Además, entre los términos enriquecidos relacionados con la pigmentación se identificaron únicamente 25 isoformas diferencialmente expresadas entre los oocitos claros y oscuros. Estas isoformas diferencialmente expresadas pertenecen a 11 genes identificados en los términos de “procesos de la biosíntesis de pigmento” (GO:0046148) y “procesos en el metabolismo del pigmento”(GO:0042440) (ANEXO 3).

Los datos de log2FC de las isoformas sobreexpresadas y subexpresadas relacionadas con el pigmento indican que existen diferencias extremadamente significativas en la expresión entre los dos tratamientos. Estos datos nos muestran que las isoformas sobreexpresadas se expresan al menos 5634 veces más en oocitos claros que en los oscuros, mientras que en las isoformas subexpresadas hay una expresión menor de al menos 5556 veces en comparación con los oocitos oscuros. Estos resultados nos sugieren que estas 25 isoformas pueden tener un rol en la determinación de la pigmentación melánica, por lo que se sugiere realizar estudios en estos genes en específicos para determinar su función en los oocitos claros y oscuros.

6.2.3 VÍAS METABÓLICAS

En la reconstrucción de los mapas de las rutas metabólicas se identificaron un total de 428 vías metabólicas en los transcritos de los oocitos.

Entre las vías metabólicas se identificó la ruta de la "maduración de oocitos mediada por la progesterona", que se encuentra completa y en la que se expresan 62 isoformas de los genes identificados en los transcritos (Figura 6). La presencia de esta ruta valida el origen de los transcritos a partir de los oocitos y nos ayuda a entender los procesos metabólicos que suceden durante la maduración de los oocitos de *Espadarana prosoblepon*.

Además, se identificó la vía metabólica de la "melanogénesis", que está parcialmente completa e involucra la expresión de 61 isoformas de los genes de oocitos. En esta ruta metabólica no está presente la hormona precursora alfa-melanocito estimulante (αMSH) y la proteína de señalización agouti (ASP) (Figura 7).

La hormona αMSH tiene su origen en la glándula pituitaria y actúa como un factor mitogénico en la producción de melanina (Flori *et al*., 2017). Su ausencia en la ruta metabólica de la melanogénesis nos sugiere que esta hormona pueda no estar involucrada de forma directa en la producción de melanina en los oocitos.

La proteína ASP cumple funciones regulatorias en la melanogénesis. ASP ha sido reportada como una hormona inhibidora de la producción de melanina (Furumura *et al*., 1996). Su ausencia en la ruta metabólica de la melanogénesis sugiere que esta proteína no tiene un papel regulatorio en los oocitos de *Espadarana prosoblepon.*

La ausencia de αMSH y ASP nos sugiere que los mecanismos de regulación de la pigmentación en oocitos pueden variar con respecto a los de *Xenopus* o también puede haber la posibilidad de que existan otras moléculas que cumplan la misma función que estas proteínas en estos oocitos.

7. CONCLUSIONES

El objetivo principal de este estudio es determinar los procesos moleculares involucrados en la pigmentación melánica de los oocitos y así tener más información sobre el rol del pigmento durante el desarrollo embrionario. Los grupos enriquecidos relacionados con el pigmento y la ruta metabólica de la melanogénesis nos indican que en los oocitos se están llevando a cabo procesos moleculares involucrados en la formación, transporte y establecimiento del pigmento. Además, se determinó que existe expresión génica diferencial entre los oocitos claros y oscuros, se identificaron 11 genes involucrados con la melanogénesis y el pigmento que se expresan diferencialmente entre ambos fenotipos.

De acuerdo a los resultados obtenidos en este estudio podemos concluir que en la pigmentación melánica normal y reducida existen controles y procesos moleculares distintos. A pesar de esto, todavía no podemos definir la función exacta del pigmento melánico en el desarrollo embrionario. Debido a esto consideramos que es necesario continuar con los estudios enfocándonos en los genes diferencialmente expresados entre ambos fenotipos y así poder determinar la función de la pigmentación melánica.

Espadarana prosoblepon ha sido propuesta como un organismo modelo para el estudio del pigmento melánico, por lo tanto, el ensamblaje de los transcritos provenientes de los oocitos contribuirá a posicionarla como una especie modelo emergente y aportará información útil para realizar comparaciones en estudios similares.

8. RECOMENDACIONES

Realizar experimentos de validación para confirmar la expresión diferencial de los genes identificados que se relacionan con la melanogénesis y el pigmento.

Realizar ensayos de re-mapeo para validar los resultados obtenidos en este estudio.

Realizar el análisis de expresión diferencial utilizando al menos dos muestras más de cada fenotipo de oocitos.

Realizar el análisis de expresión diferencial post fertilización.

Realizar estudios funcionales enfocándose en los genes diferencialmente expresados entre ambos fenotipos, para determinar la función de cada uno.

9. REFERENCIAS BIBLIOGRÁFICAS

Albagli, O., Dhordain, P., Deweindt, C., Lecocq, G., y Leprince, D. (1995). The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. *Cell Growth y Differentiation : the Molecular Biology Journal of the American Association for Cancer Research*, *6*(9), 1193-1198. https://europepmc.org/article/med/8519696

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., y Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature genetics*, *25*(1), 25. https://doi.org/10.1038/75556

Bairoch, A., y Apweiler, R. (1999). The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Research*, *27*(1), 49-54. https://doi.org/10.1093/nar/27.1.49

Bolger, A. M., Lohse, M., y Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Chen, S., Zhou, Y., Chen, Y., y Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884-i890. https://doi.org/10.1093/BIOINFORMATICS/BTY560

Consortium, T. G. O., Aleksander, S. A., Balhoff, J., Carbon, S., Cherry, J. M., Drabkin, H. J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N. L., Hill, D. P., Lee, R., Mi, H., Moxon, S., Mungall, C. J., Muruganugan, A., Mushayahama, T., Sternberg, P. W., Thomas, P. D., … Westerfield, M. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, *224*(1). https://doi.org/10.1093/GENETICS/IYAD031

Duellman, W. E., y Trueb, L. (1994). *Biology of Amphibians*. Baltimore and London: The Johns Hopkins University Press.

Elinson, R., y del Pino, E. M. (2012). Developmental diversity of amphibians. *Wiley Interdisciplinary Reviews: Developmental Biology*, *1*(3), 345-369. https://doi.org/10.1002/wdev.23

Ewing, B., Hillier, L. D., Wendl, M. C., y Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, *8*(3), 175-185. https://doi.org/10.1101/gr.8.3.175

Fedotova, A. A., Bonchuk, A. N., Mogila, V. A., y Georgiev, P. G. (2017). C2H2 zinc finger proteins: The largest but poorly explored family of higher eukaryotic transcription factors. *Acta Naturae*, *9*(2), 47. https://doi.org/10.32607/20758251-2017-9-2-47-58

Flori, E., Rosati, E., Cardinali, G., Kovacs, D., Bellei, B., Picardo, M., y Maresca, V. (2017). The α-melanocyte stimulating hormone/peroxisome proliferator activated receptor-γ pathway down-regulates proliferation in melanoma cell lines. *Journal of experimental y clinical cancer research : CR*, *36*(1). https://doi.org/10.1186/S13046-017-0611-4

Fu, L., Niu, B., Zhu, Z., Wu, S., y Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150. https://doi.org/10.1093/BIOINFORMATICS/BTS565

Furumura, M., Sakai, C., Abdel-Malek, Z., Barsh, G. S., y Hearing, V. J. (1996). The interaction of agouti signal protein and melanocyte stimulating hormone to regulate melanin formation in mammals. *Pigment Cell Research*, *9*(4), 191-203. https://doi.org/10.1111/J.1600-0749.1996.TB00109.X

Gall, J. G. (2014). Transcription in the *Xenopus* Oocyte Nucleus. *Xenopus Development*, *9781118492819*, 3-15. https://doi.org/10.1002/9781118492833.CH1

Gilbert, S. F., y Barresi, M. J. F. (2017). Developmental Biology, 11th Edition. En *American Journal of Medical Genetics Part A* (Vol. 173, Número 5). John Wiley y Sons, Ltd. https://doi.org/10.1002/ajmg.a.38166

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644-652. https://doi.org/10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature protocols*, *8*(8), 1494-1512. https://doi.org/10.1038/NPROT.2013.084

Isalan, M. (2013). Zinc Fingers. *Encyclopedia of Biological Chemistry: Second Edition*, 575-579. https://doi.org/10.1016/B978-0-12-378630-2.00027-X

Jain, B. P., y Pandey, S. (2018). WD40 Repeat Proteins: Signalling Scaffold with Diverse Functions. *The Protein Journal*, *37*(5), 391-406. https://doi.org/10.1007/S10930-018-9785-7

Kanehisa, M., y Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science*, *29*(1), 28-35. https://doi.org/10.1002/pro.3711

Kanehisa, M., Sato, Y., y Kawashima, M. (2022). KEGG mapping tools for uncovering hidden features in biological data. *Protein Science*, *31*(1), 47-53. https://doi.org/10.1002/pro.4172

Kobe, B., y Kemp, B. E. (2003). Principles of Kinase Regulation. *Handbook of Cell Signaling: Volume 1-3*, *1-3*, 539-542. https://doi.org/10.1016/B978-012124546-7/50450-2

Langmead, B., y Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357. https://doi.org/10.1038/NMETH.1923

Li, J., Mahajan, A., y Tsai, M. D. (2006). Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry*, *45*(51), 15168-15178. https://doi.org/10.1021/BI062188Q

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., y Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647-4654. https://doi.org/10.1093/molbev/msab199

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., y Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412-D419. https://doi.org/10.1093/NAR/GKAA913

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., y Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, *35*(Web Server), W182-W185. https://doi.org/10.1093/nar/gkm321

Nalefski, E. A., y Falke, J. J. (1996). The C2 domain calcium-binding motif: structural and functional diversity. *Protein Science : A Publication of the Protein Society*, *5*(12), 2375. https://doi.org/10.1002/PRO.5560051201

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., y Kingsford, C. (2017). Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nature methods*, *14*(4), 417. https://doi.org/10.1038/NMETH.4197

Piprek, R. P., Damulewicz, M., Kloc, M., y Kubiak, J. Z. (2018). Transcriptome analysis identifies genes involved in sex determination and development of *Xenopus laevis* gonads. *Differentiation*, *100*, 46-56. https://doi.org/10.1016/J.DIFF.2018.02.004

Raghavan, V., Kraft, L., Mesny, F., y Rigerte, L. (2022). A simple guide to *de novo* transcriptome assembly and annotation. *Briefings in Bioinformatics*, *23*(2), 1-30. https://doi.org/10.1093/bib/bbab563

R Core Team. (2022). *R:* A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Robinson, M. D., McCarthy, D. J., y Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139. https://doi.org/10.1093/BIOINFORMATICS/BTP616

Rudh, A., y Qvarnström, A. (2013). Adaptive colouration in amphibians. *Seminars in Cell y Developmental Biology*, *24*(6-7), 553-561. https://doi.org/10.1016/J.SEMCDB.2013.05.004

Salazar-Nicholls, M. J., Hervas, F., Muñoz-Tobar, S. I., Carrillo, A.-B., Ricaurte, H., Ron, S. R., y Romero-Carvajal, A. (2021). A polymorphism in oocyte pigmentation in natural populations of the glass frog *Espadarana prosoblepon* (Centrolenidae). *The International Journal of Developmental Biology*, *65*(4-5-6), 333-344. https://doi.org/10.1387/ijdb.200074ar

Savage, A. E., Kiemnec-Tyburczy, K. M., Ellison, A. R., Fleischer, R. C., y Zamudio, K. R. (2014). Conservation and divergence in the frog immunome: pyrosequencing and de novo assembly of immune tissue transcriptomes. *Gene*, *542*(2), 98-108. https://doi.org/10.1016/j.gene.2014.03.051

SenGupta, D. (2013). RNA-Binding Domains in Proteins. *Brenner's Encyclopedia of Genetics: Second Edition*, 274-276. https://doi.org/10.1016/B978-0-12-374984-0.01356-5

Simeoni, I., Gilchrist, M. J., Garrett, N., Armisen, J., y Gurdon, J. B. (2012). Widespread transcription in an amphibian oocyte relates to its reprogramming activity on transplanted somatic nuclei. *Stem Cells and Development*, *21*(2), 181. https://doi.org/10.1089/SCD.2011.0162

Slowikowski, K. (2023). *ggrepel: Repulsive Text and Label Geoms for «ggplot2»*. *R package version 0.9.3*. https://cran.r-project.org/package=ggrepel

Smirnov, D. N., Shekhovtsov, S. V., Shipova, A. A., Gazizova, G. R., Shagimardanova, E. I., Bulakhova, N. A., Meshcheryakova, E. N., Poluboyarova, T. V., Khrameeva, E. E., Peltek, S. E., y Berman, D. I. (2022). De novo assembly and analysis of the transcriptome of the Siberian wood frog *Rana amurensis*. *Vavilov Journal of Genetics and Breeding*, *26*(1), 109. https://doi.org/10.18699/VJGB-22-07

UniProt Consortium. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480-D489. https://doi.org/10.1093/nar/gkaa1100

Urrutia, R. (2003). KRAB-containing zinc-finger repressor proteins. *Genome Biology*, *4*(10), 1-8. https://doi.org/10.1186/GB-2003-4-10-231/FIGURES/3

Walter, W., Sánchez-Cabo, F., y Ricote, M. (2015). GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics*, *31*(17), 2912-2914. https://doi.org/10.1093/BIOINFORMATICS/BTV300

Wickham, H. (2009). ggplot2. En *ggplot2*. Springer New York. https://doi.org/10.1007/978-0-387-98141-3

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., y Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

10. FIGURAS

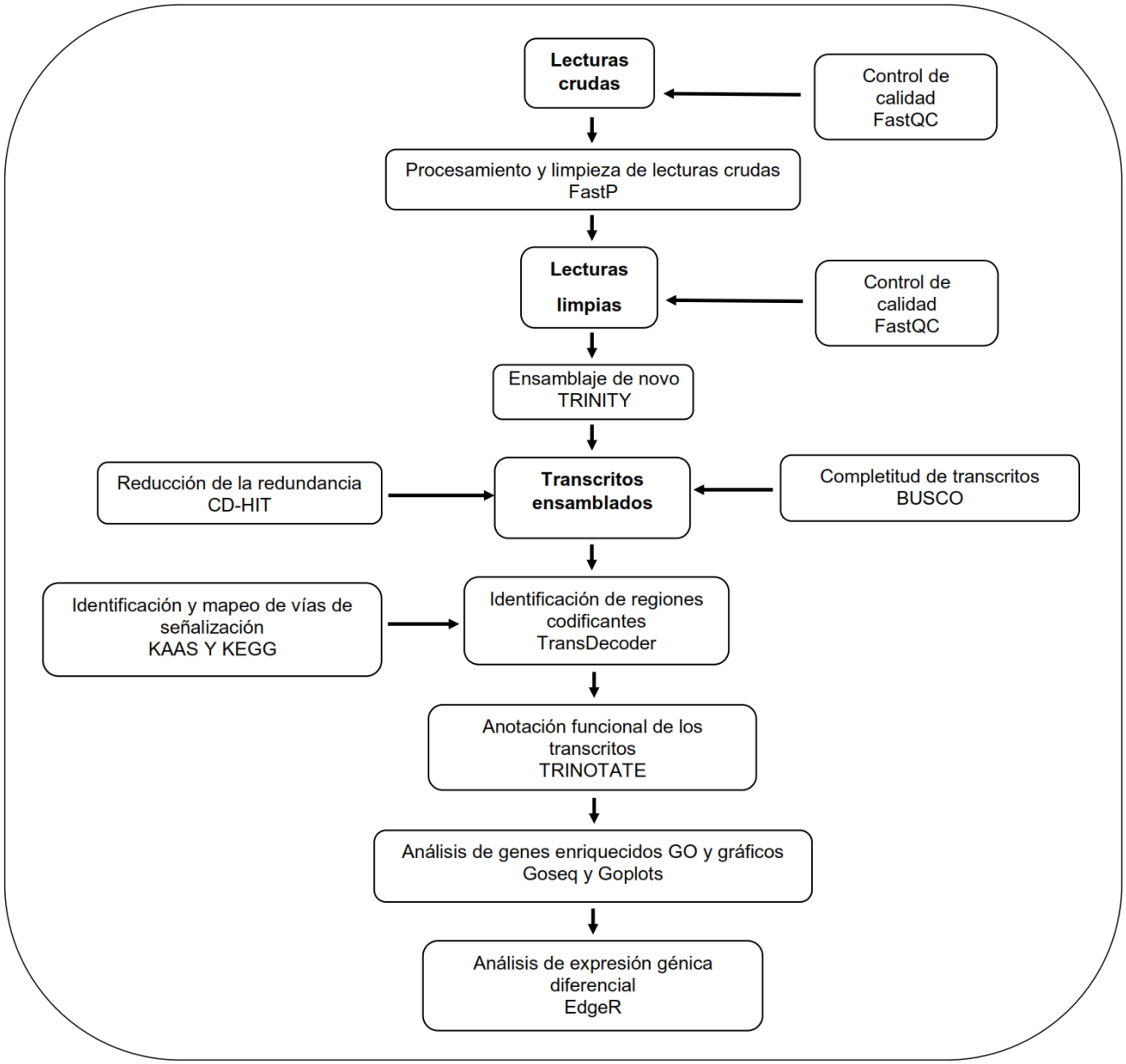


Figura 1. Flujo de trabajo del ensamblaje, anotación funcional y análisis de expresión génica diferencial.

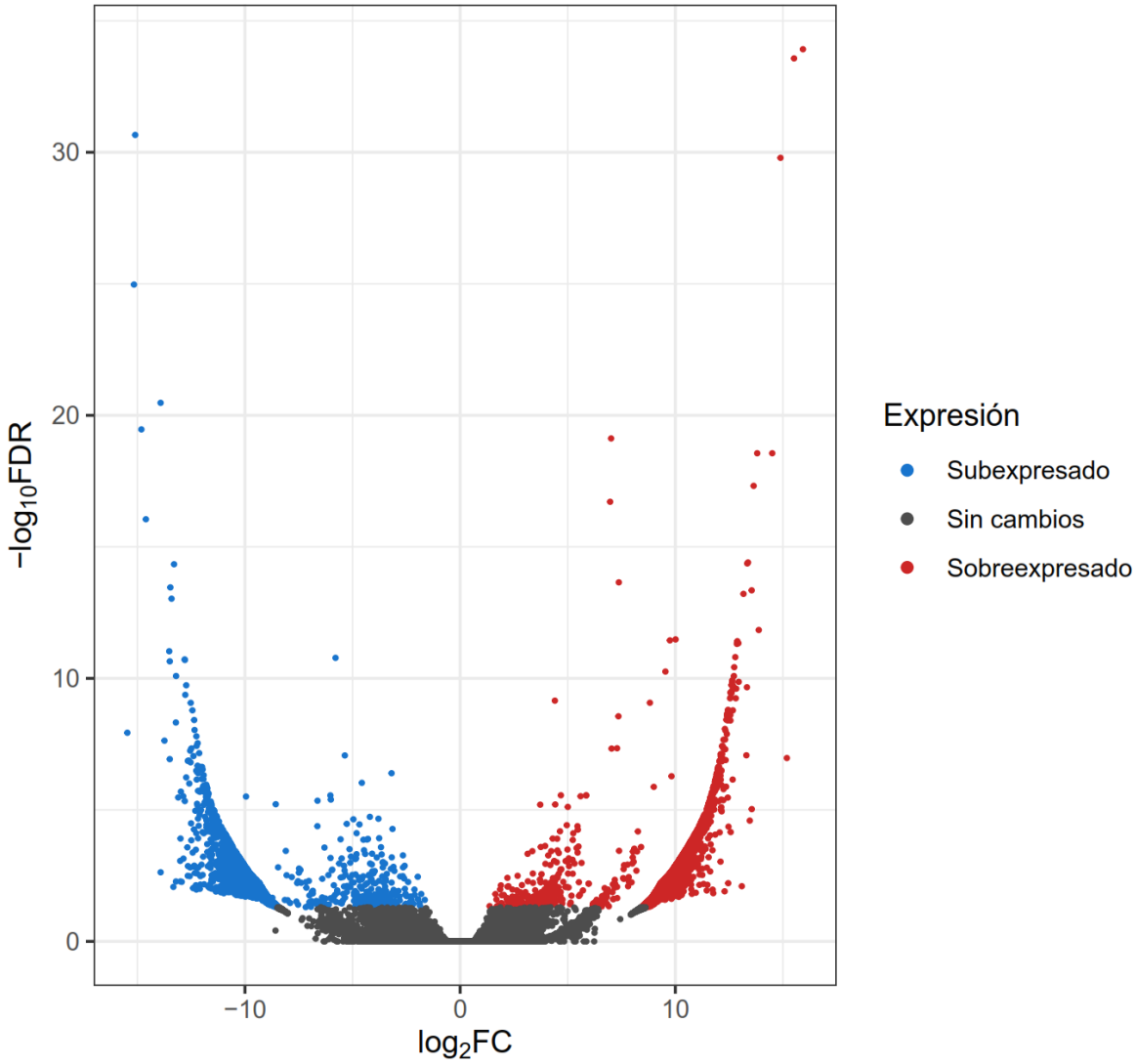


Figura 2. Gráfico de volcán de la expresión génica diferencial con genes.

El eje x representa el logaritmo en base 2 del cambio relativo o fold change (log2FC), que se refiere a la magnitud del cambio en la expresión génica entre oocitos claros y oscuros. El eje Y representa el logaritmo negativo en base 10 del valor de p ajustado (log10FDR). Los puntos de color gris representan los genes que no tienen un cambio de expresión significativo entre las dos muestras. Los puntos de color azul representan los genes sub regulados al comparar las dos condiciones. Los puntos rojos representan los genes sobreexpresados de una condición a otra.

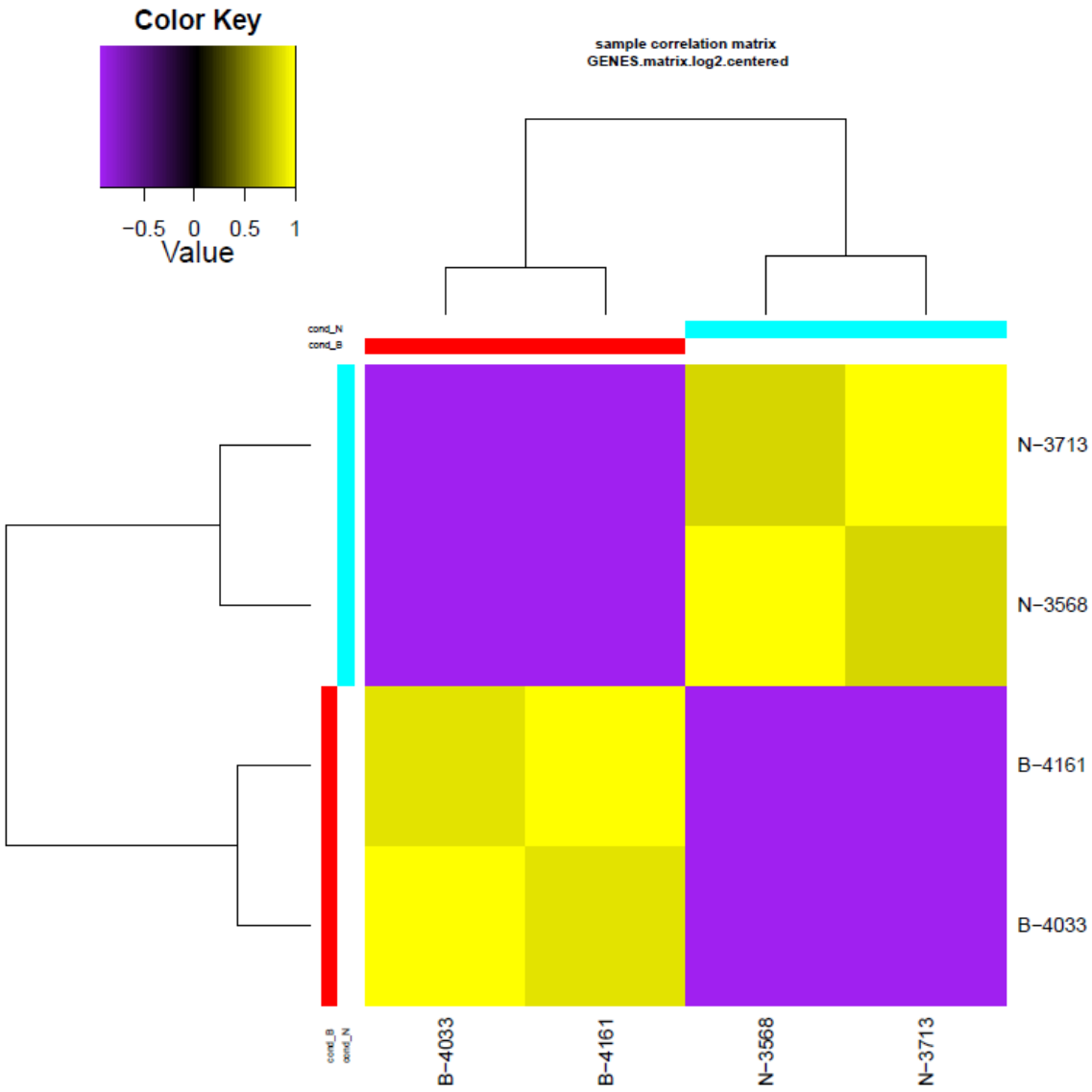


Figura 3. Mapa de calor del agrupamiento jerárquico de las muestras de oocitos de *Espadarana prosoblepon* de acuerdo a sus niveles de expresión.

En el eje x y eje y se representan todas las muestras utilizadas en el análisis de expresión génica. La clave de color indica los niveles de similitud de la expresión génica entre las muestras.

B-4033, oocitos claros; B-4161, oocitos claros; N-3568, oocitos oscuros; N-3713, oocitos oscuros.

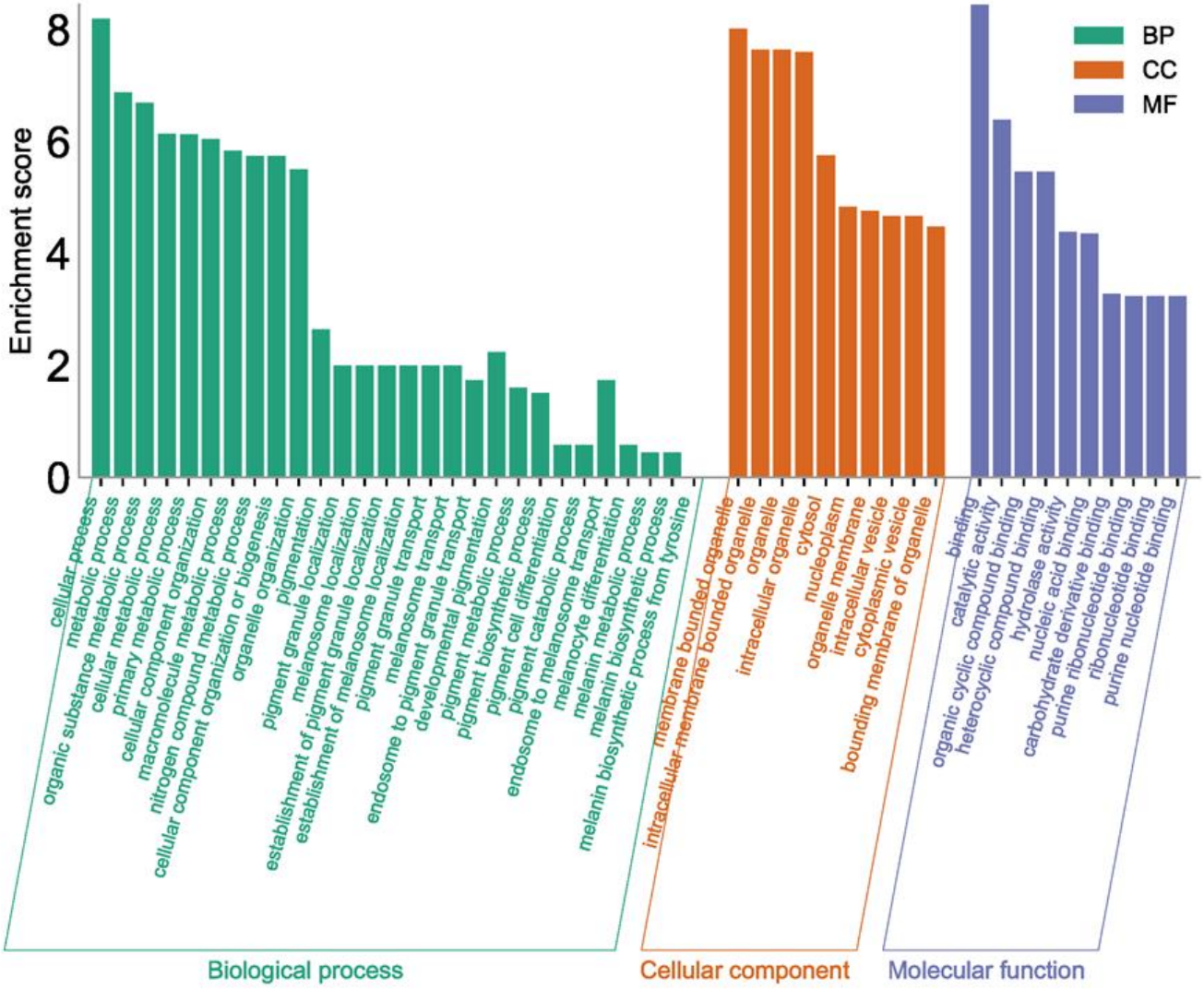
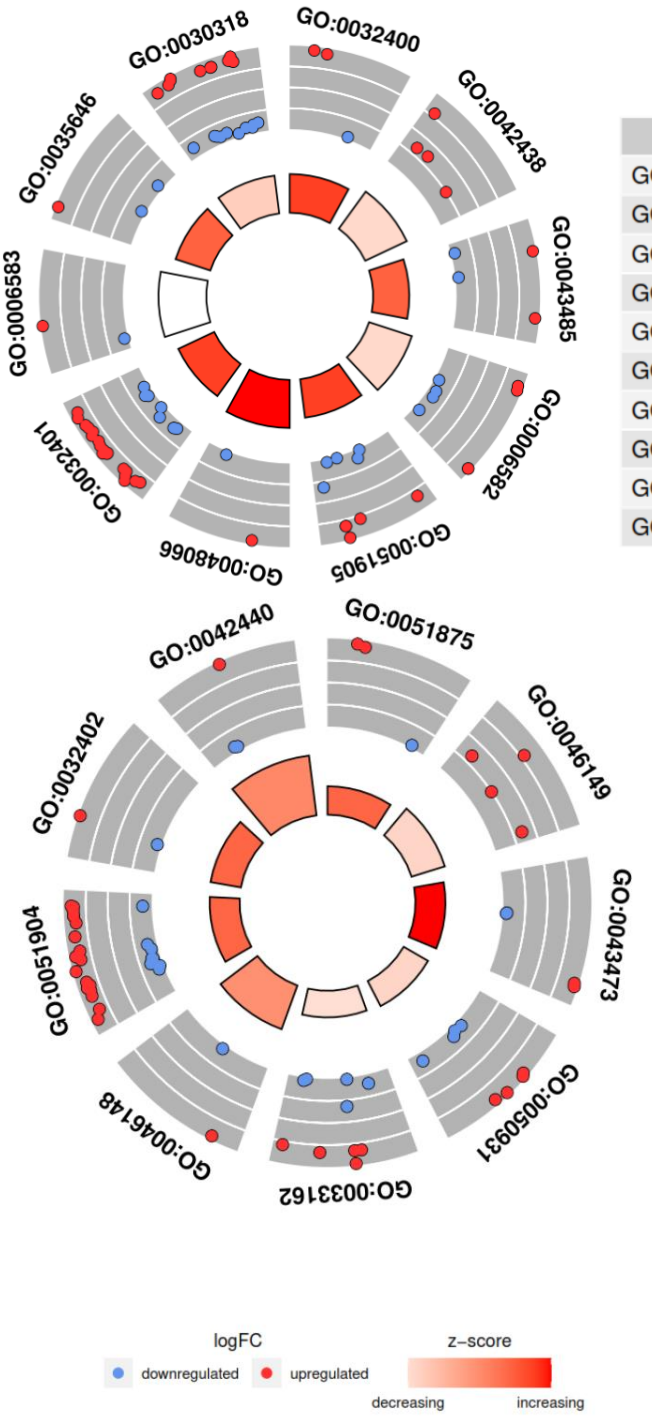


Figura 4. Clasificación de Ontología Génica (GO).

Gráfico de barras con los diez términos de GO más enriquecidos para las categorías de procesos biológicos (BP), componentes celulares (CC) y función molecular (MF). Además, se incluyeron los términos más enriquecidos relacionados a la pigmentación.



ID	Description
GO:0032400	melanosome localization
GO:0042438	melanin biosynthetic process
GO:0043485	endosome to pigment granule transport
GO:0006582	melanin metabolic process
GO:0051905	establishment of pigment granule localization
GO:0048066	developmental pigmentation
GO:0032401	establishment of melanosome localization
GO:0006583	melanin biosynthetic process from tyrosine
GO:0035646	endosome to melanosome transport
GO:0030318	melanocyte differentiation

ID	Description
GO:0051875	pigment granule localization
GO:0046149	pigment catabolic process
GO:0043473	pigmentation
GO:0050931	pigment cell differentiation
GO:0033162	melanosome membrane
GO:0046148	pigment biosynthetic process
GO:0051904	pigment granule transport
GO:0032402	melanosome transport
GO:0042440	pigment metabolic process

Figura 5. Gráfico circular de Ontología Génica (GO) con los términos relacionados a la melanogénesis.

El círculo exterior es un diagrama de dispersión que muestra los niveles de expresión de los genes relacionados con la melanogénesis y el pigmento. Los puntos rojos representan la sobreexpresión y los azules la subexpresión. El círculo central es un gráfico de barras que representa el valor de z (z-score). En el cuadro de la derecha está la lista de los términos de GO y su descripción.

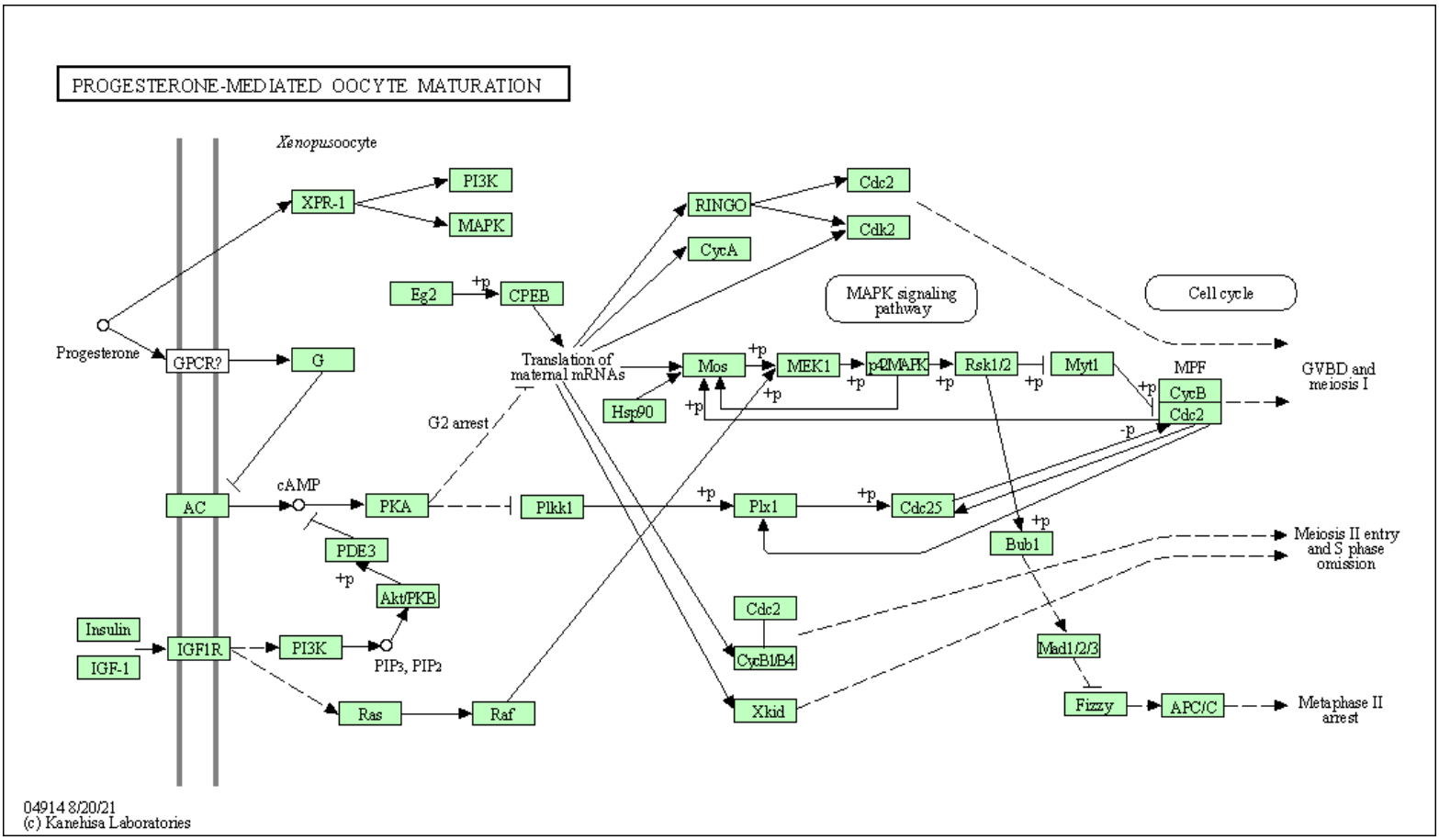


Figura 6. Mapa de la vía metabólica de la maduración de los oocitos mediada por la progesterona.

Los recuadros resaltados representan a genes que han sido identificados en los transcritos de los oocitos. Los cuadros en blanco representan a las proteínas ausentes en los transcritos.

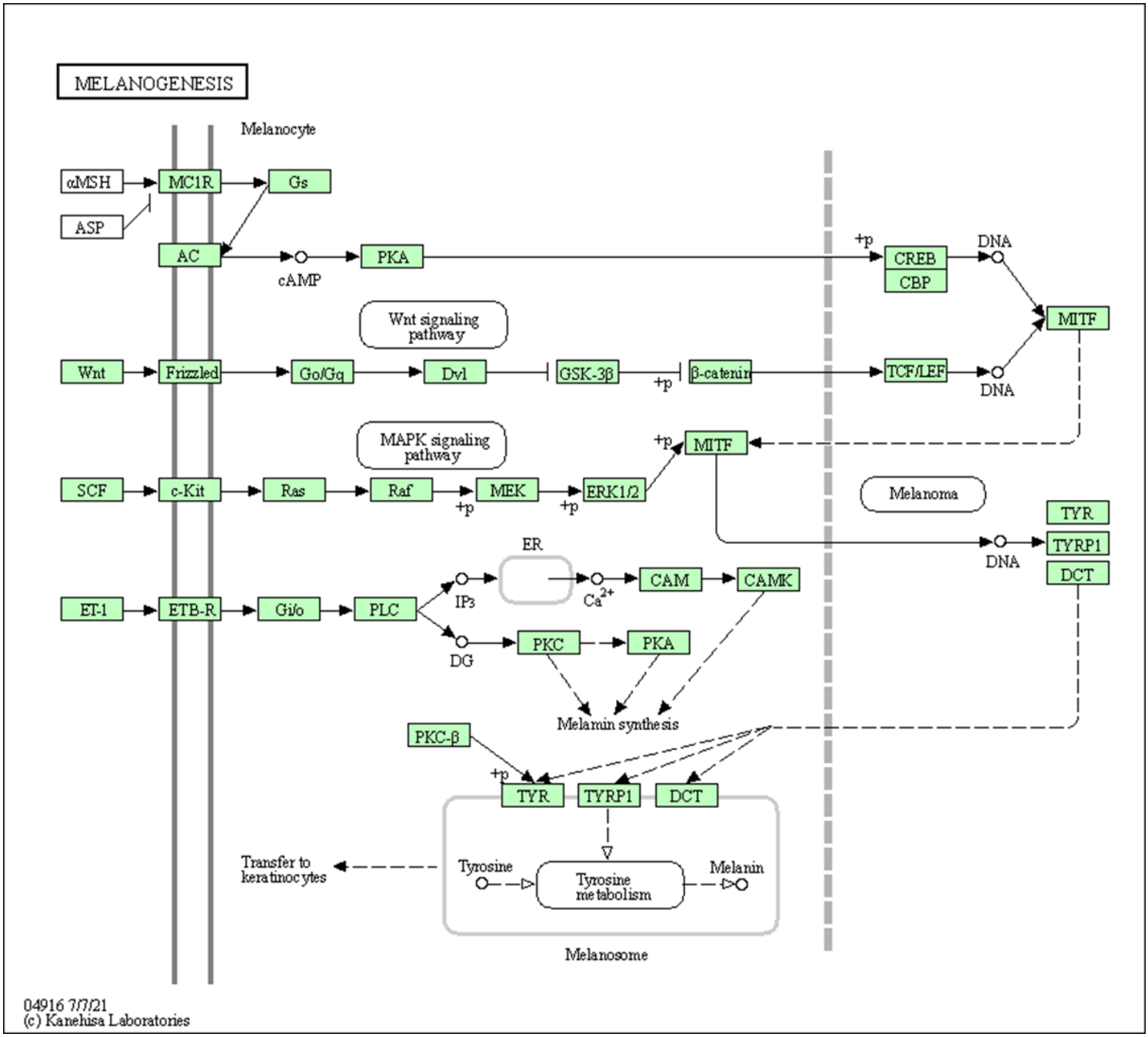


Figura 7. Mapa de la vía metabólica de la melanogénesis en oocitos.

Los recuadros resaltados representan a genes que han sido identificados en los transcritos de los oocitos. Los cuadros en blanco representan a las proteínas ausentes en los transcritos.

11. TABLAS

Tabla 1. Estadísticas del ensamblaje global de dos muestras de huevos claros y dos de huevos oscuros de *Espadarana prosoblepon.*

Estadísticas del ensamblaje	
Número de transcritos ensamblados	479370
Número de genes	239967
Longitud media de contigs	355 - 309*
Promedio de contigs	653.91 – 532.85*
Contig N50	1003- 655*
Total de bases ensambladas	127866396
Porcentaje GC	46.10

* Estadísticas basadas solo en las isoformas más largas de cada gen

Tabla 2. Datos de completitud del ensamblaje de los transcritos.

RESULTADOS BUSCO				
Estadísticas BUSCO	**Tetrápodos**	**Vertebrados**	**Metazoos**	**Eucariotas**
Total de grupos BUSCO	5310	3354	954	255
Completo	4453	3102	948	254
Único	3608	2475	742	213
Duplicado	845	627	206	41
Fragmentado	179	77	1	0
Faltante	678	175	5	1
% de completitud	83.8	92.5	99.4	99.6

Tabla 3. Datos del porcentaje de secuencias alineadas por cada muestra.

Muestras	% de lecturas alineadas
B-4033	96.72
B-4161	96.57
N-3568	96.64
N-3713	96.07
Promedio	**96.5**

Tabla 4. Datos de expresión diferencial de la comparación entre los transcritos de oocitos claros y oscuros $p < 0.05$

Expresión	n
Subexpresados	1.773
Sin cambio	61.375
Sobreexpresados	2.899
Total	**66.047**

Tabla 5. Lista de los diez dominios proteicos más comunes

Descripción del dominio proteico	Número de secuencias anotadas	Término Pfam
Zinc finger, C2H2 type	788	PF00096
KRAB box	251	PF01352
WD domain, G-beta repeat	236	PF00400
Protein kinase domain	230	PF00069
RNA recognition motif.	207	PF00076
ADP-ribosylation factor family	182	PF00025
Ankyrin repeats	122	PF12796
Zinc finger, C3HC4 type (RING finger)	116	PF13920
BTB/POZ domain	99	PF00651
C2 domain	84	PF00168

12. ANEXOS

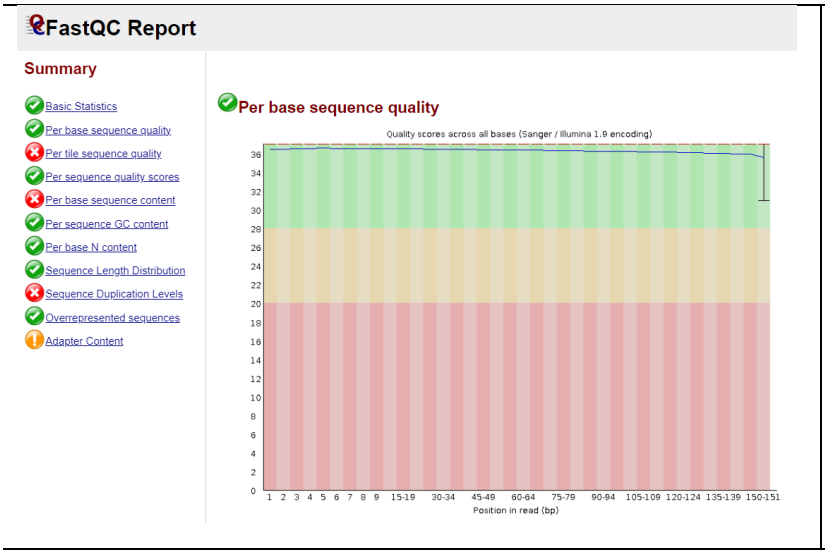
Anexo 1. Individuos de *Espadarana prosoblepon* utilizados para la extracción de oocitos.

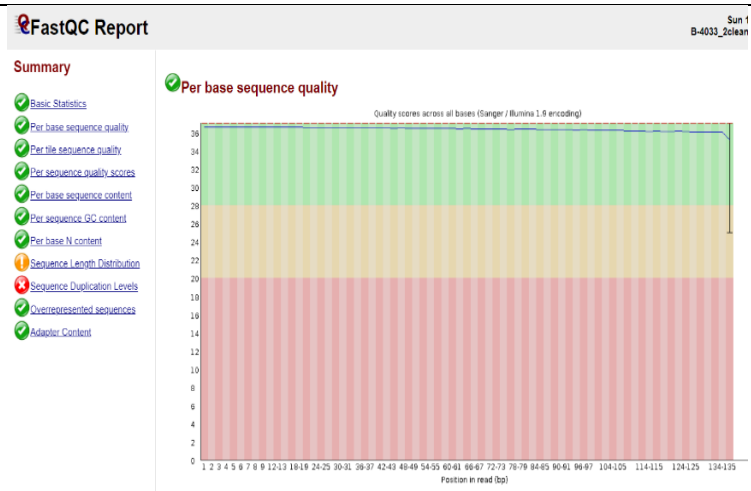
Individuo	Código de muestra	#Terrario	Fenotipo de oocitos
1	B-4033	4033	Pigmentación reducida
2	B-4161	4161	Pigmentación reducida
3	N-3568	3568	Pigmentación normal
4	N-3713	3713	Pigmentación normal

Anexo 2. Resumen de las estadísticas de calidad de las secuencias crudas y limpias.

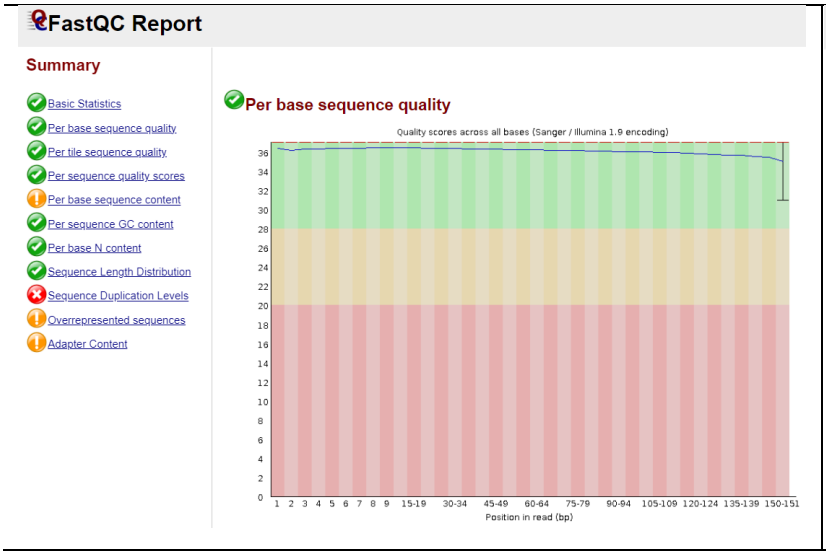
MUESTRA B-4033

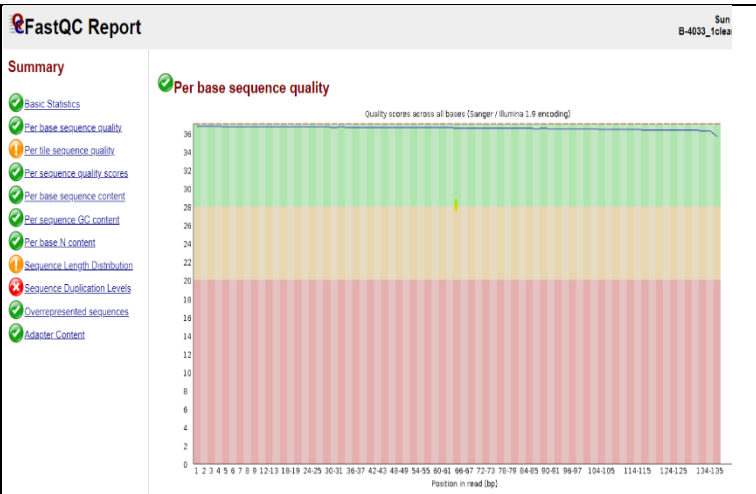
FORWARD





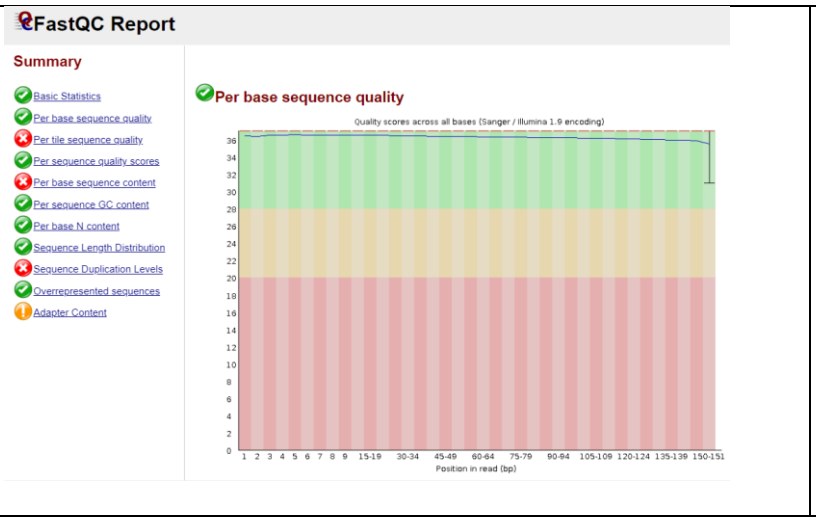
REVERSE





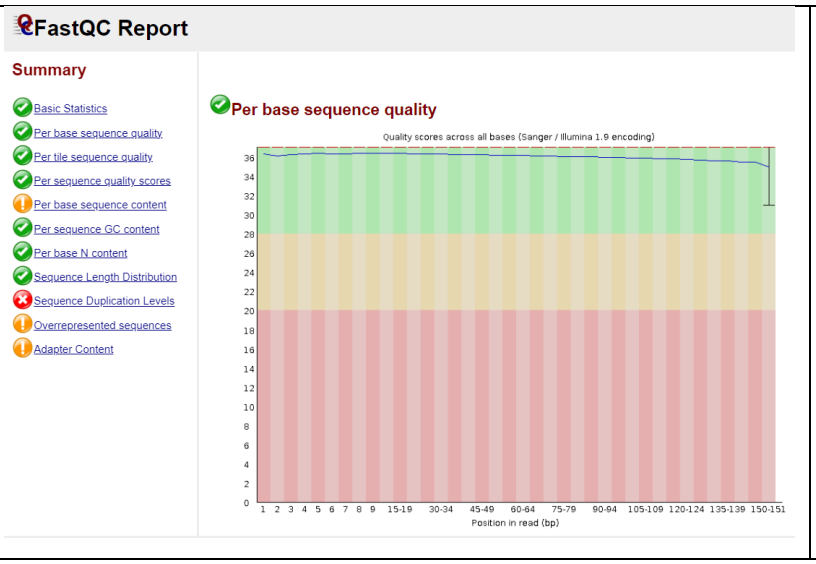
MUESTRA B-4161

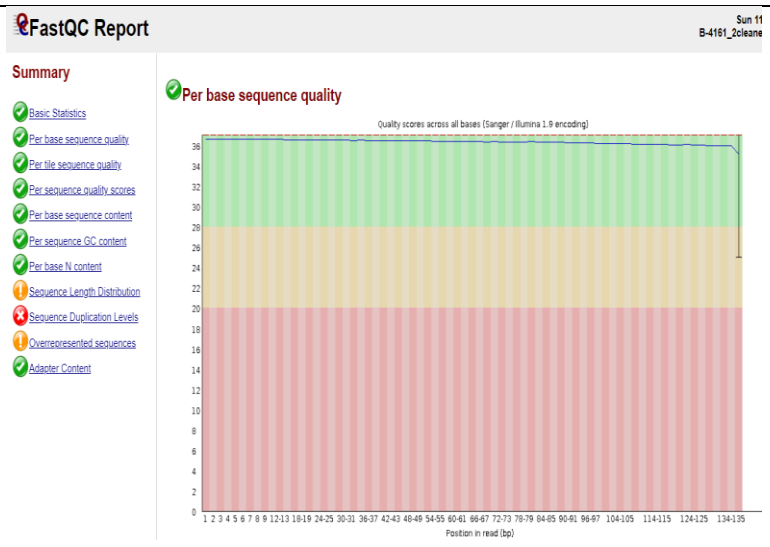
FORWARD





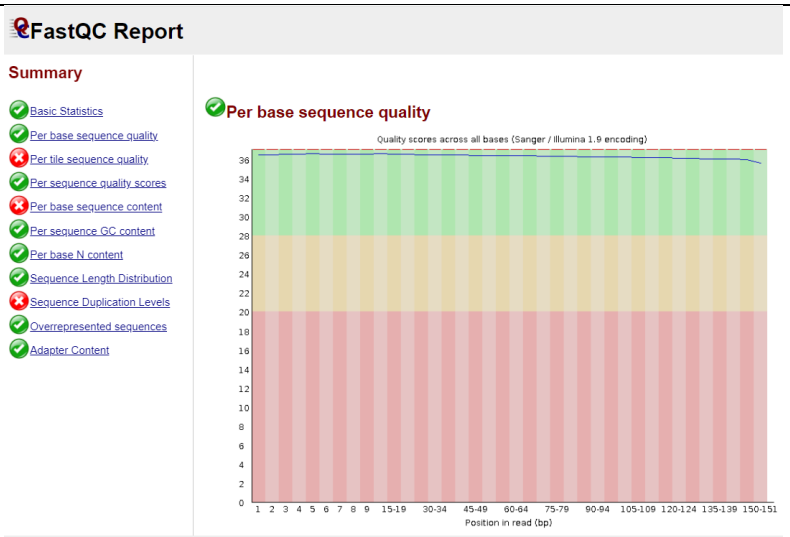
REVERSE

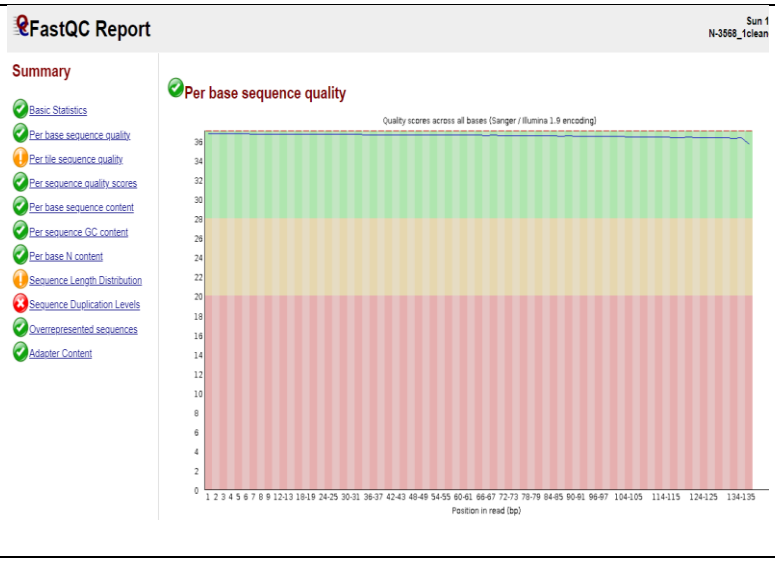




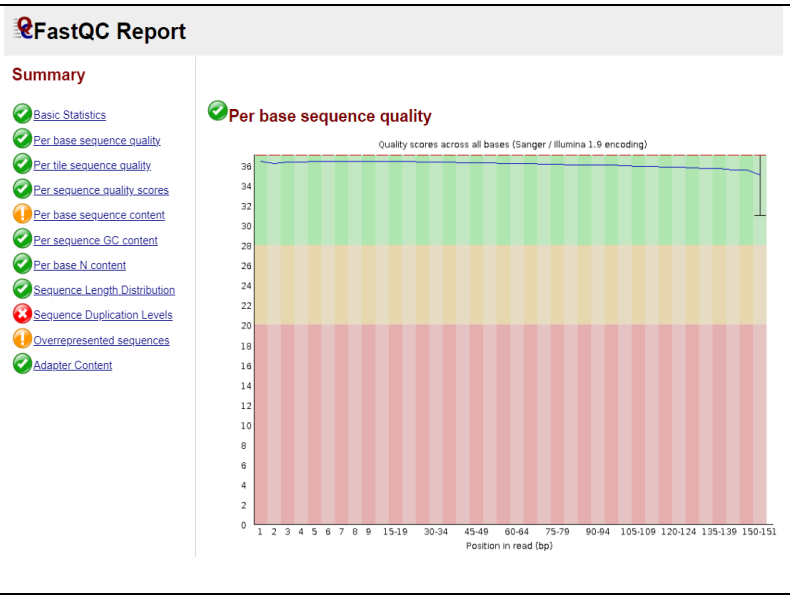
MUESTRA N-3568

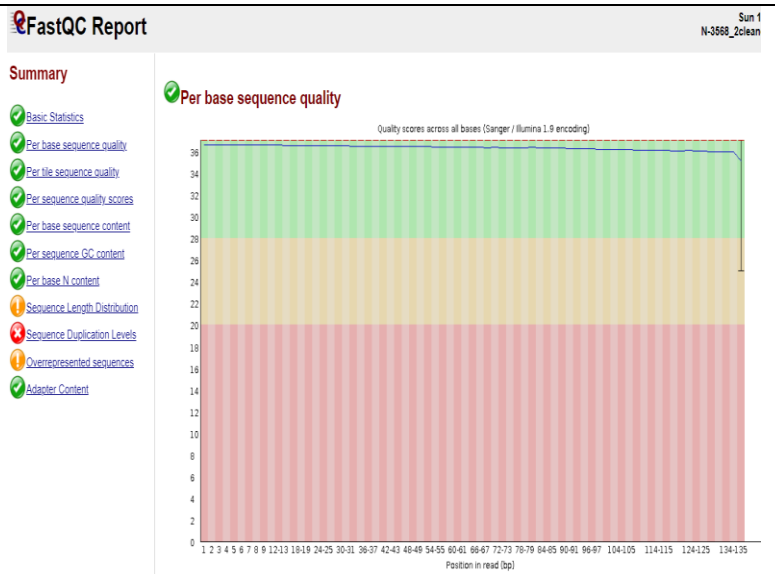
FORWARD





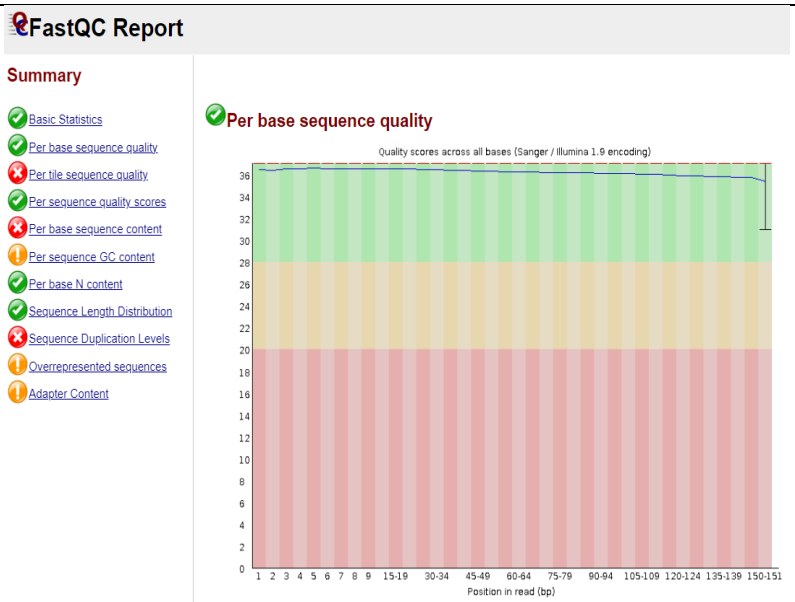
REVERSE

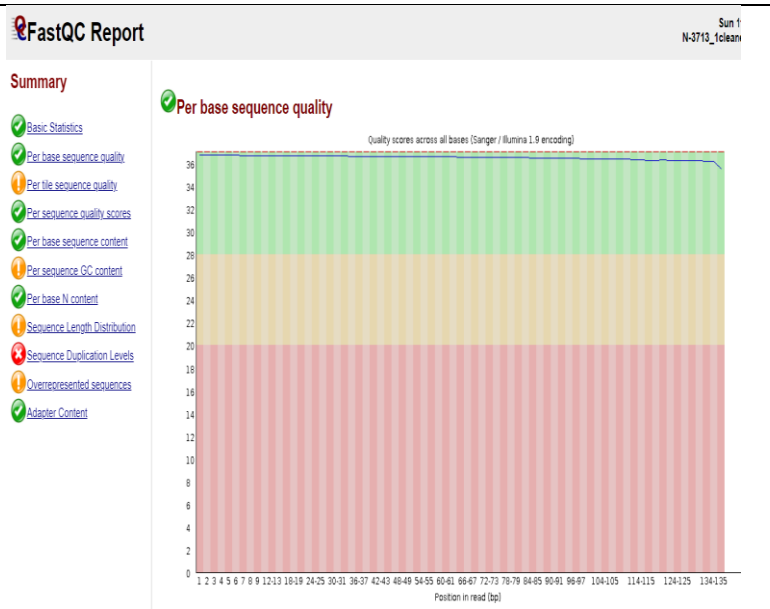




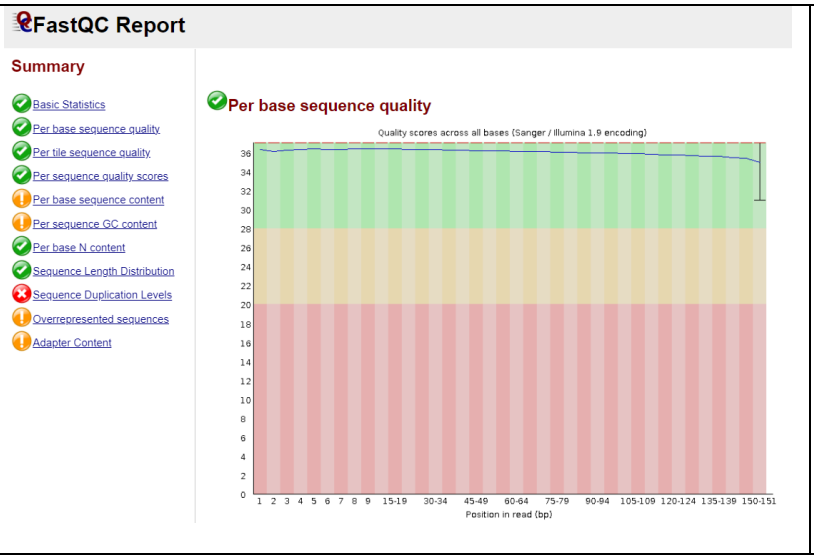
MUESTRA N-3713

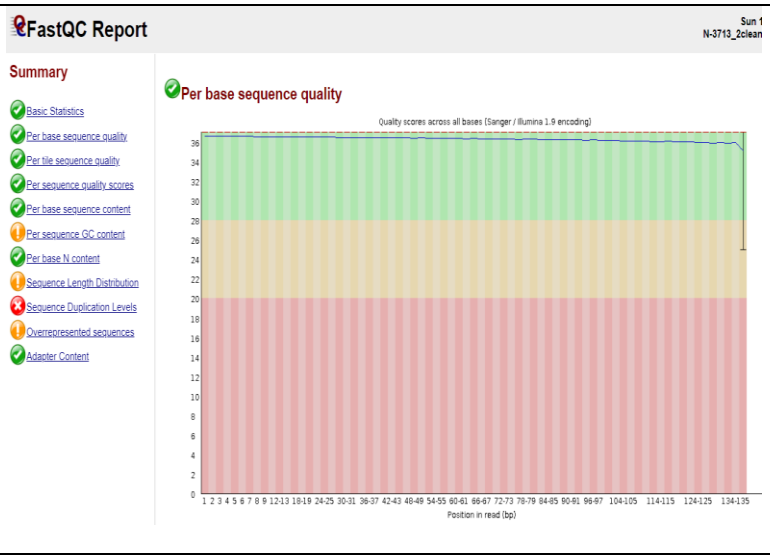
FORWARD





REVERSE





Anexo 3. Isoformas diferencialmente expresadas en los términos GO relacionados a la melanogénesis y pigmento.

Código GO	Término GO	Isoforma	log2FC	FDR	Expresión
GO:0046148	pigment biosynthetic process	TRINITY_DN587_C0_G1_I8	15.48	0.020	Sobreexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN2130_C0_G1_I1	15.14	0.020	Sobreexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN6982_C0_G1_I5	14.59	0.020	Sobreexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN1682_C0_G1_I19	13.88	0.020	Sobreexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN2650_C0_G1_I2	13.04	0.020	Sobreexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN7097_C0_G1_I18	12.96	0.020	Sobreexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN1367_C4_G1_I3	12.75	0.020	Sobreexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN1779_C0_G1_I11	12.68	0.020	Sobreexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN6240_C0_G1_I5	-12.46	0.020	Subexpresado
GO:0046148	pigment biosynthetic process	TRINITY_DN2381_C0_G1_I10	-12.50	0.020	Subexpresado

Código GO	Término GO	Isoforma	log2FC	FDR	Expresión
GO:0046148	pigment biosynthetic process	TRINITY_DN1682_C0_G1_I9	-13.63	0.020	Subexpresado
GO:0042440	pigment metabolic process	TRINITY_DN587_C0_G1_I8	15.48	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN2130_C0_G1_I1	15.14	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN6982_C0_G1_I5	14.59	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN1682_C0_G1_I19	13.88	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN2650_C0_G1_I2	13.04	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN7097_C0_G1_I18	12.96	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN3042_C0_G1_I8	12.85	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN1367_C4_G1_I3	12.75	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN1779_C0_G1_I11	12.68	0.002	Sobreexpresado
GO:0042440	pigment metabolic process	TRINITY_DN3042_C0_G1_I6	12.46	0.002	Sobreexpresado

Código GO	Término GO	Isoforma	log2FC	FDR	Expresión
GO:0042440	pigment metabolic process	TRINITY_DN3042_C0_G1_I3	-12.44	0.002	Subexpresado
GO:0042440	pigment metabolic process	TRINITY_DN6240_C0_G1_I5	-12.46	0.002	Subexpresado
GO:0042440	pigment metabolic process	TRINITY_DN2381_C0_G1_I10	-12.50	0.002	Subexpresado
GO:0042440	pigment metabolic process	TRINITY_DN1682_C0_G1_I9	-13.63	0.002	Subexpresado

DECLARACIÓN Y AUTORIZACIÓN

Yo, Henry Fabián Macías Bazante**,** con C.I. 1722148952, autor del trabajo de graduación titulado: "Expresión génica diferencial de oocitos claros y oscuros de *Espadarana prosoblepon*." previa la obtención del grado académico de MAGISTER EN BIOLOGÍA COMPUTACIONAL, en la Facultad de Ingeniería.

1.- Declaro tener pleno conocimiento de la obligación que tiene la Pontificia Universidad Católica del Ecuador, de conformidad con el artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de graduación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la Pontificia Universidad Católica del Ecuador a difundir a través del sitio web de la Biblioteca de la PUCE el referido trabajo de graduación, respetando las políticas de propiedad intelectual de la Universidad.



Quito, 05 de septiembre del 2023

Henry Fabián Macías Bazante

1722148952