



**Caracterización del transcriptoma de la araña polimórfica de color *Gasteracantha cacriformis*,
con enfoque a los genes de pigmento**

Paula Alexandra Torres Quintero

Dirección

Geimy Carolina Pardo Díaz Ph.D.

Trabajo presentado como requisito para optar por el título de Bióloga

Facultad de Ciencias Naturales, Pregrado en Biología. Universidad del Rosario

Bogotá, 2020.

RESUMEN

El polimorfismo de coloración es un sistema de gran utilidad para estudiar procesos evolutivos y de desarrollo. No obstante, el estudio de este rasgo se ha sesgado a organismos modelo y la coloración en arácnidos no ha sido bien estudiada. La araña *Gasteracantha cancriformis* es una especie neotropical que presenta polimorfismo de coloración abdominal con al menos 16 morfos conocidos distribuidos desde el sur de Estados Unidos hasta el norte de Argentina. A la fecha, las bases genéticas que controlan dicho polimorfismo son desconocidas. En este estudio implementé RNAseq para ensamblar el transcriptoma de *Gasteracantha*, y evaluar expresión génica diferencial asociada a la coloración abdominal en hembras de color blanco, amarillo y naranja. A partir de las secuencias de los transcritos hice análisis nucleotídicos y estadísticos para establecer si existe una relación entre coloración y polimorfismos de ADN (SNPs) en el transcriptoma. En general, el perfil de expresión génica fue muy similar entre los fenotipos amarillo y naranja, quienes difirieron del fenotipo blanco. En los morfos amarillo y naranja detecté mayor expresión de genes relacionados a astacina y vitelogenina, lo cual sugiere que estas coloraciones en *G. cancriformis* son producto de carotenoides. A nivel nucleotídico encontré que los transcritos diferencialmente expresados muestran señal asociada a fenotipo de color, pero esto no sucede a nivel de transcriptoma completo. Los datos obtenidos en este estudio proporcionan un recurso genético valioso para futuras investigaciones en arañas neotropicales y constituyen un paso fundamental hacia la identificación de las bases genéticas del polimorfismo de color en estos animales.

Palabras clave: *Gasteracantha cancriformis*, transcriptoma, expresión diferencial de genes, polimorfismo de color.

ABSTRACT

Color polymorphism in animals is a useful system to study evolutionary and developmental processes. However, the study of this trait is biased towards model organisms, while coloration in arachnids is less explored. *Gasteracantha cancriformis* is a neotropical spider that displays color polymorphism with at least 16 known morphs distributed from the southern United States to northern Argentina but, to date, the genetic basis underlying such polymorphism is unknown. In this study, I used RNAseq to generate a transcriptome assembly for *Gasteracantha* and assess differential gene expression between three different color morphs: white, yellow and orange. I also explored whether single nucleotide polymorphism (SNPs) across the transcriptome display any association with coloration. Overall, the gene expression profile of the yellow and orange morphs was similar, and they both differed from the white morph. Consistently, at nucleotide level I found that SNPs in the

differential expressed transcripts cluster orange and yellow morphs together while the white morph appears more differentiated. Also, I detected higher expression levels of astacin and vitellogenin genes in the yellow and orange morphs, suggesting that these colorations in *G. cancriformis* are due to carotenoids. The data obtained in this study provide a valuable genetic resource for future studies on neotropical spiders and constitute a step towards the identification of the genetic basis of color polymorphism in these animals.

Key words: *Gasteracantha cancriformis*, transcriptome, differential gene expression, color polymorphism.

INTRODUCCIÓN

El entendimiento de varios procesos evolutivos ha sido en gran parte gracias a los estudios de coloración en múltiples organismos (White & Kemp, 2015). Se sabe que la coloración cumple un papel fundamental en procesos como selección sexual, camuflaje, fotoprotección, resistencia microbiana y termorregulación (Kronforst et al., 2012; Orteu & Jiggins, 2020). En particular, los polimorfismos de coloración adaptativos han sido de gran utilidad para investigar las bases moleculares de la adaptación, ya que sus patrones de herencia suelen ser simples y sus frecuencias son fácilmente observables (Gray & McKinnon, 2007; Kronforst & Papa, 2015; Nachman et al., 2003). A pesar de esto, el estudio de los polimorfismos de color se ha sesgado a organismos modelo en evolución y desarrollo (evo-devo), especialmente insectos, ratones, y peces (Gross & Wilkens, 2013; Hoekstra et al., 2006; Linnen et al., 2013; Reed et al., 2011; Streelman et al., 2007; P. J. Wittkopp et al., 2011). Aún así, muchas especies de arañas presentan variantes polimórficas en el color, posiblemente por su actividad depredadora (Oxford & Gillespie, 1998). Sin embargo, los patrones de coloración en arácnidos no han sido bien estudiados, y la información sobre la base genética de la coloración en este grupo es muy limitada.

A la fecha se han identificado cinco clases de pigmentos en arañas: ommocromos, bilinas, guanina, carotenoides y pteridinas (Croucher et al., 2013; Oxford & Gillespie, 1998). Los ommocromos son los pigmentos más comunes y son responsables de una amplia gama de colores, desde amarillo a marrón (Croucher et al., 2013). Las bilinas son pigmentos azules o verdes que se han encontrado en la hemolinfa, los tejidos intersticiales y el vitelo de los ovocitos de algunas arañas (Hsiung et al., 2017; Oxford & Gillespie, 1998). La guanina es un producto excretor importante en todos los arácnidos pero se ha demostrado su uso como pigmento en varias familias, lo que sugiere que la capacidad de acumularlo se ha ganado (o perdido) varias veces (Oxford & Gillespie, 1998). Los carotenoides parecen estar asociados con la pigmentación de tonos verdes y rojos (Croucher et al.,

2013; Hsiung et al., 2017), pero la información al respecto aún es muy escasa. Finalmente, las pteredinas que parecen contribuir a la coloración de morfos de tonos marrón, naranja y rojo (Croucher et al., 2013). Interesantemente, a pesar de que las melaninas son pigmentos comunes en la mayoría de animales, no existen reportes de la presencia de estos pigmentos en arañas (Patricia J. Wittkopp & Beldade, 2009).

La araña *Gasteracantha cancriformis* es una especie neotropical de tela orbicular que se distribuye desde el sur de los Estados Unidos hasta el norte de Argentina abarcando así la región neotropical y subtropical del continente americano (Levi, 1978). Esta araña exhibe dimorfismo sexual (Muma, 1971), donde los machos son pequeños (0.2-0.3 cm de longitud) y poco conspicuos, mientras que las hembras son más grandes (0.5-0.7 cm de longitud), tienen un abdomen de respaldo espinoso con espinas duras y esclerotizadas y presentan polimorfismo de color con al menos 16 morfos conocidos (Figura 1; (Levi, 1978; Salgado, 2019). Sin embargo, las causas que mantienen esta variación de color en la naturaleza aún son desconocidas, aunque se hipotetiza que pueden ser coloraciones aposemáticas (i.e. coloraciones de advertencia para depredadores; (F. M. Gawryszewski & Motta, 2012). A pesar de que la coloración de *G. cancriformis* constituye un sistema interesante para estudios en evolución, biología del desarrollo y ecología, la mayoría de estudios sobre esta especie se han enfocado en su biología básica, en particular en su comportamiento (Muma, 1971; Muma & Stone, 1971), mecanismos de reproducción (Bukowski et al., 2001), y métodos de depredación (Muma & Stone, 1971; Yoshida, 1989). Si bien se han desarrollado recientes estudios moleculares en *G. cancriformis*, éstos sólo han investigado la filogeografía de la especie (Chamberland et al., 2020; Salgado-Roa et al., 2018). Así, a la fecha no existen estudios que investiguen molecularmente su polimorfismo de color (Felipe M. Gawryszewski & Motta, 2008) y, por ende, no existe información sobre las bases genéticas que controlan la coloración en la especie. El desarrollo de este tipo de estudios se ha visto limitado en gran medida por la poca disponibilidad de recursos genéticos en *G. cancriformis*. Por ejemplo, a junio de 2020 existen 526 registros nucleotídicos para esta araña en el genbank, de los cuales 247 corresponden al gen citocromo oxidasa, 87 al gen hsp90, 95 al gen ITS, 100 al gen 28S ARNr y 104 al gen 16S ARNr. Es decir, no hay ninguna secuencia asociada a genes candidatos de coloración. Por otra parte, si bien existe un transcriptoma de la especie (Prosdocimi et al., 2011), su análisis se enfoca en genes asociados con tela orbicular, y su ensamblaje no está disponible públicamente (ni siquiera bajo solicitud directa a los autores).



Figura 1. Morfotipos de color reportados en *G. cancriformis* (Salgado, 2019). En una misma localidad pueden presentarse hasta 8 morfotipos.

El presente estudio caracteriza *de novo* el transcriptoma de *G. cancriformis* y evalúa la expresión diferencial entre tres morfotipos de coloración (blanco, amarillo y naranja) simpátricos en Ibagué (Tolima, Colombia). Así, busqué identificar transcritos asociados a coloración sin tener señales de “ruido” debidas a estructura poblacional y también analicé si había diferencias a nivel nucleotídico entre los distintos morfotipos de coloración. Los datos de este estudio serán un valioso recurso genético de referencia para futuras investigaciones sobre arañas y constituyen un paso fundamental hacia la identificación de las bases genéticas del polimorfismo de color en estos animales. Esto será fundamental para desarrollar y entender futuros estudios en evolución, biología del desarrollo y ecología en arañas Neotropicales.

MÉTODOS

Recolección de muestras, extracción de ARN y secuenciación

Colecté hembras adultas de *G. cancriformis* entre los meses de mayo y junio del 2019, en Ibagué-Tolima (4°26'24"N - 75°10'48"W). En dicha población se encuentran hembras de tres colores diferentes: amarillo, naranja y blanco. La colecta se hizo siguiendo protocolos estándar de búsqueda visual (Coddington et al., 1991), y las arañas se depositaron en tubos con RNAlater que se almacenaron a -80°C.

Posteriormente extraje el ARN total de nueve individuos para tener tres réplicas biológicas por cada morfotipo de color. La extracción de ARN se realizó con el kit RNeasy (Qiagen), siguiendo las instrucciones del fabricante. La calidad e integridad del ARN se verificó por medio de electroforesis en gel de agarosa, y luego el ARN se cuantificó con Nanodrop. Una vez confirmadas la cantidad y calidad del ARN, éste se envió a Novogene (Corea) para preparación de las librerías cDNA y

secuenciación por la plataforma Hiseq de Illumina (Paired-End – PE – de 150 bp). De acuerdo a las recomendaciones de Illumina (Illumina, n.d.), se sabe que para proyectos de ensamblaje de transcriptoma y anotación usualmente se requiere entre 100 y 200 millones de lecturas. Así, secuenció 6 Gb por muestra, que corresponden a 400 millones de lecturas (cantidad más que suficiente para garantizar un buen ensamblaje).

Verificación de calidad y generación de transcriptoma de novo

Todo el procesamiento de secuencias y pasos computacionales se desarrollaron en el HPC de la Universidad del Rosario. Primero, verifiqué la calidad de las lecturas usando el programa FastQC (Andrews, 2010). Posteriormente, usé el programa Trimmomatic (Bolger et al., 2014) para remover los adaptadores remanentes y para descartar las lecturas de baja calidad (quality score cutoff = 25) o baja longitud (menores a 36bp). Usé las lecturas depuradas de los nueve individuos (3 réplicas x 3 colores) y las lecturas de transcriptomas generadas en estudios previos para *Gasteracantha arcuata* y *Gasteracantha hasselti* (Haas et al., 2013) para ensamblar un transcriptoma *de novo* que sirva como referencia para todo el género *Gasteracantha*. Este ensamblaje se construyó con el software Trinity v2.8.6 usando los parámetros por defecto (kmer = 25) (Haas et al., 2013). La calidad del ensamblaje se verificó usando el script *TrinityStats.pl*.

Mapeo de lecturas y expresión diferencial de genes

Usé el script *align_and_estimate_abundance.pl* de Trinity con dos propósitos. Primero, usar bowtie2 (Langmead et al., 2009) para mapear las lecturas de cada individuo de *G. cancriformis* contra el ensamblaje *de novo* previamente generado. Segundo, estimar la abundancia de cada transcrito (i.e. nivel de expresión) en cada individuo usando RSEM (B. Li & Dewey, 2011). Con estos datos construí una matriz de conteos y de valores de expresión normalizados por individuo implementando el script *abundance_estimates_to_matrix.pl* de Trinity.

Posteriormente, hice el análisis de expresión diferencial con tres réplicas por cada fenotipo de color en tres comparaciones: (i) amarillo vs. naranja, (ii) naranja vs. blanco, y (iii) amarillo vs. blanco. Para ello implementé el script *run_DE_analysis.pl* con edgeR (Robinson et al., 2010). Identifiqué como transcritos diferencialmente expresados (TDE) aquellos que tuvieran valores p ajustados de false discovery rate (FDR) ≤ 0.01 y $\log_2$ fold-change > 2 . Esto lo hice con el script *analyze_diff_expr.pl*. Finalmente agrupé los TDE de acuerdo a su similitud de expresión entre fenotipos de color usando la función de R *heatmap.2*.

Anotación funcional de los transcritos diferencialmente expresados

Los 10 TDE con mayores diferencias de expresión identificados en cada una de las comparaciones fueron anotados funcionalmente en el programa Blast2GO PRO (Conesa et al., 2005). Para ello, cargué en el programa un archivo fasta múltiple que contiene la secuencia de los 10 TDE top. Con este archivo hice un BLASTX contra la base de datos de proteínas no redundantes del NCBI (nr), usando los parámetros por defecto. Aquellos transcritos que encontraron hits de BLASTX fueron mapeados a términos de ontología génica (GO) y se les asignó anotación funcional (Conesa et al., 2005). Finalmente, las categorías GO identificadas (proceso biológico, función molecular y compartimento celular) fueron visualizadas gráficamente.

Identificación de SNPs asociados al fenotipo

Con objetivo de identificar si en los TDE existían diferencias a nivel nucleotídico entre los fenotipos comparados, llamé los ‘Single Nucleotide Polymorphisms’ (SNPs) totales y los SNPs de los TDE en cada individuo usando samtools v1.8 (H. Li et al., 2009). Los 102,411 SNPs extraídos se almacenaron en un archivo VCF que se filtró con vcftools v0.1.16 (Danecek et al., 2011) buscando variantes que estuvieran presentes en al menos 70% de los individuos, SNPs con recuento de alelos menor a 3 y eliminando variantes que son indeles o que no son bi-alélicas. Posterior a esto, convertí el archivo VCF a formato Phylip para crear un árbol de máxima verosimilitud para todos los SNPs concatenados y para los SNPs correspondientes sólo a los TDE. Con éstos, construí árboles con RAxML v8.2.11 bajo el modelo GTR + Gamma, usando la corrección de sesgo de adquisición de Lewis. El soporte se evaluó utilizando 1000 pseudoréplicas de bootstrap.

Adicionalmente, para identificar la relación entre individuos según su polimorfismo en ADN, hice un análisis de componentes principales (PCA) para todos los SNPs y para los SNPs correspondientes a los TDE. Este análisis se hizo usando el paquete de R v3.6.3 *adeigenet*. Del mismo modo construí un árbol de distancias, convirtiendo los archivos VCF con *plink* (Purcell et al., 2007). Usé el software ngsDist (Vieira et al., 2016) para calcular distancias genotípicas según las diferencias de frecuencias alélicas para las comparaciones: (i) color (naranja + amarillo) vs. no color (blanco) y (ii) amarillo vs naranja, con 500 pseudoréplicas bootstrap. La significancia estadística de los valores de distancia promedio se calculó con una prueba t-student. Partiendo del archivo de distancias utilicé FastME v2.1.6.1 (Lefort et al., 2015) para construir los árboles de topología nearest-neighbor interchange (NNI), bajo el método Neighbor Joining y evaluando el soporte con 1002 réplicas bootstrap.

RESULTADOS

Secuenciación y ensamblaje de novo del transcriptoma de G. cancriformis

Obtuve un total de 186'910,575 lecturas pareadas de 150bp producto de la secuenciación de nueve hembras adultas de *G. cancriformis* (Tabla 1). De estas lecturas, 186'910,522 (99%) pasaron los filtros de calidad. Estas fueron usadas para la construcción del transcriptoma *de novo*.

Tabla 1. Información los individuos respecto a su morfotipo, número de lecturas y porcentaje de lecturas mapeadas a la referencia.

Individuo (ID)	Morfotipo	Número total de lecturas	Número de lecturas que cumplen filtros de calidad	Porcentaje de lecturas mapeadas
PTUR001	Naranja	24'604,985	24'604,970	97,91%
PTUR002	Naranja	23'328,078	23'328,064	98,16%
PTUR003	Naranja	17'935,961	17'935,949	97,92%
PTUR004	Blanco	20'789,871	20'789,871	98,08%
PTUR005	Blanco	19'030,792	19'030,792	97,71%
PTUR006	Blanco	21'435,557	21'435,545	97,99%
PTUR007	Amarillo	21'073,664	21'073,664	98,19%
PTUR008	Amarillo	21'408,370	21'408,370	98,29%
PTUR009	Amarillo	17'303,297	17'303,297	98,27%
Total		186'910,575	186'910,522	

La combinación de estos datos, junto con otros dos transcriptomas previamente disponibles para el género *Gasteracantha*, produjo un transcriptoma *de novo* que se compone de 519.150 '*Trinity Transcripts*' que se agrupan en 350.375 '*Trinity Genes*'. Estos últimos tienen una longitud total de 463.665.181 pares de bases (pb), longitud promedio de 893,19 pb, longitud N50 de 1679 pb y contenido de GC de 32,06%. La longitud más común entre los transcritos identificados fue entre 100 y 500 bp (Figura 2).

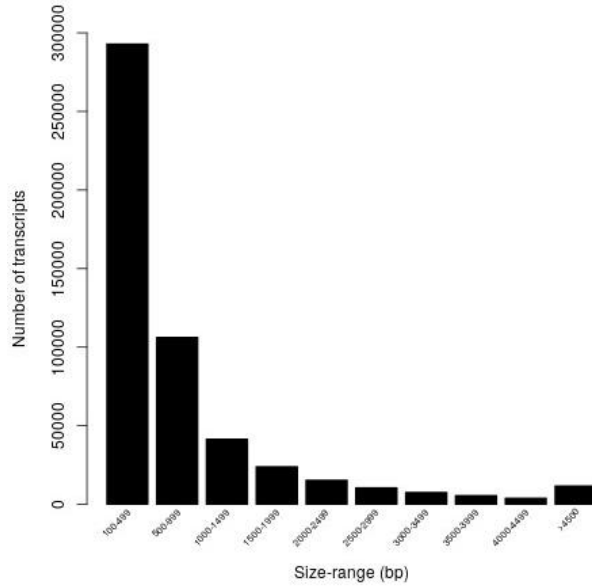


Figura 2. Número de transcritos por longitud identificados en le transcriptoma ensamblado de *novo* de *Gasteracantha*.

Mapeo y análisis de expresión diferencial de genes

El porcentaje de lecturas mapeadas por individuo al transcriptoma de referencia previamente ensamblado fue consistente entre los 9 individuos, con un promedio de $98.05\% \pm 0.2\%$ (Tabla 1). Para la comparación naranja vs. blanco encontré 2.345 TDE de los cuales 1.240 están sobreexpresados y 1.105 subexpresados en el fenotipo naranja respecto al blanco (Fig. 3a y 3b). Encontré un patrón muy similar en la comparación amarillo vs. blanco donde detecté 2.187 TDE de los cuales 1.109 están sobreexpresados y 1.078 subexpresados en el fenotipo amarillo respecto al blanco (Fig. 3e y 3f). En contraste, la comparación naranja vs. amarillo reveló un menor número de TDE (1256) de los cuales 718 están sobreexpresados y 583 subexpresados en el fenotipo naranja respecto al amarillo (Fig. 3c. y 3d). En concordancia con estos números, el PCA construido con los perfiles de expresión revela que el fenotipo blanco se diferencia transcripcionalmente de los fenotipos naranja y amarillo, quienes son mucho más similares entre ellos (Fig. 4a). Este patrón es confirmado por el heatmap donde todos los individuos de los fenotipos naranja y amarillo se agrupan en un único cluster, mientras que los individuos de fenotipo blanco aparecen lejanos a este clusuer, excepto por una muestra (Fig. 4b).

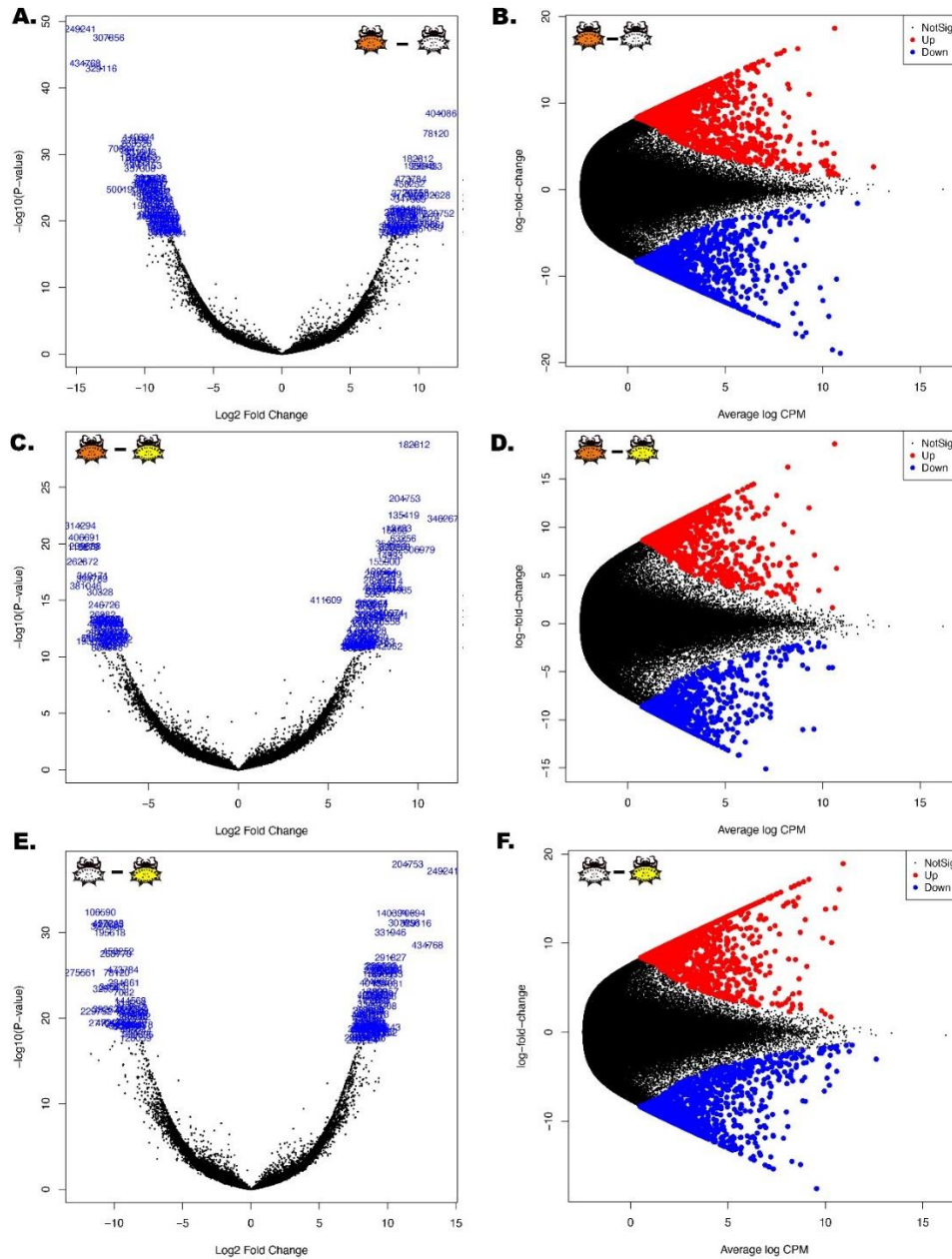


Figura 3. Transcritos Diferencialmente expresados (TDE) para cada una de las tres comparaciones entre fenotipos de color de *G. cancriformis*. A y B: naranja vs. blanco. C y D: naranja vs. amarillo E y F: amarillo vs. blanco. Los paneles A, C y E son gráficas volcano donde los TDE con un valor p ajustado $FDR \leq 0.01$ se muestran en azul. Los paneles B, D y F son gráficos de diferencia de medias (MD) donde los transcritos sobreexpresados en el primer fenotipo de la comparación están en rojo y los subexpresados en azul. Los puntos negros corresponden a los transcritos con nivel de expresión similar entre los dos fenotipos comparados.

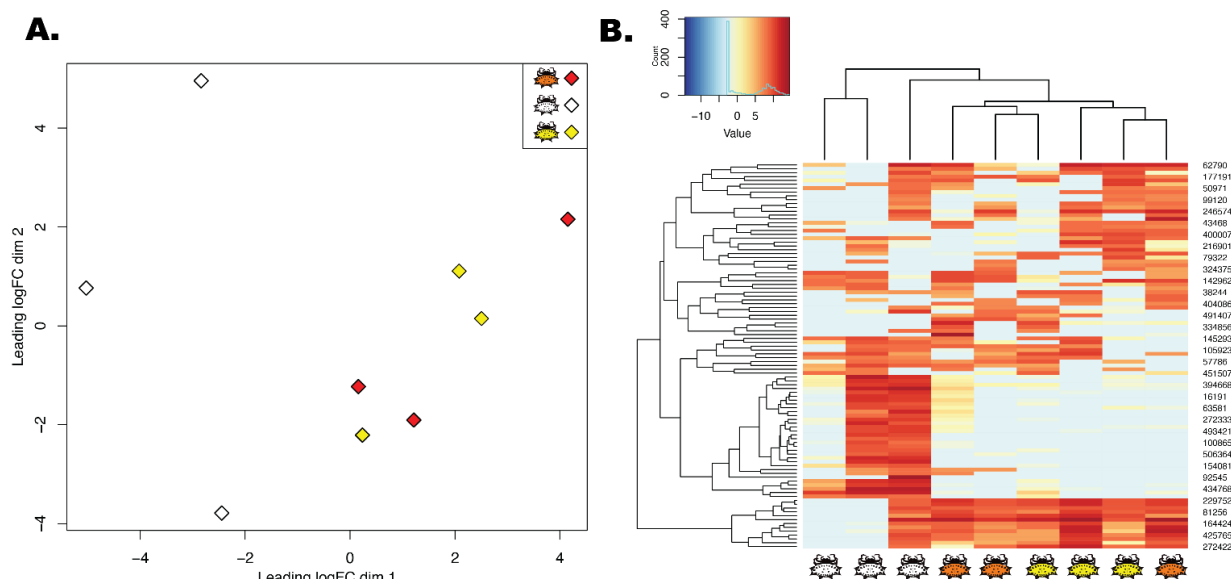


Figura 4. Similitud en perfil transcripcional entre los fenotipos de color de *G. cancriformis*. **A.** Principal Component Analysis (PCA) de los transcritos entre fenotipos de color. **B.** Heatmap que muestra la expresión relativa de los 100 transcritos con mayor diferencia de expresión (filas) entre los fenotipos y sus réplicas (columnas). Las filas y columnas están agrupadas jerárquicamente.

Anotación funcional de transcritos diferencialmente expresados

Seleccionamos los 10 TDE más diferenciados en cada una de las comparaciones y los reportamos en la tabla suplementaria 1. Interesantemente, 50% de estos TDE top eran comunes en las comparaciones naranja vs. blanco y amarillo vs. blanco, y estaban fundamentalmente anotados como Astacin-like metaloproteasas o como Vitellogenin (Tabla spp. 1). En la comparación naranja vs. blanco los TDE top están implicados en 39 términos de ontología génica (GO) que incluyen 7 procesos biológicos, 3 componentes celulares y 12 funciones moleculares (Tabla spp. 1). En la comparación amarillo vs. blanco los TDE top se clasificaron en 45 términos GO, que incluyen 8 procesos biológicos, 3 componentes celulares y 12 funciones moleculares (Tabla spp. 1). Finalmente, para la comparación naranja vs. amarillo los TDE se clasificaron en 19 términos GO, incluidos 4 procesos biológicos, 1 componente celular y 7 funciones moleculares (Tabla spp. 1).

Análisis de estructura basado en SNPs

Los valores de distancia calculados para todos los SNPs revelan que no hay diferencias significativas a nivel nucleotídico entre los fenotipos comparados ($p\text{-value} = 0,172 > 0,05$). El valor de distancia promedio para la comparación color vs. no color fue de 0,486 y para la comparación amarillo vs.

naranja de 0,476 (Fig. 6a). En contraste, las distancias calculadas con los SNPs de los TDE revelan que la distancia promedio para la comparación color vs. no color ($d=0,458$) es significativamente mayor al valor de distancia promedio obtenido para la comparación amarillo vs naranja ($d=0,416$, $p\text{-value}= 0,016 < 0,05$) (Fig. 6b).

Tanto los PCA como los árboles de máxima verosimilitud corroboran esta información. El PCA y el árbol de máxima verosimilitud contruidos para todos los SNPs revelan agrupación no asociada a fenotipo de color (Fig. 5a y 6c), mientras que el PCA y el árbol de máxima verosimilitud contruidos con los SNPs de los TDE muestran que los individuos de fenotipo blanco se agruparon monofiléticamente en un cluster aparte de los fenotipos naranja y amarillo (Fig. 5b y 6d). Estos resultados indican que para los SNPs correspondientes a los TDE existe una diferenciación mayor entre los individuos de fenotipo incoloro (blanco) y los individuos de fenotipo de color (naranja y amarillo), lo cual concuerda con el patrón observado previamente en el análisis de agrupación transcripcional (Fig. 4.).

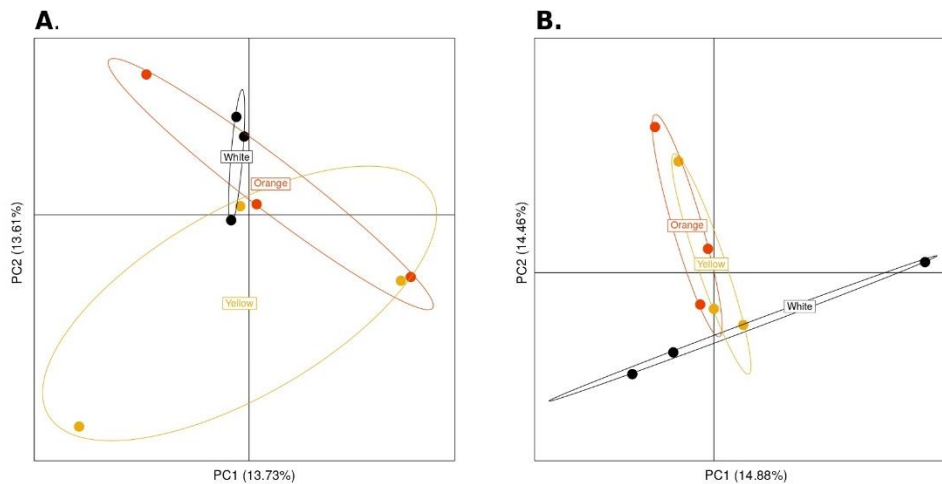


Figura 5. Análisis de agrupamiento en perfil nucleotídico entre los fenotipos de color de *G. cancriformis*. **A.** Principal Component Analysis (PCA) para los SNPs totales. **B.** Principal Component Analysis (PCA) para los SNPs correspondientes a los TDE.

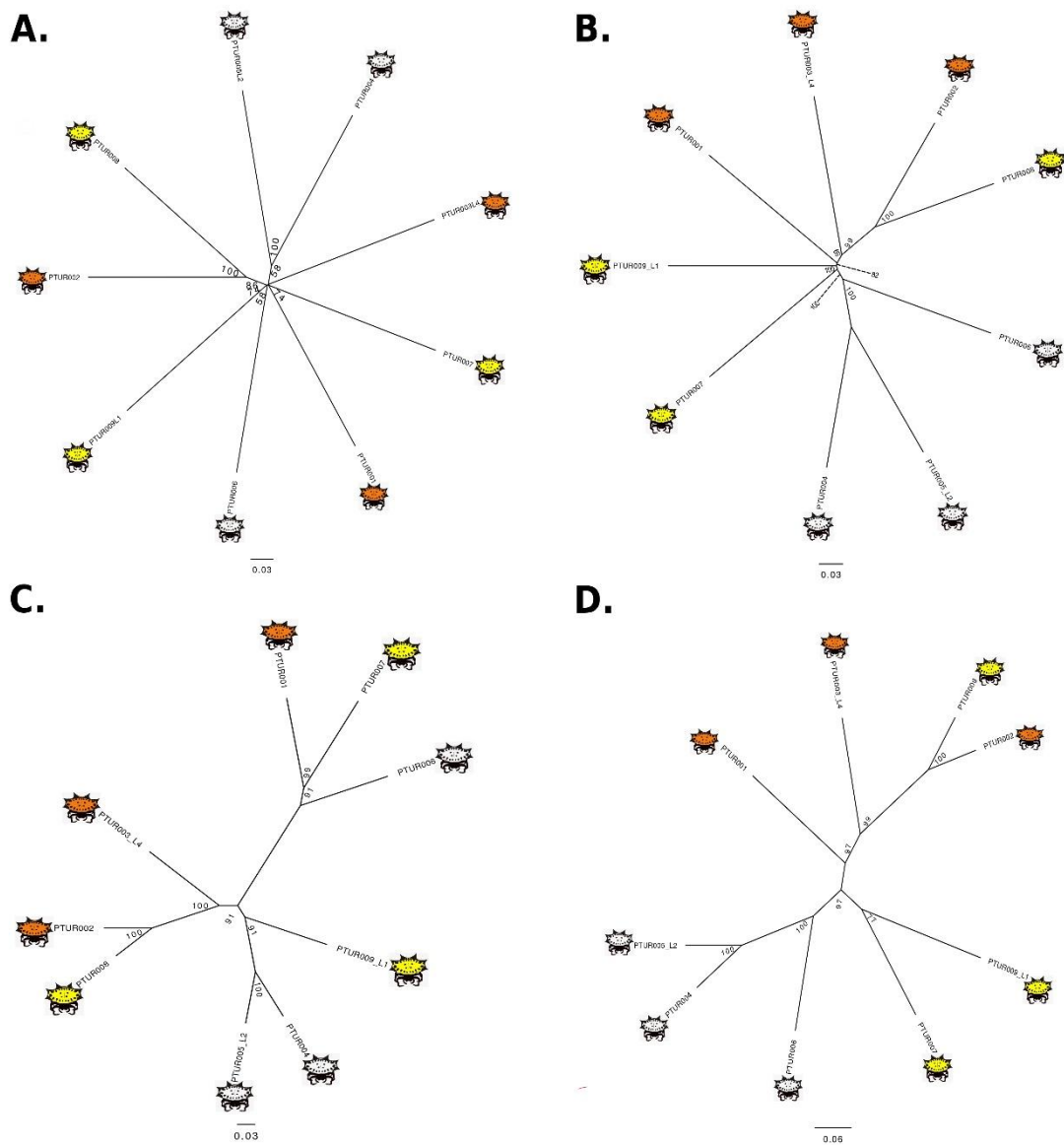


Figura 6. Árboles filogenéticos construidos con los SNPs totales y los SNPs de los TDE entre los fenotipos blanco, amarillo y naranja. A y C: SNPs totales. B y D: SNPs correspondientes a los TDE. Los paneles de A y B corresponden a los árboles por distancias y los paneles C y D corresponden a los árboles de máxima verosimilitud. Los valores de los soportes se ubican en los nodos de cada árbol.

DISCUSIÓN

En este trabajo presentó el primer análisis de transcriptoma para la araña *G. cancriformis* y reportó 519.150 transcritos y 350.375 genes. Así mismo, reportó que en hembras adultas de esta araña existe expresión diferencial de genes asociada a fenotipo de coloración, y que esta señal se confirma en términos de polimorfismos de ADN.

El análisis de expresión diferencial reveló que el perfil de expresión genética es muy similar entre fenotipos amarillo y naranja comparados con el fenotipo blanco (es decir, se encontró una diferencia de expresión entre tener color y no tenerlo). Aún así, el fenotipo blanco mostró una importante dispersión lo cual puede ser debido a variación individual. Por ende, sería ideal incluir un mayor número de individuos blancos para descartar este sesgo, o incluir réplicas técnicas para cada individuo de forma que se controle por variación individual. Además, los TDE asociados con fenotipo amarillo y naranja estaban fundamentalmente anotados como metaloproteasas Astacin-like o como Vitellogenin (Tabla spp. 1). Las astacinas son metaloproteasas dependientes de zinc que se encuentran en diversos organismos, desde bacterias a mamíferos (Bond & Beynon, 1995) y están presentes en los fluidos digestivos de las arañas, cumpliendo un papel importante en la digestión de presas extraorales (Foradori et al., 2006; Walter et al., 2017). Además, estas enzimas se han reportado como componentes fundamentales del veneno de las arañas de género *Loxosceles* (Medina et al., 2019; Prosdocimi et al., 2011) pero, más interesante aún, es su rol en artrópodos (crustáceos) y en peces donde la astacina se ha identificado como un pigmento rojo derivado de la oxidación del carotenoide astaxantina (Liang et al., 2012; Xing et al., 2018). Por otra parte, la vitelogenina es una fosfolipoglicoproteína precursor principal de la producción de vitelina (vn), proteína del vitelo de los ovocitos en animales ovíparos (Khalaila et al., 2004). Se ha demostrado que la vitelogenina está relacionada con la defensa inmune del huésped (Lourenço et al., 2009) y que posee propiedades hemaglutinantes y antibacterianas (Z. Li et al., 2008). Adicionalmente, varios estudios han demostrado que en insectos, crustáceos y peces esta proteína contiene un alto porcentaje de lípidos, conjugados invariablemente con carotenoides (Ando et al., 1986; Bazzyar Lakeh et al., 2010; García & Lara, 2013; Khalaila et al., 2004; Krishnan et al., 2008). Del mismo modo, Lubzens et al., en el 2003 demostraron que la vitelogenina está principalmente implicada en el transporte de astaxantinas en peces (Lubzens et al., 2003).

De esta manera la sobreexpresión de transcritos de tipo astacina y vitelogenina en los morfotipos naranja y amarillo respecto al morfotipo blanco pueden indicar que estas coloraciones son producto de la presencia de carotenoides, y por ende que los carotenoides son causales de las tonalidades rojas/naranjas/amarillas en *G. cancriformis*. Si este es el caso, este trabajo es el primero en reportar

cuál es la base molecular del color en esta especie. Sin embargo, existe una explicación alternativa. Debido a la función reportada de la astacina en la composición del veneno de las arañas *Loxosceles* (Medina et al., 2019; Prosdocimi et al., 2011), es posible que la mayor expresión de esta molécula en los fenotipos con color esté relacionada con aposematismo, es decir, que los fenotipos de color más conspicuo (amarillos y naranjas) son los más venenosos. Sin embargo, probar esta hipótesis implicaría caracterizar químicamente el veneno de estos morfotipos, detectar la presencia de astacina, y verificar si su concentración es mayor en fenotipos aposemáticos. Este es un tema para futuras investigaciones. Adicionalmente, se podría argumentar que la expresión de la vitelogenina esté asociada a la presencia de huevos en las hembras secuenciadas (en caso de que éstas sean grávidas, lo cual se desconoce). Sin embargo, esto no explicaría la asociación de la expresión de la vitelogenina con fenotipo de coloración y es poco probable que sólo las hembras de color naranja y amarillo hayan sido grávidas, y las blancas no.

Por otra parte, si bien al analizar la señal de polimorfismos de ADN (SNPs) a nivel de todo el transcriptoma no se observó agrupación asociada a fenotipo de color, cuando analicé únicamente los SNPs de los TDE confirmé la mayor cercanía entre fenotipos de color versus el blanco (Fig. 5 y 6). Esto indica que la diferenciación en color en *G. cancriformis* posiblemente no es únicamente a nivel transcripcional, sino que tal como se ha demostrado en otros organismos, el polimorfismo de color puede estar dado por variantes alélicas o incluso mutaciones individuales que controlen la expresión en genes de forma que se produzcan pigmentos (Orteu & Jiggins, 2020). Del mismo modo, no se descarta la posibilidad de que la variación de color en *G. cancriformis* tenga, al igual que en gran cantidad de casos, una base de arquitectura poligénica (Orteu & Jiggins, 2020). Por ello es importante realizar estudios posteriores de asociación de genoma completo (GWAS) en poblaciones más grandes y con mayor variación en polimorfismo de color para analizar con más detalle la arquitectura genética que influye en la variación de pigmentación abdominal en esta especie.

La mayoría de las investigaciones actuales sobre los patrones de coloración en arañas se centran en estudios conductuales y/o ecológicos. Sin embargo, el enfoque para comprender la base genética de la coloración que se describe en este trabajo es de gran utilidad para investigar las interacciones recíprocas entre los procesos evolutivos y de desarrollo que dan forma a la variación fenotípica. El transcriptoma ensamblado en este estudio no está completo en términos de los genes e isoformas totales de la especie, pero constituye un excelente primer recurso para posteriores estudios. La no completitud de este transcriptoma se debe a que la detección de transcritos depende de la expresión de éstos y sólo colecté individuos adultos, implicando que no se han muestreado transcritos de expresión exclusiva en estados inmaduros. Además, en este estudio solo se incluyeron hembras (las

cuales son coloridas), y por ende no se interrogó el transcriptoma de los machos. Aún así, el hallazgo de expresión del pigmento rojo astacina en los individuos de morfotipo naranja y amarillo que previamente había sido encontrado en crustáceos, podría ayudarnos a comprender el origen evolutivo de tonos de color similares en los artrópodos. Igualmente, comprender los mecanismos de producción de color también puede revelar posibles funciones no visuales como la composición del veneno, digestión, defensa inmune, etc. Es necesario realizar estudios posteriores con mayor número de muestras referentes a poblaciones más grandes y más morfotipos de coloración, lo cual será fundamental para futuros estudios en evolución, biología del desarrollo y ecología en arañas Neotropicales.

CONCLUSIONES

El ensamblaje de *nov* del transcriptoma de *G. cancriformis* me permitió identificar genes candidatos no descritos previamente que pueden estar involucrados en la pigmentación abdominal de la especie. Los resultados del análisis comparativo del transcriptoma respaldaron la idea de que hay una relación entre la expresión génica y la coloración abdominal de la araña *G. cancriformis*. Del mismo modo, los resultados del análisis de diferencias a nivel nucleotídico revelaron que el patrón de coloración puede estar dado por variantes alélicas o incluso por mutaciones individuales que controlen la expresión en genes involucrados en la producción de pigmentos, pero es necesario comprobar esto en futuras investigaciones. La posible participación de carotenoides como causales de las tonalidades rojas/naranjas/amarillas en *G. cancriformis* es un hallazgo novedoso sobre la base molecular del color en esta especie. Sin embargo, los genes encontrados aquí también pueden estar principalmente involucrados en funciones no visuales como la composición del veneno, digestión, defensa inmune, etc. Este estudio ha facilitado una descripción inicial del mecanismo molecular de la pigmentación y el polimorfismo de color en *G. cancriformis*, de la misma forma que proporciona información genómica que es de gran utilidad para futuros estudios en esta especie y especies estrechamente relacionadas.

AGRADECIMIENTOS

Agradezco de primera mano a mi directora Carolina Pardo, a Camilo Salazar y a Fabian Salgado que me guiaron y acompañaron en todo el proceso de aprendizaje, colecta de individuos, procesamiento de muestras y análisis de datos realizado en esta investigación. También agradezco a mi familia y amigos que me acompañaron y apoyaron en todo el proceso. Este estudio fue financiado por el Grupo de Genética Evolutiva de la Universidad del Rosario y todo el procesamiento de secuencias y pasos computacionales se desarrollaron en el HPC de la Universidad del Rosario.

REFERENCIAS

- Ando, S., Takeyama, T., & Hatano, M. (1986). Transport associated with serum vitellogenin of carotenoid in chum salmon (*Oncorhynchus keta*). *Agricultural and Biological Chemistry*, 50(3), 557–563. <https://doi.org/10.1080/00021369.1986.10867435>
- Andrews, S. (2010). FastQC A Quality Control tool for High Throughput Sequence Data. *Babraham Bioinformatics*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bazyar Lakeh, A. A., Ahmadi, M. R., Safi, S., Ytrestøyl, T., & Bjerkeng, B. (2010). Growth performance, mortality and carotenoid pigmentation of fry offspring as affected by dietary supplementation of astaxanthin to female rainbow trout (*Oncorhynchus mykiss*) broodstock. *Journal of Applied Ichthyology*, 26(1), 35–39. <https://doi.org/10.1111/j.1439-0426.2009.01349.x>
- Bolger, A., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bond, J. S., & Beynon, R. J. (1995). The astacin family of metalloendopeptidases. In *Protein Science* (Vol. 4, Issue 7, pp. 1247–1261). John Wiley & Sons, Ltd. <https://doi.org/10.1002/pro.5560040701>
- Bukowski, T. C., Linn, C. D., & Christenson, T. E. (2001). Copulation and sperm release in *Gasteracantha cancriformis* (Araneae: Araneidae): Differential male behaviour based on female mating history. *Animal Behaviour*, 62(5), 887–895. <https://doi.org/10.1006/anbe.2001.1834>
- Chamberland, L., Salgado-Roa, F. C., Basco, A., Crastz-Flores, A., Binford, G. J., & Agnarsson, I. (2020). Phylogeography of the widespread Caribbean spiny orb weaver *Gasteracantha cancriformis*. *PeerJ*, 8, e8976. <https://doi.org/10.7717/peerj.8976>
- Coddington, J. A., Griswold, C. E., Davila, D. S., Peftaranda, E., & Larcher, S. F. (1991). *Designing and Testing Sampling Protocols to Estimate Biodiversity in Tropical Ecosystems*. Dioscorides Press.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Croucher, P. J. P., Brewer, M. S., Winchell, C. J., Oxford, G. S., & Gillespie, R. G. (2013). De novo

- characterization of the gene-rich transcriptomes of two color-polymorphic spiders, *Theridion grallator* and *T. californicum* (Araneae: Theridiidae), with special reference to pigment genes. *BMC Genomics*, 14(1), 1–18. <https://doi.org/10.1186/1471-2164-14-862>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., Mcvean, G., Durbin, R., & Project, G. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Foradori, M. J., Tillinghast, E. K., Smith, J. S., Townley, M. A., & Mooney, R. E. (2006). Astacin family metallopeptidases and serine peptidase inhibitors in spider digestive fluid. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 143(3), 257–268. <https://doi.org/10.1016/j.cbpb.2005.08.012>
- García, M., & Lara, M. (2013). The use of carotenoid in aquaculture. *Research Journal of Fisheries and Hydrobiology*, 8(2), 38–49.
- Gawryszewski, F. M., & Motta, P. C. (2012). Colouration of the orb-web spider *Gasteracantha cancriformis* does not increase its foraging success. *Ethology Ecology and Evolution*, 24(1), 23–38. <https://doi.org/10.1080/03949370.2011.582044>
- Gawryszewski, Felipe M., & Motta, P. C. (2008). The silk tuft web decorations of the orb-weaver *Gasteracantha cancriformis*: Testing the prey attraction and the web advertisement hypotheses. *Behaviour*, 145(3), 277–295. <https://doi.org/10.1163/156853908783402911>
- Gray, S. M., & McKinnon, J. S. (2007). Linking color polymorphism maintenance and speciation. In *Trends in Ecology and Evolution* (Vol. 22, Issue 2, pp. 71–79). <https://doi.org/10.1016/j.tree.2006.10.005>
- Gross, J. B., & Wilkens, H. (2013). Albinism in phylogenetically and geographically distinct populations of *Astyanax* cavefish arises through the same loss-of-function *Oca2* allele. *Heredity*, 111(2), 122–130. <https://doi.org/10.1038/hdy.2013.26>
- 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 using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>

- Hoekstra, H. E., Hirschmann, R. J., Bunday, R. A., Insel, P. A., & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, 313(5783), 101–104. <https://doi.org/10.1126/science.1126121>
- Hsiung, B. K., Justyn, N. M., Blackledge, T. A., & Shawkey, M. D. (2017). Spiders have rich pigmentary and structural colour palettes. *Journal of Experimental Biology*, 220(11), 1975–1983. <https://doi.org/10.1242/jeb.156083>
- Khalaila, I., Peter-Katalinic, J., Tsang, C., Radcliffe, C. M., Aflalo, E. D., Harvey, D. J., Dwek, R. A., Rudd, P. M., & Sagi, A. (2004). Structural characterization of the N-glycan moiety and site of glycosylation in vitellogenin from the decapod crustacean *Cherax quadricarinatus*. *Glycobiology*, 14(9), 767–774. <https://doi.org/10.1093/glycob/cwh105>
- Krishnan, M., Bharathiraja, C., Warriar, S., Krishnan, M., Muthumeenakshi, P., Bharathiraja, C., & Subramoniam, T. (2008). A comparative study on vitellogenin receptor of a lepidopteran insect (*Spodoptera litura*) and a decapod crustacean (*Scylla serrata*): Phylogenetic implication and co-evolution with vitellogenins. In *J Endocrinol Reprod* (Vol. 12). <https://www.researchgate.net/publication/241727062>
- Kronforst, M. R., Barsh, G. S., Kopp, A., Mallet, J., Monteiro, A., Mullen, S. P., Protas, M., Rosenblum, E. B., Schneider, C. J., & Hoekstra, H. E. (2012). Unraveling the thread of nature's tapestry: The genetics of diversity and convergence in animal pigmentation. *Pigment Cell and Melanoma Research*, 25(4), 411–433. <https://doi.org/10.1111/j.1755-148X.2012.01014.x>
- Kronforst, M. R., & Papa, R. (2015). The functional basis of wing patterning in *Heliconius* butterflies: The molecules behind mimicry. In *Genetics* (Vol. 200, Issue 1, pp. 1–19). Genetics. <https://doi.org/10.1534/genetics.114.172387>
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
- Lefort, V., Desper, R., Gascuel, O., & Rosenberg, M. (2015). FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Molecular Biology and Evolution*, 32(10), 2798–2800. <https://doi.org/10.1093/molbev/msv150>
- Levi, H. (1978). The American orb-weaver genera *colphepeira*, *micrathena* and *gasteracantha* north of Mexico (Araneae, Araneidae). *Bulletin of the Museum of Comparative Zoology*, 148(9),

417–442. <https://doi.org/10.1038/193728a0>

- Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Project, G., & Subgroup, D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics Applications Note*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, Z., Zhang, S., & Liu, Q. (2008). Vitellogenin Functions as a Multivalent Pattern Recognition Receptor with an Opsonic Activity. *PLoS ONE*, 3(4), e1940. <https://doi.org/10.1371/journal.pone.0001940>
- Liang, Y., Bai, D., Yang, G., Wei, D., Guo, M., Yan, S., Wu, X., & Ning, B. (2012). Effect of Astacin on Growth and Color Formation of Juvenile Red-White Ornamental Carp (*Cyprinus carpio* var. koi L). *The Israeli Journal of Aquaculture*, 64. <http://www.aquaculturehub.org>
- Library Prep and Array Kit Selector. (n.d.). Retrieved August 3, 2020, from <https://www.illumina.com/library-prep-array-kit-selector.html>
- Linnen, C. R., Poh, Y. P., Peterson, B. K., Barrett, R. D. H., Larson, J. G., Jensen, J. D., & Hoekstra, H. E. (2013). Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science*, 339(6125), 1312–1316. <https://doi.org/10.1126/science.1233213>
- Lourenço, A. P., Martins, J. R., Bitondi, M. M. G., & Simões, Z. L. P. (2009). Trade-off between immune stimulation and expression of storage protein genes. *Archives of Insect Biochemistry and Physiology*, 71(2), 70–87. <https://doi.org/10.1002/arch.20301>
- Lubzens, E., Lissauer, L., Levavi-Sivan, B., Avarre, J. C., & Sammar, M. (2003). Carotenoid and retinoid transport to fish oocytes and eggs: What is the role of retinol binding protein? In *Molecular Aspects of Medicine* (Vol. 24, Issue 6, pp. 441–457). Elsevier Ltd. [https://doi.org/10.1016/S0098-2997\(03\)00040-2](https://doi.org/10.1016/S0098-2997(03)00040-2)
- Medina, R., Guerra, C., de Almeida, S., Costal, F., Alves, P., Carmo, O., Ferreyra, A., Bonilla, C., Gonzalez, E. E., Kalapothakis, E., & Chávez, C. (2019). Diversity of astacin-like metalloproteases identified by transcriptomic analysis in Peruvian *Loxosceles laeta* spider venom and in vitro activity characterization. *Biochimie*, 167, 81–92.

<https://doi.org/10.1016/j.biochi.2019.08.017>

- Muma, M. H. (1971). Biological and Behavioral Notes on *Gasteracantha cancriformis* (Arachnida: Araneidae). In *The Florida Entomologist* (Vol. 54, Issue 4, p. 345).
<https://doi.org/10.2307/3493600>
- Muma, M. H., & Stone, K. J. (1971). Predation of *Gasteracantha cancriformis* (Arachnidae: Araneidae) Eggs in Florida Citrus Groves by *Phalacrotophora epeirae* (Insecta: Phoridae) and *Arachnophaga ferruginea* (Insecta: Eupelmidae). In *The Florida Entomologist* (Vol. 54, Issue 4, p. 305). <https://doi.org/10.2307/3493590>
- Nachman, M. W., Hoekstra, H. E., & D'Agostino, S. L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, 100(9), 5268–5273. <https://doi.org/10.1073/pnas.0431157100>
- Orteu, A., & Jiggins, C. D. (2020). The genomics of coloration provides insights into adaptive evolution. *Nature Reviews Genetics*. <https://doi.org/10.1038/s41576-020-0234-z>
- Oxford, G. S., & Gillespie, R. G. (1998). Evolution and Ecology of Spider Coloration. *Annual Review of Entomology*, 43(1), 619–643. <https://doi.org/10.1146/annurev.ento.43.1.619>
- Prosdocimi, F., Bittencourt, D., da Silva, F. R., Kirst, M., Motta, P. C., & Rech, E. L. (2011). Spinning Gland Transcriptomics from Two Main Clades of Spiders (Order: Araneae) - Insights on Their Molecular, Anatomical and Behavioral Evolution. *PLoS ONE*, 6(6), e21634. <https://doi.org/10.1371/journal.pone.0021634>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet*, 81, 559–575. <https://doi.org/10.1086/519795>
- Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., Jiggins, C. D., Chamberlain, N. L., Kronforst, M. R., Chen, R., Halder, G., Nijhout, H. F., & McMillan, W. O. (2011). Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, 333(6046), 1137–1141. <https://doi.org/10.1126/science.1208227>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics Applications Note*, 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>

- Salgado-Roa, F. C., Pardo-Diaz, C., Lasso, E., Arias, C. F., Solferini, V. N., & Salazar, C. (2018). Gene flow and Andean uplift shape the diversification of *Gasteracantha cancriformis* (Araneae: Araneidae) in Northern South America. *Ecology and Evolution*, 8(14), 7131–7142. <https://doi.org/10.1002/ece3.4237>
- Streelman, J. T., Peichel, C. L., & Parichy, D. M. (2007). Developmental Genetics of Adaptation in Fishes: The Case for Novelty. *Annual Review of Ecology, Evolution, and Systematics*, 38(1), 655–681. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095537>
- Vieira, F. G., Lassalle, F., Korneliussen, T. S., & Fumagalli, M. (2016). Improving the estimation of genetic distances from Next-Generation Sequencing data. *Biological Journal of the Linnean Society*, 117(1), 139–149. <https://doi.org/10.1111/bij.12511>
- Walter, A., Bechsgaard, J., Scavenius, C., Dyrland, T. S., Sanggaard, K. W., Enghild, J. J., & Bilde, T. (2017). Characterisation of protein families in spider digestive fluids and their role in extra-oral digestion. *BMC Genomics*, 18(1), 600. <https://doi.org/10.1186/s12864-017-3987-9>
- White, T. E., & Kemp, D. J. (2015). Technicolour deceit: A sensory basis for the study of colour-based lures. *Animal Behaviour*, 105, 231–243. <https://doi.org/10.1016/j.anbehav.2015.04.025>
- Wittkopp, P. J., Smith-Winberry, G., Arnold, L. L., Thompson, E. M., Cooley, A. M., Yuan, D. C., Song, Q., & McAllister, B. F. (2011). Local adaptation for body color in *Drosophila americana*. *Heredity*, 106(4), 592–602. <https://doi.org/10.1038/hdy.2010.90>
- Wittkopp, Patricia J., & Beldade, P. (2009). Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy. *Seminars in Cell and Developmental Biology*, 20(1), 65–71. <https://doi.org/10.1016/j.semcdb.2008.10.002>
- Xing, L., Sun, L., Liu, S., Wan, Z., Li, X., Miao, T., Zhang, L., Bai, Y., & Yang, H. (2018). Growth, histology, ultrastructure and expression of MITF and astacin in the pigmentation stages of green, white and purple morphs of the sea cucumber, *Apostichopus japonicus*. *Aquaculture Research*, 49(1), 177–187. <https://doi.org/10.1111/are.13446>
- Yoshida, M. (1989). Predatory behavior of *Gasteracantha mammosa* C. Koch (Araneae: Araneidae). *Acta Arachnologica*, 37(2), 57–67. <https://doi.org/10.2476/asjaa.37.57>

ANEXOS

Tabla suplementaria 1. Listado de los 10 Transcritos Diferencialmente expresados (TDE) con mayor diferencia de expresión en cada una de las comparaciones. Se reporta su identidad de acuerdo con hit con BLAST y los términos de ontología génica (GO) identificados para cada transcrito. Los GO pueden ser de proceso biológico (P), Función (F), o Componente celular (C).

Comparación	Transcrito diferencialmente expresado	p ajustado	Descripción según BLAST	ID de ontología génica (GO)	Nombre de GO
Naranja vs. blanco	TRINITY_DN1405_c0_g2_i2	8,07e-112	Hypothetical protein AVEN_259965_1	---NA---	---NA---
	TRINITY_DN3491_c0_g1_i1	3,16e-56	Polyprotein	P:GO:0001172; P:GO:0006351; P:GO:0006508; P:GO:0039694; F:GO:0003723; F:GO:0003724; F:GO:0003968; F:GO:0005198; F:GO:0005524; F:GO:0008234; C:GO:0016020; C:GO:0019028; C:GO:0033644	P:transcription, RNA-templated; P:transcription, DNA-templated; P:proteolysis; P:viral RNA genome replication; F:RNA binding; F:RNA helicase activity; F:RNA-directed 5'-3' RNA polymerase activity; F:structural molecule activity; F:ATP binding; F:cysteine-type peptidase activity; C:membrane; C:viral capsid; C:host cell membrane
	TRINITY_DN4547_c0_g1_i9	1,12e-67	Vitellogenin-like protein	P:GO:0006351; P:GO:0006869; F:GO:0003899; F:GO:0005319; F:GO:0046983; C:GO:0016020; C:GO:0016021	P:transcription, DNA-templated; P:lipid transport; F:DNA-directed 5'-3' RNA polymerase activity; F:lipid transporter activity; F:protein dimerization activity; C:membrane; C:integral component of membrane
	TRINITY_DN5822_c0_g2_i1	6,57e-86	---NA---	---NA---	---NA---
	TRINITY_DN6584_c1_g1_i1	9,01e-63	Hypothetical protein AVEN_28772_1	---NA---	---NA---
	TRINITY_DN7077_c0_g2_i1	6,01e-66	Astacin-like metalloprotease toxin 1	P:GO:0006259; P:GO:0006974; P:GO:0051276; F:GO:0000166; F:GO:0004386; F:GO:0008237; F:GO:0046872	P:DNA metabolic process; P:cellular response to DNA damage stimulus; P:chromosome organization; F:nucleotide binding; F:helicase activity; F:metallopeptidase activity; F:metal ion binding
	TRINITY_DN11959_c0_g1_i1	4,68e-131	Protein P1	F:GO:0003824	F:catalytic activity

Amarillo vs. blanco	TRINITY_DN12039_c0_g1_i1	8,22e-65	Vitellogenin-like protein	P:GO:0006351; P:GO:0006869; F:GO:0003899; F:GO:0005319; F:GO:0046983	P:transcription, DNA-templated; P:lipid transport; F:DNA- directed 5'-3' RNA polymerase activity; F:lipid transporter activity; F:protein dimerization activity
	TRINITY_DN12039_c0_g1_i8	1,84e-56	Vitellogenin-like protein	P:GO:0006351; P:GO:0006869; F:GO:0003899; F:GO:0005319; F:GO:0046983	P:transcription, DNA-templated; P:lipid transport; F:DNA- directed 5'-3' RNA polymerase activity; F:lipid transporter activity; F:protein dimerization activity
	TRINITY_DN15094_c0_g1_i7	1,07e- 116	Putative polyprotein	F:GO:0016787	F:hydrolase activity
	TRINITY_DN1111_c0_g1_i1	9,59e-78	RNA-dependent RNA polymerase	P:GO:0008152; F:GO:0016787	P:metabolic process; F:hydrolase activity
	TRINITY_DN1405_c0_g2_i2	1,35e-86	hypothetical protein AVEN_259965_1	---NA---	---NA---
	TRINITY_DN4547_c0_g1_i9	4,26e-73	vitellogenin-like protein	P:GO:0006351; P:GO:0006869; F:GO:0003899; F:GO:0005319; F:GO:0046983; C:GO:0016020; C:GO:0016021	P:transcription, DNA-templated; P:lipid transport; F:DNA- directed 5'-3' RNA polymerase activity; F:lipid transporter activity; F:protein dimerization activity; C:membrane; C:integral component of membrane
	TRINITY_DN5822_c0_g2_i1	2,42e-60	---NA---		
	TRINITY_DN7077_c0_g2_i1	8,81e- 107	Astacin-like metalloprotease toxin 1	P:GO:0006259; P:GO:0006974; P:GO:0051276; F:GO:0000166; F:GO:0004386; F:GO:0008237; F:GO:0046872	P:DNA metabolic process; P:cellular response to DNA damage stimulus; P:chromosome organization; F:nucleotide binding; F:helicase activity; F:metallopeptidase activity; F:metal ion binding
	TRINITY_DN11959_c0_g1_i1	4,46e- 130	protein P1	F:GO:0003824	F:catalytic activity
	TRINITY_DN12039_c0_g1_i1	1,61e-66	Vitellogenin-like protein	P:GO:0006351; P:GO:0006869; F:GO:0003899; F:GO:0005319; F:GO:0046983	P:transcription, DNA-templated; P:lipid transport; F:DNA- directed 5'-3' RNA polymerase activity; F:lipid transporter activity; F:protein dimerization activity
	TRINITY_DN12039_c0_g1_i8	2,42e-60	Vitellogenin-like protein	P:GO:0006351; P:GO:0006869; F:GO:0003899; F:GO:0005319; F:GO:0046983	P:transcription, DNA-templated; P:lipid transport; F:DNA- directed 5'-3' RNA polymerase activity; F:lipid transporter activity; F:protein dimerization activity

	TRINITY_DN12039_c0_g1_i13	3,45e-57	Vitellogenin-like protein	P:GO:0006351; P:GO:0006869; F:GO:0003899; F:GO:0005319; F:GO:0046983	P:transcription, DNA-templated; P:lipid transport; F:DNA- directed 5'-3' RNA polymerase activity; F:lipid transporter activity; F:protein dimerization activity
	TRINITY_DN12062_c1_g1_i1	1,21e-54	Polyprotein	P:GO:0001172; P:GO:0006351; P:GO:0006508; P:GO:0039694; F:GO:0003723; F:GO:0003724; F:GO:0003968; F:GO:0005198; F:GO:0005524; F:GO:0008234; C:GO:0016020; C:GO:0019028; C:GO:0033644	P:transcription, RNA-templated; P:transcription, DNA-templated; P:proteolysis; P:viral RNA genome replication; F:RNA binding; F:RNA helicase activity; F:RNA-directed 5'-3' RNA polymerase activity; F:structural molecule activity; F:ATP binding; F:cysteine- type peptidase activity; C:membrane; C:viral capsid; C:host cell membrane
	TRINITY_DN864_c0_g1_i20	6,91e-34	Adult-specific rigid cuticular protein 15.7	F:GO:0016740; F:GO:0042302	F:transferase activity; F:structural constituent of cuticle
	TRINITY_DN3719_c0_g1_i23	4,11e-30	Adult-specific rigid cuticular protein 15.7	F:GO:0042302	F:structural constituent of cuticle
	TRINITY_DN5242_c0_g1_i8	6,12e-30	hypothetical protein AVEN_250603_1	---NA---	---NA---
	TRINITY_DN49334_c0_g1_i14	6,33e-29	---NA---	---NA---	---NA---
Naranja vs. amarillo	TRINITY_DN1132_c0_g2_i1	2,24e-47	polyprotein	P:GO:0006807; P:GO:0043170; P:GO:0044238; F:GO:0000166; F:GO:0004386; F:GO:0008233; F:GO:0016740; F:GO:0140098; C:GO:0019012	P:nitrogen compound metabolic process; P:macromolecule metabolic process; P:primary metabolic process; F:nucleotide binding; F:helicase activity; F:peptidase activity; F:transferase activity; F:catalytic activity, acting on RNA; C:virion
	TRINITY_DN11004_c0_g2_i1	2,83e-32	polyprotein	F:GO:0016787	F:hydrolase activity
	TRINITY_DN3343_c0_g1_i1	5,72e-50	putative polyprotein	F:GO:0016787	F:hydrolase activity
	TRINITY_DN15094_c0_g1_i7	3,69e- 116	putative polyprotein	F:GO:0016787	F:hydrolase activity
	TRINITY_DN2648_c0_g3_i1	2,85e-41	RNA-dependent RNA polymerase	P:GO:0008152; F:GO:0016787	P:metabolic process; F:hydrolase activity
	TRINITY_DN1111_c0_g1_i1	6,35e-49	RNA-dependent RNA polymerase	P:GO:0008152; F:GO:0016787	P:metabolic process; F:hydrolase activity