

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

ESCUELA DE CIENCIAS BIOLÓGICAS

ENSAMBLAJE Y CARACTERIZACIÓN DEL GENOMA MITOCONDRIAL DE
DROSOPHILA AMAGUANA (DIPTERA: DROSOPHILIDAE)

**DISERTACIÓN PREVIA A LA OBTENCIÓN DEL TÍTULO DE
LICENCIADO EN CIENCIAS BIOLÓGICAS**

Neal Gonzalo Machado Balladares

Quito, febrero 2023

CERTIFICACION

Ph.D. Doris Vela

Directora de la disertación

CERTIFICACIÓN

Certifico que la Disertación de Licenciatura en Ciencias Biológicas del Sr. Neal Gonzalo Machado Balladares ha sido concluida de conformidad con las normas establecidas; por lo tanto, puede ser presentada para la calificación correspondiente.

Dra. Doris Vela

Directora de la Disertación

Quito, 15 de febrero de 2023

DEDICATORIA

Primeramente, quiero dedicar esta tesis fruto del mayor logro que he hecho hasta la fecha a dos personas especiales en mi vida.

A Carlos Ordoñez, mi amigo del colegio, Carlitos con mucho cariño te agradezco por ser parte de mi vida, por enseñarme a valorar las emociones, gracias por las risas y por los llantos por las experiencias y enseñanzas, por ti valoro más al mundo y a las personas, por ti voy a seguir viviendo hasta que el cuerpo aguante. per aspera ad Astra.

Bradley Machado tal como te prometí, te dedico mi tesis hermana, gracias por permitirme estudiar, gracias por enseñarme a valorar a mi familia, gracias por apoyarme en mis caprichos; tu ñaña siempre estuviste para mí, en todas mis caídas y en mis logros, para conste que si en algún momento llego a ser grande en la vida, todo te lo deberé a ti ñaña. Se que hemos tenidos nuestros roses, pero quiero que sepas que guardo en mi corazón todas las noches que nos desvelamos viendo series o capítulos de Dragon ball y todas las veces que fuimos a comer sushi, gracias por ser parte de mi vida.

AGRADECIMIENTOS

Por último, quiero agradecer a mi directora de tesis Doris Vela y a mis tutores Romel Montufar, Simón Orozco y Romain Guyot, que gracias a sus conocimientos y sobre todo a su paciencia pude aprender, e iniciar mi carrera como investigador, son un ejemplo personal y espero en el futuro llegar a ser su colega. También quiero agradecer a la institución y sobre todo a mis amigos, amigos que conocí a lo largo de toda la carrera de diferentes semestres y diferentes profesiones. También quiero agradecer a mi profesor del colegio Roberto Giler, gracias por nunca perder la fe en mí; e igualmente agradecer a mi novia Katy, gracias por siempre apoyarme nena.

Asimismo, quiero agradecer a mi familia, mi padre Gonzalo Machado, mi madre Marthy Balladares, mis hermanas Maby y Brad, mi sobrino Stefano, mi mejor amigo Carlos Iglesias y mi hermano Donovan; quiero que sepan que sus esfuerzos dieron frutos, gracias por ayudarme tanto económica como emocionalmente, y por apoyarme en todo, los amo.

TABLA DE CONTENIDO

RESUMEN	1
ABSTRACT	2
1. INTRODUCCIÓN	3
2. OBJETIVOS.....	4
2.1. OBJETIVO GENERAL.....	5
2.2. OBJETIVOS ESPECÍFICOS	5
3. MATERIALES Y MÉTODOS	6
3.1. ESPECIES INCLUIDAS EN EL ANÁLISIS	6
3.2. ENSAMBLAJE GENÓMA MITOCONDRIAL DE <i>Drosophila amaguana</i>	6
3.3. ÁRBOL DE DISTANCIA DE LOS GENES MITOCONDRIALES.....	8
4 RESULTADOS	9
4.1 ENSAMBLAJE	9
4.2 COMPARACIÓN GRÁFICA CON GEPARD	9
4.3 IDENTIFICACIÓN DE LOS GENES MITOCONDRIALES CON MITOS	9
4.4 ÁRBOL DE DISTANCIA DE GENES	10
5. DISCUSIÓN.....	11
6. CONCLUSIONES	15
7. REFERENCIAS BIBLIOGRÁFICAS	16
8. FIGURAS, TABLAS Y ANEXOS	20
8.1. FIGURAS	20
8.2. TABLAS	23
8.3. ANEXOS	25

ÍNDICE DE FIGURAS

FIGURA 1. DIAGRAMA DE FLUJO DEL ENSAMBLAJE	20
FIGURA 2. ALINEAMIENTO GRAFICO DE LOS GENOMAS MITOCONDRIALES. 20	20
FIGURA 3. GENES MITOCONDRIALES.....	211
FIGURA 4. ALINEAMIENTO DE LOS GENOMAS MITOCONDRIALES.....	211
FIGURA 5. ÁRBOL DE DISTANCIA DE GENES.....	222

ÍNDICE DE TABLAS

TABLA 1. TABLA DE GENES MITOCONDRIALES.....	23
TABLA 2. TABLA DE ESPECIES DE <i>Drosophila</i>	24

ÍNDICE DE ANEXOS

ANEXO 1. A) GENOMA MITOCONDRIAL DE <i>D. amaguana</i> ENSAMBLADO: ...	25
B) LOS 37 GENES MITOCONDRIALES DEL GENOMA ENSAMBLADO DE <i>D. amaguana</i> EN FAA FILE:	25
ANEXO 2. WEB SERVER MITOS, GENES MITOCONDRIALES DE <i>D. amaguana</i> :	25

RESUMEN

Drosophila amaguana, es una especie andina perteneciente al grupo *mesophragmatica*, que a su vez pertenece a la radiación virilis-repleta del subgénero *Drosophila* esta especie es de gran tamaño y se caracteriza por habitar bosques andinos con vegetación alta, baja luminosidad y mucha humedad. *D. amaguana* es un organismo de interés, debido a que tiene uno de los genomas mitocondriales más grandes del género *Drosophila*, se estima que el tamaño de su genoma podría alcanzar unas 600 Mb (datos no publicados). En este estudio se realizó el ensamblaje del genoma mitocondrial de *Drosophila amaguana*. La información de este genoma es importante para esclarecer las relaciones filogenéticas entre las especies del grupo *mesophragmatica*. Utilizando los datos de secuenciación illumina, mediante herramientas bioinformáticas como NOVOPlasty 4.3, bowtie2, IGV, BBMap y GEPARD, se ensambló y caracterizó el genoma mitocondrial de *D. amaguana*. Además, se comparó con el genoma de la especie de referencia *Drosophila melanogaster*. Los resultados demostraron que existe diferencias de tamaño en los genomas mitocondriales de *D. melanogaster* 19523 pb y *D. amaguana* 15078 pb. Además de encontrarse que *D. melanogaster* tiene muchos satélites en el bloque final de su genoma mitocondrial. A través del web server MITOS, se identificaron los genes mitocondriales del genoma de *D. amaguana*: 37 genes (2 ARNr, 22 ARNt y 13 polipéptidos implicados en la fosforilación oxidativa), número similar a los genes mitocondriales encontrados en otras especies de este género. Adicionalmente con los genes mitocondriales *COX 1* y *COX 2* se construyó en árbol de distancia genética con especies neotropicales.

ABSTRACT

Drosophila amaguana is an Andean species belonging to the *mesophragmatica* group, which in turn belongs to the virilis-repleta radiation of the subgenus *Drosophila*, this species is large and is characterized by inhabiting Andean forests with high vegetation, low luminosity and high humidity. *D. amaguana* is an organism of interest because it has one of the largest mitochondrial genomes of the genus *Drosophila*, it is estimated that the size of its genome could reach 600 Mb (unpublished data). In this study, the assembly of the mitochondrial genome of *Drosophila amaguana* was performed. The information from this genome is important for elucidating the phylogenetic relationships among species of the *mesophragmatica* group. Using illumina sequencing data, using bioinformatics tools such as NOVOPlasty 4.3, bowtie2, IGV, BBMap and GEPARD, the mitochondrial genome of *D. amaguana* was assembled and characterized. In addition, it was compared with the genome of the reference species *Drosophila melanogaster*. The results showed that there are size differences in the mitochondrial genomes of *D. melanogaster* 19523 bp and *D. amaguana* 15078 bp. In addition, it was found that *D. melanogaster* has many satellites in the final block of its mitochondrial genome. Through the MITOS web server, the mitochondrial genes of the *D. amaguana* genome were identified: 37 genes (2 rRNAs, 22 tRNAs and 13 polypeptides involved in oxidative phosphorylation), a number similar to the mitochondrial genes found in other species of this genus. Additionally, with the mitochondrial genes *COX 1* and *COX 2*, a genetic distance tree with neotropical species was constructed.

1. INTRODUCCIÓN

La familia Drosophilidae se encuentra conformada por dos subfamilias: Drosophilinae y Steganinae. La subfamilia Drosophilinae, cuyas especies son empleados con frecuencia en estudios moleculares, es considerada la más abundante en términos de géneros (Remsen & O'Grady, 2002). La subfamilia Steganinae, es monofilética, es decir, este grupo incluye a un mismo ancestro en común, y hasta la fecha su taxonomía y filogenia ha sido estudiada de manera superficial, especialmente de aquellas especies distribuidas en los Andes, además, muchos aspectos como su historia evolutiva, la sistemática y genética de esta subfamilia se desconocen (Dias et al., 2020; Otranto et al., 2008). Las especies del género *Drosophila* (Drosophilidae) son organismos modelo ampliamente utilizados en las ciencias de la vida como la biología del desarrollo, biomedicina, genética molecular, biotecnología, ecología, biogeografía, entre otras, por su diversidad y distribución en diferentes hábitats alrededor del planeta (Remsen & O'Grady, 2002).

Estudios realizados en Ecuador, revelan que es uno de los países biológicamente más diversos del planeta, debido a la incidencia de la Cordillera de los Andes en su geografía, puesto que su gradiente altitudinal y factores meteorológicos recrean gran variabilidad de ecosistemas y microclimas (Larsen et al., 2011; Rahbek et al., 2019), estos factores contribuyen en la variabilidad genética entre las especies. Becker en 1919, realizó el primer registro de una especie del género *Drosophila* en el Ecuador; habiéndose descrito en los últimos 20 años un gran número de nuevas especies (Figuroa & Rafael, 2017). El grupo *mesophragmatica*, que está conformado por 19 especies y más de la mitad se localiza en el Ecuador (Figuroa, 2017), sin embargo, posee relaciones filogenéticas ambiguas hasta la fecha.

Drosophila amaguana, es una especie andina perteneciente al grupo *mesophragmatica*, que a su vez pertenece a la radiación *virilis-repleta* del subgénero *Drosophila* (Figuroa, 2017; Saletti-cuesta et al., 2020; Vela &

Rafael, 2004), esta especie es de gran tamaño y se caracteriza por habitar bosques andinos con vegetación alta, baja luminosidad y mucha humedad (Céspedes & Rafael, 2017). *D. amaguana* es un organismo de interés, debido a que tiene uno de los genomas más grandes del género *Drosophila*, se estima que el tamaño del genoma mitocondrial de esta especie podría alcanzar unas 600 Mb (datos aún no publicados). No existe información sobre el genoma mitocondrial de *Drosophila amaguana*, por lo que resulta fundamental iniciar con el análisis de este genoma con el propósito de contribuir en la generación de información que permita esclarecer las relaciones filogenéticas entre las especies del grupo *mesophragmatica* y describir el genoma mitocondrial de *Drosophila amaguana*.

El genoma mitocondrial de todos los organismos eucariotas tiene características particulares como por ejemplo la herencia materna y la heteroplasmia. Esta molécula se caracteriza por tener un alto número de copias en cada célula y una alta tasa de mutación en comparación con el genoma nuclear, además de su papel fundamental en el metabolismo aeróbico asociado a la respiración celular y la generación de energía (Montooth et al., 2009).

Se conoce que el tamaño del genoma mitocondrial de *Drosophila melanogaster* varía entre 19 532 y 19 537 pares de bases (pb) de longitud (Wolff et al., 2016), presentando 37 genes (2 ARNr, 22 ARNt y 13 polipéptidos implicados en la fosforilación oxidativa), además, el contenido genómico y estructural es similar al de otros genomas mitocondriales de otras especies de este género (Zhang & Jin, 2020a). Algunos genes mitocondriales relevantes son el *COX 1* y el gen *COX 2*, los cuales forman parte de la enzima citocromo c oxidasa, enzima de múltiples subunidades y eslabón final en la cadena de transporte de electrones mitocondriales de la membrana interna, impulsando la fosforilación oxidativa (Beckenbach & Borkent, 2003; Otranto et al., 2008).

Esta investigación pretende ensamblar y caracterizar el genoma mitocondrial de *Drosophila amaguana*, generar un árbol de distancia de genes de las especies de *Drosophila* (*D. acanthoptera*, *D. nannopectera*, *D. macroptera*, *D. ornatipennis*, *D. incompta*, *D. mercatorum*, *D. canalinea*) a partir de los genes mitocondriales *COX 1* y *COX 2* utilizando herramientas bioinformáticas como

NOVOplasty, GUEPARD, FigTree & IGV (Cummings, 2004b; Dierckxsens et al., 2017; Krumsiek et al., 2007b; Thorvaldsdóttir et al., 2013a).

2. OBJETIVOS

2.1. 2.1. OBJETIVO GENERAL

- Ensamblar y analizar la composición del genoma mitocondrial de la especie *Drosophila amaguana*.

2.2. OBJETIVOS ESPECÍFICOS

- Ensamblar el genoma mitocondrial de *Drosophila amaguana*, a partir de datos de secuenciación Illumina.
 - Elaborar un árbol de distancia genética basado en los genes *COX 1* y *COX 2*, con diez especies de *Drosophila* (*D. acanthoptera*, *D. nannoptera*, *D. macroptera*, *D. ornatipennis*, *D. incompta*, *D. mercatorum*, *D. canalinea*, *D. tripunctata*, *D. amaguana*, *D. melanogaster*).
 - Comparar la secuencia del genoma mitocondrial entre *Drosophila amaguana* y *Drosophila melanogaster*.
 - Identificar los genes del genoma mitocondrial de *Drosophila amaguana*.

3. MATERIALES Y MÉTODOS

3.1. ESPECIES INCLUIDAS EN EL ANÁLISIS

Se analizaron dos especies, *Drosophila amaguana*, cuyas secuencias fueron obtenidas mediante secuenciación Illumina en investigaciones previas, y *Drosophila melanogaster* cuyo genoma se obtuvo de las bases de datos del Centro Nacional de Información de Biotecnología de los Estados Unidos de Norteamérica (NCBI), (número de acceso de GenBank Nc_024511.2)

3.2 ENSAMBLAJE GENÓMA MITOCONDRIAL DE *Drosophila amaguana*

Para el ensamblaje del genoma mitocondrial completo de la especie de interés, se requiere seguir al menos tres fases: obtener la secuencia, alineamiento con depuración y el ensamblaje. La secuencia se obtuvo mediante Illumina Inc. y se usó como referencia el genoma mitocondrial de *D. melanogaster* para su uso en NOVOplasty. El protocolo de ensamblaje se detalla en el diagrama de flujo que se ilustra en la Figura 1.

El archivo que contiene el genoma mitocondrial completo de la especie en estudio, proporcionado por Illumina Inc. es provisto en formato “.fastq”. El ensamblaje requiere que este archivo sea reestructurado en formato “.sam” en el terminal del sistema operativo Ubuntu LTS 18.04.06, empleando el comando “bowtie2-build” a fin de que el programa Bowtie 2 pueda alinear las lecturas de secuenciación usando como referencia la secuencia de igual manera previamente formateada del genoma mitocondrial de *D. melanogaster*. Según Dierckxsens et al. (2017) el ensamblaje y caracterización del genoma mitocondrial debe cumplir con las siguientes fases:

La primera fase consistió en el mapeo de las secuencias archivo mediante bowtie2, que analizó segmentos de genoma cada 8 millones de pb de lectura forward y 8 millones de pb de lectura reverse, en el genoma mitocondrial de *D. melanogaster* empleada como referencia; la velocidad de ejecución del ensamblaje se pudo optimizar reduciendo los datos del genoma no ensamblado, éstos fueron eliminados de la lectura del genoma, el recorte de los segmentos del genoma se obtuvo calculando el número de secuencias ejecutando el comando “zgrep-c” en el terminal del sistema operativo Ubuntu. El genoma recortado, resultado del mapeo, se obtuvo en formato “.sam”, como ya se explicó anteriormente.

La segunda fase consistió en filtrar e identificar del genoma resultante las lecturas que mapean en el mismo sitio de la secuencia, éstas son secuencias repetitivas o idénticas que no son informativas para el ensamblaje, para ello se empleó la herramienta BBMap, que filtro las secuencias del genoma mitocondrial y elimino las secuencias del genoma nuclear, que es un alineador global con reconocimiento de empalme para lecturas de secuenciación de ADN.

Una vez filtradas y eliminadas las lecturas duplicadas, en la tercera fase se ejecutó el ensamblaje empleando el software NOVOplasty (Figura 1), finalizado este procedimiento, se obtuvo una secuencia almacenada en formato “.fasta”, a la cual se le asignó por defecto la identificación: “Uncirculazired_assemblies_1.Secuencia.fasta”. Finalmente, se determinan los contigs que permiten caracterizar la calidad del ensamblaje.

La comparación genómica mitocondrial o alineamiento, se facilita con la aplicación del software GEPARD, Genome Pair Rapid Dotter, esta herramienta permite el cálculo de diagramas de puntos incluso para secuencias grandes como cromosomas o genomas bacterianos (Krumstiek et al., 2007a). La comparación local de dos secuencias de nucleótidos se puede sintetizar en gráficas comparativas entre genomas mitocondriales, en este caso, los pertenecientes a *D. amaguana* y *D. melanogaster*, este procedimiento permite esclarecer las diferencias estructurales entre genomas comparados. La comprobación del ensamblaje se realizó mediante la aplicación del software Integrative Genomics Viewer (IGV), herramienta interactiva de alto rendimiento y fácil aplicación para la exploración visual de datos genómicos, cuyo entorno de programación se

desarrolla en JavaScript protocolo descrito por Thorvaldsdóttir et al. (2013). Que facilita el alineamiento con mejor detalle a nivel de nucleótidos. Para la identificación de los genes del genoma mitocondrial se empleó el Webserver “MITOS” (Tabla 1). URL: <http://mitos.bioinf.uni-leipzig.de/index.py>

3.3 **ÁRBOL DE DISTANCIA DE LOS GENES MITOCONDRIALES.**

Se obtuvieron las secuencias nucleotídicas del gen mitocondrial *COX 1* y *COX 2*, de diez especies del género *Drosophila*, *D. melanogaster*, *D. tripunctata*, *D. acanthoptera*, *D. nannoptera*, *D. macroptera*, *D. ornatipennis*, *D. incompta*, *D. mercatorum*, *D. canalinea*, mediante las bases de datos del Centro Nacional de Información de Biotecnología de los Estados Unidos de Norteamérica (NCBI), el Archivo Europeo de Nucléotidos (ENA) y UniProt como se observa en la (Tabla 2). Para *D. amaguana* se usó los genes del ensamblaje; las secuencias conseguidas se alinearon mediante el algoritmo “muscle”.

La depuración del archivo y la selección del modelo de árbol filogenético de máxima verisimilitud se realizó con el Webserver Gblock. Finalmente, para la modelización del árbol filogenético se empleó el software libre FigTree como recomienda Cummings (2004).

4 RESULTADOS

4.1 ENSAMBLAJE

Se ensambló y alineó el genoma mitocondrial de *D. amaguana* mediante la utilización de la herramienta informática NOVOplasty.

Del análisis realizado en Bowtie 2 se determinó que el 100 % de las lecturas están pareadas (8 000 000 de pb), el 99,85 % de pb no se pudieron alinear con el genoma (7 988 365 pb), y 11 538 pb, es decir el 0,14 % se pudieron alinear al menos una vez (97 pb), por lo tanto no se identificaron irregularidades en el proceso de ensamblaje y se determinó alta calidad en la secuencia, por lo que se puede considerar apto para el alineamiento, puesto que el análisis ejecutado por NOVOplasty determinó la existencia de un conting.

4.2 COMPARACIÓN GRÁFICA CON GEPARD

Con el software GEPARD se alineó gráficamente el genoma mitocondrial ensamblado de *Drosophila amaguana* con el de *Drosophila melanogaster* como referencia, facilitándose así la comparación local de las secuencias de nucleótidos. En la Figura 2, se observa una diagonal de línea entrecortada que representa falta de alineamiento en dichas zonas.

4.3 IDENTIFICACIÓN DE LOS GENES MITOCONDRIALES CON MITOS

Se realizó la identificación de genes que codifican proteínas del genoma mitocondrial detallando sus posiciones, nombres, secuencias, longitudes, calidad y estructura mediante el Webserver MITOS. Se hallaron 37 genes, de los cuales 13 codifican ARNs mensajeros, equivalentes a la síntesis de 13 proteínas, 22 genes que codifican para tARNs y 2 genes que codifican rRNAs mitocondriales (Tabla 1)

4.4 ÁRBOL DE DISTANCIA DE GENES

La Figura 8 comprende el árbol de distancia de genes de máxima verosimilitud de las especies *D. acanthoptera*, *D. nannoptera*, *D. acróptera*, *D. ornatipennis*, *D. incompta*, *D. mercatorum*, *D. canalinea*, *D. amaguana*, *D. tripunctata*, *D. melanogaster*; con los genes *COX 1* y *COX 2*.

5. DISCUSIÓN

En los archivos crudos secuenciado por Illumina viene unido tanto genoma mitocondrial como el genoma nuclear, motivo por el cual la cantidad de secuencias del archivo “.fastq” en bruto es inmensa, a razón de esto y para que corran de manera adecuada los softwares, se recortó el número de secuencias del archivo, utilizándose 8 millones de secuencias. Se hizo esto ya que, existen varias copias del genoma mitocondrial en una célula, adicionalmente, de que el ensamblaje corresponde a un genoma mitocondrial, que es una molécula pequeña con menos secuencias que un genoma nuclear al no poseer regiones repetidas, asimismo estos 8 millones de secuencias son las comúnmente utilizadas para genomas mitocondriales, variando generalmente el número de secuencias de entre 8 millones a 16 millones como referencia (Bernt et al., 2013; Kefi-Ben Atig et al., n.d.; Zhang & Jin, 2020)

Se optó por el software NOVOplasty idóneo para códigos genéticos cortos. No se encontraron irregularidades en el proceso de ensamblaje puesto que se determinó la existencia de un conting, lo que indica una alta calidad en la secuencia ensamblada por lo que se puede circularizar el genoma (Dierckxsens et al., 2017; Zhang & Jin, 2020). De igual manera, los resultados del software Bowtie 2 aportaron información del genoma mitocondrial ensamblado, como que el porcentaje de lecturas pareadas en el ensamblaje es del 100 %; y que el 99,85 % de pb no se alinearon, resultando esto debido a que este porcentaje corresponde a las secuencias nucleares, en cambio el 0,14% de pb que se alinearon correctamente se atribuye a las secuencias mitocondriales; esta diferencia en los porcentajes a su vez representa la diferencia natural que existe en el tamaño de los genomas nucleares y mitocondriales (Ballard & Kreitman, 1994; González L et al., 2016; Langmead & Salzberg, 2012). Una vez se corrobora la buena calidad del ensamblaje, el genoma mitocondrial de *D. amaguana* está listo para ser alineado y comparado con el de otras especies.

De igual manera en cuanto a la cantidad y calidad de los genes, por medio del server MITOS se obtuvieron los 37 genes mitocondriales junto con lo que estos transcriben y la dirección de su lectura, constando de 13 genes codificantes

de proteínas, 22 genes de ARNt y 2 genes de ARNr; la organización genética de *D. amaguana* es idéntica a la disposición que se encuentra en la mayoría de los insectos (Bernt et al., 2013; González L et al., 2016; Montooth et al., 2009b; Zhang & Jin, 2020).

La Tabla 1 indica los 37 genes junto con sus respectivos codónes de inicio y de stop, muestra también el tamaño de estos. Los genes importantes utilizados para el árbol de distancia genética son el *COX 1* y el *COX 2* (citocromo oxidasa) (Figura 5). Estos genes son clave para la fosforilación oxidativa y para otros procesos celulares. No obstante, su principal función recae en su utilidad para estudios filogenéticos. Como se aprecia en la figura 5 al estar representados con flechas rojas se indica que producen proteínas (Beckenbach & Borkent, 2003; Otranto et al., 2008; Remsen & O'Grady, 2002).

El tamaño del genoma mitocondrial de *D. amaguana* es grande para su especie, teniendo un total de 15078 pares de bases (pb) reflejado en la Figura 4. Como referencia el tamaño del genoma mitocondrial de *D. melanogaster* varía entre 19 532 y 19 537 pares (pb) de longitud (Wolff et al., 2015). En la Figura 4 se halla el alineamiento gráfico de los genomas mitocondriales, la línea representa la homología en los nucleótidos de ambos genomas, el genoma de *D. melanogaster* es más grande por lo que se observa una diagonal entrecortada, representando la falta de alineamiento en dichas zonas (Krumtsiek et al., 2007). En caso de que la diagonal fuera continua, significaría una mayor similitud entre ambas moléculas, a razón de esta diferencia de tamaño, la parte final de la diagonal y el alineamiento están incompletos; los puntos y manchas negras visibles al final del alineamiento son satélites y repeticiones del genoma más grande, el de *D. melanogaster* (Krumtsiek et al., 2007; Montooth et al., 2009).

También se optó por utilizar otra herramienta bioinformática para analizar más a detalle la parte final del par de genomas, corroborándose en el proceso la calidad del ensamblaje, esto se logró mediante la aplicación del software Integrative Genomics Viewer (IGV), herramienta que permite ver el alineamiento más a detalle a nivel de nucleótidos, para identificar las bases

nitrogenadas homólogas entre ambas especies (Thorvaldsdóttir et al., 2013a). Con el fin de esclarecer si la falta de alineamiento de la zona terminal entre ambos genomas mitocondriales consiste en un error del procedimiento, o si resulta de una diferencia propia de las especies, siendo este último caso el correcto. Analizándose la zona final del alineamiento se encontró como resultado a los mismos satélites vistos con el programa GUEPARD (Krumnsiek et al., 2007b).

Por otro lado, cada nucleótido es un carácter para el árbol de distancia genes, permitiendo esto identificar la similitud y diferencia de las bases nitrogenadas en los genes *COX1* y *COX2* de las 10 *Drosophilas* seleccionadas (Beckenbach & Borkent, 2003; Otranto et al., 2008). Nueve de estos diez insectos son neotropicales y más de la mitad de las especies se encuentran en el Ecuador, los mismos que poseen relaciones filogenéticas que siguen sin estar claras del todo hasta la fecha (Figuero & Rafael, 2013; Loma et al., 1984; Vela & Rafael, 2004). No obstante, los resultados obtenidos no pretenden trazar las ramas de un árbol filogenético completo ni la historia evolutiva de las especies, puesto que para el estudio solo se utilizó de un par de genes y no de genomas mitocondriales completos. Sin embargo, se puede ver la distancia genética entre las especies y por ende la homología a nivel nucleotídico de los genes seleccionados en los insectos del análisis, mostrando que, estas moscas comparten en común la misma ubicación de sus bases nitrogenadas que sus antepasados (Ayala et al., 1974; Zhang & Jin, 2020c).

En la filogenia se pueden identificar ocho eventos de cladogénesis en las especies. La especie de interés *D. amaguana* (que conforma parte del grupo *mesophragmatica*) pertenece a la misma rama filogenética de *D. canalinea* al no haber una distancia genética alta de los genes *COX 1* y *COX 2* en estos insectos. Así mismo, especies con las que *D. amaguana* mantiene bases similares, pero con mayor distancia genética son *D. melanogaster*, *D. incompta*, *D. mercatorum*; otras especies que integran una rama filogenética cercana son *D. acanthoptera*, *D. nannoptera*. Por otro lado, las moscas que conforma una rama alejada del nodo de *D. amaguana* son: *D. ornatipennis*, *D. tripunctat*; finalmente *D. macroptera*, es igualmente representada perteneciendo a una de las ramas más alejadas del centro del árbol (de Vienne, 2016; O'Grady & DeSalle, 2018).

En contraste con estos resultados, la literatura basada en el web server “lifemap” de NCBI, nos indica que, *D. nannoptera* y *D. acanthoptera* están relacionadas al pertenecer ambas al grupo nanoptera, *D. mercatorum* pertenece al grupo repleta, *D. tripunctata* y *D. incompta* están agrupadas al formar parte del grupo *flavopilosa*, *D. canalinea* pertenece al grupo *canalinea*, *D. ornatipennis* pertenece al grupo (calloptera), *D. macróptera* pertenece al grupo marcoptera, y *D. melanogaster* pertenece al grupo melanogaster. Como se evidencia en la bibliografía las agrupaciones resultantes de los genes *COX 1* y *COX 2* contrastan, con las que las se encuentran en la literatura, siendo esta diferencia algo esperada por la variación en la cantidad de información que aporta uno o dos genes comparados a todo un genoma completo (Brncic & Santibanez, 1957; de Vienne, 2016; Pruitt et al., 2005).

6. CONCLUSIONES

En conclusión, *D. amaguana* tiene 37 genes mitocondriales; la diferencia de tamaño de los genomas mitocondriales de *D. amaguana* y *D. melanogaster* se evidencio en el alineamiento, teniendo *D. melanogaster* 19523 pb y *D. amaguana* 15078 pb.

El árbol de distancia de genes permitió identificar que las especies *D. amaguana* y *D. canalinea* comparten una mayor cantidad de nucleótidos homólogos basado en los genes de *COX 1* y *COX 2*.

7. REFERENCIAS BIBLIOGRÁFICAS

- Atig, R. K. Ben, Hsouna, S., Beraud-Colomb, E., & Abdelhak, S. (2009). Mitochondrial DNA: properties and applications. *Archives de l'Institut Pasteur de Tunis*, 86(1–4).
- Ayala, F. J., Tracey, M. L., Hedgecock, D., & Richmond, R. C. (1974). Genetic Differentiation During the Speciation Process in *Drosophila*. *Evolution*, 28(4), 576. <https://doi.org/10.2307/2407283>
- Ballard, J. W. O., & Kreitman, M. (1994). Unraveling selection in the mitochondrial genome of *Drosophila*. *Genetics*, 138(3). <https://doi.org/10.1093/genetics/138.3.757>
- Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., da Silva, A., de Giorgi, M., Dogan, T., Fazzini, F., ... Zhang, J. (2015). UniProt: a hub for protein information. *Nucleic Acids Research*, 43(D1), D204–D212. <https://doi.org/10.1093/NAR/GKU989>
- Beckenbach, A. T., & Borkent, A. (2003). Molecular analysis of the biting midges (Diptera: Ceratopogonidae), based on mitochondrial cytochrome oxidase subunit 2. *Molecular Phylogenetics and Evolution*, 27(1), 21–35. [https://doi.org/10.1016/S1055-7903\(02\)00395-0](https://doi.org/10.1016/S1055-7903(02)00395-0)
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2), 313–319. <https://doi.org/10.1016/J.YMPEV.2012.08.023>
- Bofarull, B. N. de. (2005). Caracterización Genómica y Funcional de las Reorganizaciones del Complejo de Genes Hox en *Drosophila*. Genomic and Functional Characterization of Hox Gene Complex Rearrangements in *Drosophila*. *Nature*, 255.
- Brcic, D., & Santibanez, S. K. (1957). The Mesophragmatica Group of Species of *Drosophila*. *Evolution*, 11(3), 300. <https://doi.org/10.2307/2405794>
- Céspedes, D., & Rafael, V. (2017). Diversidad del género *Drosophila* (Diptera, Drosophilidae) en la quebrada de Cruz Loma, Pichincha, Ecuador. *Revista Ecuatoriana de Medicina y Ciencias Biológicas*, 34(1–2), 215–221. <https://doi.org/10.26807/remcb.v34i1-2.245>
- Cummings, M. P. (2004). FigTree. In *Dictionary of Bioinformatics and Computational Biology*. <https://doi.org/10.1002/9780471650126.dob0904>
- de Vienne, D. M. (2016). Lifemap: Exploring the Entire Tree of Life. *PLoS Biology*, 14(12). <https://doi.org/10.1371/JOURNAL.PBIO.2001624>
- Dias, G. R., Dupim, E. G., Vanderlinde, T., Mello, B., & Carvalho, A. B. (2020). A phylogenomic study of Steganinae fruit flies (Diptera: Drosophilidae): strong gene

- tree heterogeneity and evidence for monophyly. *BMC Evolutionary Biology*, 20(1).
<https://doi.org/10.1186/s12862-020-01703-7>
- Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, 45(4), e18–e18.
<https://doi.org/10.1093/NAR/GKW955>
- Figuroa, M. (2017). *Filogenia molecular de especies ecuatorianas del grupo Drosophila mesophragmatica (Diptera, Drosophilidae)*.
- Figuro, M. L., & Rafael, V. (2013). Diversidad del género *Drosophila* (Diptera, Drosophilidae) en el páramo de Papallacta, Pichincha, Ecuador. *Revista Ecuatoriana de Medicina y Ciencias Biológicas*, 34(1–2), 151–165.
<https://doi.org/10.26807/REMCB.V34I1-2.241>
- Figuro, M. L., & Rafael, V. (2017). Diversidad del género *Drosophila* (Diptera, Drosophilidae) en el páramo de Papallacta, Pichincha, Ecuador. *Revista Ecuatoriana de Medicina y Ciencias Biológicas*, 34(1–2), 151–165.
<https://doi.org/10.26807/remcb.v34i1-2.241>
- García Mejjide, J. A., & Gómez-Reino Carnota, J. J. (2000). Fisiopatología de la ciclooxigenasa-1 y ciclooxigenasa-2. *Revista Espanola de Reumatologia*, 27(1).
- González L, C. D., Castillo D, C. P., Angiolillo R, G., Fernández S, D. J., & Ascanio Rojas, A. (2016). Herramientas bioinformáticas para el análisis de secuencias en el Instituto Nacional de Higiene “Rafael Rangel”. *Revista Del Instituto Nacional de Higiene Rafael Rangel*, 47(1–2), 143–149.
http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0798-04772016000100011&lng=es&nrm=iso&tlng=es
- Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., van den Broek, A., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Gamble, J., Garcia Diez, F., Harte, N., Kulikova, T., Lin, Q., Lombard, V., ... Apweiler, R. (2005). The EMBL Nucleotide Sequence Database. *Nucleic Acids Research*, 33(suppl_1), D29–D33. <https://doi.org/10.1093/NAR/GKI098>
- Kefi-Ben Atig, R., Hsouna, S., Beraud-Colomb, E., & Abdelhak, E. S. (n.d.). *ADN MITOCHONDRIAL : PROPRIETES ET APPLICATIONS*.
- Krumsiek, J., Arnold, R., & Rattei, T. (2007). Gepard: A rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*, 23(8).
<https://doi.org/10.1093/bioinformatics/btm039>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012 9:4, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Larsen, T. H., Escobar, F., & Armbrrecht, I. (2011). Insects of the tropical Andes: diversity patterns, processes and global change. In *Climate Change and Biodiversity in the Tropical Andes*.

- Lietti, M., Gramajo, M., Budai, N., Mondino, M., & Balaban, D. (2016). Moscas de las riberas (Diptera: Ephydriidae) asociadas a plantines de lechuga en invernaderos. *Agromensajes*, 45(Agosto), 53–60.
- Loma, C., Céspedes, D., Rafael, V., & Ecuador, D. (1984). Diversidad del género *Drosophila* (Diptera, Drosophilidae) en la quebrada de Cruz Loma, Pichincha, Ecuador. *Revista Ecuatoriana de Medicina y Ciencias Biológicas: REMCB, ISSN-e 2477-9148, ISSN 2477-9113, Vol. 20, N.º. 2, 1984, 20(2)*, 114.
<https://dialnet.unirioja.es/servlet/articulo?codigo=8575456&info=resumen&idioma=SPA>
- Montooth, K. L., Abt, D. N., Hofmann, J. W., & Rand, D. M. (2009). Comparative genomics of drosophila mtDNA: Novel features of conservation and change across functional domains and lineages. *Journal of Molecular Evolution*, 69(1).
<https://doi.org/10.1007/s00239-009-9255-0>
- Negre de Bofarull, B. (2006). *Caracterización genómica y funcional de las reorganizaciones del complejo de genes hox en Drosophila*.
<https://ddd.uab.cat/record/38321>
- O'Grady, P. M., & DeSalle, R. (2018). Phylogeny of the Genus *Drosophila*. *Genetics*, 209(1), 1–25. <https://doi.org/10.1534/GENETICS.117.300583>
- Otranto, D., Stevens, J. R., Testini, G., Cantacessi, C., & MácA, J. (2008). Molecular characterization and phylogenesis of Steganinae (Diptera, Drosophilidae) inferred by the mitochondrial cytochrome c oxidase subunit 1. *Medical and Veterinary Entomology*, 22(1). <https://doi.org/10.1111/j.1365-2915.2008.00714.x>
- Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 33(suppl_1), D501–D504.
<https://doi.org/10.1093/NAR/GKI025>
- Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalgaard, B., Holt, B. G., Morueta-Holme, N., Nogues-Bravo, D., Whittaker, R. J., & Fjeldså, J. (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity? *Science*, 365(6458).
<https://doi.org/10.1126/science.aax0149>
- Remsen, J., & O'Grady, P. (2002). Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. *Molecular Phylogenetics and Evolution*, 24(2). [https://doi.org/10.1016/S1055-7903\(02\)00226-9](https://doi.org/10.1016/S1055-7903(02)00226-9)
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/BIB/BBS017>
- Vela, D., & Rafael, V. (2004). Three new andean species of *Drosophila* (Diptera, Drosophilidae) of the mesophragmatica group. *Iheringia. Série Zoologia*, 94(3).
<https://doi.org/10.1590/s0073-47212004000300012>
- Wolff, J. N., Camus, M. F., Clancy, D. J., & Dowling, D. K. (2015). Complete mitochondrial genome sequences of thirteen globally sourced strains of fruit fly

(*Drosophila melanogaster*) form a powerful model for mitochondrial research. <https://doi.org/10.3109/19401736.2015.1106496>, 27(6), 4672–4674.
<https://doi.org/10.3109/19401736.2015.1106496>

Wolff, J. N., Camus, M. F., Clancy, D. J., & Dowling, D. K. (2016). Complete mitochondrial genome sequences of thirteen globally sourced strains of fruit fly (*Drosophila melanogaster*) form a powerful model for mitochondrial research. *Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis*, 27(6). <https://doi.org/10.3109/19401736.2015.1106496>

Yuzuki, K., & Tidon, R. (2020). Identification key for drosophilid species (Diptera, Drosophilidae) exotic to the Neotropical Region and occurring in Brazil. *Revista Brasileira de Entomologia*, 64(1), 1–9. <https://doi.org/10.1590/1806-9665-RBENT-2019-100>

Zhang, C., & Jin, J. (2020). Characterization of the complete mitochondrial genome of *Drosophila busckii* (Diptera: Drosophilidae). *Mitochondrial DNA Part B: Resources*, 5(3), 2188–2190. <https://doi.org/10.1080/23802359.2020.1768959>

8. FIGURAS, TABLAS Y ANEXOS

8.1. FIGURAS

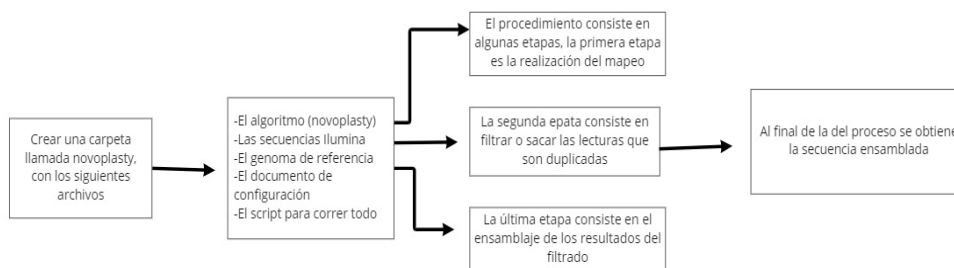


Figura 1. Diagrama de flujo del ensamblaje. Esquema del ensamblaje, mediante el programa NOVOplasty.

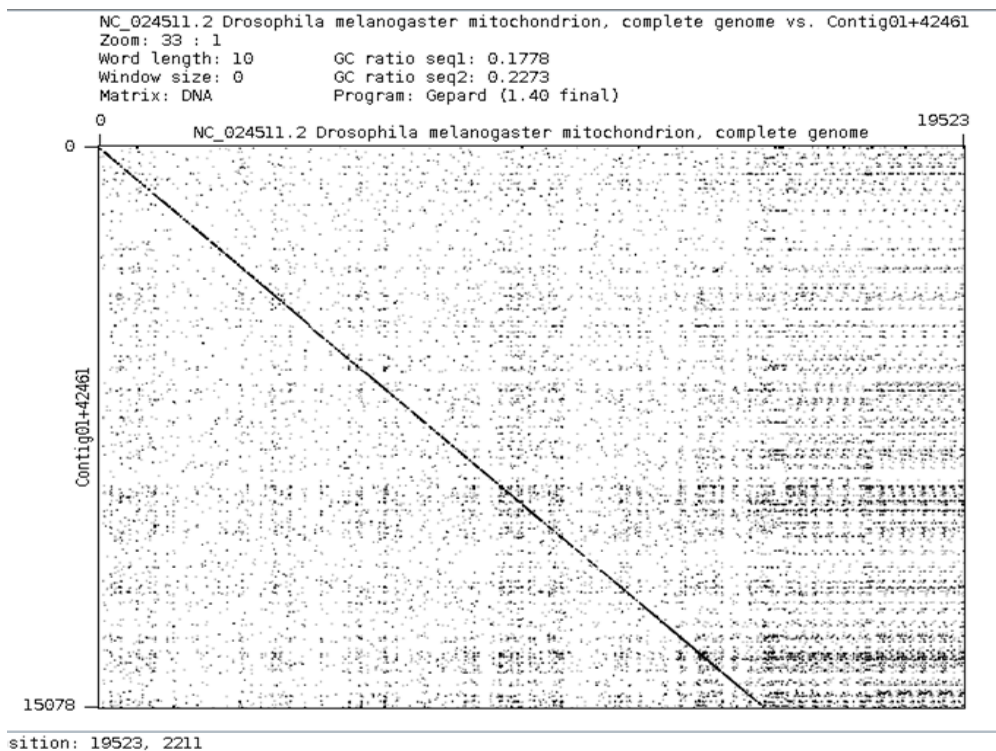


Figura 2. Alineamiento gráfico de los genomas mitocondriales. Alineamiento gráfico de los genomas mitocondriales de *D. melanogaster* y *D. amaguana* con el programa GUEPARD.

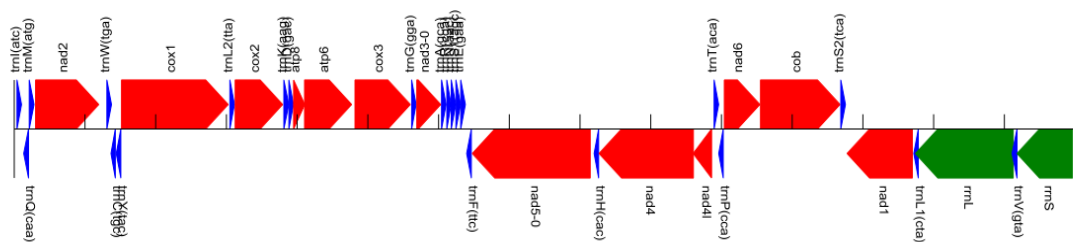


Figura 3. Genes mitocondriales. Genes mitocondriales y la lectura de su dirección, mediante el web server MITOS.



Figura 4. Alineamiento de los genomas mitocondriales. Alineamiento a nivel de nucleótidos los genomas mitocondriales mediante el software IGV.

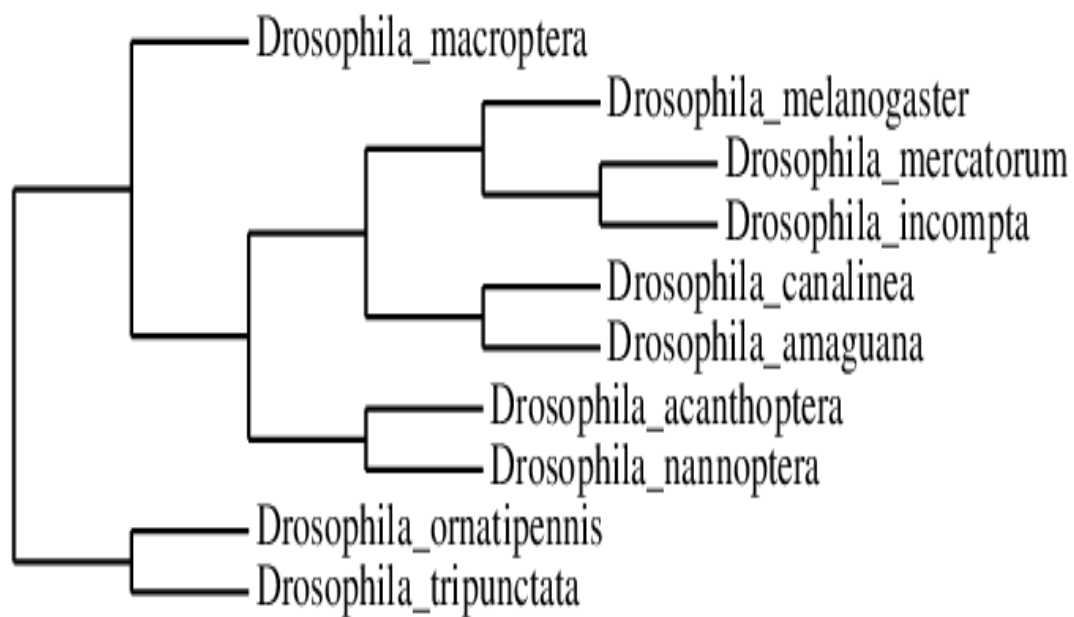


Figura 5. Árbol de distancia de genes. Genes *COX 1* y *COX 2* de las especies *D. acanthoptera*, *D. nannoptera*, *D. 22acróptera*, *D. ornatipennis*, *D. incompta*, *D. mercatorum*, *D. canalinea*, *D. amaguana*, *D. tripunctata*

8.2. TABLAS

Tabla 1. Tabla de genes mitocondriales. Genes mitocondriales de *D. amaguana*

Jobid: Mytos prueba (Contig01+42461)

Name	Start	Stop	Strand	Length	Structure
trnI(atc)	40	104	+	65	svg ps
trnQ(caa)	135	203	-	69	svg ps
trnM(atg)	213	281	+	69	svg ps
nad2	303	1196	+	894	
trnW(tga)	1311	1376	+	66	svg ps
trnC(tgc)	1369	1430	-	62	svg ps
trnY(tac)	1443	1508	-	66	svg ps
cox1	1518	3026	+	1509	
trnL2(tta)	3050	3115	+	66	svg ps
cox2	3125	3796	+	672	
trnK(aag)	3813	3883	+	71	svg ps
trnD(gac)	3884	3950	+	67	svg ps
atp8	3951	4109	+	159	
atp6	4106	4771	+	666	
cox3	4819	5601	+	783	
trnG(gga)	5620	5683	+	64	svg ps
nad3-0	5693	6031	+	339	
trnA(gca)	6043	6107	+	65	svg ps
trnR(cga)	6117	6179	+	63	svg ps
trnN(aac)	6180	6245	+	66	svg ps
trnS1(agg)	6246	6313	+	68	svg ps
trnE(gaa)	6314	6380	+	67	svg ps
trnF(ttc)	6398	6464	-	67	svg ps
nad5-0	6478	8145	-	1668	
trnH(cac)	8200	8264	-	65	svg ps
nad4	8270	9604	-	1335	
nad4l	9601	9861	-	261	
trnT(aca)	9897	9961	+	65	svg ps
trnP(cca)	9962	10026	-	65	svg ps
nad6	10041	10544	+	504	
cob	10553	11677	+	1125	
trnS2(tca)	11688	11754	+	67	svg ps
nad1	11777	12700	-	924	
trnL1(cta)	12720	12784	-	65	svg ps
rrnL	12743	14126	-	1384	svg ps
trnV(gta)	14109	14180	-	72	svg ps
rrnS	14180	14964	-	785	svg ps

■ tRNA gene ■ rRNA gene ■ protein coding gene

Tabla 2. Tabla de las especies de *Drosophila*. Las nueve especies de *Drosophila* con los números de acceso de NCBI de los genes *COX 1* y *COX 2*.

Especies	<i>COX 1</i> (número de acceso)	<i>COX 2</i> (número de acceso)
<i>D. melanogaster</i>	KY310615.1	KY310615.1
<i>D. tripunctata</i>	AB932759.1	AB932804.1
<i>D. acanthoptera</i>	KF632601.1	AF183968.1
<i>D. nanoptera</i>	DQ471531.1	AF478425.1
<i>D. macróptera</i>	EU493597.1	EU493727.1
<i>D. ornatipennis</i>	EF570010.1	EU493704.1
<i>D. incompta</i>	JX993098.1	AY847764.1
<i>D. mercatorum</i>	DQ202068.1	DQ202028.1
<i>D. canalinea</i>	EU493575.1	KX275255.1

8.3. ANEXOS

Anexo 1.

A) Genoma mitocondrial de *D. amaguana* ensamblado:

https://puceeduec-my.sharepoint.com/:w:/g/personal/ngmachado_puce_edu_ec/EULGVythDEVBmfiICNd8S04B5Ss9eudJA6RbxfQaQbID5w?e=Q142QF

B) Los 37 genes mitocondriales del genoma ensamblado de *D. amaguana* en FAA file:

https://puceeduec-my.sharepoint.com/:w:/g/personal/ngmachado_puce_edu_ec/EUP7SPmVUX9AuXPyqTD7VD4BpVihg4dV6KTxlm4W08Y9Sg?e=iH8H70

Anexo 2. Web server MITOS, genes mitocondriales de *D. amaguana*:

<http://mitos.bioinf.uni-leipzig.de/result.py?hash=wVRjdANU>