

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y CREATIVIDAD

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

Análisis genómico de *Drosophila guayllabambae* (Diptera,
Drosophilidae) considerada especie endémica del Ecuador utilizando
biología computacional

Autora: Keyla Monserrath Villacís López

LABORATORIO DE GENÉTICA EVOLUTIVA

Doris Jimena Vela Peralta, PhD.

Directora del Trabajo de Titulación

DEDICATORIA

"One day, you'll leave this world behind

So, live a life you will remember"

Avicii - The Nights

A mi familia,

*A mi mami, **María Mercedes López J.**, por ser mi inspiración, mi mayor apoyo y la mujer más increíble en el mundo.*

*A mi papi, **José Miguel Villacís J.**, por ser mi héroe, mi ejemplo de lucha constante y mi significado de fuerza y valentía.*

*A mi hermana mayor, **Evelin Natalí**, por ser mi cómplice, mi mejor amiga y la persona que me regaló un motivo más para seguir adelante. A mi sobrina **Antonella Monserrath**, mi princesa.*

*A mi hermano menor, **Joseph Steveen**, por ser mi mejor amigo, mi apoyo, mi alegría, mi diversión. Creo firmemente en ti y estoy segura de que puedes alcanzar todo aquello que tu corazón desee.*

Dios los cuide y los proteja siempre.

Los amo con todo mi corazón.

Villacís López Keyla Monserrath

AGRADECIMIENTOS

Agradezco a **Dios** por su bendición, por guiarme y darme la fortaleza para seguir adelante cumpliendo mis sueños.

A mis **padres**, por su amor infinito, esfuerzo y dedicación, porque gracias a ellos soy la mujer que soy. A mis **hermanos**, por su complicidad, motivación y felicidad. A mi **sobrino**, por su ternura e inocencia que me inspira a ser mejor cada día. A **mis abuelitos**, por ser un apoyo incondicional, por su cálido cariño y por ser mis segundos padres, con quienes puedo contar siempre. A **Holguer**, a quien conocí al iniciar la maestría: gracias por tu amor, por tu paciencia y por hacerme tan feliz cada día. Gracias por sacar lo mejor de mí. Te amo.

A **Jess Alba**, por su amistad sincera, apoyo mutuo y por ser una de las mejores personas que he conocido y quien me recordará esta etapa.

A la **Pontificia Universidad Católica del Ecuador**, por su ilustre calidad en educación y profesionales que imparten su conocimiento y profesionales que forman para beneficio de la sociedad. Especialmente, al **Tnlgo. Paúl Guachamin Andrango**, por su apoyo incondicional y, a mi tutora, **Doris Jimena Vela Peralta, PhD**, por darme la oportunidad de crecer como investigadora. Su ejemplo como científica, junto con su orientación experta, han sido de un valor incalculable para mi formación y para el avance de este proyecto, un profundo y eterno agradecimiento.

Dios los bendiga siempre.

Los quiero mucho.

Villacís López Keyla Monserrath

ÍNDICE DE CONTENIDO

DERECHOS DE AUTOR	ii
APROBACIÓN DEL DIRECTOR DEL TRABAJO DE TITULACIÓN	iii
HOJA DE EVIDENCIA ANTIPLAGIO.....	iv
DEDICATORIA.....	v
AGRADECIMIENTOS.....	vi
ÍNDICE DE FIGURAS	ix
ÍNDICE DE TABLAS.....	ix
ÍNDICE DE ANEXOS	ix
RESUMEN	x
ABSTRACT	xi
1. INTRODUCCIÓN	1
1.1. La familia Drosophilidae y su diversidad en Ecuador como parte de la región neotropical.....	1
1.2. <i>D. guayllabambae</i> y el grupo <i>repleta</i>	2
1.3. Información genómica de especies en la región neotropical	4
1.4. Importancia de la biología computacional en estudios genómicos para <i>Drosophila</i>	6
1.5. Elementos Transponibles (TEs) en <i>Drosophila</i>	8
2. OBJETIVOS.....	9
2.1. Objetivo General.....	9
2.2. Objetivos Específicos.....	9
3. METODOLOGÍA	10
3.1. Análisis de la calidad del ensamblaje de <i>novi</i> del genoma de <i>D. guayllabambae</i>	10
3.2. Alineamiento de la estructura del genoma de <i>D. guayllabambae</i> con las especies <i>D. mojavensis</i> , <i>D. melanogaster</i> y <i>D. rucux</i>	11
3.3. Identificación de los componentes del genoma de <i>D. guayllabambae</i>	11

4. RESULTADOS Y DISCUSIÓN.....	13
4.1. Análisis de la calidad de las secuencias y ensamblaje de <i>novο</i> del genoma de <i>D. guayllabambae</i>	13
4.2. Alineamiento de la estructura del genoma de <i>D. guayllabambae</i> con las especies <i>D. mojavensis</i> , <i>D. melanogaster</i> y <i>D. rucux</i>	17
4.3. Componentes del genoma de <i>D. guayllabambae</i> : genes y elementos transponibles (TEs).....	20
5. CONCLUSIONES Y RECOMENDACIONES.....	26
5.1. Conclusiones.....	26
5.2. Recomendaciones	27
6. REFERENCIAS	28
7. ANEXOS.....	38

ÍNDICE DE FIGURAS

Figura 1 Representación esquemática de los resultados de Benchmarking Universal Single-Copy Orthologs (BUSCO) de <i>D. guayllabambae</i>	16
Figura 2 Dotplot del alineamiento del genoma ensamblado de <i>D. guayllabambae</i> vs <i>D. mojavensis</i> , <i>D. melanogaster</i> y <i>D. rucux</i>	19
Figura 3 Estadísticas de AGAT de la predicción de genes realizada por AUGUSTUS para la especie <i>D. guayllabambae</i>	24

ÍNDICE DE TABLAS

Tabla 1 Resultados del procesamiento de las secuencias de <i>D. guayllabambae</i> realizado con BBTools.....	13
Tabla 2 Estadísticas de las lecturas del secuenciamiento de Illumina en <i>D. guayllabambae</i>	14
Tabla 3 Estadísticas de la calidad del ensamblado final de <i>D. guayllabambae</i>	15
Tabla 4 Resultados de los elementos transponibles (ET) del genoma ensamblado de <i>D. guayllabambae</i> generados por RepeatModeler	21
Tabla 5 Resultados de los elementos transponibles (ET) del genoma ensamblado de <i>D. guayllabambae</i> generados por RepeatMasker	22
Tabla 6 Comparación de las características entre el genoma ensamblado de <i>D. guayllabambae</i> frente a los genomas de <i>D. mojavensis</i> , <i>D. melanogaster</i> y <i>D. rucux</i> ...	25

ÍNDICE DE ANEXOS

Figura A1 Contenido de bases en las secuencias de <i>D. guayllabambae</i>	38
--	----

RESUMEN

El género *Drosophila* es un modelo clave para la genética evolutiva, particularmente en la región neotropical, que alberga una biodiversidad excepcional. Este estudio se centra en *Drosophila guayllabambae* (*D. guayllabambae*), una especie endémica de Ecuador perteneciente al grupo *repleta*. El ADN genómico fue secuenciado utilizando la tecnología de secuenciación Illumina, mediante la librería TruSeq DNA Nano y el secuenciador Illumina HiSeq, obteniéndose un ensamblaje del genoma de 160.42 Mb con 9,909 contigs mediante MaSuRCA. La evaluación con BUSCO reveló un 98.90% de ortólogos completos, lo que indica un ensamblaje robusto. Los análisis genómicos comparativos con *D. mojavensis*, *D. melanogaster* y *D. rucux* destacaron tanto regiones conservadas como divergencias significativas, reflejando adaptaciones específicas a ambientes áridos y andinos. El genoma comprende 14,364 genes determinados por AUGUSTUS, 69,628 exones y 48,084 intrones, con un 20.06% de elementos transponibles (TEs) enmascarados por RepeatMasker. Estos hallazgos subrayan la importancia de *D. guayllabambae* como modelo para estudiar procesos evolutivos en especies neotropicales endémicas. Este estudio no solo proporciona un recurso genómico valioso para análisis comparativos, sino que también resalta la necesidad de investigar más a fondo los mecanismos ecológicos y evolutivos que impulsan la biodiversidad en la región. La integración de datos genómicos con estudios ecológicos mejorará las estrategias de conservación para especies endémicas frente a los desafíos ambientales.

Palabras clave: *Drosophila guayllabambae*, grupo *repleta*, ensamblaje de genoma completo, genómica comparativa, componentes genómicos, elementos transponibles, biología computacional.

ABSTRACT

The genus *Drosophila* is a key model for evolutionary genetics, particularly in the Neotropical region, which harbors exceptional biodiversity. This study focuses on *Drosophila guayllabambae* (*D. guayllabambae*), an endemic species from Ecuador belonging to the *repleta* group. Genomic DNA was sequenced using Illumina technology, specifically the TruSeq DNA Nano library and the Illumina HiSeq platform, resulting in a genome assembly of 160.42 Mb with 9,909 contigs using MaSuRCA. Evaluation with BUSCO revealed 98.90% complete orthologs, indicating a robust assembly. Comparative genomic analyses with *D. mojavensis*, *D. melanogaster*, and *D. rucux* highlighted both conserved regions and significant divergences, reflecting specific adaptations to arid and Andean environments. The genome comprises 14,364 genes predicted by AUGUSTUS, 69,628 exons, and 48,084 introns, with 20.06% of the genome masked as transposable elements (TEs) by RepeatMasker. These findings underscore the importance of *D. guayllabambae* as a model for studying evolutionary processes in endemic Neotropical species. This study not only provides a valuable genomic resource for comparative analyses but also emphasizes the need for further research into the ecological and evolutionary mechanisms driving biodiversity in the region. In addition, the integration of genomic data with ecological studies will enhance conservation strategies for endemic species in the face of environmental challenges.

Keywords: *Drosophila guayllabambae*, *repleta* group, whole genome assembly, comparative genomics, genomic components, transposable elements, computational biology.

1. INTRODUCCIÓN

1.1. La familia Drosophilidae y su diversidad en Ecuador como parte de la región neotropical

La región neotropical, compuesta por América del Sur y Central, es un Hot-spot de biodiversidad, particularmente evidente en el género *Drosophila*, que exhibe una notable diversidad adaptativa y evolutiva (Silva *et al.*, 2020). Este género pertenece a la familia Drosophilidae al orden de los dípteros, que se distingue no solo por su vasta cantidad de especies, aproximadamente 3.962 distribuidas en 73 géneros y su distribución cosmopolita, sino también por su rápida radiación, lo cual la convierte en un modelo excelente para estudios de genética evolutiva (O'Grady & DeSalle, 2018; Rafael, 2023a).

La clasificación dentro de esta familia a veces puede ser objeto de debate, no obstante, por lo general se reconoce que se organiza en dos subfamilias principales: Steganinae y Drosophilinae (Brake & Bächli, 2008). Dentro del amplio género *Drosophila*, se distinguen ocho subgéneros, destacando *Sophophora*, que incluye a la conocida *D. melanogaster* y el subgénero *Drosophila*, el más abundante (Figuro, 2017). Además, se identifican dos importantes radiaciones evolutivas en este género: la radiación *immigrans-Hirtodrosophila* y la radiación *virilis-repleta*, ambas originarias de zonas tropicales (Throckmorton, 1975; O'Grady & DeSalle, 2018). En la región neotropical, el subgénero *Drosophila* predomina dentro de la familia, agrupando al menos 46 grupos de especies (O'Grady & DeSalle, 2018).

En Ecuador, que figura entre los países más biodiversos del planeta, la Cordillera de los Andes ha propiciado la formación de numerosos microecosistemas con condiciones ambientales diversas, estos microecosistemas fomentan la diversificación biológica, incluyendo a las especies de Drosophilidae (Bussmann, 2011; Rafael, 2023b). El género *Drosophila*, que constituye aproximadamente el 53% de las especies de la familia, es el más estudiado en el país. En 1987, Arcos y Rafael iniciaron los primeros estudios del género *Drosophila* en Ecuador (Acurio & Rafael, 2009). Estos estudios fueron

realizados en zonas semiáridas, como el valle de Guayllabamba (**Rafael et al., 2000**). Así es como, desde finales de la década de 1980, se han identificado más de 139 especies de este género en el país y la cifra sigue en aumento cada año.

En cada muestreo, realizado en Ecuador, se encuentran numerosos individuos del género *Drosophila*, desconocidas para la ciencia. Por lo que, ha sido indispensable el análisis taxonómico riguroso y muy detallado para determinar la identidad de cada espécimen (**Llangarí et al., 2022**). Varias investigaciones han contribuido al conocimiento de la diversidad del género *Drosophila* en el Ecuador; no obstante, aún no se conoce la totalidad de especies del género (**Acurio & Rafael, 2009**). Sin embargo, esta diversidad subraya la importancia de Ecuador como un laboratorio natural para la investigación en adaptación y evolución de *Drosophila* y su potencial para ofrecer pistas en genética evolutiva.

1.2. *D. guayllabambae* y el grupo *repleta*

Durante más de un siglo, el grupo de especies *repleta* del subgénero *Drosophila* Fallén ha sido utilizado como un sistema modelo en investigaciones sobre adaptación ecológica, evolución y especiación (**Oliveira et al., 2012; Stefanini et al., 2021**). Según la definición de **Throckmorton, (1975)**, este grupo pertenece a la radiación *virilis-repleta*, con 106 especies (**O'Grady & DeSalle, 2018**), el grupo *repleta* es uno de los más diversos dentro del género. Muchas de estas especies son cactófilas, lo que significa que habitan en zonas áridas y desiertos a lo largo de América. Se reconocen seis subgrupos dentro del grupo de especies *repleta*: *fasciola*, *hydei*, *mercatorum*, *repleta*, *mulleri* e *inca*, siendo este último el más reciente en ser definido, que incluye tres especies, todas endémicas de Ecuador y Perú (**Rafael & Arcos, 1989**).

Históricamente, la Región Trans-volcánica de México fue considerada el centro de diversificación del grupo *repleta* (**Patterson & Stone, 1952; Throckmorton, 1975**). Sin embargo, en el estudio realizado por **Oliveira et al., (2012)**, sugieren que el centro de diversificación del grupo *repleta* podría haber sido Sudamérica y estar asociado con la

radiación de sus huéspedes cactus, aunque los autores no proporcionaron soporte estadístico para esta hipótesis. En otro estudio realizado por **Morales & Vieira, (2012)**, se analizaron los patrones de diversificación en el subgénero *Drosophila*, que respalda la monofilia del grupo *repleta*. A pesar de la contribución significativa de estos dos estudios, el origen geográfico del grupo sigue siendo desconocido.

Es importante mencionar que, un factor que ha influido en la comprensión de la diversificación del grupo *repleta* ha sido el sesgo en las localidades muestreadas para las colecciones de *Drosophila* (**Acurio, 2024**). Durante varios años, los esfuerzos para recolectar muestras de *Drosophila* se centraron en regiones de Norteamérica y Centroamérica, con un énfasis en las zonas áridas de México (**Sturtevant, 1921; Patterson & Mainland, 1945; Oliveira et al., 2005**). En contraste, existen relativamente pocas colecciones provenientes de los países andinos de Sudamérica (**Oliveira et al., 2012**). Como resultado, la fauna de *Drosophila* es bien conocida en Norteamérica, mientras que en Sudamérica aún se siguen describiendo nuevas especies y registros.

Este grupo incluye varias especies, como *D. mojavensis*, *Drosophila hydei* y *Drosophila buzzatii*, que han desarrollado adaptaciones únicas para sobrevivir en ambientes áridos, utilizando recursos como cactus y frutas fermentadas. En 1940, Patterson, describió que, *D. mojavensis* se destaca por su carácter cactófilo, una característica compartida con otras especies de este grupo, como *D. buzzatii*, descrita en 1942, por el mismo Patterson junto con Wheeler (**Puppato et al., 2023**). Su distribución geográfica abarca principalmente las regiones desérticas del suroeste de Estados Unidos y México. Esta especie ha captado considerable atención científica debido a su notable capacidad de coevolución con diversas especies de cactus, lo que refleja una rápida adaptación evolutiva (**Benowitz et al., 2020**). Estos atributos convierten a *D. mojavensis*, en un modelo valioso para estudios sobre interacciones ecológicas y procesos evolutivos.

Asimismo, los estudios en especies como *D. buzzatii* han revelado variaciones genómicas significativas que subrayan la importancia de la selección natural en la conformación de su fisiología y comportamientos adaptativos (**Hasson et al., 2019**). De igual forma, referente a miembros del grupo *repleta* con distribuciones restringidas a los Andes son *Drosophila huaylasi* del subgrupo *mulleri*, endémica de Ecuador y Perú (**Fontdevila et**

al., 1990) y *D. guayllabambae* del subgrupo *hydei*, endémica de Ecuador (**Rafael & Arcos, 1989**). Esta última, en particular se alimenta del cactus *Opuntia*, compartiendo muchas características evolutivas con sus congéneres del mismo grupo. La biología, ecología y genética de *D. guayllabambae* siguen siendo relativamente desconocidas en comparación con otras especies del grupo *repleta*, misma especie que será analizada en este proyecto de investigación.

1.3. Información genómica de especies en la región neotropical

La región Neotropical es reconocida por su biodiversidad, impulsada por la variedad de ambientes que han fomentado la evolución de una amplia gama de especies de plantas y animales, como especies de la familia Drosophilidae. Muchos estudios han intentado descubrir los factores que contribuyen a generar y sostener la biodiversidad, pero existe una notable escasez de investigación centrada en la evolución genómica específica de esta región (**Prediger, 2024**). Sin embargo, el análisis genómico de especies de esta zona se ha vuelto esencial para entender la rica biodiversidad de estas áreas.

Investigaciones en este campo han descubierto, una amplia gama de adaptaciones genéticas que estas especies han desarrollado para prosperar en condiciones tropicales, como la resistencia a enfermedades, la tolerancia a temperaturas y humedades extremas, así como, métodos de reproducción y dispersión especializados (**O’Grady & DeSalle, 2018**). Adicionalmente, estudios comparativos de genomas en Diptera:Drosophilidae han proporcionado información valiosa sobre su evolución y su historia evolutiva, realzando su importancia en los esfuerzos de conservación y en la evaluación de los efectos del cambio climático y la deforestación en estas comunidades biológicas (**Doré et al., 2022**).

Aunque la información genómica de las especies neotropicales sigue siendo limitada, se han logrado avances significativos en la caracterización de ciertas especies, como es el caso de *Drosophila willistoni*. Esta especie de *Drosophila* fue una de las primeras en la región neotropical cuyo genoma fue completamente secuenciado (**Bergman, 2002**). Este estudio pionero ha proporcionado pistas cruciales sobre los mecanismos genómicos de

adaptación y evolución en *D. willistoni*, marcando un hito importante para futuras investigaciones en la genética de especies neotropicales. Asimismo, en una investigación realizada por **Moreyra et al., (2019)**, indican que, dentro del grupo *repleta*, el cluster *D. buzzatii* es un clado sudamericano de siete especies estrechamente relacionadas en distintas etapas de divergencia, siendo un modelo valioso para estudios evolutivos. Aunque se han realizado esfuerzos para clarificar sus relaciones filogenéticas, los resultados han sido ambiguos, variando según los marcadores moleculares utilizados. A pesar de la utilidad del ADN mitocondrial en estudios evolutivos, ninguno de los más de 20 mitogenomas de *Drosophila* secuenciados hasta 2019 incluía este cluster. Este estudio presentó el ensamblaje de seis mitogenomas completos de cinco especies: *Drosophila antonietae*, *Drosophila borborema*, *D. buzzatii*, dos cepas de *Drosophila koepferae* y *Drosophila seriema*, con el fin de revisar las relaciones filogenéticas y los tiempos de divergencia mediante análisis mitogenómicos. Los resultados apoyaron la monofilia del cluster *D. buzzatii*, dividiéndolo en dos clados principales: uno con *D. buzzatii* y *D. koepferae* (ambas cepas) y otro con las especies restantes. Estos hallazgos coinciden con estudios previos basados en genes mitocondriales y nucleares, pero contradicen una filogenia nuclear reciente, sugiriendo que los genomas nuclear y mitocondrial reflejan historias evolutivas distintas (**Rohland et al., 2007; Pacheco et al., 2011**). Asimismo, la investigación realizada por **Rodríguez, (2020)**, donde realizó la secuenciación parcial del genoma mitocondrial de *D. buzzatii*, mediante la técnica de amplificación de PCR (del inglés, *Polymerase Chain Reaction*) de fragmentos de ADN mitocondrial con el fin de comparar las secuencias obtenidas con una secuencia de referencia de *D. buzzatii*.

También, se encuentra la investigación realizada por **Guillén et al., (2014)**, donde describe que, *D. buzzatii* y *D. mojavensis* son dos especies cactófilas que pertenecen al grupo *repleta*; no obstante, tienen distribuciones geográficas y plantas hospederas primarias muy diferentes. Por lo que, para investigar la base genómica de la adaptación ecológica, secuenciaron el genoma y el transcriptoma de *D. buzzatii* y compararon su contenido génico con el de *D. mojavensis* y otras dos especies de *Drosophila* no cactófilas del mismo subgénero. Hoy en día, la investigación genómica continúa siendo una herramienta esencial para la conservación y gestión de especies neotropicales, además de ser crucial en el desarrollo de estrategias para adaptarse y mitigar los retos ambientales contemporáneos.

1.4. Importancia de la biología computacional en estudios genómicos para *Drosophila*

Esta falta de información genómica de las especies neotropicales representa una barrera para entender completamente su ecología y evolución, por lo que, la biología computacional desempeña un papel esencial en los estudios genómicos de eucariotas, como en *Drosophila*, organismos que generalmente presentan genomas más grandes y complejos que los procariotas. Esta complejidad se refleja en características como múltiples copias de genes, extensos intrones, elaborada regulación transcripcional y una gran cantidad de ADN no codificante. Estos elementos representan tanto desafíos como oportunidades para el análisis genómico, haciendo indispensable el uso de herramientas computacionales avanzadas para el ensamblaje y la anotación de genomas (**Jung et al., 2020**).

El desarrollo exponencial de las tecnologías de secuenciación de próxima generación (NGS, del inglés *Next Generation Sequencing*), junto con herramientas bioinformáticas eficientes para el análisis de la información genómica, ha permitido el ensamblaje eficiente de genomas mitocondriales, dando lugar a la aparición de la era de la mitogenómica (**Smith, 2016**). Así es como se ha podido identificar los ADN satélites (satDNAs, del inglés *satellite DNAs*), estos son secuencias altamente repetitivas y organizadas en tándem, que generalmente constituyen una gran parte (>20%) del genoma eucariota (**Mora et al., 2023**). La mayoría de los satDNAs evolucionan rápidamente y los cambios en su abundancia y composición nucleotídica, pueden estar relacionados con incompatibilidades genéticas entre especies. Este tema es abordado por **Laborne et al., (2024)**, donde utilizaron datos crudos de secuenciación pareada de Illumina y agrupamiento de lecturas basado en grafos con la herramienta bioinformática TAREAN (**Novák et al., 2017; Silva et al., 2019**) para analizar los satDNAs en dos especies crípticas neotropicales cactófilas de *Drosophila* del cluster *buzzatii*, *Drosophila serido* y *D. antonietae*, provenientes de cinco localidades de Brasil. Dentro del grupo *repleta*, los satDNAs han sido estudiados principalmente en el clado monofilético *buzzatii*, sus siete especies son: *D. buzzatii*, *D. koepferae*, *D. serido*, *D. antonietae*, *D. seriema*, *D. gouveai* y *D. borborema* (**Manfrin & Sene, 2006**). Estas especies son predominantemente endémicas de América del Sur, con la excepción de *D. buzzatii*, que ha logrado una

distribución subcosmopolita (**Manfrin & Sene, 2006; Santos et al., 2023**). Igualmente, en su investigación **Kuhn et al., (2021)**, resaltan que, los satDNAs representan una pieza importante en la arquitectura del genoma y que los cambios en su estructura, organización y abundancia pueden afectar la evolución de los genomas y las especies de múltiples maneras. Desde su descubrimiento hace más de 50 años, las especies del género *Drosophila* han sido utilizadas continuamente como modelos para estudiar diversos aspectos de la biología de los satDNAs mediante herramientas bioinformáticas.

Así es como, la biología computacional se ha establecido como un pilar fundamental en el estudio de la genómica, proporcionando herramientas esenciales para el análisis y la interpretación de grandes conjuntos de datos biológicos. En particular, su aplicación al estudio de especies del género *Drosophila*, ha revelado patrones de diversidad genética y adaptaciones evolutivas con precisión (**Stefanini et al., 2021**). La capacidad de secuenciar y analizar genomas completos a bajo costo y con alta eficiencia, ha permitido investigar a fondo la biología evolutiva y ecológica de estas especies. Un ejemplo destacado de esto, es el estudio realizado por **Ometto et al., (2013)**, donde la secuenciación del genoma de *D. mojavensis*, miembro del grupo *repleta* subgrupo *mulleri*, proporcionó indicios significativos sobre los mecanismos genéticos subyacentes a la adaptación al nicho ecológico específico de descomposición de cactus.

Además, la integración de datos de múltiples fuentes y su análisis mediante software especializado de biología computacional posibilita una comprensión más profunda de la estructura poblacional y la dinámica evolutiva. Puesto que, comprender la genética de especies endémicas no solo contribuye al conocimiento científico global, sino que también tiene implicaciones prácticas para la conservación y la gestión de los recursos naturales, especialmente en países megadiversos como Ecuador (**Mestanza et al., 2020**). Estos estudios y muchos más, realizados subrayan la importancia de la biología computacional en la investigación genómica, demostrando su capacidad para descifrar complejas interacciones genéticas y adaptaciones en el género *Drosophila*. Por lo que, con la finalidad de preservar la especie *D. guayllabambae* en Ecuador, se realizará una caracterización del genoma completo mediante profundo análisis utilizando biología computacional.

1.5. Elementos Transponibles (TEs) en *Drosophila*

Los elementos transponibles (TEs, del inglés *Transposable Elements*) de clase I, conocidos como retrotransposones, se clasifican en cinco categorías principales: retrotransposones con repeticiones terminales largas (LTR), elementos similares a las secuencias intermedias repetidas de *Dictyostelium* (DIRS), elementos tipo Penelope (PLEs), elementos nucleares intercalados largos (LINEs) y elementos nucleares intercalados cortos (SINEs) (Sessegolo *et al.*, 2016). No obstante, Hill (2019), menciona que, aunque todos estos tipos están presentes en *Drosophila*, los retrotransposones LTR y los LINEs son los más comunes. Lindsley & Zimm, (2012), mencionan que, en el caso de *Drosophila*, los retrotransposones LTR suelen tener una longitud de entre 5 y 7 kb, tal dato fue presentado previamente por investigaciones de Wicker *et al.*, (2007) y Mount & Rubin, (1985).

Asimismo, Lindsley & Zimm, (2012) y Wicker *et al.*, (2007), indican que, los elementos LINEs tienen una longitud de 3 a 5 kb y generalmente, contienen dos marcos abiertos de lectura (ORFs, del inglés *Open Reading Frame*), misma información que años antes había sido descrita por Bucheton *et al.*, (1984) y Fawcett *et al.*, (1986). Como se ha mencionado previamente, además de los retrotransposones LTR y los LINEs, que son muy comunes en los genomas de *Drosophila*, la clase I también abarca otros tres grupos: DIRS, PLEs y SINEs. Hasta el 2016, no se detectaron DIRS ni SINEs en *Drosophila* (Sessegolo *et al.*, 2016). Según Evgen'ev, (2013), los PLEs fueron identificados por primera vez en *Drosophila virilis* y están relacionados con el fenómeno de disgenesia híbrida. Estos elementos transponibles están presentes, al menos, en el grupo *virilis* y en *D. willistoni*. Los PLEs comparten similitudes con los LINEs, ya que también codifican una endonucleasa y una transcriptasa inversa. No obstante, se diferencian por tener repeticiones terminales que pueden presentarse en orientación directa o inversa.

Por su parte, los TE de clase II, son transposones de ADN, según Mérel *et al.*, (2020), existen cuatro órdenes: transposones con repeticiones invertidas terminales (TIR), Crypton, Helitron y Maverick. Los TIR y los Helitrones son los más abundantes en *Drosophila*. De acuerdo con Wicker *et al.*, (2007), los elementos transponibles TIR

suelen tener un tamaño de 1.5 a 3 kb en *D. melanogaster* y se caracterizan por sus repeticiones invertidas terminales (TIR) de longitud variable. El orden Helitron, representado por la única superfamilia Helitron, ha dado lugar a elementos transponibles bastante pequeños en *D. melanogaster* (< 1 kb) (**Kapitonov & Jurka, 2007**).

Es importante resaltar que, aunque la mayoría de los TEs se consideran dañinos o neutrales, algunas inserciones pueden ser adaptativas. Estas inserciones pueden afectar la evolución del genoma al inactivar o duplicar genes, añadir o eliminar regiones reguladoras, introducir empalmes alternativos y afectar la expresión de genes cercanos. Además, se han detectado inserciones de TEs en transcripciones maduras de genes en genomas de *D. melanogaster* de poblaciones silvestres, lo que sugiere que las secuencias de TEs pueden contribuir a la novedad genética y a la estructura de los genes (**Han, 2021**).

2. OBJETIVOS

2.1. Objetivo General

Analizar el genoma de *Drosophila guayllabambae*, una especie endémica del Ecuador, utilizando biología computacional.

2.2. Objetivos Específicos

- Ensamblar el genoma de *D. guayllabambae* a partir de lecturas crudas obtenidas a través de secuenciación Illumina.
- Comparar la estructura del genoma de *D. guayllabambae* con la especie de referencia *Drosophila mojavensis*, *Drosophila melanogaster* y *Drosophila rucux*.
- Identificar los componentes del genoma de *D. guayllabambae*: genes y elementos transponibles (TEs).

3. METODOLOGÍA

La presente investigación, llevó a cabo el análisis del genoma completo de *D. guayllabambae*, considerada una especie endémica del Ecuador, cuyas secuencias se recopilaron de investigaciones previas, las cuales se generaron utilizando la tecnología Illumina, mediante la librería TruSeq DNA Nano y el secuenciador Illumina HiSeq.

3.1. Análisis de la calidad del ensamblaje de *nov*o del genoma de *D. guayllabambae*

Con los datos crudos obtenidos mediante la secuenciación Illumina de la especie endémica *D. guayllabambae*, se utilizó la herramienta FastQC (v0.12.1) (**Babraham Bioinformatics, 2018**) para determinar la calidad de las secuencias en formato .fastq. Tal herramienta identificó una diferencia significativa en las secuencias totales, para lo cual, se empleó la herramienta bioinformática BBTools (**Graham et al., 2024**), a través del comando: `repair.sh`; los resultados del procesamiento se muestran en la **Tabla 1**. Nuevamente, se utilizó FastQC, con la que se determinó que ya no era necesario realizar ningún tipo de ajuste, ya que el programa infirió una buena calidad de secuencias para iniciar con el proceso de ensamblaje.

Para el ensamblaje de las secuencias de *D. guayllabambae*, se utilizó la herramienta MaSuRCA (v4.1.1) (del inglés, *Maryland Super-Read Celera Assembler*) (**Bronski et al., 2020**), el archivo de configuración para la especie contenía la media y la desviación estándar del tamaño de inserción de la biblioteca de secuenciación correspondiente, así como los siguientes parámetros: `GRAPH_KMER_SIZE = 31`, `USE_LINKING_MATES = 1`, `NUM_THREADS=20`, `LIMIT_JUMP_COVERAGE = 300` y `CA_PARAMETERS = cgwErrorRate = 0.15`. El tamaño del hash de Jellyfish (`JF_SIZE`) se estableció como el producto del tamaño estimado del genoma $\times 20$ (de acuerdo al manual de MaSuRCA en GitHub). Por su parte, los gaps que se encontraban entre los contigs fueron sellados por GapFiller (v1.1.0) (**Boetzer & Pirovano, 2012**). Las estadísticas del ensamblaje inicial y final fueron generadas por QUAST (v5.3.0) (del inglés, *Quality Assessment Tool for Genome Assemblies*) (**Gurevich et al., 2013**). Complementariamente, se utilizó las herramientas Minimap2 (**Li, 2018**) y BUSCO (v5.8.0) (del inglés, *Benchmarking*

Universal Single-Copy Orthologs) (Simão *et al.*, 2015; Seppey *et al.*, 2019; Manni *et al.*, 2021), para identificar genes ortólogos, haciendo uso del dataset del linaje *díptera_odb10* para evaluar la cobertura y calidad del genoma.

3.2. Alineamiento de la estructura del genoma de *D. guayllabambae* con las especies *D. mojavensis*, *D. melanogaster* y *D. rucux*

Para realizar la comparación del genoma de *D. guayllabambae* con *D. mojavensis* (accesion: GCF_018153725.1) y *D. melanogaster* (accesion: GCF_000001215.4), se obtuvo sus respectivos ensamblajes mediante la base de datos del NCBI (del inglés, *National Center for Biotechnology Information*). Por su parte, el ensamblaje de *D. rucux*, fue proporcionado por el grupo de investigación del Laboratorio de Genética Evolutiva de la Pontificia Universidad Católica del Ecuador. Con lo cual, se utilizó MUMmer's NUCmer (v3.1), mediante el comando: `nucmer [options] <Reference> <Query>` (Marçais *et al.*, 2018). Esto permitió alinear fragmentos de genomas similares de manera rápida y eficiente, generando el archivo de salida de nucmer “.delta”. Seguidamente, se utilizó la herramienta Dot, para generar el plot de los alineamientos (Nattestad, 2020), ejecutando el script de Python “DotPrep.py”, donde los archivos de salida incluyeron los archivos *.coords y *.coords.idx que fueron usados con Dot para la visualización de cada alineamiento junto con el genoma de *D. guayllabambae*.

3.3. Identificación de los componentes del genoma de *D. guayllabambae*

Para la identificación de elementos transponibles de *novo*, se empleó la herramienta RepeatModeler (v2.0.5) (Hubley & Smit, 2023), en la cual se encuentran tres programas: RECON, RepeatScout y LtrHarvest/Ltr_retriever, utilizando métodos computacionales distintos para identificar y delinear los límites de elementos repetidos en secuencias de ADN. Al finalizar con éxito la ejecución, RepeatModeler generó 3 archivos: 1) families.fa : Consensus sequences, 2) families.stk : Seed alignments y 3) rmod.log : A summarized log of the run. Así es como, el formato fasta resultó útil para emplear en RepeatMasker

(v4.1.5) a través del comando: `<RepeatMaskerPath>/RepeatMasker -lib <database_name>-families.fa mySequence.fa.`

En cuando a la predicción de genes, se utilizó AUGUSTUS (v3.5.0) (**Hoff & Stanke, 2019**), como es un programa *ab initio* basó sus predicciones exclusivamente en la secuencia genómica del archivo *Dguayllabambae.fasta*. Además, se empleó el conjunto predeterminado "fly" de AUGUSTUS para llevar a cabo las comparaciones, el cual está basado en un conjunto de entrenamiento aplicable a dípteros en general, utilizando el comando: `augustus --gff3=on --species=fly Dguayllabambae.fasta > output.gff3`. Por último, se empleó la herramienta AGAT (del inglés, *Annotation Genome Analysis Toolkit*) (**Dainat et al., 2023**), la cual es perfectamente capaz de analizar y manipular archivos en formato GFF3, esta herramienta generó estadísticas detalladas sobre los elementos contenidos en el archivo GFF3 producido por AUGUSTUS, como número de genes, genes predictores de proteínas, exones, intrones, entre otros componentes.

4. RESULTADOS Y DISCUSIÓN

4.1. Análisis de la calidad de las secuencias y ensamblaje de *novο* del genoma de *D. guayllabambae*

Al iniciar el análisis del genoma completo de *D. guayllabambae*, se observó una variación considerable entre las secuencias, la cual fue corregida mediante el uso de la herramienta BBTools. De acuerdo con **Wang (2023)**, esta discrepancia en la longitud de las lecturas puede deberse tanto a errores técnicos como a variaciones biológicas propias de la muestra, lo que resalta la importancia de una evaluación detallada. En este contexto, el uso de herramientas bioinformáticas como BBTools resulta especialmente adecuado para abordar y corregir tales inconsistencias. Los resultados del procesamiento indicaron que se analizó un total de 129.324.881 lecturas, equivalentes a 19.528.057.031 bases, logrando un proceso completo del 100% de las lecturas y bases. De estas, el 95.65% fueron identificadas como pares (pairs), mientras que el 4.35% se clasificaron como únicas (singletons), dichos datos se detallan en la **Tabla 1**. De acuerdo con **Hanussek et al., (2021)**, indican que, estos datos reflejan una eficiencia alta en la separación y procesamiento de las secuencias, lo que es fundamental para análisis posteriores en estudios genómicos o transcriptómicos.

Tabla 1

Resultados del procesamiento de las secuencias de D. guayllabambae realizado con BBTools

	Input	Result	Pairs	Singletons
Reads	129324881	129324881	123703226	5621655
Bases	19528057031	19528057031	18679187126	848869905
Percentage (%)		100	95.65	4.35

La secuenciación de extremos emparejados (paired-end) de Illumina del ADN genómico (gDNA) de *D. guayllabambae* generó 9.3 Gbp de datos de secuenciación, donde se trabajó con 123.708.226 lecturas identificadas como pares, las estadísticas de las lecturas se proporcionan en la **Tabla 2**. El andamiaje (scaffolding) y el llenado de huecos (gap-

filling) produjeron un borrador final del genoma de 160,421,720 pb, compuesto por 9,909 contigs, las estadísticas del ensamblaje se presentan en la **Tabla 3**.

Tabla 2

Estadísticas de las lecturas del secuenciamiento de Illumina en D. guayllabambae

Read statistics		Raw reads		Processed reads (BBTools)	
		Forward	Reverse	Forward	Reverse
	Read count	61,851,613	67,473,268	61,851,613	61,851,613
Nucleotide percentages (%)	A	32.14	32.14	31.94	32.14
	T	30.40	30.40	30.10	30.40
	G	20.20	20.20	20.30	20.20
	C	17.26	17.26	17.66	17.26
Dinucleotide percentages (%)	AT	62.54	62.54	62.04	62.54
	GC	37.46	37.46	37.96	37.46

Nota. Las gráficas de resultados de las lecturas crudas se muestran en el apartado de **Figura A1**.

La **Tabla 3** presenta un análisis comparativo de las estadísticas de ensamblaje genómico antes y después de la aplicación de la herramienta GapFiller, utilizando los resultados generados por QUASt. En ambas etapas, tanto en el ensamblaje inicial (Initial contig assembly), como en el ensamblaje final (Final draft assembly), se observa que el número de bases ensambladas (160,421,720 pb) y el número de contigs (9,909) permanecen constantes. Esto sugiere que la herramienta GapFiller no alteró significativamente la estructura general del ensamblaje en términos de cantidad de secuencias o su longitud total. Por su parte, el valor de N50, que es una medida comúnmente utilizada para evaluar la calidad del ensamblaje, también se mantiene en 110,832 pb en ambas etapas. El N50 indica que la mitad de todo el ensamblaje está contenido en contigs de al menos este tamaño, lo que refleja una consistencia en la calidad del ensamblaje antes y después de la corrección de gaps. Además, el contenido de GC (40.56%) no muestra variación, lo que es esperable, según **Kyriacou et al., (2024)**, ya que mencionan que, esta característica es inherente al genoma en estudio y no debería verse afectada por procesos de corrección de

gaps. Además, **Boetzer & Pirovano, (2012)**, describen que, la similitud de los datos del ensamblaje inicial y final, podría sugerir que el ensamblaje inicial ya era de alta calidad o que las correcciones realizadas por GapFiller no impactaron de manera notable en las estadísticas generales.

Tabla 3

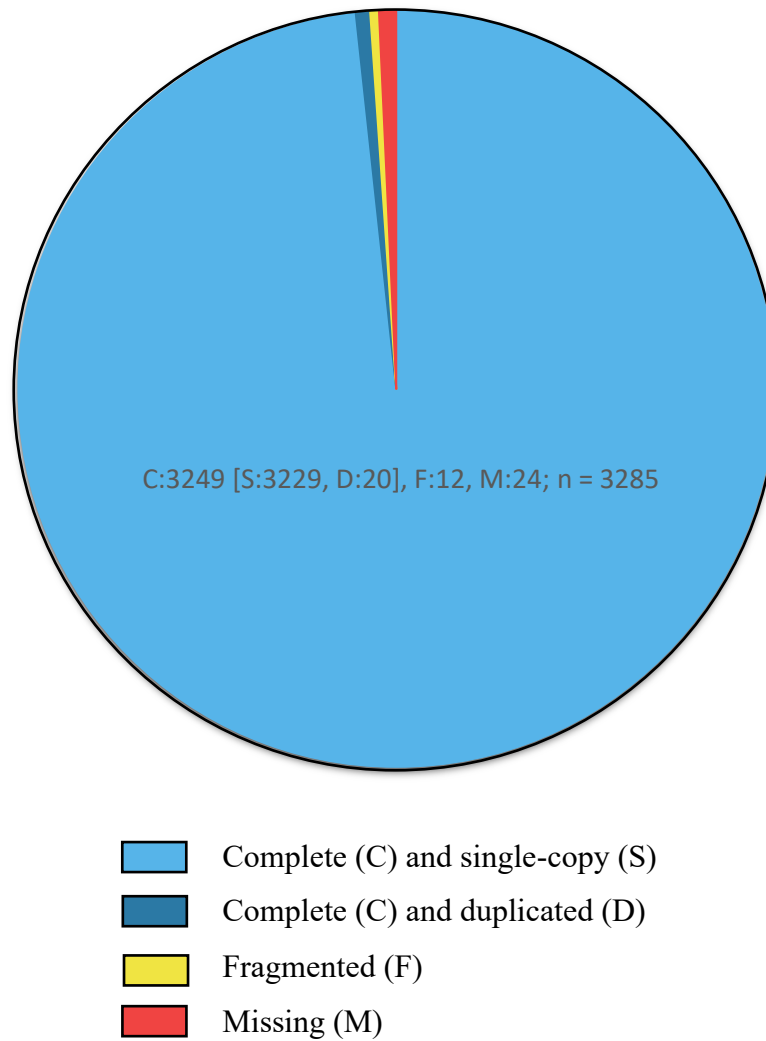
Estadísticas de la calidad del ensamblado final de D. guayllabambae

	Initial contig assembly	Final draft assembly
Assembled bases (pb)	160,421,720	160,421,720
Number of contigs	9,909	9,909
N50 (pb)	110,832	110,832
GC Content (%)	40.56	40.56

Los resultados de BUSCO para determinar la integridad del ensamblaje del genoma de *D. guayllabambae*, identificó que, el 98.90% de los genes ortólogos eran completos. De estos, el 98.29% eran genes completos de una única copia y el 0.61% correspondía a genes completos duplicados. Adicionalmente, se encontró que el 0.37% de los genes ortólogos estaban fragmentados y el 0.73% no fueron identificados. Al contrastar estos resultados con los genomas de *Drosophila* reportados en el NCBI, se notó una alta similitud del genoma de *D. guayllabambae* con *D. mojavensis* (99% de genes completos) (NCBI, 2021) y *D. melanogaster* (99.3% de genes completos) (NCBI, 2014), lo que indica valores similares entre las especies. Asimismo, en la investigación de **DSouza et al., (2021)**, el ADN genómico de la cepa de *Drosophila nasuta nasuta* fue secuenciado en pares de extremos utilizando la tecnología Illumina HiSeq 2500, obteniéndose un ensamblaje preliminar del genoma de 145.64 Mb, este ensamblaje recuperó el 93.6% de los ortólogos con BUSCO, lo que indica que el análisis comparativo final de *D. guayllabambae* reveló una mayor presencia de 3,249 genes eucarióticos conservados de un total de 3,285 genes conocidos, según el conjunto de datos *diptera_odb10* (ver **Figura 1**), en comparación a otras especies. Sin embargo, en la investigación realizada por **Casal, (2024)**, se identificó un 98.6% de genes ortólogos completos en el genoma ensamblado de *D. rucux*, siendo un valor más cercano al valor de 98.90% obtenido por *D. guayllabambae*.

Figura 1

Representación esquemática de los resultados de Benchmarking Universal Single-Copy Orthologs (BUSCO) de *D. guayllabambae*



Nota. Representación esquemática de los resultados generados por BUSCO para determinar la calidad del ensamblaje del genoma de *D. guayllabambae*. Los colores indican los grados de integridad de los genes predichos en el ensamblaje. 'n' indica el número total de genes en el conjunto de datos *diptera_odb10*.

4.2. Alineamiento de la estructura del genoma de *D. guayllabambae* con las especies *D. mojavensis*, *D. melanogaster* y *D. rucux*

El análisis de alineamiento genómico entre *D. guayllabambae* y *D. mojavensis* (ver **Figura 2-A**) muestra una correlación lineal en la mayoría del genoma, indicando una conservación significativa en sus secuencias genómicas, lo cual podría sugerir un ancestro común reciente o presiones selectivas similares. Las diferencias observadas en el alineamiento pueden deberse a divergencias en la selección natural, deriva genética y eventos de especiación. No obstante, la presencia de puntos dispersos y regiones no alineadas refleja divergencias genómicas como inserciones, deleciones y rearrreglos cromosómicos post-divergencia, esenciales para la adaptación y evolución de las especies (**Markow & O'Grady, 2007; Markow, 2019**).

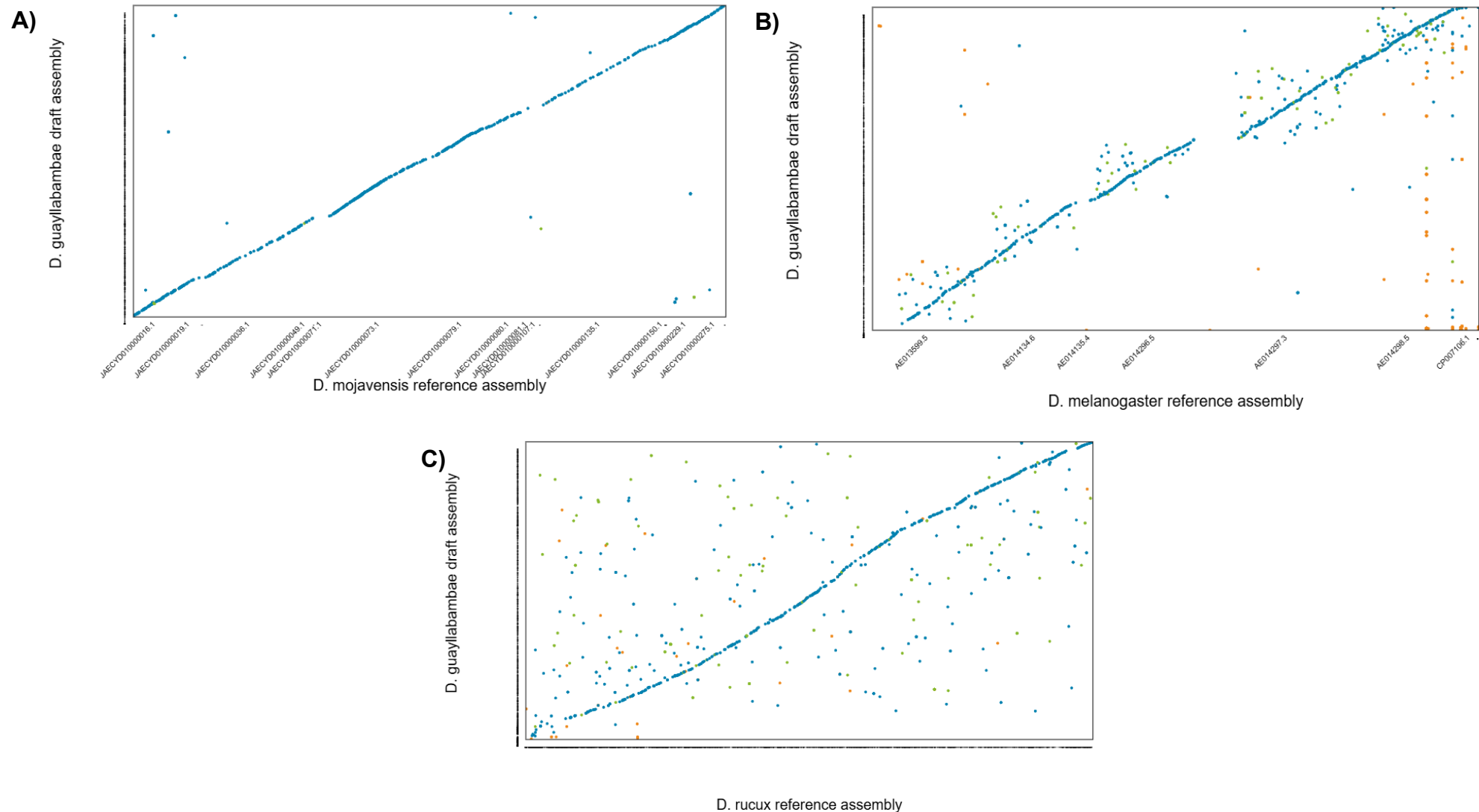
El alineamiento genómico entre *D. guayllabambae* y *D. melanogaster* (ver **Figura 2-B**) muestra una relación menos lineal y una mayor dispersión de puntos que con *D. mojavensis*, reflejando una divergencia evolutiva significativa y diferencias profundas en sus historias evolutivas y adaptaciones a sus respectivos ambientes. Según **Coronado et al., (2019)**, menciona que, los puntos dispersos en el gráfico probablemente representan variaciones genéticas como Polimorfismos de Nucleótido Único (SNPs, del inglés *Single Nucleotide Polymorphisms*) o Variaciones en el Número de Copias (CNVs, del inglés *Copy Number Variations*), que son cruciales para entender los mecanismos adaptativos y evolutivos de estas especies. Estas diferencias genómicas son fundamentales para las respuestas adaptativas ante los cambios ambientales y desafíos ecológicos, impactando directamente en la biodiversidad y la supervivencia de las especies (**Li et al., 2022**). Además, la amplia caracterización del genoma de *D. melanogaster*, un organismo modelo adaptado a una variedad de ambientes incluidos los urbanos, contrasta con las adaptaciones específicas de *D. guayllabambae* a su hábitat. Esto se refleja en diferencias en genes relacionados con la respuesta al estrés, el metabolismo y la interacción con el entorno. Las discrepancias y regiones no conservadas en el alineamiento sugieren acumulación de mutaciones, eventos de especiación y adaptación divergente (**Borne et al., 2021**).

El alineamiento genómico entre *D. guayllabambae* y *D. rucux* (ver **Figura 2-C**), revelando tanto similitudes como divergencias genómicas clave. La diagonal principal,

dominada por puntos azules que representan alineaciones directas únicas, indica una conservación significativa de grandes segmentos genómicos, sugiriendo una relación evolutiva cercana entre ambas especies. Estas regiones conservadas probablemente incluyen genes esenciales y secuencias regulatorias que han mantenido su función a lo largo de la evolución. Por otro lado, los puntos verdes, que indican alineaciones inversas únicas, y los puntos naranjas, que representan alineaciones repetitivas, están más dispersos. Las inversiones genómicas podrían reflejar eventos de reorganización cromosómica posteriores a la divergencia, mientras que las regiones repetitivas resaltan la presencia de elementos transponibles y secuencias repetidas, cruciales para la variabilidad genética pero desafiantes para el ensamblaje genómico. Este análisis sugiere que, aunque *D. guayllabambae* y *D. rucux* comparten una base genómica común, existen diferencias significativas en la organización y estructura de sus genomas, posiblemente relacionadas con adaptaciones específicas a sus respectivos entornos.

Figura 2

Dotplot del alineamiento del genoma ensamblado de *D. guayllabambae* vs *D. mojavensis*, *D. melanogaster* y *D. rucux*



Nota. Dotplot del alineamiento del genoma ensamblado de *D. guayllabambae* (a lo largo del eje Y) en comparación a los genomas de referencia (a lo largo del eje X) con **A)** *D. mojavensis*, **B)** *D. melanogaster* y **C)** *D. rucux*. Los puntos de colores representan alineaciones directas únicas (azul), alineaciones inversas únicas (verde) y alineaciones repetitivas (naranja).

4.3. Componentes del genoma de *D. guayllabambae*: genes y elementos transponibles (TEs).

Los datos presentados por RepeatModeler, detallados en la **Tabla 4**, se pueden contrastar con la investigación realizada por **Rius *et al.*, (2016)**, donde se menciona que, los elementos transponibles representan el 8.43%, el 4.15% y el 15.35% de los ensamblajes de los genomas de *D. buzzatii* st-1, *D. buzzatii* j-19 (dos cepas de esta especie) y *D. mojavensis*, respectivamente. Como se evidencia, los genomas de *D. buzzatii* st-1 y j-19 tienen un porcentaje menor de elementos transponibles que *D. mojavensis*, incluso después de corregir un sesgo de subestimación en la secuenciación NGS. Dado que, al realizar la corrección en la cobertura de lecturas (read coverage), reveló 4.4 Mb adicionales de TEs en *D. buzzatii* st-1 (aumentando su contenido al 11 %) y 5 Mb más en j-19 (7 % del genoma). A pesar de esto, el contenido de TEs en las cepas de *D. buzzatii* continuaba siendo inferior al de *D. mojavensis*. Finalmente, se determinaron que, los genomas de *D. buzzatii* contenían TEs entre un 21 % (st-1) y un 25 % (j-19) más pequeños que los de *D. mojavensis*. Por su parte, según **Hasson *et al.*, (1992)** y **Oliveira *et al.*, (2012)**, detallan que, las diferencias en el contenido de TEs podrían deberse a presiones selectivas vinculadas a los recursos utilizados por cada especie. Es decir, *D. buzzatii* se desarrolla en cactus *Opuntia* más pequeños y de descomposición rápida, lo que favorecería un desarrollo más acelerado y un genoma compacto, similar a *D. guayllabambae*. Mientras que, *D. mojavensis* utiliza cactus columnares más grandes y persistentes, lo que podría relajar la presión selectiva sobre el tamaño genómico, permitiendo una mayor acumulación de TEs. Esto sugiere que las diferencias ecológicas entre especies podrían influir en la evolución de sus genomas.

El análisis comparativo de los elementos transponibles (TEs) en los genomas de *D. buzzatii* (cepas st-1 y j-19) y *D. mojavensis* reveló patrones distintivos en la contribución relativa de los diferentes órdenes definidos por **Wicker *et al.*, (2007)**. Si bien ambas cepas de *D. buzzatii* presentan una composición similar de TEs (Helictrones, LINEs, retrotransposones LTR, transposones TIR y Mavericks/Polintones), esta difiere significativamente de la observada en *D. mojavensis*. Puesto que, los helictrones constituyen el orden más abundante en las tres especies, aunque su proporción varía notablemente: representan el 40.61% del contenido total de TEs en *D. buzzatii* st-1, frente al 30.65% en *D. buzzatii* j-19 y 33.90% en *D. mojavensis*; donde en *D. guayllabambae*

con tan solo el 1.29%. En contraste, los retrotransposones LTR, que ocupan el segundo lugar en abundancia en *D. mojavensis* (33.46%), muestran una contribución significativamente menor en ambas cepas de *D. buzzatii* (17.38% en st-1 y 19.54% en j-19), donde los LINEs emergen como el segundo orden más representativo; pues para *D. guayllabambae*, los LTR representan el 16.13% (valor cercano a *D. buzzatii* st-1) y los LINEs el 7.74%. Asimismo, los transposones TIR son más frecuentes en *D. buzzatii* (14.81% en st-1; 14.46% en j-19) que en *D. mojavensis* (9.24%), patrón que se repite para las secuencias repetitivas no clasificadas (7.15% en st-1 y 9.11% en j-19 vs. 2.42% en *D. mojavensis*), siendo para *D. guayllabambae* un total de 61.94% de secuencias desconocidas. Por lo tanto, estas diferencias cuantitativas podrían reflejar distintos procesos evolutivos o presiones selectivas entre estas especies estrechamente relacionadas.

Tabla 4

Resultados de los elementos transponibles (ET) del genoma ensamblado de D. guayllabambae generados por RepeatModeler

	Transposable Elements	Number	Percentage (%)
Retroelements (Class I)	LINE total	12	7.74
	LTR total	25	16.13
	SINE total	4	2.58
DNA transposons (Class II)	Helitron total	2	1.29
	Maverick total	2	1.29
	Others (DNA/TcMar-Tc1 y DNA/TcMar-Mariner)	9	5.81
Simple repeat		2	1.29
Others	rRNA, DNA, ARTEFACT	3	1.95
Unknown		96	61.94

Nota. De 155 secuencias analizadas se muestra la prevalencia de los elementos transponibles indicados por la herramienta RepeatModeler.

Los análisis realizados con RepeatMasker y presentados en la **Tabla 5** muestran que el 20.06% del genoma consiste en elementos repetitivos enmascarados, debido a que, este porcentaje ilustra la extensión del genoma influenciada por dichos elementos. De acuerdo con **Goubert et al., (2015)** y **Hill (2019)** indica que el genoma de *D. melanogaster* contiene aproximadamente un 20% de TEs. No obstante, cabe mencionar que existe una variación significativa en estas estimaciones, como indican **Kaminker et al., (2002)** y **Drosophila 12 Genomes Consortium et al., (2007)**, pues es probable que estas diferencias se expliquen, al menos en parte, por el ensamblaje del genoma o la parte del ensamblaje del genoma que se analiza, o ambos. Por lo que, existen variaciones en los informes sobre *D. melanogaster*, como lo indican **Kapitonov et al., (2002)**, quienes sugieren que entre el 22% y el 25% del genoma de esta especie está compuesto por elementos móviles. Por otro lado, estudios recientes en especies como *D. mojavensis* y *D. arizonae*, realizados por **Banho et al., (2021)**, indican que los elementos transponibles constituyen el 25% y el 26% del genoma, respectivamente. Estos datos sugieren que la presencia de elementos transponibles en *D. guayllabambae* podría exhibir patrones similares a los observados en estas especies estrechamente relacionadas.

Tabla 5

Resultados de los elementos transponibles (ET) del genoma ensamblado de D. guayllabambae generados por RepeatMasker

	Number of elements	Lenght Occupied (bp)	Percentage of sequence (%)
Class I	54049	5748695	3.55
SINEs	710	39049	0.02
PLEs	0	0	0.00
LINEs	26050	2288102	1.41
LTR elements	27289	3421544	2.11
Gypsy/DIRS1	17345	1935327	1.19
Clas II	22540	3999946	2.47
Rolling-circles	32720	4637043	2.86
Unclassified	83047	7485082	4.62

	Number of elements	Lenght Occupied (bp)	Percentage of sequence (%)
Total interspersed repeats		17233723	10.64
Small RNA	1122	281197	0.17
Satellites	2289	228212	0.14
Simple repeats	168851	9109679	5.62
Low complexity	19417	1004023	0.62

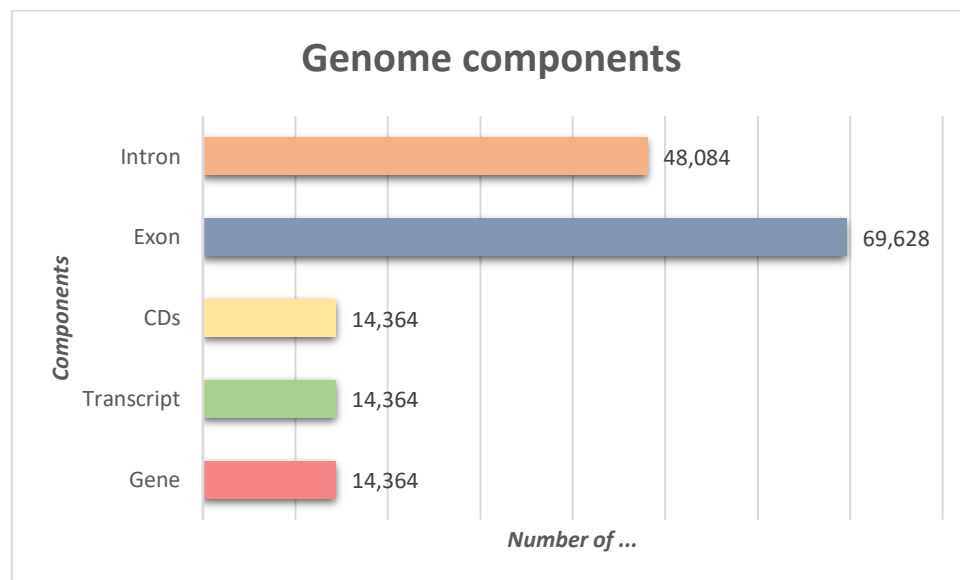
Nota. La mayoría de las repeticiones fragmentadas por inserciones o deleciones se han contado como un solo elemento. Se excluyeron las ejecuciones de ≥ 20 X/Ns en la consulta en los cálculos de %. *Total length: 162181721 bp (161981076 bp excl N/X-runs). *Bases masked: 32490503 bp (20.06 %).

El análisis desarrollado por AUGUTUS determinó que, 14,364 genes se encontraban en el genoma de *D. guayllabambae*, además del número de exones e intrones, lo cual se visualiza en la **Figura 3**. Este dato se puede contrastar con el genoma de *D. melanogaster*, en el cual se han identificado alrededor de 17,872 genes, y es uno de los modelos genéticos más detalladamente mapeados entre los organismos no humanos (**Misra et al., 2009**). Por último, la **Tabla 6** muestra un resumen comparativo de puntos relevantes de los genomas de *D. guayllabambae* frente a *D. mojavensis*, *D. melanogaster* y *D. rucux*. Estos datos se pueden contrastar con la investigación realizada por **Guillén et al., (2014)**, donde describe que, las especies de *Drosophila* cactófilas, como *D. buzzatii* y *D. mojavensis*, son modelos importantes para estudiar las interacciones gen-ambiente y la adaptación ecológica. Aunque ambas pertenecen al grupo *repleta*, difieren en sus distribuciones geográficas y plantas hospederas. Para explorar la base genómica de estas adaptaciones, se secuenciaron el genoma y el transcriptoma del desarrollo de *D. buzzatii* y se comparó su contenido génico con el de *D. mojavensis* y otras dos especies no cactófilas del mismo subgénero. El genoma recién secuenciado de *D. buzzatii* (161.5 Mb) contenía 13,657 genes codificantes de proteínas anotados. Al comparar estos datos con los de *D. guayllabambae* con 160.42 Mb, se infiere que, los resultados obtenidos son similares, pues es evidente, ya que estas especies se encuentran estrechamente relacionadas, principalmente, por pertenecer al grupo *repleta*. No obstante, la

discrepancia en el número de genes codificantes de proteínas de 0,707 genes, entre *D. guayllabambae* (14,364 genes) y *D. buzzatii* (13,657 genes) puede atribuirse a diferencias en los métodos de anotación genética, la calidad del ensamblaje genómico y las herramientas bioinformáticas utilizadas en cada estudio. Además, factores biológicos como la evolución específica de cada especie y la presencia de elementos genéticos móviles pueden influir en la variación del contenido génico (Blumenstiel, 2011). Es importante mencionar que, estos factores que generan diferencias genómicas entre estas dos especies también se las puede atribuir a las diferencias existentes entre todas las especies del grupo *repleta* y, por tanto, a las especies en general, del género *Drosophila* (Droge et al., 2012; Monier et al., 2024).

Figura 3

Estadísticas de AGAT de la predicción de genes realizada por AUGUSTUS para la especie D. guayllabambae



<i>Genome components</i>	<i>Number</i>	<i>Total length (bp)</i>
<i>Gene</i>	14,364	95,033,674
<i>Transcript</i>	14,364	95,033,674
<i>CDs</i>	14,364	30,025,643
<i>Exon</i>	69,628	45,082,236
<i>Intron</i>	48,084	36,546,040

Tabla 6

Comparación de las características entre el genoma ensamblado de D. guayllabambae frente a los genomas de D. mojavensis, D. melanogaster y D. rucux

Categories	<i>D. guayllabambae</i>	<i>D. mojavensis</i>	<i>D. melanogaster</i>	<i>D. rucux</i>
Sequencing Technology	Illumina TrueSeq	Oxford Nanopore MinION; Illumina HiSeq	PacBio; Illumina HiSeq	Illumina TrueSeq
Genome size (Mb)	160.4	163.2	143.7	182.7
Percentage GC (%)	40.56	39.50	41.90	37.73
AUGUSTUS	14, 364	15,130	17,872	15,920
Encode proteins	14, 364	13,270	13,962	19.805
BUSCO (%)	98.29	99.30	99.40	98.20

Nota. Los datos de *D. mojavensis* y *D. melanogaster* se los obtuvo mediante los números del NCBI, códigos de acceso GCF_018153725.1; GCF_000001215.4 respectivamente. La información de *D. rucux*, fue obtenida de **Casal, (2024)**.

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- El análisis computacional del genoma completo de *D. guayllabambae* permitió caracterizar de manera integral su estructura genómica, mostrando una alta calidad de ensamblaje y un contenido genético comparable con otras especies del grupo *repleta*, como la especie de referencia estrechamente relacionada, *D. mojavenensis* y la especie comúnmente conocida, *D. melanogaster*, lo que subraya un ancestro común y posibles presiones selectivas similares a lo largo de su evolución. La aplicación de herramientas de biología computacional fue fundamental para obtener resultados precisos y relevantes sobre su diversidad genética y adaptaciones evolutivas, pues este análisis *in silico* se logró a través de la utilización de diferentes herramientas de biología computacional y bioinformáticas, como BBTools, FASTQC, MaSuRCA, QUAST, GapFiller, MUMmer's NUCmer, Dot, RepeatModeler, RepeatMasker, AUGUSTUS y AGAT. Por lo tanto, el ensamblaje del genoma de *D. guayllabambae* fue exitoso, con métricas de calidad que lo posicionan como un recurso valioso para estudios genómicos comparativos y evolutivos.
- Las secuencias del genoma completo de *D. guayllabambae*, considerada especie endémica del Ecuador, obtenidas mediante la tecnología de secuenciación Illumina fueron exitosamente ensambladas a través de MaSuRCA, el ensamblaje *de novo* del genoma de *D. guayllabambae* mostró un tamaño de 160,421,720 pb, con un N50 de 110,832 pb y un contenido de GC del 40.56%. Complementariamente, la calidad del ensamblaje fue validada mediante BUSCO, que identificó un 98.90% de genes ortólogos completos, lo que confirma la integridad y precisión del genoma ensamblado.
- La comparación del genoma completo de *D. guayllabambae* con la especie de referencia *D. mojavenensis*, *D. melanogaster* y *D. rucux*, resultó altamente eficiente, debido a que, el alineamiento genómico mostró una alta similitud *entre D. guayllabambae* y *D. mojavenensis*, indicando una relación evolutiva cercana. Sin embargo, se observaron diferencias significativas en la estructura genómica, como inserciones, deleciones y rearrreglos cromosómicos, que reflejan adaptaciones

específicas a sus respectivos entornos. En comparación con *D. melanogaster*, se evidenció un mejor alineamiento de genomas. Por su parte, con *D. rucux*, se identificaron regiones conservadas y divergencias genómicas, lo que resalta la complejidad de la evolución dentro del género *Drosophila*.

- El 20.06% del genoma de *D. guayllabambae* está compuesto por elementos transponibles, los cuales fueron determinados por parte de RepeatMasker, lo que sugiere un papel activo de estos elementos en la diversificación genética y adaptativa de la especie. Por su parte AUGUSTUS, determinó 14,364 genes siendo el mismo número de transcriptos y CDs, adicionalmente se identificó el número de exones e intrones siendo igual a 69,628 y 48,084, respectivamente. Estos valores son consistentes con los observados en otras especies de *Drosophila*, lo que sugiere una estructura génica conservada. Por lo tanto, las especies de referencia indican valores cercanos a la especie endémica estudiada, lo que proporciona una base sólida para futuros estudios funcionales y comparativos.

5.2. Recomendaciones

- Complementar la secuenciación Illumina con tecnologías de lecturas largas (PacBio u Oxford Nanopore) para mejorar la resolución de regiones repetitivas y reducir la fragmentación del ensamblaje.
- Ampliar los estudios comparativos con otras especies del género *Drosophila* para profundizar en el entendimiento de la evolución genética y adaptaciones específicas de *D. guayllabambae*.
- Realizar un análisis más profundo de los elementos transponibles (TEs) para identificar su impacto en la estructura y función del genoma, incluyendo su papel en la regulación génica y la variabilidad genética.
- Utilizar la información genómica de *D. guayllabambae* para desarrollar estrategias de conservación que protejan su hábitat y biodiversidad en Ecuador.

6. REFERENCIAS

- Acurio, A. & Rafael, V. (2009). Diversity and geographical distribution of *Drosophila* (Diptera, Drosophilidae) in Ecuador. *Drosoph Inf Serv.* 92:20-25
- Acurio, A. (2024). Evidence of a South American origin for the *Drosophila repleta* group (Diptera: Drosophilidae). *European Journal Of Entomology*, 121, 124-133
- Babraham Bioinformatics. (2018). *FastQC*. Babraham Institute. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Banho, C., Oliveira, D., Haudry, A., Fablet, M., Vieira, C., & Carareto, C. (2021). Transposable element expression and regulation profile in gonads of interspecific hybrids of *Drosophila arizonae* and *Drosophila mojavensis wrightleyi*. *Cells*, 10(12), 3574.
- Benowitz, K., Coleman, J., Allan, C., & Matzkin, L. (2020). Contributions of cis- and trans-Regulatory Evolution to Transcriptomic Divergence across Populations in the *Drosophila mojavensis* Larval Brain. *Genome Biology and Evolution*, 12(8), 1407–1418. <https://doi.org/10.1093/gbe/evaa145>
- Blumenstiel, J. (2011). Evolutionary dynamics of transposable elements in a small RNA world. *Trends in genetics, TIG*, 27(1), 23–31. <https://doi.org/10.1016/j.tig.2010.10.003>
- Boetzer, M., & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology*, 13(6). <https://doi.org/10.1186/gb-2012-13-6-r56>
- Borne, F., Kulathinal, R., & Courtier, V. (2021). Glue Genes Are Subjected to Diverse Selective Forces during *Drosophila* Development. *Genome biology and evolution*, 13(12), evab248. <https://doi.org/10.1093/gbe/evab248>
- Brake, I., & Bächli, G. (2008). *World catalogue of Insects*. Apollo Books.
- Bronski, M., Martinez, C., Weld, H., & Eisen, M. (2020). Whole genome sequences of 23 species from the *Drosophila montium* species group (Diptera: Drosophilidae): A resource for testing evolutionary hypotheses. *G3: Genes, Genomes, Genetics*, 10(5). <https://doi.org/10.1534/g3.119.400959>
- Bucheton, A., Paro, R., Sang, H., Pelisson, A., & Finnegan, D. (1984). The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification,

- cloning, and properties of the I factor. *Cell*, 38(1), 153–163.
[https://doi.org/10.1016/0092-8674\(84\)90536-1](https://doi.org/10.1016/0092-8674(84)90536-1)
- Bussmann, R. (2011). Vegetation ecology and regeneration of tropical mountain forests. *Mod Trends Appl Terr Ecol.*:195-223. doi:10.1007/978-1-4615-0223-4_11.
- Casal, K. (2024). Ensamble de *novo* y caracterización del genoma en *Drosophila rucux* (Diptera, Drosophilidae). Pontificia Universidad Católica del Ecuador.
- Chan, P. & Lowe, T. (2019). tRNAscan-SE: searching for tRNA genes in genomic sequences. In: Kollmar M. (eds) *Gene Prediction. Methods Mol. Biol.*, 1962, 1–14. Humana, New York, NY.10.1007/978-1-4939-9173-0_1
- Coronado, M., Salvador, I., Castellano, D., Barbadilla, A. & Salazar, I. (2019). Adaptation and Conservation throughout the *Drosophila melanogaster* Life-Cycle. *Genome biology and evolution*, 11(5), 1463–1482.
<https://doi.org/10.1093/gbe/evz086>
- Dainat, J., Hereñú, D., Davis, E., Crouch, K., LucileSol, Agostinho, N., Pascal, Zollman, Z., & Tayyrov. (2023). *AGAT. Another Gtf/Gff Analysis Toolkit*. GitHub.
<https://github.com/NBISweden/AGAT/blob/master/README.md>
- Doré, M., Willmott, K., Leroy, B., Chazot, N., Mallet, J., Freitas, A., ... Elias, M. (2022). Anthropogenic pressures coincide with Neotropical biodiversity hotspots in a flagship butterfly group. *Diversity and Distributions*, 28(12).
<https://doi.org/10.1111/ddi.13455>
- Droge E., Manier, M., Lüpold, S., Belote, J., & Pitnick, S. (2012). Covariance among premating, post-copulatory and viability fitness components in *Drosophila melanogaster* and their influence on paternity measurement. *Journal of evolutionary biology*, 25(8), 1555–1563. <https://doi.org/10.1111/j.1420-9101.2012.02540.x>
- Drosophila* 12 Genomes Consortium, Clark, A., Eisen, M., Smith, D., Bergman, C., Oliver, B., Markow, T., Kaufman, T., Kellis, M., Gelbart, W., Iyer, V., Pollard, D., Sackton, T., Larracuent, A., Singh, N., Abad, J., Abt, D., Adryan, B., Aguade, M., Akashi, H., ... MacCallum, I. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450(7167), 203–218.
<https://doi.org/10.1038/nature06341>

- DSouza, S., Ponnanna, K., Chokkanna, A., & Ramachandra, N. (2020). Illumina short-read sequencing data, de novo assembly and annotations of the *Drosophila nasuta nasuta* genome. *Data in brief*, 34, 106674. <https://doi.org/10.1016/j.dib.2020.106674>
- Evgen'ev, M. (2013). What happens when Penelope comes? *Mob Genet Elem*, 3:e24542.
- Fawcett, D., Lister, C., Kellett, E., & Finnegan, D. (1986). Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. *Cell*, 47(6), 1007–1015. [https://doi.org/10.1016/0092-8674\(86\)90815-9](https://doi.org/10.1016/0092-8674(86)90815-9)
- Figuro, M. (2017). Filogenia molecular de especies ecuatorianas del grupo *Drosophila mesophragmatica* (Diptera, Drosophilidae). Pontificia Universidad Católica del Ecuador.
- Fontdevila, A., Wasserman, M., Pila, C., Pilaes, L., De Armengol, R., Suyo, M., Del, P., Sanchez, A., Vasquez, J., Ruiz, A. & Garcia, J. (1990). Description and evolutionary relationships of two species of the *Drosophila mulleri* cluster (Diptera: Drosophilidae). *Ann. Entomol. Soc. Am*, 83, 444–452.
- Goubert, C., Modolo, L., Vieira, C., ValienteMoro, C., Mavingui, P., & Boulesteix, M. (2015). De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome biology and evolution*, 7(4), 1192–1205. <https://doi.org/10.1093/gbe/evv050>
- Graham, B., Hipfner, J., Wellband, K., Ito, M., & Burg, T. (2024). Genetic–environment associations explain genetic differentiation and variation between western and eastern North Pacific rhinoceros auklet (*Cerorhinca monocerata*) breeding colonies. *Ecology and Evolution*, 14(7), e11534.
- Guillén, Y., Rius, N., Delprat, A., Williford, A., Muyas, F., Puig, M., Casillas, S., Ràmia, M., Egea, R., Negre, B., Mir, G., Camps, J., Moncunill, V., Ruiz, F., Cabrero, J., de Lima, L., Dias, G., Ruiz, J., Kapusta, A., Garcia, J., ... Ruiz, A. (2014). Genomics of ecological adaptation in cactophilic *Drosophila*. *Genome biology and evolution*, 7(1), 349–366. <https://doi.org/10.1093/gbe/evu291>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8). <https://doi.org/10.1093/bioinformatics/btt086>

- Han, S. (2021). *Novel computational strategies for the analysis of transposable elements in Drosophila cell culture genomes*. University of Georgia ProQuest Dissertations & Theses.
- Hanussek, M., Bartusch, F., & Krüger, J. (2021). Performance and scaling behavior of bioinformatic applications in virtualization environments to create awareness for the efficient use of compute resources. *PLoS Computational Biology*, *17*(7), e1009244.
- Hasson E, Naveira H, Fontdevila A. (1992). The breeding sites of Argentinian cactophilic species of the *Drosophila mulleri* complex. *Revista Chilena de Historia natural*, *65*(3), 319–26.
- Hasson, E., De Panis, D., Hurtado, J., & Mensch, J. (2019). Host Plant Adaptation in Cactophilic Species of the *Drosophila buzzatii* Cluster: Fitness and Transcriptomics. *Journal of Heredity*, *110*(1). <https://doi.org/10.1093/jhered/esy043>
- Hill, T. (2019). Transposable element dynamics are consistent across the *Drosophila* phylogeny, despite drastically differing content. *bioRxiv*. 651059. <https://doi.org/10.1101/651059>.
- Hoff, K., & Stanke, M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, *65*(1). <https://doi.org/10.1002/cpbi.57>
- Hubley, R. & Smit, A. (2023). *Repeat Modeler 2.0.5*. Institute for Systems Biology.
- Jung, H., Ventura, T., Sook, J., Kim, W., Nam, B., Kong, H., ... Eyun, S. (2020). Twelve quick steps for genome assembly and annotation in the classroom. *PLoS Computational Biology*, *16*(11). <https://doi.org/10.1371/journal.pcbi.1008325>
- Kaminker, J., Bergman, C., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D., Lewis, S., Rubin, G., Ashburner, M., & Celniker, S. (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome biology*, *3*(12), RESEARCH0084. <https://doi.org/10.1186/gb-2002-3-12-research0084>
- Kapitonov, V., & Jurka, J. (2003). Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proceedings of the National Academy of Sciences*, *100*(11), 6569-6574.

- Kapitonov, V., & Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends in genetics: TIG*, 23(10), 521–529. <https://doi.org/10.1016/j.tig.2007.08.004>
- Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic acids research*, 51(W1), W207-W212.
- Kuhn, G., Heringer, P. & Dias, G. (2021). Structure, Organization, and Evolution of Satellite DNAs: Insights from the *Drosophila repleta* and *D. virilis* Species Groups. *Progress in molecular and subcellular biology*, 60, 27–56. https://doi.org/10.1007/978-3-030-74889-0_2
- Kyriacou, R., Mulhair, P., & Holland, P. (2024). GC content across insect genomes: phylogenetic patterns, causes and consequences. *Journal of Molecular Evolution*, 92(2), 138-152.
- Laborne, A., Barrios, D., Heslop, J., Manfrin, M., & Kuhn, G. (2024). Genome location, evolution and centromeric contribution of satellite DNAs shared between the two closely related species *Drosophila serido* and *D. antonietae* (*repleta* group, *buzzatii* cluster). *Chromosoma*, 134(1), 1. <https://doi.org/10.1007/s00412-024-00827-9>
- Li, F., Rane, R., Luria, V., Xiong, Z., Chen, J., Li, Z., Catullo, R., Griffin, P., Schiffer, M., Pearce, S., Lee, S., McElroy, K., Stocker, A., Shirriffs, J., Cockerell, F., Coppin, C., Sgrò, C., Karger, A., Cain, J., Weber, J., ... Zhang, G. (2022). Phylogenomic analyses of the genus *Drosophila* reveals genomic signals of climate adaptation. *Molecular ecology resources*, 22(4), 1559–1581. <https://doi.org/10.1111/1755-0998.13561>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18). <https://doi.org/10.1093/bioinformatics/bty191>
- Lindsley, D. & Zimm, G. (2012). *The Genome of Drosophila melanogaster*. Academic Press.
- Llangarí, L., Tamayo, I., Peñafiel, A. & Rafael, V. (2022). Description of three new species in Chamanapamba Natural Reserve of *Drosophila* (Diptera: Drosophilidae), Tungurahua, Ecuador. *Revista Ecuatoriana de Medicina y Ciencias Biologicas*, 43(2), 37-53.

- Manfrin, M. & Sene, F. (2006). Cactophilic *Drosophila* in South America: a model for evolutionary studies. *Genética* 126(1–2):57–75. <https://doi.org/10.1007/s10709-005-1432-5>
- Manni, M., Berkeley, M., Seppey, M., & Zdobnov, E. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols*, 1(12). <https://doi.org/10.1002/cpz1.323>
- Marçais, G., Delcher, A., Phillippy, A., Coston, R., Salzberg, S. & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, 14(1). <https://doi.org/10.1371/journal.pcbi.1005944>
- Markow, T. (2019). Ecological and Evolutionary Genomics: The Cactophilic *Drosophila* Model System. *Journal of Heredity*, 110(1), 1–3, <https://doi.org/10.1093/jhered/esy062>
- Markow, T., & O'Grady, P. (2007). *Drosophila* biology in the genomic age. *Genetics*, 177(3), 1269–1276. <https://doi.org/10.1534/genetics.107.074112>
- Mérel, V., Boulesteix, M., Fablet, M., & Vieira, C. (2020). Transposable elements in *Drosophila*. *Mobile DNA*, 11, 23. <https://doi.org/10.1186/s13100-020-00213-z>
- Mestanza, C., Henkanaththegedara, S., Duchicela, P., Tierras, Y., Capa, M., Mejía, D., ... Ramón, P. (2020). In-situ and ex-situ biodiversity conservation in Ecuador: A review of policies, actions and challenges. *Diversity*. <https://doi.org/10.3390/D12080315>
- Misra, S., Crosby, M., Mungall, C., Matthews, B., Campbell, K., Hradecky, P., ... & Ashburner, M. (2002). Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biology*, 3(12).
- Monier, M., Nuez, I., Borne, F., & Courtier V. (2024). Higher evolutionary dynamics of gene copy number for *Drosophila* glue genes located near short repeat sequences. *BMC ecology and evolution*, 24(1), 18. <https://doi.org/10.1186/s12862-023-02178-y>
- Mora, P., Pita, S., Montiel, E., Rico, J., Palomeque, T., Panzera, F. & Lorite, P. (2023). Making the genome huge: the case of *Triatoma delpontei*, a Triatominae species with more than 50% of its genome full of Satellite DNA. *Genes* 14(2):371. <https://doi.org/10.3390/genes14020371>

- Morales, R. & Vieira, J. (2012). Phylogenetic patterns of geographical and ecological diversification in the subgenus *Drosophila*. *PLoS ONE* 7(11): e49552, 15 pp.
- Moreyra, N., Mensch, J., Hurtado, J., Almeida, F., Laprida, C., & Hasson, E. (2019). What does mitogenomics tell us about the evolutionary history of the *Drosophila buzzatii* cluster (*repleta* group)? *PLoS ONE* 14(11): e0220676. <https://doi.org/10.1371/journal.pone.0220676>
- Mount, M. & Rubin, G (1985). Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol Cell Biol*;5,1630–8.
- Nattestad, M (2020). *Dot: An interactive dot plot viewer for comparative genomics*. GitHub. <https://github.com/MariaNattestad/dot/blob/master/README.md>
- NCBI. (2014). *Genome assembly Release 6 plus ISO1 MT*. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001215.4/
- NCBI. (2021). *Genome assembly ASM1815372v1*. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_018153725.1/
- Novák, P., Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P. & Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res*. <https://doi.org/10.1093/nar/gkx257>
- O'Grady, P., & DeSalle, R. (2018). Phylogeny of the genus *Drosophila*. *Genetics*, 209(1), 1-25.
- Oliveira, D., Almeida, F., O'Grady, P., Armella, M., DeSalle, R., & Etges, W. (2012). Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. *Molecular phylogenetics and evolution*, 64(3), 533–544. <https://doi.org/10.1016/j.ympev.2012.05.012>
- Ometto, L., Cestaro, A., Ramasamy, S., Grassi, A., Revadi, S., Siozios, S., Moretto, M., Fontana, P., Varotto, C., Pisani, D., Dekker, T., Wrobel, N., Viola, R., Pertot, I., Cavalieri, D., Blaxter, M., Anfora, G., & Rota, O. (2013). Linking genomics and ecology to investigate the complex evolution of an invasive *Drosophila* pest. *Genome biology and evolution*, 5(4), 745–757. <https://doi.org/10.1093/gbe/evt034>

- Pacheco, M., Battistuzzi, F., Lentino, M., Aguilar, R., Kumar, S., Escalante, A. (2011). Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Mol Biol Evol*, 28: 1927–1942. <https://doi.org/10.1093/molbev/msr014>
- Patterson, J. & Mainland, G. (1945). *The Drosophilidae of Mexico*. University of Texas, Austin, TX, 101 pp.
- Patterson, J. & Stone, W. (1952). *Evolution in the Genus Drosophila*. Macmillan, New York, 610 pp.
- Prediger, C. (2024). *Evolutionary genomics of the Neotropical Drosophila saltans species group (Diptera: Drosophilidae)*. UNESP.
- Puppato, S., Fiorenza, G., Carraretto, D., Gomulski, L., Gasperi, G., Caceres, C., Grassi, A., Mancini, M., De Cristofaro, A., Ioriatti, C., Guilhot, R., & Malacrida, A. (2023). High promiscuity among females of the invasive pest species *Drosophila suzukii*. *Molecular ecology*, 32(22), 6018–6026. <https://doi.org/10.1111/mec.17161>
- Rafael, V. & Arcos, G. (1989). Subgrupo inca, un nuevo subgrupo del grupo *repleta*, con descripción de *Drosophila huancavilcae* n. sp (Diptera, Drosophilidae). *Evolución Biológica*, 3, 233–243.
- Rafael, V. (2023a). *Análisis morfológico del edeago en poblaciones naturales de Drosophila amaguana y D. neoamaguana de la provincia de Napo-Ecuador*. Pontificia Universidad Católica del Ecuador.
- Rafael, V. (2023b). *Seis especies nuevas del género Drosophila (diptera: drosophilidae), en la localidad de río Guango, provincia de Napo, Ecuador*. Pontificia Universidad Católica del Ecuador.
- Rafael, V., Arcos, G. & Terán, L. (2000). Ecología y distribución del género *Drosophila* en Guayllabamba y El Quinche, provincia de Pichincha- Ecuador. Federico Ma. Sanfeliu Vilar. Jesús Aguinaga Zumárraga Elena Roldán Baca HRLLG del CFM-G, editor. *Rev la Univ Católica del Ecuador*, 65, 130-155.
- Rius, N., Guillén, Y., Delprat, A., Kapusta, A., Feschotte, C., & Ruiz, A. (2016). Exploration of the *Drosophila buzzatii* transposable element content suggests underestimation of repeats in *Drosophila* genomes. *BMC genomics*, 17, 344. <https://doi.org/10.1186/s12864-016-2648-8>

- Rodríguez, E. (2020). *Partial sequencing of the mitochondrial genome of Drosophila buzzatii*. Universidad de Coruña.
- Rohland, N., Malaspinas, S., Pollack, J., Slatkin, M., Matheus, P., Hofreiter, M. (2007). Proboscidean mitogenomics: chronology and mastodon as outgroup. *PLoS Biol*, 5. <https://doi.org/10.1371/journal.pbio.0050207>
- Santos, M., Barrios, D. & Manfrin, M. (2023). Phylogeography of *Drosophila buzzatii* (Diptera, Drosophilidae): responses of the species to quaternary climates in tropical and subtropical South America. *Acad Bras Cienc* 95(suppl 2):e20220846. <https://doi.org/10.1590/0001-3765202320220846>
- Seppy M., Manni, M. & Zdobnov, E. (2019) BUSCO: Assessing genome assembly and annotation completeness. In M. Kollmar (Ed.), *Gene Prediction, 1962*, 227–245. Springer New York. https://doi.org/10.1007/978-1-4939-9173-0_14
- Sessegolo, C., Burlet, N., & Haudry, A. (2016). Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. *Biology letters*, 12(8), 20160407. <https://doi.org/10.1098/rsbl.2016.0407>
- Silva, B., Heringer, P., Dias, G., Svartman, M., & Kuhn, G. (2019). De novo identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines. *PLoS ONE*, 14(12). <https://doi.org/10.1371/journal.pone.0223466>
- Silva, D., Schmitz, H., de Medeiros, H., Rohde, C., Montes, M. & Garcia, A. (2020). Geographic expansion and dominance of the invading species *Drosophila nasuta* (Diptera, Drosophilidae) in Brazil. *Journal of Insect Conservation*, 24(3). <https://doi.org/10.1007/s10841-020-00219-1>
- Simão, F., Waterhouse, R., Ioannidis, P., Kriventseva, E., & Zdobnov, E. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxford, England)*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Smith, D. (2016). The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs?. *Briefings in functional genomics*, 15(1), 47–54. <https://doi.org/10.1093/bfgp/elv027>

- Stefanini, M., Gottschalk, M., Calvo, N. & Soto, I. (2021). Evolution of male genitalia in the *Drosophila repleta* species group (Diptera: Drosophilidae). *Journal of evolutionary biology*, 34(9), 1488–1502. <https://doi.org/10.1111/jeb.13913>
- Sturtevant, A. (1921). *The North American Species of Drosophila*. Carnegie Institute, Washington, iv + 150 pp. + 3 pl.
- Throckmorton, L. (1975). The phylogeny, ecology and geography of *Drosophila*. *Handbook of genetics*, 3(17), 422-469.
- Wang, X. (2023). *Next-generation sequencing data analysis*. CRC Press.
- Wicker, T., Sabot, F., Hua, A., Bennetzen, J., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. (2007). A unified classification system for eukaryotic transposable elements. *Nature reviews. Genetics*, 8(12), 973–982. <https://doi.org/10.1038/nrg2165>
- Wyngaard, G., Rasch, E., Manning, N., Gasser, K., & Domangue, R. (2005). The relationship between genome size, development rate, and body size in copepods. *Hydrobiologia*, 532(1-3), 123–137. doi:10.1007/s10750-004-9521-5

7. ANEXOS

Figura A1

Contenido de bases en las secuencias de D. guayllabambae. A) Forward y B) Reverse.

