

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR**

**FACULTAD DE INGENIERIA**

**Identificación de elementos genéticos transponibles de clase II en el genoma de  
*Drosophila amaguana* (Diptera, Drosophilidae)**

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

**MANUEL ALEJANDRO COBA MALES**

Quito, 2023

## Derechos de Autor

Yo, **Manuel Alejandro Coba Males** en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación “ **Identificación de elementos genéticos transponibles de clase II en el genoma de *Drosophila amaguana* (Diptera, Drosophilidae)**”, modalidad proyecto de investigación aplicada, de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN, concedo a favor de la Pontificia Universidad Católica del Ecuador una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra con fines estrictamente académicos. Conservo a mi favor todos los derechos de autor sobre la obra, establecidos en la normativa citada.

Así mismo, autorizo a la Pontificia Universidad Católica del Ecuador para que realice la digitalización y publicación de este trabajo de titulación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

El autor declara que la obra objeto de la presente autorización es original en su forma de expresión y no infringe el derecho de autor de terceros, asumiendo la responsabilidad por cualquier reclamación que pudiera presentarse por esta causa y liberando a la Universidad de toda responsabilidad.

---

Manuel Alejandro Coba Males

C.C. 1723904098

majano004@hotmail.com

## **Aprobación del Director del Trabajo de Titulación**

Certifico que la Disertación de Maestría en Biología Computacional del Sr. Manuel Alejandro Coba Males ha sido concluida de conformidad con las normas establecidas; por lo tanto, puede ser presentada para la calificación correspondiente.

---

Doris Jimena Vera Peralta, PhD.

Directora de la disertación

Quito, 3 de Abril del 2023

## **Dedicatoria**

*Con mucho cariño a mis padres, Wilson Coba y Anita Lucía Males, su apoyo incondicional me impulsa cada día a ser mejor persona y profesional para alcanzar mis más grandes anhelos. Gracias por educarme con todas sus enseñanzas, guiándome a lo largo de mi vida para forjarme como la persona que soy hoy en día; una vez más, su esfuerzo también se ve plasmado en este proyecto que hoy ha culminado exitosamente y es el inicio para nuevos retos en el futuro. Los amo tanto!*

*A mi hermano Wilson Coba Males, quien ha sido desde siempre un ejemplo para mí. Gracias por tu apoyo en las grandes y pequeñas decisiones que tomo y que permiten avanzar hacia mis metas en la vida.*

*A todos quienes se adelantaron en este camino de la vida y desde el cielo ahora son la luz que me da fuerzas para continuar, siempre los llevo en el corazón.*

***Manuel Alejandro Coba Males***

## **Agradecimientos**

*Mi eterno agradecimiento a Dios y a la Virgen de la Dolorosa por guiar mi camino con sus santas bendiciones para continuar en el día a día incluso en los momentos más difíciles.*

*A toda mi familia, gracias a cada uno de ustedes por su preocupación, consejos y palabras de aliento que han sido importantes durante mi formación personal y académica.*

*Cordialmente agradezco a la Dra. Doris Vela, mi directora en este trabajo. Gracias por confiar en mí y permitirme desarrollar esta investigación para adquirir valiosos conocimientos; gracias por el cariño, paciencia, dedicación y apoyo incondicional en cada instante en el que requerí acudir para recibir observaciones, comentarios o sugerencias, que me ayudaron a mejorar la calidad en el contenido de mi trabajo.*

*Un sincero agradecimiento al Dr. Romain Guyot quien siempre estuvo presto poniendo a disponibilidad su valioso tiempo, para brindarme desinteresadamente ayuda con sus conocimientos que respondieron a cuestiones particulares que surgían durante el desarrollo de este trabajo, y que sin duda alguna permitieron el avance y finalización de mi trabajo.*

*En una mención especial y honorífica, agradezco a la Dra. Ana Poveda, por ser quien me ha motivado, impulsado y ha despertado mi interés para continuar académicamente dentro del ámbito de la investigación científica. Gracias por su amistad, confianza y por darme la oportunidad de demostrar las habilidades de las que soy capaz de desarrollar.*

*Expreso mi gratitud a todos los miembros que forman parte del Instituto de Investigación en Zoonosis (CIZ) de la Universidad Central del Ecuador, por ser el lugar que me abrió las puertas para mi primera experiencia laboral, en el cual he aprendido y adquirido muchos conocimientos de personas que han hecho que llegue realmente a amar la investigación.*

*Agradezco a mi primo Alex Estrella, quién por segunda ocasión me facilitó un ordenador personal con los recursos computacionales suficientes para la ejecución de algunos programas bioinformáticos que fueron necesarios para la realización de mi trabajo.*

*Agradezco también a las personas que tuve la dicha de conocer durante esta maestría, especialmente a aquellas con las que entablé una sincera amistad desde el primer día, mis amigas Yadi y Gaby, con quienes he podido compartir momentos tanto dentro como fuera de la parte académica, demostrando la excelente calidad personas que ellas son, al estar presentes en los momentos buenos y malos.*

*Para concluir, dejo sentado el más sincero reconocimiento a la Pontificia Universidad Católica del Ecuador, para todo el equipo que forma parte de la Maestría en Biología Computacional, desde la coordinación hasta la planta docente; gracias por permitirme continuar con mi formación académica profesional adquiriendo nuevos conocimientos.*

***Manuel Alejandro Coba Males***

## Índice General

<b>Derechos de Autor</b> .....	<b>ii</b>
<b>Aprobación del Director del Trabajo de Titulación</b> .....	<b>iii</b>
<b>Dedicatoria</b> .....	<b>iv</b>
<b>Agradecimientos</b> .....	<b>v</b>
<b>Índice General</b> .....	<b>vii</b>
<b>Índice de Figuras</b> .....	<b>x</b>
<b>Índice de Tablas</b> .....	<b>xii</b>
<b>Índice de Anexos</b> .....	<b>xiii</b>
<b>Lista de Abreviaturas</b> .....	<b>xiv</b>
<b>Resumen</b> .....	<b>xvi</b>
<b>Abstract</b> .....	<b>xvii</b>
<b>Capítulo I: Introducción</b> .....	<b>1</b>
Planteamiento del Problema.....	1
Formulación del Problema .....	3
Preguntas de Investigación.....	3
Objetivos .....	4
Objetivo General.....	4
Objetivos Específicos .....	4
Justificación e Importancia.....	4
<b>Capítulo II: Revisión de la Literatura</b> .....	<b>1</b>
Antecedentes de la Investigación .....	1
Fundamentación Teórica .....	2
DNA Repetitivo .....	2
Elementos Transponibles.....	3
Retrotransposones.....	7
Retrotransposones LTR.....	8
Retrotransposones DIRS.....	9
Retrotransposones non-LTR (LINEs).....	10
Retrotransposones de tipo Penélope (PLE).....	11
SINEs.....	12
Transposones de DNA.....	13
TIR.....	14

<i>Tc1/Mariner</i> .....	15
<i>hAT</i> .....	15
<i>Mutator</i> .....	15
<i>Merlin</i> .....	16
<i>Transib</i> .....	16
<i>P elements</i> .....	17
<i>PiggyBac</i> .....	18
<i>PIF/Harbinger</i> .....	18
<i>CACTA</i> .....	18
<i>Crypton</i> .....	19
<i>Helitron</i> .....	19
<i>Maverick/Polinton</i> .....	20
<i>MITEs</i> .....	21
<i>Drosophila</i> .....	21
Elementos transponibles en <i>Drosophila</i> .....	25
<b>Capítulo III: Metodología</b> .....	<b>26</b>
Diseño de la Investigación.....	26
Enfoque de la Investigación.....	26
Alcance de la Investigación.....	26
Tipo de Investigación.....	26
Metodología.....	27
Evaluación de la calidad del ensamblaje del genoma.....	28
Preparación del entorno de trabajo.....	28
Pipeline EDTA.....	28
Caracterización de elementos transponibles.....	29
Identificación de elementos transponibles completos.....	30
Identificación de elementos transponibles activos.....	31
Identificación manual de elementos transponibles.....	31
<b>Capítulo IV: Resultados</b> .....	<b>33</b>
Calidad del ensamblaje del genoma en <i>Drosophila amaguana</i> .....	33
Composición del genoma en <i>Drosophila amaguana</i> .....	33
Elementos transponibles en el genoma de <i>Drosophila amaguana</i> .....	34
Caracterización de elementos transponibles en <i>Drosophila amaguana</i> .....	36
Elemento <i>P</i> .....	37

<i>Mutator</i> .....	39
<i>CACTA</i> .....	41
<i>hAT</i> .....	43
<i>PIF/Harbinger</i> .....	45
<i>Tc1/Mariner</i> .....	47
Contenido de elementos transponibles en otras especies de <i>Drosophila</i> .....	49
<b>Capítulo V: Análisis y Discusión de Resultados</b> .....	<b>52</b>
<b>Capítulo VI: Conclusiones y Recomendaciones</b> .....	<b>63</b>
Conclusiones .....	63
Recomendaciones .....	65
<b>Referencias Bibliográficas</b> .....	<b>66</b>
<b>Anexos</b> .....	<b>96</b>

## Índice de Figuras

<b>Figura 1.</b> Porcentaje de elementos transponibles en genomas de diferentes especies. ....	5
<b>Figura 2.</b> Mecanismos de transposición de los elementos móviles. ....	6
<b>Figura 3.</b> Sistema de clasificación jerárquica para elementos transponibles. ....	7
<b>Figura 4.</b> Estructura de los retrotransposones LTR. ....	9
<b>Figura 5.</b> Estructura del retrotransposón <i>DIRS-1</i> . ....	10
<b>Figura 6.</b> Estructura del retrotransposón <i>LINE-1</i> . ....	11
<b>Figura 7.</b> Estructura de los retrotransposones tipo Penélope. ....	12
<b>Figura 8.</b> Estructura de los retrotransposones SINEs. ....	13
<b>Figura 9.</b> Estructura de los transposones de DNA de repeticiones terminales invertidas (TIR). ....	14
<b>Figura 10.</b> Estructura del transposón <i>Tc1/Mariner</i> . ....	15
<b>Figura 11.</b> Estructura del transposón <i>MuDR</i> . ....	16
<b>Figura 12.</b> Estructura de un <i>elemento P</i> completo. ....	17
<b>Figura 13.</b> Estructura del elemento <i>Crypton-Cn1</i> . ....	19
<b>Figura 14.</b> Estructura de transposones Helitron. ....	20
<b>Figura 15.</b> Estructura de los transposones MITEs. ....	21
<b>Figura 16.</b> Estructura cromosómica de <i>Drosophila melanogaster</i> ....	23
<b>Figura 17.</b> Genes reportados en el genoma de <i>D. melanogaster</i> FB2016_05 (R6.13) ....	24
<b>Figura 18.</b> Resumen de anotaciones de los genomas para 12 especies de <i>Drosophila</i> . ....	24
<b>Figura 19.</b> Workflow en el pipeline EDTA. ....	29
<b>Figura 20.</b> Visualización parcial de un archivo de anotación para ETs intactos. ....	30
<b>Figura 21.</b> Composición del genoma de <i>Drosophila amaguana</i> . ....	34
<b>Figura 22.</b> Totalidad de elementos transponibles en el genoma de <i>Drosophila amaguana</i> . ....	36
<b>Figura 23.</b> <i>Dot plot</i> de un transposón de la superfamilia <i>elemento P</i> en <i>D. amaguana</i> . ....	38
<b>Figura 24.</b> Anotación funcional del <i>elemento P</i> en <i>D. amaguana</i> . ....	39
<b>Figura 25.</b> <i>Dot plot</i> de un transposón de la superfamilia <i>Mutator</i> en <i>D. amaguana</i> . ....	40
<b>Figura 26.</b> Anotación funcional del transposón <i>Mutator</i> en <i>D. amaguana</i> . ....	41
<b>Figura 27.</b> <i>Dot plot</i> de un transposón de la superfamilia <i>CACTA</i> en <i>D. amaguana</i> . ....	42
<b>Figura 28.</b> Anotación funcional del transposón <i>CACTA</i> en <i>D. amaguana</i> . ....	43
<b>Figura 29.</b> <i>Dot plot</i> de un transposón de la superfamilia <i>hAT</i> en <i>D. amaguana</i> . ....	44
<b>Figura 30.</b> Anotación funcional del transposón <i>hAT</i> en <i>D. amaguana</i> . ....	45
<b>Figura 31.</b> <i>Dot plot</i> de un transposón de la superfamilia <i>PIF/Harbinger</i> en <i>D. amaguana</i> . ..	46
<b>Figura 32.</b> Anotación funcional del transposón <i>PIF/Harbinger</i> en <i>D. amaguana</i> . ....	47

<b>Figura 33.</b> <i>Dot plot</i> de un transposón de la superfamilia <i>Tc1/Mariner</i> en <i>D. amaguana</i> . .....	48
<b>Figura 34.</b> Anotación funcional del transposón <i>Tc1/Mariner</i> en <i>D. amaguana</i> .....	49
<b>Figura 35.</b> Contenido de ETs anotados por EDTA en <i>Drosophila</i> .....	50
<b>Figura 36.</b> Contenido de ETs identificado para especies de <i>Drosophila</i> .....	51

## Índice de Tablas

<b>Tabla 1.</b> Resumen del reporte de QUASt para el genoma de <i>D. amaguana</i> .....	33
<b>Tabla 2.</b> Elementos transponibles en el genoma de <i>Drosophila amaguana</i> .....	35
<b>Tabla 3.</b> Porcentaje de ETs identificados en <i>D. melanogaster</i> y <i>D. amaguana</i> . .....	49
<b>Tabla 4.</b> Contenido de ETs identificados en especies de <i>Drosophila</i> .....	51

## Índice de Anexos

<b>Anexo A.</b> Árbol de Problemas .....	96
<b>Anexo B.</b> Anotación del <i>elemento P</i> en <i>D. amaguana</i> en formato GenBank.....	97
<b>Anexo C.</b> Anotación del transposón <i>Mutator</i> en <i>D. amaguana</i> en formato GenBank.....	98
<b>Anexo D.</b> Anotación del transposón <i>CACTA</i> en <i>D. amaguana</i> en formato GenBank.....	99
<b>Anexo E.</b> Anotación del transposón <i>hAT</i> en <i>D. amaguana</i> en formato GenBank.....	100
<b>Anexo F.</b> Anotación del transposón <i>PIF/Harbinger</i> en <i>D. amaguana</i> en formato GenBank.....	101
<b>Anexo G.</b> Anotación del transposón <i>Tc1/Mariner</i> en <i>D. amaguana</i> en formato GenBank.....	102

## Lista de Abreviaturas

**AA:** *Amino acids* (aminoácidos)

**cDNA:** *Complementary deoxyribonucleic acid* (Ácido desoxirribonucleico complementario)

**CDS:** *Coding sequence* (Región de codificación)

**DDD:** *Aspartic acid - Aspartic acid - Aspartic acid* (aminoácidos)

**DDE:** *Aspartic acid - Aspartic acid - Glutamic acid* (aminoácidos)

**DIRS:** *Dictyostelium Intermediate Repeat Sequence* (Secuencia de repetición intermedia de *Dictyostelium*)

**DNA:** *Deoxyribonucleic acid* (Ácido Desoxirribonucleico)

**dsDNA:** *Double strand deoxyribonucleic acid* (Ácido desoxirribonucleico de doble cadena)

**EN:** *Endonuclease* (Endonucleasa)

**ETs:** Elementos transponibles

**IN:** *Integrase* (Integrasa)

**LINEs:** *Long Interspersed Nuclear Elements* (Elementos nucleares intercalados largos)

**lncRNA:** *Long noncoding ribonucleic acid* (Ácido ribonucleico largo no codificante)

**LTR:** *Long Terminal Repeat* (Repeticiones terminales largas)

**m.s.n.m:** metros sobre el nivel del mar

**miRNA:** *Micro ribonucleic acid* (Micro ácido ribonucleico)

**MITEs:** *Miniature Inverted-repeat Transposable Elements* (Elementos transponibles de repetición invertida en miniatura)

**mRNA:** *Messenger ribonucleic acid* (Ácido ribonucleico mensajero)

**MT:** *Methylase* (Metilasa)

**NGS:** *Next Generation Sequencing* (Secuenciación de próxima generación)

**non-LTR:** *Non-Long Terminal Repeat* (Repeticiones terminales no largas)

**ORFs:** *Open Reading Frames* (Marcos abiertos de lectura)

**PBS:** *Primer binding site* (Sitio de unión de primer)

**PCR:** *Polymerase Chain Reaction* (Reacción en cadena de la polimerasa)

**PLE:** *Penelope-like Element* (Elementos de tipo Penélope)

**PR:** *Protease* (Proteasa)

**RH:** *Ribonuclease H* (Ribonucleasa H)

**RNA *pol*:** *RNA polymerase* (RNA polimerasa)

**RNA:** *Ribonucleic acid* (Ácido Ribonucleico)

**rRNA:** *Ribosomal ribonucleic acid* (Ácido ribonucleico ribosomal)

**RT:** *Reverse transcriptase* (Reverso transcriptasa)

**SINES:** *Short Interspersed Nuclear Elements* (Elementos nucleares intercalados cortos)

**snoRNA:** *Small nucleolar ribonucleic acid* (Pequeño ácido ribonucleico nucleolar)

**snRNA:** *Small nuclear ribonucleic acid* (Pequeño ácido ribonucleico nuclear)

**SO:** *Sequence Ontology* (Ontología de secuencias)

**ssDNA:** *Single strand deoxyribonucleic acid* (Ácido desoxirribonucleico monocatenario)

**TIR:** *Terminal Inverted Repeats* (Repeticiones terminales invertidas)

**tRNA:** *Transfer ribonucleic acid* (Ácido ribonucleico de transferencia)

**TSD:** *Target Site Duplication* (Duplicación del sitio objetivo)

**UTR:** *Untranslated region* (Región no traducida)

**YR:** *Tyrosine recombinase* (Tirosina recombinasa)

**TÍTULO:** Identificación de elementos genéticos transponibles de clase II en el genoma de *Drosophila amaguana* (Diptera, Drosophilidae)

**Autor:** Manuel Alejandro Coba Males

**Directora:** Doris Jimena Vela Peralta, PhD

### Resumen

La diversidad y abundancia de elementos genéticos móviles en la mayoría de los genomas eucariotas han hecho que estas secuencias de DNA moderadamente repetitivo sean importantes para el estudio y comprensión de algunos mecanismos de regulación génica y variabilidad genética en muchas especies. La investigación tiene la finalidad de identificar *de novo*, clasificar y caracterizar los elementos transponibles (ETs) clase II presentes en el genoma de *Drosophila amaguana*, una especie endémica neotropical que habita en los bosques andinos del Ecuador. Las secuencias de elementos móviles se anotaron estructuralmente a partir del genoma ensamblado (455.5 Mb) de *D. amaguana* utilizando el pipeline *The Extensive De Novo TE Annotator* (EDTA), el cual agrupa programas con distintos algoritmos para una identificación *de novo* de elementos transponibles en todo el genoma. La caracterización de un elemento completo en cada superfamilia de transposones TIR (una subclase de ETs de DNA de repetición terminal invertida) se realizó a través de *dot plots* con DOTTER, y la anotación funcional se llevó a cabo con Artemis. Se encontró que el 11.87 % del genoma de *D. amaguana* corresponde de elementos transponibles, del cual el 7.87 % pertenece a transposones de DNA, incluyendo ocho superfamilias como *Mutator*, *CACTA*, y *Helitron*, siendo estas tres las que mostraron mayor abundancia en el genoma de *D. amaguana*, a ellas se suma la presencia de superfamilias *PIF/Harbinger*, *Tc1/Mariner*, *hAT*, *Maverick*, y *elementos P*. Estos resultados constituyen el primer reporte sobre el contenido de elementos móviles de una especie neotropical en el Ecuador, y serán el punto de partida de análisis posteriores para comprender los procesos adaptativos y evolutivos del grupo *mesophragmatica* en los bosques andinos.

**Palabras clave:** Elementos transponibles / *Drosophila* / *Drosophila amaguana* / grupo *mesophragmatica* / DNA repetitivo / EDTA

**TITLE:** Identification of class II transposable genetic elements in the genome of *Drosophila amaguana* (Diptera, Drosophilidae)

**Author:** Manuel Alejandro Coba Males

**Director:** Doris Jimena Vela Peralta, PhD

### **Abstract**

The abundance and diversity of transposable elements in most eukaryotic genomes have made these moderately repetitive sequences essential tools for studying and comprehending the mechanisms of gene regulation and genetic variability in many species. The aim of this investigation was to *de novo* identify, classify, and characterize class II transposable elements (TEs) in the genome of *Drosophila amaguana*, a neotropical endemic species that inhabits the Andean forests of Ecuador. Mobile element sequences were structurally annotated from the assembly genome (455.5 Mb) of *D. amaguana* using The Extensive De Novo TE Annotator (EDTA) pipeline, which integrates programs with different algorithms for *de novo* identification of transposable elements in the whole genome. The characterization of a complete element for each superfamily of TIR transposons (a subclass of terminal inverted repeat DNA TEs) was performed through dot plots with DOTTER, and functional annotation was conducted using Artemis. We identified that 11.87 % of the *D. amaguana* genome corresponds to transposable elements, of which 7.87 % belong to DNA transposons, including eight superfamilies such as *Mutator*, *CACTA*, and *Helitron*. These three superfamilies exhibited high abundance in the genome of *D. amaguana*, and other sequences from *PIF/Harbinger*, *Tc1/Mariner*, *hAT*, *Maverick*, and *P element* were also identified. These results represent the first report on the content of mobile elements in a neotropical species of Ecuador, and they will serve as a starting point for further in-depth analyses to gain an understanding of the adaptative and evolutionary processes of the *mesophragmatica* group in the Andean forests.

**Keywords:** Transposable elements / *Drosophila* / *Drosophila amaguana* / *mesophragmatica* group / repetitive DNA / EDTA

## Capítulo I: Introducción

### Planteamiento del Problema

La presencia de elementos transponibles (ETs) en los genomas de un gran número de organismos pertenecientes a los tres dominios de la vida: Bacteria, Archaea, y Eukarya, ha conducido a realizar diversas investigaciones para conocer los efectos funcionales significativos que pueden influir en los rasgos de importancia evolutiva de las especies, sobre todo en organismos eucariotas (Tiedeman & Signor, 2021).

Los ETs son considerados componentes importantes de los genomas, pues el impacto que han tenido desde su descubrimiento en 1950 por Barbara McClintock ha abierto la investigación de una manera más profunda (Almojil et al., 2021). La capacidad que tienen estos elementos para propagarse y colonizar nuevas regiones del genoma mediante la transferencia horizontal (M. Kidwell, 1983), sugiere hipótesis en las que se habla de su participación en procesos celulares importantes que tienen lugar en el anfitrión, como por ejemplo, la regulación génica y la síntesis de proteínas (Chuong et al., 2016).

Por tanto, es primordial comprender los mecanismos de transposición a través de los cuales los ETs se mueven e inducen mutaciones a lo largo de todo el genoma, interrumpiendo secuencias funcionales (Kaminker et al., 2002). Dado que estas modificaciones generadas por elementos móviles, ayudan a entender la evolución al favorecer la variabilidad genética inter e intraespecífica a lo largo del tiempo (Biémont et al., 1999).

Entender la evolución desde los puntos de vistas de diversas teorías es un hecho interesante. Si bien la teoría de evolución darwiniana, propone una evolución por selección natural, siendo la más aceptada, ésta aún mantiene cuestiones que son debatibles, y que al ser contrastadas con conceptos planteados en otras teorías abren muchas interrogantes que pueden ser respondidas a través de nuevas hipótesis evolutivas (Colonna Romano & Fanti, 2022).

Y es precisamente la hipótesis de McClintock la que biológicamente mejor encaja con las ideas que plantean que la evolución se evidencia con apariciones abruptas de otras variantes después de periodos de estasis, donde no ha existido cambios en las características fenotípicas de una especie, convirtiéndose en un mecanismo de especiación (Colonna Romano & Fanti, 2022).

Los ETs representan una porción considerable del genoma de organismos eucariontes y en algunos casos su presencia está asociada al tamaño del genoma debido a su abundancia (Almojil et al., 2021). Así, los ETs tienen un impacto sobre la función y estructura del genoma en la adaptación y evolución de especies y poblaciones (Lerat et al., 2019). Pues algunas inserciones o deleciones en secuencias de DNA que se generan producto de la transposición de estos elementos, pueden conducir a efectos perjudiciales para el huésped (Bourque et al., 2018).

Uno de los modelos biológicos más estudiados, en cuanto se refiere a elementos transponibles desde hace más de 40 años, es el género *Drosophila* con aproximadamente 1 500 especies descritas (Markow, 2015). Las especies de *Drosophila* han permitido durante varios años aportar al conocimiento sobre los mecanismos de regulación y transposición de ETs (Mérel et al., 2020). Por consiguiente, *Drosophila* es un modelo que puede ayudar a comprender la aparición de algunas enfermedades humanas y variantes fenotípicas en plantas y animales, asociadas a la movilización de estos elementos (McCullers & Steiniger, 2017).

Los estudios sobre ETs intentan aclarar el panorama del curso que sigue la dinámica temporal evolutiva, explicando cómo los seres vivos están en la capacidad de adaptarse a entornos cambiantes para evitar su migración a otras áreas geográficas o extinguirse (Colonna Romano & Fanti, 2022). Ya que gracias a la expresión y transposición de ETs es posible inducir la variabilidad genética que permite al organismo responder a condiciones ambientales

estresantes inicialmente para la especie endémica; mostrando diferentes respuestas morfológicas, fisiológicas o conductuales (Colonna Romano & Fanti, 2022).

Por lo que, es indispensable la identificación de estas secuencias repetitivas en los genomas de cada una de las especies que se describen. En este sentido, el presente proyecto busca anotar, clasificar y caracterizar los ETs clase II presentes en el genoma de *Drosophila amaguana*, una especie descrita en el año 2004 cuya distribución es andina (Vela & Rafael, 2000, 2003, 2004).

### **Formulación del Problema**

¿Qué superfamilias de elementos transponibles clase II están presentes en el genoma de *Drosophila amaguana*?

### **Preguntas de Investigación**

- ¿Cómo es la estructura completa y funcionalidad de los elementos transponibles clase II de las superfamilias que se identificaron en el genoma de *Drosophila amaguana*?
- ¿Cuál es el porcentaje de los transposones de DNA que fueron encontrados en el genoma de *Drosophila amaguana*?
- ¿Cuál es la diferencia en el contenido de elementos transponibles entre el genoma de *Drosophila amaguana* y el modelo de estudio por excelencia, *Drosophila melanogaster*?

## **Objetivos**

### ***Objetivo General***

Identificar las superfamilias de elementos transponibles clase II en el genoma ensamblado de *Drosophila amaguana* mediante la anotación estructural *de novo*.

### ***Objetivos Específicos***

- Caracterizar estructural y funcionalmente los elementos transponibles clase II de la subclase TIR identificados en el genoma de *Drosophila amaguana*.
- Determinar el porcentaje de los elementos transponibles de clase II que forman parte del genoma de *Drosophila amaguana*.
- Comparar el contenido de elementos transponibles clase II identificados en *Drosophila amaguana* respecto a los reportados en *Drosophila melanogaster*.

## **Justificación e Importancia**

En el Ecuador, el “Refugio de Vida Silvestre Pasochoa” es uno de los bosques relictos de los Andes ubicado desde los 3 100 a 4 000 m.s.n.m., que permanece en condiciones poco alteradas (Céspedes & Rafael, 2017). En este bosque se estudió la diversidad de *Drosophila* y en el año 2004 se describieron tres especies pertenecientes al grupo *mesophragmatica* (Vela & Rafael, 2004).

El grupo *mesophragmatica* fue propuesto en 1957 por Brncic y Koref (Brncic & Koref, 1957). Es un grupo endémico de la región neotropical que actualmente cuenta con un número considerable de especies endémicas que se encuentran distribuidas en los bosques andinos y páramos sobre los 2 000 m.s.n.m (Figuro, 2017). En estos ecosistemas, las especies de *Drosophila* ocupan nichos como descomponedores, polinizadores, depredadores, parásitos, o vectores de organismos patógenos (Dajoz, 2000); todas ellas de tipo alopátricas, al ser

derivadas de un escenario evolutivo en el cual los sistemas montañosos permitieron su rápida especiación, originando poblaciones locales aisladas (Brncic, 1957; Nina R. Mota et al., 2008).

*Drosophila amaguana* es una especie de gran tamaño principalmente presente en lugares donde prima la alta vegetación, poca luminosidad y alta humedad (Céspedes & Rafael, 2017), perteneciente al grupo de *Drosophila mesophragmatica* (Vela & Rafael, 2004). Morfológicamente, *D. amaguana* presenta particularidades que la diferencia del resto de especies del grupo *mesophragmatica*, como el surestilo con dientes primarios y secundarios en los machos, y una espermateca pequeña no quitinizada y semicircular en las hembras (Ramos & Rafael, 2017; Vela & Rafael, 2004). Estas características son importantes para su identificación morfológica pero insuficientes para permitir abordar análisis evolutivos en este grupo (Romero, 2018).

Debido a que la abundancia y regulación de los ETs en los genomas es muy diferente incluso entre especies cercanas (Lerat et al., 2019), es necesario hacer un análisis de la diversidad de los elementos transponibles en cada genoma, pues estas variaciones pueden resultar en cambios significativos en la organización y mecanismos de regulación (genética y epigenética) de los genomas, dando como resultado la evolución de las especies (Lerat et al., 2019).

Y precisamente, la transposición de los elementos móviles es una fuente importante de nueva variabilidad genética que impulsa los procesos adaptativos a los cambios ambientales (Lerat et al., 2019), en algunos casos favoreciendo la aparición de nuevos fenotipos que podrían contribuir a una rápida evolución (García, 2012).

Lo que finalmente se pretende con el presente proyecto, es contribuir con información sobre ETs para entender los mecanismos de diversificación y adaptación de *Drosophila*, así como las relaciones evolutivas de estas especies que habitan los bosques andinos del Ecuador.

## Capítulo II: Revisión de la Literatura

### Antecedentes de la Investigación

Para el desarrollo del presente proyecto fue necesario la realización previa de una investigación bibliográfica, en esta etapa se encontró trabajos similares al que proponemos y que se detallan brevemente a continuación:

Tiedeman & Signor, en su estudio “**The transposable elements of the *Drosophila serrata* reference panel**” publicado en 2021 por *Genome Biology and Evolution*, reportaron 676 elementos transponibles en el genoma referencial de *D. serrata* distinguiendo elementos MITEs, Helitron, y TIR; encontraron correspondencias importantes en al menos 27 elementos relacionados con *D. melanogaster*, y 10 relacionados con *D. simulans*. Además, no se halló evidencia de *elementos P* en la especie de *D. serrata*.

Jeong et al., en 2018 publicaron en *Pacific Symposium on Computing* su estudio “**An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data**”, en el cual principalmente, desarrollaron un pipeline denominado SalmonTE como herramienta rápida para cuantificar elementos transponibles a partir de datos de RNA-seq, obtenidos por NGS. Su objetivo, superar limitaciones comunes que las herramientas de anotación comparten.

Sassi et al., en su estudio “**Transposable elements *P* and gypsy in natural populations of *Drosophila willistoni***” que fue publicado por la revista *Genetics and Molecular Biology* en el año 2005, reportaron *elementos P* y *gypsy* en el genoma de poblaciones de *D. willistoni* generados por enzimas de restricción mediante ensayos de Southern Blot, con la finalidad de comprender la variabilidad intrínseca sujeta a un proceso de selección natural, donde los ETs juegan un papel importante.

Kaminker et al., identificaron 1 572 elementos transponibles con secuencias completas o parciales en *D. melanogaster*, donde cada brazo cromosómico se compone de al menos 3.11 y 4.29 % de elementos transponibles. El estudio publicado bajo el título “**The transposable elements of the *Drosophila melanogaster*: euchromatin a genomics perspective**” en *Genome Biology* en el año 2002, fue realizado gracias a la disponibilidad de la versión 3 para el genoma eucromático de *D. melanogaster* por parte del *Berkeley Drosophila Genome Project*, los autores determinaron que al menos 3.86 % de esta secuencia corresponde a ETs.

Germanos et al., en 2006 publican en *Genetics and Molecular Biology* su investigación “**Transposable elements from the *mesophragmatica* group of *Drosophila*”**, en la cual mediante análisis Dot Blot y PCR, estudiaron la distribución de elementos móviles como: *I, mariner, hobo, gypsy, Tom/17.6, micropia*, y *elementos P*, en especies de *Drosophila* que conforman el grupo *mesophragmatica* tales como: *Drosophila pavani, Drosophila viracochi, Drosophila brncici, y Drosophila gasici*. Encontrando elementos *gypsy* en todas ellas, elementos *mariner* ausentes solo en *D. pavani*, y, secuencias homólogas a *elementos P* exclusivamente en *D. pavani* y *D. gasici*.

## **Fundamentación Teórica**

### ***DNA Repetitivo***

Las secuencias repetitivas son fragmentos de DNA que se encuentran a lo largo de todo el genoma en múltiples copias, exhibiendo un alto grado de polimorfismo como consecuencia de las mutaciones dentro de cada unidad repetida (Pathak & Ali, 2012). Este tipo de secuencias fueron identificadas con las primeras investigaciones llevadas a cabo por Britten y Kohne cuyo trabajo fue publicado en 1968 en la revista *Science* (Britten & Kohne, 1968). Su investigación manifiesta la presencia de grandes cantidades de DNA repetitivo en los genomas eucariotas de forma moderada o altamente repetitiva, logrando estimar que en el genoma de una especie es

posible encontrar hasta 90 % de DNA repetitivo y que la frecuencia de estas secuencias puede alcanzar hasta 1 millón de veces (López & Garrido, 2012).

Estas secuencias en su mayoría de carácter no codificante tienen la capacidad de remodelar el genoma a través de varios mecanismos, por tanto, tienen un impacto en la regulación de la expresión génica, mantenimiento de la integridad cromosómica y la estabilidad genómica, lo cual puede conducir a una divergencia evolutiva (Kuo et al., 2021; Pathak & Ali, 2012). Sin embargo, cuando fueron descubiertos se desconocía la función de regulación genómica que cumplían estos segmentos repetitivos de DNA dentro del genoma (Gemmell, 2021), y fueron catalogados como “DNA basura” (Kuo et al., 2021).

El DNA repetitivo es considerado como un recurso de variabilidad importante en la evolución de los genomas, y ha sido asociado con eventos de reordenamiento cromosómico que han dado lugar a variaciones de cariotipos, lo cual se ha evidenciado a través de estudios evolutivos de varias especies (Fablet et al., 2017; Zattera et al., 2020). Por ejemplo, en mamíferos el gen *syncytin*, esencial para la formación de placenta, se originó de un elemento transponible (Sha et al., 2000); en *Drosophila* la inserción del elemento transponible *Doc* en el gen *CHKov* le proporcionó la resistencia a pesticidas (Aminetzach et al., 2005); en ratones el gen *FvI* derivado del gen *gag* de un ET le confiere resistencia frente a un retrovirus al bloquear el avance del ciclo viral (Yan et al., 2009).

La clasificación de DNA repetitivo principalmente se lo hace en función de su organización estructural, la composición de su secuencia y su frecuencia, de esta forma encontramos dos grupos: repeticiones en tándem (secuencias altamente repetitivas) y elementos transponibles (secuencias medianamente repetitivas) (Kuo et al., 2021).

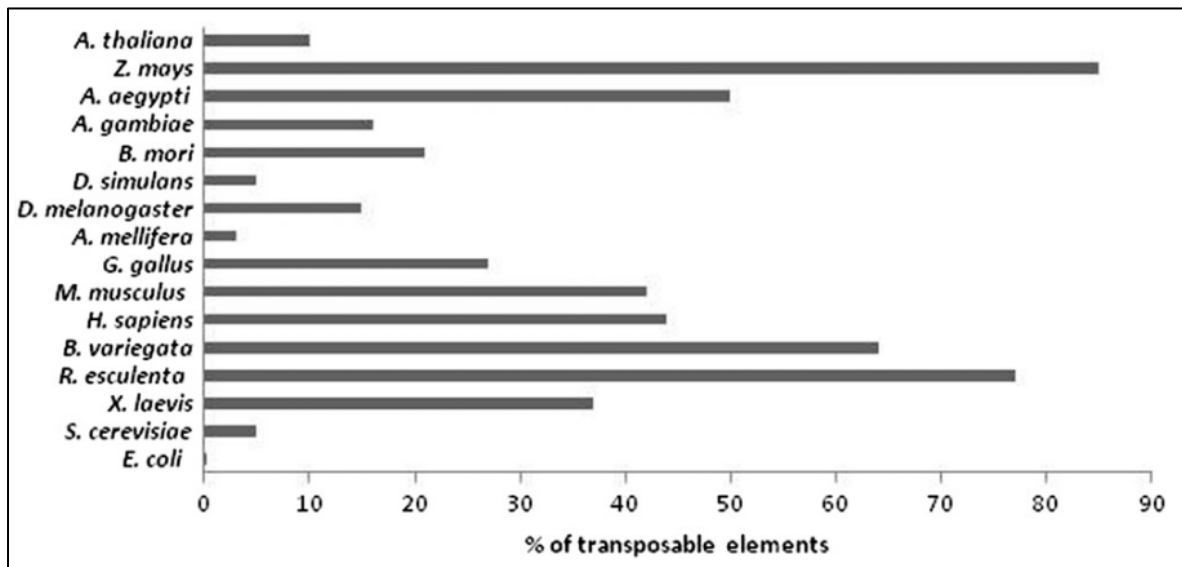
**Elementos Transponibles.** Son considerados secuencias de DNA que tienen autonomía para movilizarse, por tanto pueden cambiar de posición a lo largo de todo el genoma,

lo cual evidentemente tiene un efecto mutagénico (Bourque et al., 2018). Sin embargo, también los ETs son fuente de variabilidad genética que pueden llevar a una diversificación y evolución en las especies (Bourque et al., 2018). Bajo este concepto también reciben la denominación de elementos móviles, y aunque si bien su auto propagación genera consecuencias en el genoma, es importante comprender su impacto biológico (Saleh et al., 2019), pues algunos de estos elementos podrían ser utilizados como valiosas herramientas dentro del campo de la ingeniería genética (Kapitonov & Jurka, 2006).

El descubrimiento de estos elementos se lo atribuye a Barbara McClintock durante sus experimentos realizados con maíz desde el año 1944, incluso los bautizó bajo el nombre de elementos controladores, pues la evidencia sugería que estos elementos influían en rasgos fenotípicos (McClintock, 1950). A pesar de ello, su trabajo no fue bien recibido por personas que formaban parte de la comunidad genética, e incluso fue despreciado sino hasta los primeros años de la década de los 70 cuando el grupo Szybalski descubre secuencias de inserción en genomas bacterianos (Malamy et al., 1972). Con este trabajo, la comunidad científica reconoció el aporte que Barbara McClintock había hecho inicialmente (Makalowski et al., 2019).

Su presencia en la mayoría de organismos vivos los han vuelto en un componente principal del DNA repetitivo en los genomas de animales y plantas, en algunos casos representan una fracción importante del genoma con múltiples números de copias, por ejemplo, en el genoma del maíz se estima más del 50% de ETs (M. G. Kidwell & Lisch, 1997). En *Drosophila melanogaster* aproximadamente el 15% de todo el genoma corresponde a elementos transponibles, siendo su mayor distribución en las regiones de heterocromatina centromérica, donde el 80% de las mutaciones espontáneas son producidas por la inserción de elementos móviles o por reordenamientos cromosómicos asociados a estos (Deloger et al., 2009).

**Figura 1.** Porcentaje de elementos transponibles en genomas de diferentes especies.

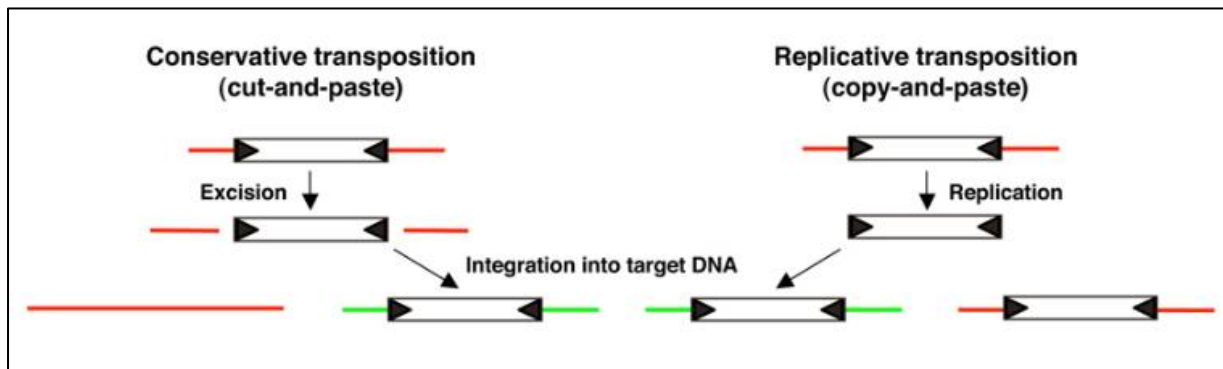


*Nota.* Adaptado de “A comparative analysis of the amounts and dynamics of transposable elements in natural populations of *Drosophila melanogaster* and *Drosophila simulans*” (p. 84), por Vieira et al., 2012, *Journal of Environmental Radioactivity*, 113.

Históricamente, la clasificación de ETs ha sido bien aceptada desde 1989 hasta la presente fecha, siendo Finnegan quien propuso por primera vez un sistema de clasificación diferenciando dos clases según el mecanismo intermedio de transposición (Finnegan, 1989). En este sistema, los ETs de clase I incluyen retrotransposones cuyo mecanismo de tipo “copiar y pegar” requieren de un intermediario de RNA a través de una reverso transcriptasa; mientras que, aquellos elementos de clase II abarcan transposones de DNA con un mecanismo “cortar y pegar” que no requieren de intermediarios de RNA para su transposición (Kojima, 2019).

Esta clasificación ha recibido algunas objeciones y diversas críticas que han desafiado este sistema de dos clases (Muñoz & Garcia, 2010). Con el descubrimiento de ETs bacterianos y eucariotas que no requieren un intermediario de RNA, y cuya transposición se lo realiza a través de un modelo de “copiar y pegar” junto con elementos transponibles de repetición invertida en miniatura (MITEs), se abrió la posibilidad de considerar una tercera clase de elementos no autónomos pero que presumiblemente requieren transposones autónomos para su movilidad (Wicker et al., 2007).

**Figura 2.** Mecanismos de transposición de los elementos móviles.



*Nota.* Adaptado de “Transposon-mediated Genome Manipulations in Vertebrates” (p. 14), por Ivics et al., 2009, *Nature methods*, 6 (6).

Esta situación particular llevo a Wicker y colaboradores a proponer una clasificación jerárquica desde niveles de clase hasta subfamilia, basándose en el modo de replicación y las características estructurales de los ETs aplicando algunos criterios mecanicistas y enzimáticos (Makalowski et al., 2019; Wicker et al., 2007).

Como se puede apreciar en la **Figura 3**, el objetivo de esta nueva clasificación fue diseñar un sistema unificado con criterios más profundos a los utilizados inicialmente por la idea de Finnegan, conservando su estructura básica (Piégu et al., 2015).

**Figura 3.** Sistema de clasificación jerárquica para elementos transponibles.

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
<b>Class I (retrotransposons)</b>					
LTR	<i>Copia</i>	→ GAG AP INT RT RH →	4-6	RLC	P, M, F, O
	<i>Gypsy</i>	→ GAG AP RT RH INT →	4-6	RLG	P, M, F, O
	<i>Bel-Pao</i>	→ GAG AP RT RH INT →	4-6	RLB	M
	<i>Retrovirus</i>	→ GAG AP RT RH INT ENV →	4-6	RLR	M
	<i>ERV</i>	→ GAG AP RT RH INT ENV →	4-6	RLE	M
DIRS	<i>DIRS</i>	→ GAG AP RT RH YR ←	0	RYD	P, M, F, O
	<i>Ngaro</i>	→ GAG AP RT RH YR →	0	RYN	M, F
	<i>VIPER</i>	→ GAG AP RT RH YR →	0	RYV	O
PLE	<i>Penelope</i>	← RT EN →	Variable	RPP	P, M, F, O
LINE	<i>R2</i>	— RT EN —	Variable	RIR	M
	<i>RTE</i>	— APE RT —	Variable	RIT	M
	<i>Jockey</i>	— ORF1 — APE RT —	Variable	RIJ	M
	<i>L1</i>	— ORF1 — APE RT —	Variable	RIL	P, M, F, O
	<i>I</i>	— ORF1 — APE RT RH —	Variable	RII	P, M, F
SINE	<i>tRNA</i>	— — —	Variable	RST	P, M, F
	<i>7SL</i>	— — —	Variable	RSL	P, M, F
	<i>5S</i>	— — —	Variable	RSS	M, O
<b>Class II (DNA transposons) - Subclass 1</b>					
TIR	<i>Tc1-Mariner</i>	→ Tase* ←	TA	DTT	P, M, F, O
	<i>hAT</i>	→ Tase* ←	8	DTA	P, M, F, O
	<i>Mutator</i>	→ Tase* ←	9-11	DTM	P, M, F, O
	<i>Merlin</i>	→ Tase* ←	8-9	DTE	M, O
	<i>Transib</i>	→ Tase* ←	5	DTR	M, F
	<i>P</i>	→ Tase ←	8	DTP	P, M
	<i>PiggyBac</i>	→ Tase ←	TTAA	DTB	M, O
	<i>PIF-Harbinger</i>	→ Tase* — ORF2 ←	3	DTH	P, M, F, O
	<i>CACTA</i>	→ Tase — ORF2 ←	2-3	DTC	P, M, F
Crypton	<i>Crypton</i>	— YR —	0	DYC	F
<b>Class II (DNA transposons) - Subclass 2</b>					
Helitron	<i>Helitron</i>	— RPA // Y2 HEL —	0	DHH	P, M, F
Maverick	<i>Maverick</i>	→ C-INT — ATP // CYP — POL B ←	6	DMM	M, F, O

**Structural features**

→ Long terminal repeats    ← Terminal inverted repeats    — Coding region    — Non-coding region

— Diagnostic feature in non-coding region    — // — Region that can contain one or more additional ORFs

**Protein coding domains**

AP, Aspartic proteinase    APE, Apurinic endonuclease    ATP, Packaging ATPase    C-INT, C-integrase    CYP, Cysteine protease    EN, Endonuclease

ENV, Envelope protein    GAG, Capsid protein    HEL, Helicase    INT, Integrase    ORF, Open reading frame of unknown function

POL B, DNA polymerase B    RH, RNase H    RPA, Replication protein A (found only in plants)    RT, Reverse transcriptase

Tase, Transposase (\* with DDE motif)    YR, Tyrosine recombinase    Y2, YR with YY motif

**Species groups**

P, Plants    M, Metazoans    F, Fungi    O, Others

*Nota.* Adaptado de “A unified classification system for eukaryotic transposable elements” (p. 974), por Wicker et al., 2007, *Nature Reviews Genetics*, 8 (12).

**Retrotransposones.** Los resultados obtenidos de la secuenciación del genoma completo en organismos eucariotas mostraron que estos genomas tienen más genes que codifican para una reverso transcriptasa en comparación a cualquier otra proteína (Finnegan, 2012).

Adicionalmente, las secuencias más grandes que están presentes en los genomas, se han generado gracias a un mecanismo de transcripción inversa; un descubrimiento sorprendente, pues estudios anteriores limitaban la capacidad de copiar RNA en DNA únicamente a los retrovirus (Finnegan, 2012).

La forma en que estos elementos móviles se integran en el genoma es mediante un intermediario de RNA, que gracias a un mecanismo de “copiar y pegar” logran conservar su ubicación original, al mismo tiempo que las copias creadas por la reverso transcriptasa permite la acumulación de estas secuencias en una o varias copias en otros sitios del genoma (Saleh et al., 2019).

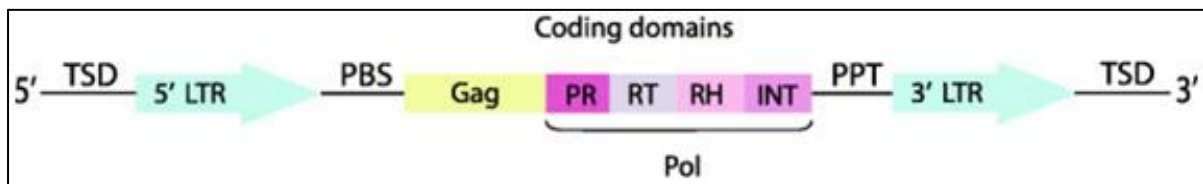
López & Garrido (2012), basados en la clasificación de Wicker et al. (2007), mencionan que los retrotransposones de eucariotas se dividen en 5 órdenes según la presencia de repeticiones terminales: (1) retrotransposones LTR; (2) retrotransposones de secuencias de repetición intermedia de *Dictyostelium* (DIRS); (3) retrotransposones non-LTR o LINEs; (4) retrotransposones de tipo Penélope (PLE); y, (5) SINEs. De los cuales, los retrotransposones LTR y non-LTR son los más abundantes en el reino Eukarya (López & Garrido, 2012).

***Retrotransposones LTR.*** Los retrotransposones de repetición terminal larga (LTR) están presentes en la mayoría de genomas eucarióticos, y proliferan en el genoma del huésped a través de múltiples mecanismos, convirtiéndose en uno de los principales grupos abundantes de retrotransposones (Havecker et al., 2004).

En su estructura contienen dos repeticiones terminales largas en ambos extremos con una longitud que va desde 250 a 600 pb flanqueadas por *target site duplication* (TSD), contiene un dominio interno entre 5 - 7 kb que abarca un tracto de polipurina (PPT), un *primer binding site* (PBS) y una zona de codificación de proteínas estructurales y enzimáticas con dos *open reading frames* (ORFs) a partir de dos genes (L. Zhang et al., 2014). El gen *gag* que codifica

proteínas estructurales que forman una partícula similar a un virus (VLP), y el gen *pol* codificante de cuatro dominios proteicos: una proteasa (PR) que corta la poliproteína Pol, una RNasa H (RH) que escinde el RNA de la unión DNA-RNA, una transcriptasa inversa (RT) que sintetiza cDNA a partir del RNA del retrotransposón y una integrasa (IN) que incluye el cDNA al genoma (Aroh & Halanych, 2021).

**Figura 4.** Estructura de los retrotransposones LTR.



*Nota.* Adaptado de “Genome-wide characterization of LTR retrotransposons in the non-model deep-sea annelid *Lamellibrachia luymesii*” (p. 2), por Aroh & Halanych, 2021, *BMC Genomics*, 22 (1).

El contenido de estos retrotransposones es muy alto en los genomas de plantas, reportando más del 50 % de retrotransposones LTR en el genoma de *Zea mays* con predominancia de las superfamilias *Copia* y *Gypsy* en comparación a los genomas animales (Neumann et al., 2019). En *D. melanogaster* se han reconocido las superfamilias *Gypsy*, *Copia* y *BEL/Pao* con al menos 35 familias, siendo el grupo *Gypsy* el más abundante (McCullers & Steiniger, 2017).

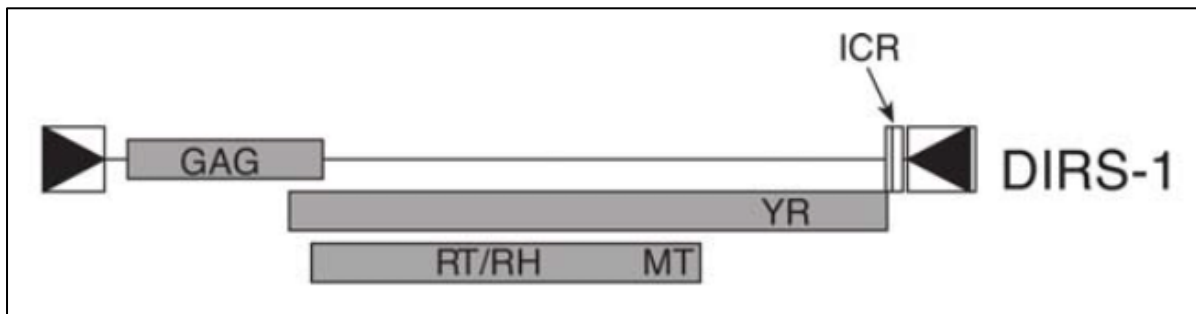
**Retrotransposones DIRS.** El grupo de retrotransposones DIRS o también llamados retrotransposones de tirosina recombinasa se encuentra en genomas fúngicos, y recibe su nombre por su miembro fundador, *Dictyostelium discoideum* (Muszewska et al., 2013). Estructuralmente carece de una proteasa y reemplazan la integrasa por una tirosina recombinasa (YR), la cual cumple la función de integración genómica (Malicki et al., 2020).

Este orden de retrotransposones es el menos comprendido de la clase I, ya que apenas se han descrito tres superfamilias: *DIRS1* en el moho mucilaginoso *Dictyostelium discoideum*,

*PAT* del nemátodo *Panagrellus redivivus* y *Prt1* del hongo zigomiceto *Phycomyces blakesleeanus* (Goodwin & Poulter, 2001).

Se caracterizan por que en los extremos las secuencias LTR que flanquean estos elementos no son idénticas y son repeticiones invertidas, que internamente abarcan una región complementaria interna (ICR) y las regiones de codificación proteica con tres ORFs largos y atípicos (Poulter & Butler, 2015). El primero de los ORFs corresponde al gen *gag* similar al de otros retrotransposones, el segundo ORF se superpone al primero extendiéndose hasta el extremo 3' del elemento transponible codificando una YR; y finalmente, su último ORF que se encuentra totalmente superpuesto al segundo ORF, codifica para una RT, RNasaH, y una metilasa (MT) con función desconocida (Duncan et al., 2002).

**Figura 5.** Estructura del retrotransposón *DIRS-1*.



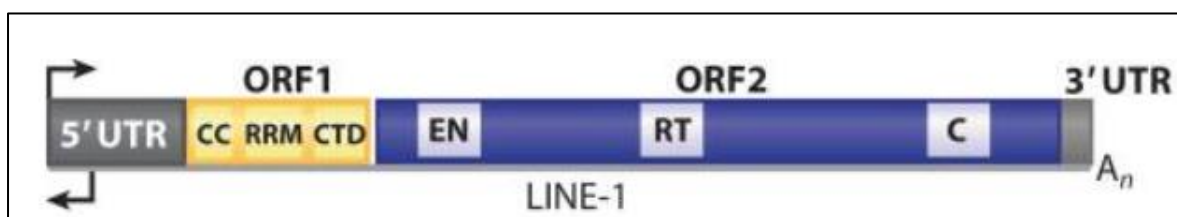
*Nota.* Adaptado de “Tyrosine Recombinase Retrotransposons and Transposons” (p. 3), por Poulter & Butler, 2015, *Microbiology Spectrum*, 3 (2).

**Retrotransposones non-LTR (LINEs).** La retrotransposición de LINEs han diversificado el genoma humano de tal manera que ahora comprenden hasta un 17 % de todo el genoma (Lander et al., 2001). Además, sus dos ORFs permiten la codificación de proteínas para la movilización de elementos móviles no autónomos y secuencias de RNAs no codificantes, generando algunos pseudogenes (Beck et al., 2011).

Los retrotransposones de repetición terminal no larga (non-LTR) conocidos como LINEs (elementos nucleares largos intercalados) tienen una longitud completa entre 4 a 7 kb y

están flanqueados por regiones variables de UTR 5' y 3' donde se incluyen dos ORFs (Qi & Sandmeyer, 2012). La 5'-UTR actúa como un promotor interno que es reconocido por la RNA *pol II* para la síntesis del transcrito (Beck et al., 2011). El primer ORF codifica una proteína que permite la unión a RNA inespecíficos, mientras que el siguiente ORF codifica una reverso transcriptasa (RT) y una proteína endonucleasa (EN) de DNA además de presentar un dominio rico en cisteína, todos ellos esenciales para la retrotransposición (Richardson et al., 2015).

**Figura 6.** Estructura del retrotransposón *LINE-1*.



*Nota.* Adaptado de “LINE-1 Elements in Structural Variation and Disease” (p. 189), por Beck et al., 2011, *Annual review of genomics and human genetics*, 12.

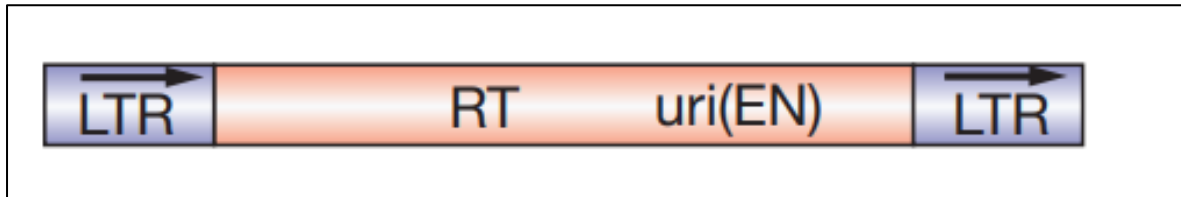
Cabe mencionar que la importancia de estos elementos, radica en la capacidad que tienen para movilizar otros retrotransposones no autónomos y mRNA hacia nuevas posiciones del genoma, dando lugar a pseudogenes (Esnault et al., 2000). Esto sucede debido a que los elementos no autónomos pueden ser reconocidos por su estructura, ya que terminan con secuencias ricas en poli A, están flanqueados por TSD variables y se pueden integrar a sitios consensos de la endonucleasa (Richardson et al., 2015).

***Retrotransposones de tipo Penélope (PLE).*** Pertenecen a retrotransposones poco estudiados que incluyen a un único representante, el elemento *Penélope*, al que se le ha atribuido las mutaciones visibles en la disgenesia híbrida en *Drosophila virilis* (Arkhipova, 2006).

Comparten un ancestro común con las transcriptasas inversas de la telomerasa, pueden detectarse en genomas de protistas, hongos, animales y plantas, en este último presentan una distribución irregular (Arkhipova et al., 2013). Su estructura es un tanto inusual y cuenta con

un único ORF que codifica una RT altamente especializada no móvil responsable del mantenimiento de los extremos cromosómicos, y una endonucleasa (EN) (Arkhipova et al., 2013); cuyo dominio pertenece a la superfamilia GIY-YIG que se descubrieron en el grupo I de los intrones móviles de bacterias (Dunin et al., 2006).

**Figura 7.** Estructura de los retrotransposones tipo Penélope.



*Nota.* Adaptado de “*Nonhomologous Recombination: Retrotransposons*” (p. 285), por Qi & Sandmeyer, 2012, *Encyclopedia of Biological Chemistry*, 4.

**SINES.** Estos retrotransposones corresponden a secuencias de elementos nucleares intercalados cortos que se los puede identificar en un abundante número de copias en la mayoría de los genomas de mamíferos; por ejemplo, en el ser humano alrededor del 10 % del genoma (equivalente a 300 Mb) se compone de la familia de SINEs denominada *Alu s* (Wang & Kirkness, 2005).

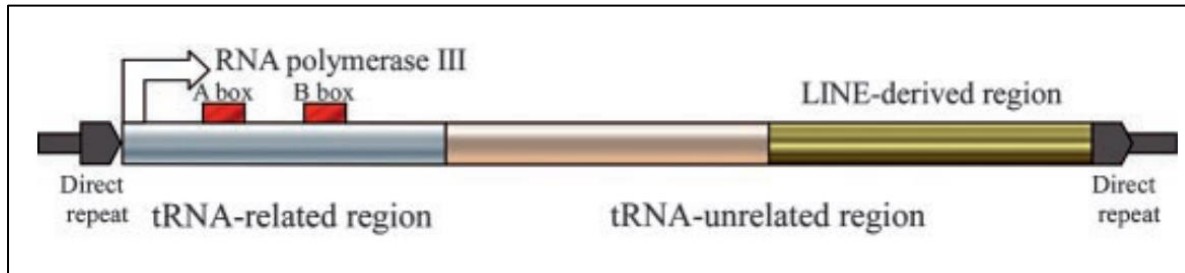
Muchos elementos de esta familia pueden inducir mutagénesis por inserción o sobrecruzamiento desigual (Kramerov & Vassetzky, 2011). Aunque de igual manera pueden generar beneficios al influir en procesos como la expresión génica, la poliadenilación y el splicing alternativo (Deininger, 2011).

Su tamaño varía de 100 a 600 pb y son catalogados como elementos no autónomos, ya que dependen de enzimas codificadas por los genes de los elementos LINEs para permitir su retrotranscripción y retrotransposición (Kanhayuwa & Coutts, 2016).

En su estructura se evidencian tres partes: 5' - terminal “*head*”, “*body*” y 3' - terminal “*tail*”; la secuencia en la cabeza 5' - terminal muestra similitud con uno de los RNA sintetizados por la RNA *pol III* cuyo origen proviene de *tRNA*, *7SL RNA* o *5S rRNA* (Kramerov &

Vassetzky, 2011). En su estructura se ha identificado regiones originarias de LINEs, mientras finalmente la cola 3' - terminal consta de secuencias de repeticiones simples de variable longitud (Kramerov & Vassetzky, 2011).

**Figura 8.** Estructura de los retrotransposones SINEs.



*Nota.* Adaptado de “*SINEs and LINEs: symbionts of eukaryotic genomes with a common tail*” (p. 478), por Ohshima & Okada, 2005, *Cytogenetic and Genome Research*, 110 (1-4).

**Transposones de DNA.** Esta clase II de ETs son el grupo más antiguos de elementos móviles que pueden encontrarse en el genoma de los organismos de todos los reinos (M. C. Thomas et al., 2010), salvo ciertas excepciones, como en el caso particular de *Saccharomyces cerevisiae* donde su genoma carece de transposones de DNA (López & Garrido, 2012).

Estos ETs se reorganizan a lo largo del genoma a través del mecanismo “cortar y pegar”, donde se escinde el elemento del DNA para posteriormente integrarlo en un nuevo entorno cromosómico, este proceso no replicativo, es realizado gracias a la intervención de una proteína transposasa que se encuentra flanqueada por *terminal inverted repeats* (TIRs) (Ivics et al., 2009).

Se los clasifica en 3 subclases en función de la relación de la transposasa y las características estructurales compartidas, abarcando las secuencias TIRs y la longitud de sus TSDs, siendo estos: (1) DNA transposones “cortar y pegar”; (2) DNA transposones de círculo rodante; y, (3) DNA transposones de autosíntesis (López & Garrido, 2012). Esta última clase se encuentra comúnmente más extendida en protistas, hongos y animales, siendo los

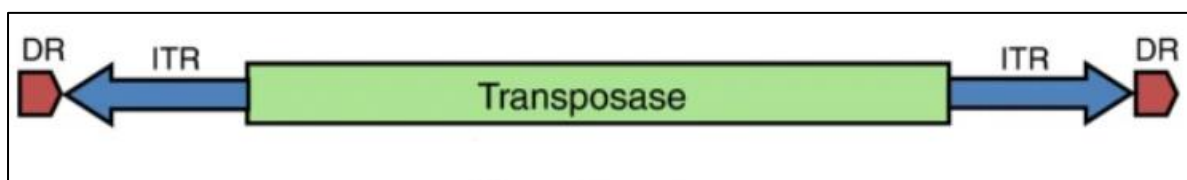
transposones de DNA más complejos, pues tienen la capacidad de codificar hasta 10 proteínas diferentes (Han et al., 2013).

**TIR.** Son elementos que contienen repeticiones terminales invertidas y pertenecen a la subclase I de los transposones de DNA propios de genomas eucariotas, a su vez comprenden 10 superfamilias de acuerdo con la similitud de sus transposasas (Su et al., 2019), mientras que en plantas existen apenas 5 grandes superfamilias (Feschotte & Pritham, 2007). Estos elementos tienen efectos significativos en la regulación epigenética, expresión génica, producción de genes quiméricos y variaciones en el número de copias (Su et al., 2019).

Estructuralmente presentan TIRs que van de 9 a 40 pb de longitud en ambos extremos, aunque esto depende de cada superfamilia, también comprenden internamente un ORF principalmente encargado de codificar una transposasa (Pray, 2008). Además, toda su estructura se encuentra flanqueada por cortas repeticiones directas idénticas denominadas TSD, que se originan debido a la inserción de un transposón en el sitio de destino (Shen et al., 2021).

Esta proteína según la triada de residuos ácidos en su dominio catalítico puede dar lugar a una transposasa DDE o DDD, que estructuralmente tienen un pliegue similar a la RNasa H (Nesmelova & Hackett, 2010), siendo precisamente, la encargada de reconocer aquellas secuencias TIRs de los ETs para mediar la movilidad hacia otra posición del genoma a través de un mecanismo conservador (Shen et al., 2021).

**Figura 9.** Estructura de los transposones de DNA de repeticiones terminales invertidas (TIR).

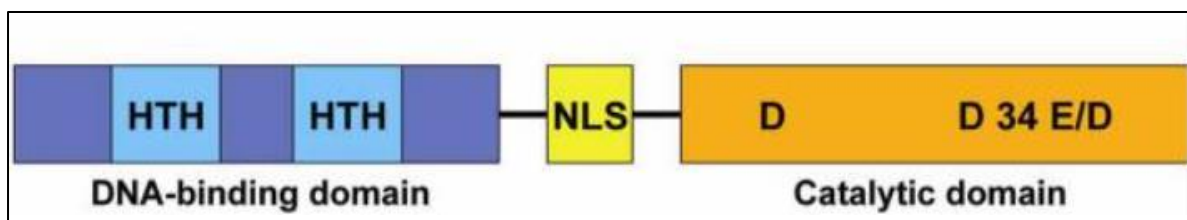


*Nota.* Adaptado de “*Mobile DNA transposition in somatic cells*” (p. 2), por Kazazian, 2011, *BMC Biology*, 9 (11).

***Tc1/Mariner***. Tienen una longitud entre 1 a 5 kb y está flanqueado con repeticiones terminales invertidas que pueden ir desde los 17 a 1 100 pb (Muñoz & Garcia, 2010). Codifica una transposasa de 282 - 345 AA con dos dominios principales: la región *N*-terminal donde se encuentra un motivo *helix - turn - helix* (HTH) para reconocer y unirse a los TIRs, y una región *C*-terminal donde se encuentra el sitio catalítico conformado por 3 residuos ácidos, DDD para elementos *mariner* o DDE para elementos *Tc1* (Muñoz & Garcia, 2010). Además, cuenta con una región de señal de localización nuclear (NLS) que permite el transporte de la transposasa a través de la membrana nuclear (Brillet et al., 2007).

Esta superfamilia es considerada como la mayormente distribuida en taxones como rotíferos, hongos, plantas, peces y mamíferos, apenas cuenta con alrededor de diez elementos activos (Plasterk et al., 1999).

**Figura 10.** Estructura del transposón *Tc1/Mariner*.



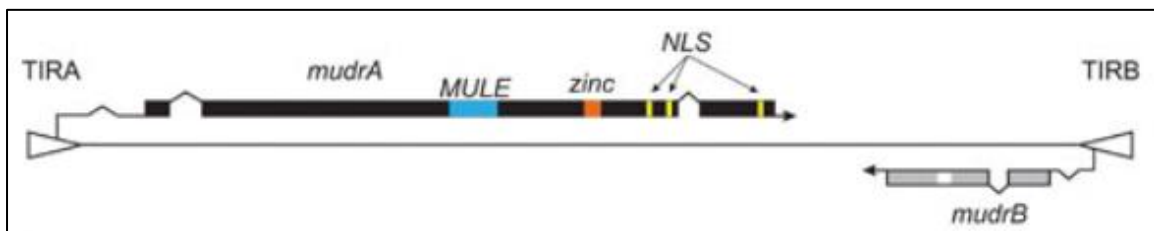
*Nota.* Adaptado de “*DNA Transposons: Nature and Applications in Genomics*” (p. 117), por Muñoz & Garcia, 2010, *Current Genomics*, 11 (2).

***hAT***. Con una longitud media entre 2.5 a 5 kb los elementos que forman parte de esta superfamilia presentan TIRs cortos entre 11 - 24 pb, y codifican una transposasa multidominio con 600 a 800 residuos (Muñoz & Garcia, 2010). Algunos de estos elementos reportados en especies lejanamente relacionadas como en plantas y moscas, por ejemplo, *Tam3* en *Antirrhinum majus*, *hobo* en *D. melanogaster*, y *Activator* en *Z. mays*, sugieren que existiría una transferencia horizontal de genes (Rubin et al., 2001).

***Mutator***. Es una familia altamente mutagénica en especies de maíz, los elementos autónomos *MuDR* son los más representativos de esta superfamilia que incluso permite la

transposición de algunos elementos no autónomos (Lisch, 2002). Presentan TIRs de aproximadamente de 215 pb, aunque en cada clase de transposón, internamente se puede identificar secuencias únicas que los diferencian unos de otros (Lisch, 2015). El gen *mudrA* codifica para una transposasa de 823 AA con un dominio catalítico DDE, mientras que el gen *mudrB* codifica un polipéptido de 207 AA del que aún se desconoce su función (Lisch, 2015).

**Figura 11.** Estructura del transposón *MuDR*.



*Nota.* Adaptado de “*Mutator and MULE Transposons*” (p. 2), por Lisch, 2015, *Microbiology Spectrum*, 3 (2).

**Merlin.** Esta superfamilia se ha identificado en genomas eubacterianos y animales, incluyendo nemátodos, platelmintos, mosquitos, ascidias, peces cebras y ranas, existiendo muy pocos reportes en genomas de plantas u hongos (Lopes et al., 2021). Poseen TIRs variables que van desde 24 a 462 pb, además, particularmente muestran similitud en las secuencias dentro y entre especies, y también codifican proteínas de 300 residuos que han mostrado similitudes con algunas transposasas codificadas por el grupo *IS 1016*, que pertenece a secuencias de inserción bacterianas (Feschotte, 2004).

**Transib.** Los elementos de esta superfamilia se han reconstruido a través de metodologías *in silico* partiendo de elementos fósiles degenerados de antiguas secuencias en genomas de *D. melanogaster* y *Anopheles gambiae* (Kapitonov & Jurka, 2003). Se los logró caracterizar bajo una longitud de 3 a 4 kb con TIRs entre 9 a 60 pb, que a su vez están flanqueadas por TSD de 5 pb con alto contenido de guanina-citosina (Chen & Li, 2008).

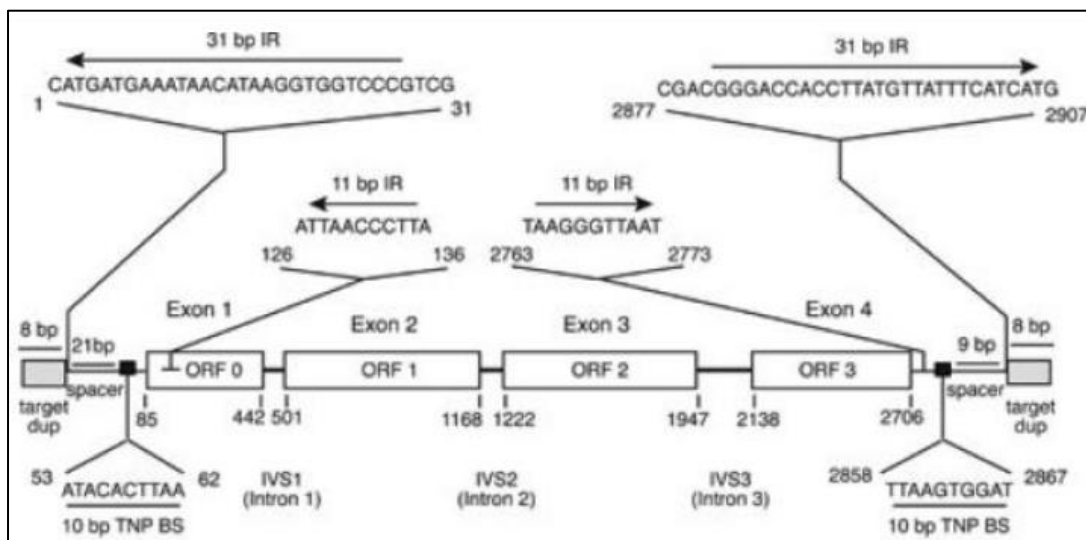
Se han reportado elementos autónomos y no autónomos en organismos eucariotas, como por ejemplo, en otras especies de *Drosophila*, anquilostoma del perro, gusano de seda,

erizo de mar, entre otras (Kapitonov & Jurka, 2005). Su transposasa tiene entre 595 a 708 AA pero particularmente no mantiene ninguna similitud con otras transposasas de las demás superfamilias (Chen & Li, 2008). Aunque se ha encontrado una semejanza con la proteína RAG1 de la recombinasa V(D)J, esencial en la diversidad inmunológica en los reordenamientos de los genes receptores de antígenos durante el desarrollo de los linfocitos (Fugmann, 2001).

**P elements.** Elementos estudiados con mucha intensidad en poblaciones naturales de *D. melanogaster*, estos transposones tienen un tamaño variable entre 0.5 a 2.9 kb, reportándose dos clases en particular: elementos completos con tamaño de 2 907 pb, y, elementos incompletos con secuencias internas de menor tamaño; en ambos casos las TIRs están dadas por una longitud de 31 pb (Simmons, 2013).

Codifican una transposasa compleja de seis dominios que abarca 751 AA únicamente cuando en los tejidos de la línea germinal se tiene al menos un elemento P completo, pues en los elementos incompletos, la proteína codificada no es funcional y por lo tanto el elemento pierde su capacidad de transposición (Ghanim et al., 2020; Simmons, 2013).

**Figura 12.** Estructura de un *elemento P* completo.



*Nota.* Adaptado de “P transposable elements in *Drosophila* and other eukaryotic organisms” (p. 2), por Majumdar & Rio, 2015, *Microbiology Spectrum*, 3 (2).

**PiggyBac.** Transposón identificado en el genoma de la polilla Cabbage Looper (*Trichoplusia ni*), aunque se ha encontrado también en genomas de plantas, hongos, animales e incluso en el ser humano; sin embargo, en estas especies hay una alta probabilidad de que los elementos transponibles de estas familias carezcan de actividad por mutaciones que han presentado (Muñoz & Garcia, 2010). Tiene un tamaño de 2.4 kb, con 13 pb correspondientes a TIRs y 19 pb que corresponden a repeticiones internas asimétricas adicionales, además presenta un ORF de 1.8 kb que codifica para una transposasa funcional (Zhao et al., 2016). Este es un elemento muy versátil que ha ganado campo para su uso en estudios genéticos, ya que a partir de la modificación de su transposasa se han direccionado investigaciones hacia la transgénesis, mutagénesis y terapia génica (Li et al., 2013).

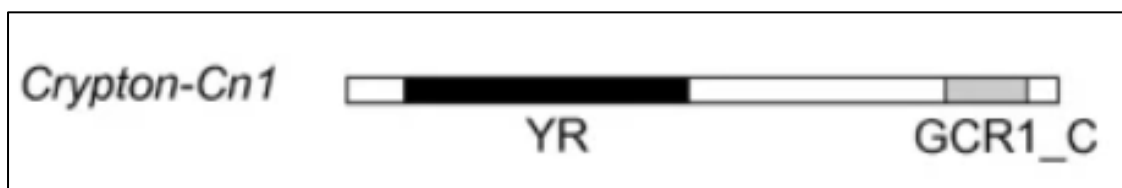
**PIF/Harbinger.** Los elementos de esta superfamilia predominan en el genoma del maíz, *Arabidopsis thaliana*, *Oryza sativa*, zanahoria y algunas otras especies del reino vegetal y animal (Grzebelus et al., 2007). Se trata de elementos autónomos con TIRs de 14 a 25 pb que se encuentran flanqueadas por TSDs de 3 pb (Sinzelle et al., 2008), internamente tiene dos ORFs que codifican hacia una transposasa catalítica semejante a la misma proteína de la secuencia de inserción bacteriana *IS5* (Grzebelus et al., 2007), y una segunda proteína de función hasta ahora desconocida pero de características similares a la proteína *Myb*, importante en los mecanismos transcripcionales de procesos celulares (Sinzelle et al., 2008). El estudio de estos elementos es importante para entender la función de genes domesticados que han derivado de esta clase de transposones, contribuyendo a la aparición de nuevas secuencias codificantes (Casola et al., 2007).

**CACTA.** Estos elementos han sido descritos en los últimos 20 años como parte de una de las superfamilias más abundantes de transposones clase II, identificados con demasía en los reinos vegetal, fungi y animal; observando que han contribuido moderadamente a la expansión de tamaño del genoma del huésped y su evolución (Belyayev et al., 2022). En su estructura

muestran dos TIRs de 10 a 28 pb, en donde en sus extremos tiene terminaciones de un motivo conservado CACTA o CACTG, además de repeticiones subterminales muy próximas entre sí, con 10 a 20 nucleótidos en orientación directa e invertida (Wicker et al., 2003). Su primer ORF codifica una transposasa con un centro catalítico muy conservado, mientras un segundo ORF lo hace para una proteína de función desconocida (Buchmann et al., 2014).

**Crypton.** Superfamilia detectada en varios hongos patógenos como *Cryptococcus neoformans*, *Coccidioides posadasii* e *Histoplasma capsulatum*, muchos de ellos altamente virulentos asociados a infecciones sistémicas graves, por lo cual se diferencian del resto de transposones (Goodwin et al., 2003). Estructuralmente resulta difícil caracterizarlos debido a que no presentan TIRs ni tampoco repeticiones largas directas, sino tan solo se ha podido identificar repeticiones directas cortas entre 4 a 6 pb que se cree que pueden actuar como sustratos para el reconocimiento de la tirosina recombinasa (YR), una proteína de aproximadamente 691 pb (Kojima & Jurka, 2011). Además, presentan un dominio proteico GCR1\_C conservado de unión al DNA de varios factores de transcripción fúngicos (Goodwin et al., 2003; Kojima & Jurka, 2011).

**Figura 13.** Estructura del elemento *Crypton-Cn1*.

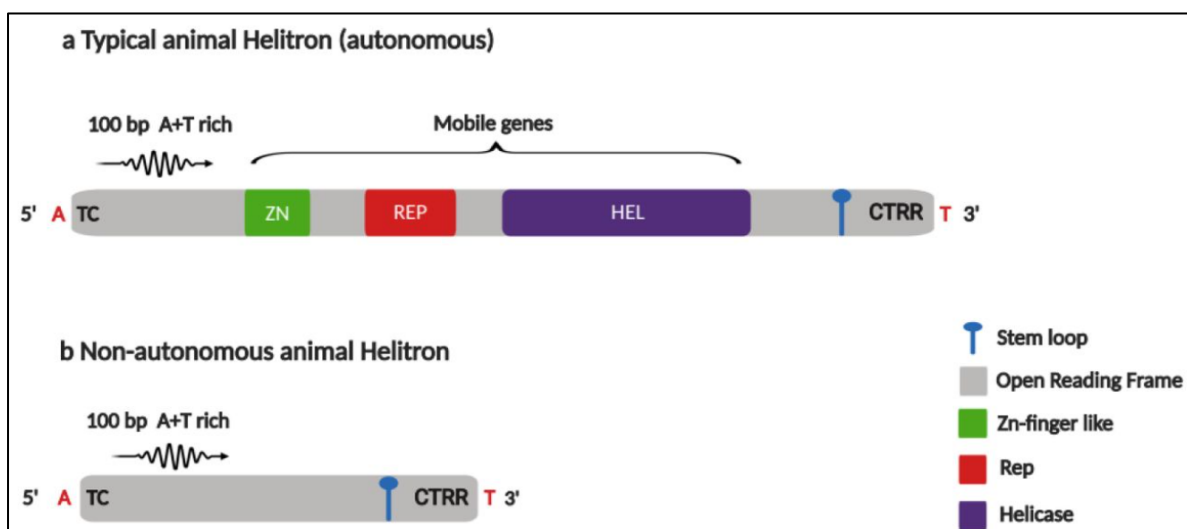


*Nota.* Adaptado de “*Crypton transposons: Identification of new diverse families and ancient domestication events*” (p. 3), por Kojima & Jurka, 2011, *Mobile DNA*, 2 (1).

**Helitron.** Carecen de repeticiones terminales invertidas y tienen una longitud desde 1 a 3 kb, con motivos bien conservados en 5'-TC y CTRR-3' que se transpone gracias a un mecanismo denominado replicación de círculo rodante (Kapitonov & Jurka, 2001). En genomas de plantas se han reportado algunos elementos Helitron autónomos, con tamaños que van desde 8 a 15 kb en especies como por ejemplo *A. thaliana*, *O. sativa* y *Caenorhabditis*

*elegans* (K. Hu et al., 2019). Esta superfamilia codifica una proteína multidominio de hasta aproximadamente 1 800 AA carente de un dominio catalítico semejante a una RNasa H (Grabundzija et al., 2018). Sino más bien, se observa un dominio Rep que contiene dos residuos de tirosina en el sitio activo cerca de la región *N*-terminal (Grabundzija et al., 2018), que da la actividad de endonucleasa HUH para permitir su unión al ssDNA. Seguidamente también se aprecia una helicasa encargada del desenrollamiento del DNA (J. Thomas & Pritham, 2015).

**Figura 14.** Estructura de transposones Helitron.



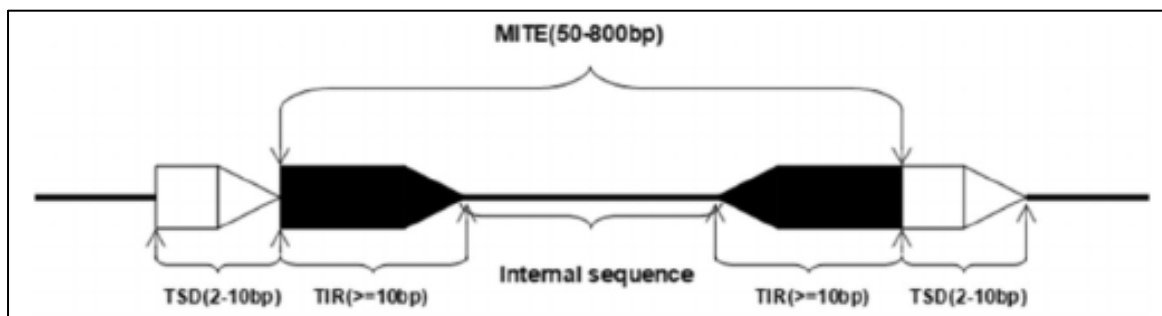
*Nota.* Adaptado de “Comparative analysis of morabine grasshopper genomes reveals highly abundant transposable elements and rapidly proliferating satellite DNA repeats” (p. 5), por Palacios-Gimenez et al., 2020, *BMC Biology*, 18 (1).

**Maverick/Polinton.** Son elementos muy grandes con un tamaño que puede ir desde 15 a 20 kb y con TIRs largas desde 100 hasta 1 000 pb, pueden codificar hasta 10 proteínas, algunas de ellas, homólogas a la de diversos virus de dsDNA (López & Garrido, 2012). Su importancia está en la capacidad que tienen para transferirse horizontalmente entre especies no relacionadas (Haapa et al., 2014). Se encuentran muy distribuidos en invertebrados y vertebrados no mamíferos, no siendo detectados en plantas (Pritham et al., 2007). Sin embargo, llama la atención su relación evolutiva con los virus, donde análisis filogenéticos han colocado a estos elementos dentro del linaje PRD1-adenovirus (Niño & Katzourakis, 2021), mismos que

se encuentran representados en la virosfera asociado a los tres dominios: arqueovirus, bacteriovirus y eucariovirus (Woo et al., 2019).

**MITEs.** Considerados como derivados truncados de transposones de DNA autónomos (Feschotte & Mouchès, 2000), se distribuyen principalmente en los brazos cromosómicos y son los más abundantes en los genomas de plantas (Macko et al., 2021). Tienen una longitud aproximada entre 100 a 800 pb, poseen TIRs menores a 10 pb y apenas cortos TSDs de tamaño desde 2 a 10 pb (Venkatesh & Nandini, 2020). Estos elementos se activan naturalmente cuando el huésped está expuesto a condiciones ambientales severas como respuesta para mitigar el peligro (McClintock, 1984).

**Figura 15.** Estructura de los transposones MITEs.



*Nota.* Adaptado de “MiteFinderII: A novel tool to identify miniature inverted-repeat transposable elements hidden in eukaryotic genomes” (p. 52), por J. Hu et al., 2018, *BCM Medical Genomics*, 11 (5).

### ***Drosophila***

Género que abarca más de 1 500 especies de pequeñas moscas conocidas comúnmente bajo el nombre de moscas de la fruta, pertenecen a la familia Drosophilidae y se distribuyen a lo largo de regiones tropicales, zonas alpinas, ciudades, desiertos, pantanos y selva tropical (Perveen, 2018). Constituye uno de los organismos modelos más versátiles con importancia en la investigación científica y médica (Animal Research, s.f.). *Drosophila* es un género en el que se han realizado diversos trabajos sobre diferentes aspectos que involucran su biología, para el

estudio de varios fenómenos por su rápido ciclo de vida, rentabilidad y tecnologías avanzadas disponibles (Harnish et al., 2021; Jennings, 2011).

Durante un siglo, la participación de *Drosophila* en el desarrollo de la biología ha contribuido notablemente para estudiar y probar teorías evolutivas, sentando las bases de la genética y permitiendo la comprensión de algunos mecanismos moleculares (Martinez, 2008). Desde hace más de cien años se le atribuye a Thomas Hunt Morgan como el padre de la investigación de *Drosophila*, al ser quién refinó la teoría cromosómica de la herencia planteada por Mendel, al demostrar que el gen *white* se localizaba en el cromosoma X (Tolwinski, 2017).

Particularmente *D. melanogaster* se ha convertido en el modelo común para estudios genéticos en la biología moderna, ya que se la considera como un organismo sumamente importante en biología, genética, medicina, entre otras ciencias del área de la salud y la vida (Perveen, 2018).

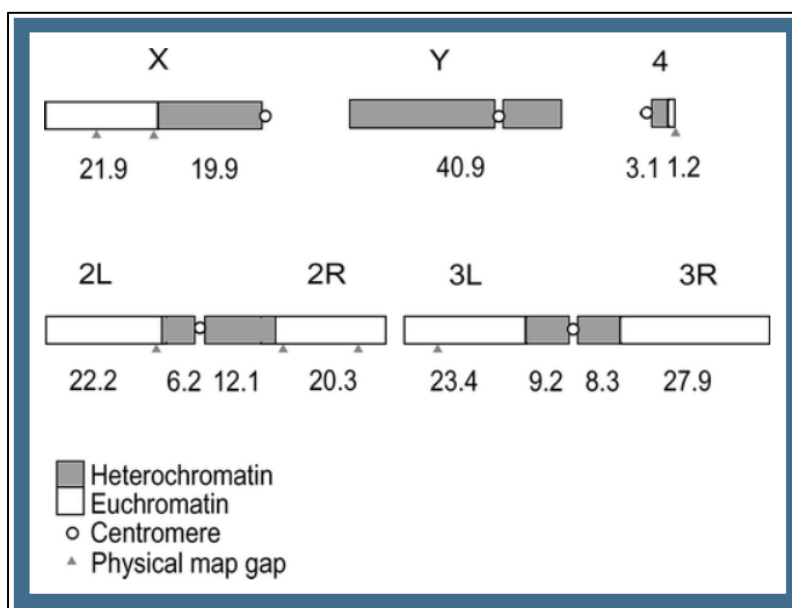
La secuenciación del genoma de *D. melanogaster* a inicios del siglo XXI fue uno de los resultados importantes que se han podido obtener (Adams et al., 2000; Myers et al., 2000). Años más tarde al secuenciarse el genoma humano (Craig Venter et al., 2001), las comparaciones de las secuencias nucleotídicas entre ambos organismos mostraron una identidad global de casi el 40% entre *Drosophila* y mamíferos, con una identidad probablemente mayor al 80% en cuanto se refiere a la conservación de dominios proteicos funcionales (Yamaguchi & Yoshida, 2018).

El genoma de *D. melanogaster* comprende cuatro pares de cromosomas: los cromosomas sexuales X/Y, dos cromosomas autosómicos II y III y un cromosoma muy pequeño IV, siendo común encontrar la mayor cantidad del contenido genético en los cromosomas X, II y III (Fernández et al., 2007). El genoma fue publicado en marzo del año

2000 convirtiéndose en el segundo y más largo genoma animal que había sido secuenciado (Kornberg & Krasnow, 2000).

El tamaño del genoma en *D. melanogaster* es de aproximadamente 180 Mb, del cual dos tercios corresponde a regiones eucromáticas (120 Mb) y el tercio restante (60 Mb) se trata de heterocromatina, que principalmente está formada por repeticiones de secuencias simples (Celniker & Rubin, 2003).

**Figura 16.** Estructura cromosómica de *Drosophila melanogaster*



*Nota.* Adaptado de “*The Drosophila Melanogaster Genome*” (p. 91), por Celniker & Rubin, 2003, *Annual Review of Genomics and Human Genetics*, 4.

La anotación funcional del genoma R6.13 secuenciado de *Drosophila* logró identificar 17 728 genes, de donde 13 907 se traducen en proteínas que codifican 21 953 polipéptidos únicos; mientras que los 3 821 genes restantes codifican diversos RNA no codificantes, entre ellos rRNA, tRNA, snRNA, snoRNA, miRNA, lncRNA y algunos pseudogenes (Kaufman, 2017).

**Figura 17.** Genes reportados en el genoma de *D. melanogaster* FB2016\_05 (R6.13)

Gene Type	Number
Protein coding	13,907
rRNA	147
tRNA	313
snRNA	31
snoRNA	288
miRNA	256
LncRNA	2,470
Pseudogenes	315

*Nota.* Adaptado de “*A Short History and Description of Drosophila melanogaster Classical Genetics: Chromosome Aberrations, Forward Genetic Screens, and the Nature of Mutations*” (p. 674), por Kaufman, 2017, *Genetics*, 206 (2).

La diversidad fenotípica (morfología, ecología y comportamiento) de las especies de *Drosophila* se puede analizar a través de la genómica comparativa, ya que permite observar las funciones conservadas frente a la divergencia de secuencias producidas por los factores ambientales a los cuales cada especie debe adaptarse (Clark et al., 2007). La **Figura 18** nos permite apreciar la anotación de los genomas en 12 especies diferentes de *Drosophila*.

**Figura 18.** Resumen de anotaciones de los genomas para 12 especies de *Drosophila*.

	Protein-coding gene annotations		tRNA (pseudo)	Non-coding RNA annotations				Repeat coverage (%) <sup>*</sup>	Genome size (Mb; assembly†/flow cytometry‡)
	Total no. of protein-coding genes (per cent with <i>D. melanogaster</i> homologue)	Coding sequence/intron (Mb)		snoRNA	miRNA	rRNA (5.8S + 5S)	snRNA		
<i>D. melanogaster</i>	13,733 (100%)	38.9/21.8	297 (4)	250	78	101	28	5.35	118/200
<i>D. simulans</i>	15,983 (80.0%)	45.8/19.6	268 (2)	246	70	72	32	2.73	111/162
<i>D. sechellia</i>	16,884 (81.2%)	47.9/21.9	312 (13)	242	78	133	30	3.67	115/171
<i>D. yakuba</i>	16,423 (82.5%)	50.8/22.9	380 (52)	255	80	55	37	12.04	127/190
<i>D. erecta</i>	15,324 (86.4%)	49.1/22.0	286 (2)	252	81	101	38	6.97	134/135
<i>D. ananassae</i>	15,276 (83.0%)	57.3/22.3	472 (165)	194	76	134	29	24.93	176/217
<i>D. pseudoobscura</i>	16,363 (78.2%)	49.7/24.0	295 (1)	203	73	55	31	2.76	127/193
<i>D. persimilis</i>	17,325 (72.6%)	54.0/21.9	306 (1)	199	75	80	31	8.47	138/193
<i>D. willistoni</i>	15,816 (78.8%)	65.4/23.5	484 (164)	216	77	76	37	15.57	187/222
<i>D. virilis</i>	14,680 (82.7%)	57.9/21.7	279 (2)	165	74	294	31	13.96	172/364
<i>D. mojavensis</i>	14,849 (80.8%)	57.8/21.9	267 (3)	139	71	74	30	8.92	161/130
<i>D. grimshawi</i>	15,270 (81.3%)	54.9/22.5	261 (1)	154	82	70	32	2.84	138/231

<sup>\*</sup> Repeat coverage calculated as the fraction of scaffolds >200kb covered by repeats, estimated as the midpoint between BLASTER-tx + PILER and RepeatMasker + ReAS (Supplementary Information section 3). †Total genome size estimated as the sum of base pairs in genomic scaffold >200,000bp. ‡Genome size estimates based on flow cytometry<sup>28</sup>.

*Nota.* Adaptado de “*Evolution of genes and genomes on the Drosophila phylogeny*” (p. 205), por Clark et al., 2007, *Nature*, 450 (8).

**Elementos transponibles en *Drosophila*.** Comprender la variabilidad genética inducida por elementos transponibles en poblaciones naturales, ha sido posible gracias a estudios de dinámica poblacional de ETs llevados a cabo durante más de 25 años, utilizando *D. melanogaster* como organismo modelo (Barrón et al., 2014). Los reportes del contenido de ETs en *Drosophila* han sido obtenidos a través de estudios interespecíficos, pues la abundancia y actividad de estos elementos repetitivos están relacionadas al genotipo del huésped y al medio ambiente en el cual se desarrolla (Mérel et al., 2021).

La activación de ETs es inducida por dos principales causas: el estrés ambiental y/o shocks genómicos, como por ejemplo, la hibridación, esto permite una divergencia en la secuencias reguladoras y/o en el contenido de ETs (Banho et al., 2021). Estos elementos móviles sufren patrones evolutivos complejos a través de mecanismos como: transferencia horizontal, pérdida estocástica, inactivación y degeneración de secuencias, y, reactivación por recombinación de secuencias (Germanos et al., 2006), pese a esto, se han reportado 96 familias de ETs para *D. melanogaster* (Kaminker et al., 2002).

Las repercusiones que tienen los ETs a nivel de especie son importantes para considerar, ya que del 80 % de mutaciones fenotípicas generadas por elementos transponibles en *Drosophila*, algunas resultan perjudiciales (Mérel et al., 2020). Cerca del 5 – 10 % de inserciones por *elementos P* activos genera mutaciones recesivas letales en *D. melanogaster* (Cooley et al., 1988), así como, la transposición somática de elementos *mariner* en *D. simulans* puede disminuir la esperanza de vida en la especie (Nikitin & Woodruff, 1995).

Sin embargo, los ETs siguen mostrando un papel relevante en la adaptación ambiental (Casacuberta & González, 2013), pues, las mutaciones de gran variedad y magnitud generadas por estos elementos, tienen en algunos casos la capacidad de responder a los cambios ambientales, fijándose rápidamente en las poblaciones (Biémont & Vieira, 2006).

## Capítulo III: Metodología

### Diseño de la Investigación

#### *Enfoque de la Investigación*

El trabajo investigativo se realizó dentro del enfoque cualitativo, ya que llevando a cabo un proceso inductivo y haciendo uso de herramientas bioinformáticas se logró de forma objetiva anotar estructuralmente y caracterizar funcionalmente los elementos genéticos móviles clase II en el genoma de *D. amaguana*.

#### *Alcance de la Investigación*

El alcance de este proyecto de desarrollo fue catalogado bajo un nivel exploratorio, porque se requirió de un acercamiento inicial a la temática debido a que se ha investigado muy poco en cuanto se refiere a la identificación de elementos transponibles para genomas de especies de *Drosophila* reportadas en Ecuador.

Por otra parte, también cayó dentro de un nivel descriptivo, con el cual se logró eficazmente identificar elementos transponibles de clase II en *D. amaguana*. Esto permitió establecer una clasificación de los ETs por superfamilias, reportando el porcentaje de ETs en el genoma de *D. amaguana*. Posteriormente estos resultados fueron comparados con el contenido de los elementos móviles clase II reportados en estudios previos para la especie de *D. melanogaster*.

#### *Tipo de Investigación*

Principalmente este estudio recurrió a los tipos de investigación que se enlistan a continuación:

- Investigación básica, misma que permitió adquirir información y ampliar el conocimiento teórico sobre los elementos móviles, el genoma de *Drosophila* y la relación entre ambos.

- Investigación documental, aquella que brindó información bibliográfica sobre los algoritmos para identificación de ETs clase II, así como, los softwares apropiados para estos análisis de identificación y anotación estructural específicos.
- Investigación aplicada, principalmente porque se logró hacer uso de conocimientos previamente adquiridos en el campo de la biología computacional, complementando con la información teórica recopilada en las primeras etapas del desarrollo de este trabajo.

## Metodología

El genoma ensamblado de *D. amaguana* fue proporcionado con anterioridad por la directora de este trabajo en colaboración con otros investigadores. El genoma se obtuvo mediante secuenciación Illumina, y su ensamblaje *de novo* se realizó con la herramienta *Maryland Super-Read Celera Assembler* (MaSuRCA) (Zimin et al., 2013). Además, integridad del ensamblaje del genoma se evaluó a través de la herramienta *Benchmarking Universal Single-Copy Orthologs* (BUSCO) versión 2.0 (Simão et al., 2015)

La anotación estructural para identificar los elementos transponibles clase II en el genoma ensamblado de *D. amaguana* fue realizada a través de la herramienta *The Extensive de novo TE Annotator* (EDTA) v.2.1.0. (<https://github.com/oushujun/EDTA>) (Ou et al., 2019) basado en el lenguaje de programación Perl. EDTA consiste en un pipeline para la anotación *de novo* de elementos transponibles en genomas eucariotas recién ensamblados (Ou et al., 2019), esto es realizado a través de múltiples softwares de código abierto específicos para una subclase particular de ETs, produciendo una biblioteca no redundante de estas secuencias para anotarlas estructuralmente intactas y fragmentadas, con un conjunto de programas de alta sensibilidad, especificidad y precisión (Ou et al., 2019). Esto permitió posteriormente llevar a cabo la caracterización de un elemento completo en cada superfamilia de la subclase TIR y la

anotación de su secuencia, utilizando DOTTER v.4.22.1. (Sonnhammer & Durbin, 1995) y Artemis v.17.0.1. (Carver et al., 2012) respectivamente.

### ***Evaluación de la calidad del ensamblaje del genoma***

El archivo fasta proporcionado con el genoma ensamblado de *D. amaguana*, fue sometido a una evaluación de calidad de su ensamblaje a través de la herramienta *Quality Assessment Tool for Genome Assemblies* (QUAST) v.5.2.0. (Gurevich et al., 2013).

### ***Preparación del entorno de trabajo***

Se realizó la creación de un entorno conda en el servidor de la Pontificia Universidad Católica del Ecuador que cuenta con un sistema operativo CentOS Linux 7 (Core), en él se instaló el paquete EDTA (<https://github.com/oushujun/EDTA.git>) que incluye varios programas de anotación de elementos transponibles, tales como: LTRharvest (Ellinghaus et al., 2008), LTR\_FINDER (Xu & Wang, 2007), LTR\_FINDER\_parallel (Ou & Jiang, 2019), LTR\_retriever (Ou & Jiang, 2018), TIR-Learner (Su et al., 2019), Generic Repeat Finder (Shi & Liang, 2019), HelitronScanner (Xiong et al., 2014), TESorter (R. G. Zhang et al., 2022), y, RepeatModeler (<http://www.repeatmasker.org/RepeatModeler/>).

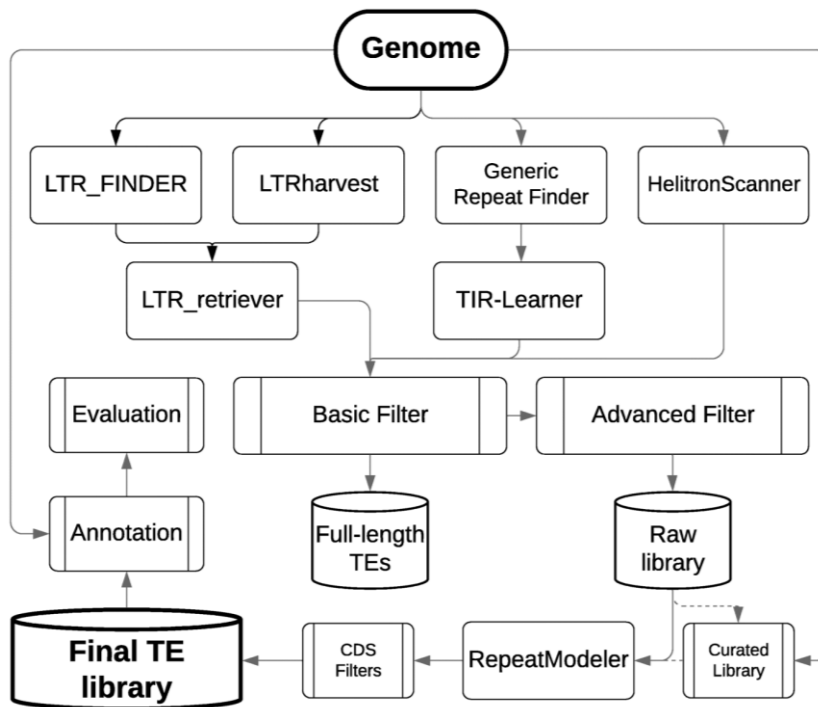
### ***Pipeline EDTA***

Se comenzó creando un directorio de trabajo dentro del cual se encontraba el archivo en formato FASTA del genoma ensamblado de *D. amaguana*, obtenido partir de los datos de secuenciación con lecturas *paired-end* provenientes de Illumina.

A continuación, se ejecutó el script EDTA.pl con el genoma ensamblado de *D. amaguana* en formato FASTA, el script contiene un conjunto de comandos que secuencialmente filtra los outputs de cada programa para reducir descubrimientos falsos positivos (Ou et al., 2019). El flujo de trabajo de este pipeline se evidencia en la **Figura 19** y brevemente se describe a continuación.

El pipeline inició con un filtro básico para la eliminación de repeticiones en tándem y secuencias cortas en los candidatos a transposones identificados por cada una de las herramientas que lo conforman. Posteriormente, otros filtros de mayor nivel son aplicados para la eliminación de secuencias falsas positivas, esto permite la generación de una biblioteca de elementos transponibles completa y de alta calidad. A partir de los ETs no redundantes, el genoma es enmascarado y la porción desenmascarada es analizada por RepeatModeler para identificar ETs que no fueron detectados por su estructura. Finalmente, todos estos elementos restantes se agregan a la biblioteca anterior para la obtención de una biblioteca final.

**Figura 19.** Workflow en el pipeline EDTA.



*Nota.* Adaptado de “*Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline*” (p. 11), por (Ou et al., 2019), *Genome Biology*, 20 (275).

### **Caracterización de elementos transponibles**

Esta etapa se realizó bajo una distribución de Ubuntu 20.04 LTS de Linux. La caracterización de ETs fue llevada a cabo a partir del archivo de anotación en formato GFF3 para los elementos intactos reconocidos por el pipeline de EDTA. Se inició con la separación y ordenamiento descendente de todas las anotaciones para cada superfamilia en función de la

longitud de la secuencia para cada transposón de forma independiente, según el código ID asignado a cada superfamilia acorde la ontología de secuencias para ETs ([https://github.com/oushujun/EDTA/blob/master/util/TE\\_Sequence\\_Ontology.txt](https://github.com/oushujun/EDTA/blob/master/util/TE_Sequence_Ontology.txt)). Esto se realizó a través de la terminal utilizando el lenguaje de programación AWK y la salida del resultado fue dirigida hacia un archivo GFF3 nuevo.

Posteriormente dentro de cada archivo para cada superfamilia se procedió a separar las secuencias clasificadas como MITEs en un nuevo archivo de anotación, para dejar únicamente aquellas secuencias que podrían corresponder a elementos transponibles intactos en un solo archivo de tipo GFF3 que contenía información en 7 columnas (ver **Figura 20**).

**Figura 20.** Visualización parcial de un archivo de anotación para ETs intactos.

(A)	(B)	(C)	(D)	(E)	(F)	(G)
CACTA_TIR_transposon	4942	7180002881543	Name=TE_00001609	Classification=DNA/DTC	Identity=0.9	TIR=CACTGTAAGT_ACTTACAGCG
CACTA_TIR_transposon	4939	7180002876874	Name=7180002876874:22455..27394	Classification=DNA/DTC	Identity=1	TIR=CACTAATACA_TGTATTAGTG
CACTA_TIR_transposon	4939	7180002852805	Name=TE_00000970	Classification=DNA/DTC	Identity=0.9	TIR=CACTAAGAG_CTTTTAGTG
CACTA_TIR_transposon	4931	7180002869268	Name=TE_00001621	Classification=DNA/DTC	Identity=0.9	TIR=CACTATCAGA_TCTGATAGAG
CACTA_TIR_transposon	4924	7180002787722	Name=TE_00001959	Classification=DNA/DTC	Identity=0.9	TIR=CACTAGGGGA_TCGCCTTGTG
CACTA_TIR_transposon	4895	7180002874058	Name=TE_00001650	Classification=DNA/DTC	Identity=0.9	TIR=CACTGAAAGA_TCATTCAGTG
CACTA_TIR_transposon	4875	7180002868810	Name=TE_00000944	Classification=DNA/DTC	Identity=0.9	TIR=CACTGGGGTA_TACCCCATTTG
CACTA_TIR_transposon	4828	7180002880582	Name=TE_00001463	Classification=DNA/DTC	Identity=1	TIR=CACTGCAGCA_TGCTGCAGTG
CACTA_TIR_transposon	4828	7180002868961	Name=TE_00001463	Classification=DNA/DTC	Identity=1	TIR=CACTGCAGCA_TGCTGCAGTG
CACTA_TIR_transposon	4785	7180002818701	Name=TE_00001174	Classification=DNA/DTC	Identity=0.9	TIR=CACTGCTAAA_TTCAGCAGTG
CACTA_TIR_transposon	4760	7180002883895	Name=TE_00000781	Classification=DNA/DTC	Identity=0.9	TIR=CACTGCAAAA_TTTGAAGTG

*Nota.* Archivo de anotación en formato GFF3 modificado para la superfamilia CACTA que contiene información principal que permitirá caracterizar y anotar manualmente los elementos con estructura completa. (A) Nombre de la superfamilia del ET; (B) Longitud del elemento transponible; (C) Contig del genoma donde se identificó el ET; (D) Nombre del elemento transponible; (E) Código ID asignado al ET; (F) Valor de identidad que representa la divergencia entre sus TIRs; (G). Secuencia de los TIRs para el ET. **Elaborado por:** Coba, M.

A continuación, desde el archivo resultante de la librería que contenía todas las secuencias de los elementos transponibles bajo el formato EDTA.TElib.fa identificados en el genoma por el pipeline, se extrajo en formato FASTA una por una la secuencia de cada elemento transponible cuyos nombres constaban en el último archivo de anotación modificado (**Figura 20**, columna D).

**Identificación de elementos transponibles completos.** A partir de cada secuencia en formato FASTA extraída previamente y en orden descendente de longitud, se llevó a cabo un

*dot plot* de cada una de ellas contra sí misma, esto se realizó utilizando el software DOTTER (Sonnhammer & Durbin, 1995). El análisis de cada *dot plot* consistió en identificar visualmente en el gráfico los TIRs a los extremos de la secuencia del transposón, estos tenían que mostrarse como diagonales de sentido contrario al del dominio interno conservado de la secuencia. Este procedimiento se detuvo una vez que se identificaron TIRs en una secuencia que eran congruentes con el tamaño reportado por la literatura para cada superfamilia analizada. Una vez que se encontró una secuencia completa, se obtuvo una imagen de su *dot plot* y se etiquetó asignando cada estructura que forma parte del elemento.

**Identificación de elementos transponibles activos.** La anotación funcional de cada elemento transponible que fue identificado anteriormente como un elemento estructuralmente completo se realizó a través del programa Artemis (Carver et al., 2012). Se inició cargando la secuencia en formato FASTA del elemento a anotar, y se reportó cada componente estructural del elemento indicando sus posiciones en la secuencia. Luego, se marcó automáticamente con una de las funcionalidades del programa los ORFs vacíos a partir de 100 pb, o de ser necesario aquellos ORFs desde 75 pb. El siguiente paso consistió en alinear localmente cada secuencia de los ORFs mostrados en ese elemento utilizando el servidor de BLAST (Altschul et al., 1990) bajo el algoritmo *blastx* para la identificación principalmente de una proteína transposasa. También se realizó la búsqueda de dominios conservados en las proteínas identificadas utilizando el servicio de CD-Search (Marchler-Bauer & Bryant, 2004) disponible también en el mismo servidor de BLAST. Así, se concluyó anotando la funcionalidad de estos elementos de la secuencia en el transposón mediante Artemis. Para concluir, se exportó el archivo de anotación en formato GenBank y se extrajo la imagen con todos los elementos anotados.

### ***Identificación manual de elementos transponibles***

Esta sección fue llevada a cabo como alternativa para identificar ETs por homología utilizando secuencias de elementos transponibles de referencia recuperados de especies de

*Drosophila*, en su mayoría de *D. melanogaster*. Las secuencias referenciales en formato FASTA fueron recuperadas desde bases de datos que contenían elementos transponibles naturales como *Berkeley Drosophila Genome Project* (BDGP) (<https://www.fruitfly.org/>) (Kaminker et al., 2002), y, *FlyBase* (<http://flybase.org/>) (Drysdale, 2008; Gelbart et al., 1998). El objetivo de esta identificación manual fue verificar algún elemento activo de cada superfamilia de transposones de TIRs reportados por el pipeline EDTA en el genoma de *D. amaguana*.

Su metodología se basó en el alineamiento local del elemento de referencia y de su transposasa contra el genoma ensamblado de *D. amaguana* y con el archivo de salida de la biblioteca no redundante de elementos transponibles en formato EDTA.TElib.fa, a través de los algoritmos *blastx* y *tblastn* respectivamente utilizando BLAST+ 2.8.1. (Camacho et al., 2009). Para los resultados del alineamiento solo se consideraron aquellos que mostraron porcentajes de identidad mayor al 85% y longitud de alineamiento similar al tamaño del elemento referencial. El proceso exclusivamente continuó con los resultados que cumplieron estas dos condiciones, caso contrario la identificación manual de ETs no fue llevada a cabo.

Para los resultados favorables se procedió a extraer en formato FASTA el contig específico del genoma de *D. amaguana* con el cual se obtuvo el mejor alineamiento con el elemento referencial utilizado, esto se realizó a través del paquete de EMBOSS (Rice et al., 2000). Posteriormente mediante DOTTER se realizó un *dot plot* del contig extraído frente al elemento referencial para identificar las posiciones de sus TIRs y la longitud completa del elemento en el contig. Finalmente se extrajo la secuencia FASTA del elemento transponible según sus posiciones en el contig del genoma utilizando nuevamente EMBOSS, y se realizó la caracterización de este utilizando la metodología ya descrita previamente en la sección anterior.

## Capítulo IV: Resultados

La anotación estructural de los elementos transponibles ejecutada con la herramienta EDTA permitió la obtención de resultados iniciales que posteriormente sirvieron para la caracterización de las superfamilias, información con la cual se logró determinar con mejor precisión los elementos completos que podrían tener capacidad de transposición en el genoma.

### Calidad del ensamblaje del genoma en *Drosophila amaguana*

Los resultados de la evaluación de calidad del ensamblaje del genoma de *D. amaguana* obtenidos en el reporte a través de *Quality Assessment Tool for Genome Assemblies* (QUAST) muestran algunas métricas importantes que se reportan en la **Tabla 1**.

**Tabla 1.** Resumen del reporte de QUAST para el genoma de *D. amaguana*.

<i>D. amaguana</i>	
Número de contigs ( $\geq 0$ pb)	75 476
Número de contigs	55 860
Longitud total ( $\geq 0$ pb)	455 546 830
Longitud total	448 275 502
% GC	38.36
N50	29 434
N90	4 151
L50	4 094
L90	18 082
Número de N's por 100 kb	140.60

*Nota.* Métricas basadas en contigs de longitud  $\geq 500$  pb, exceptuando aquellas en las que propiamente se indica la longitud del contig. **Elaborado por:** Coba, M.

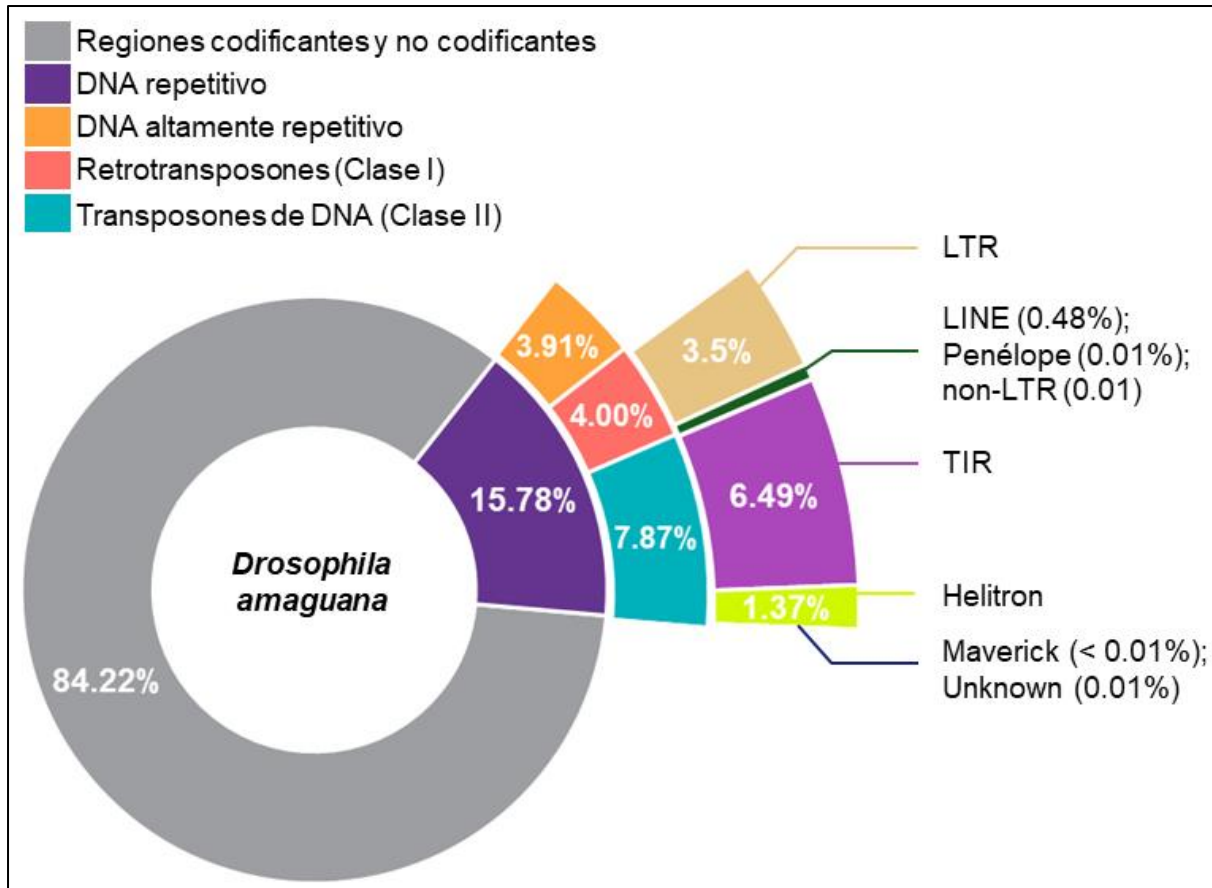
Con esto se ha estimado que el tamaño del genoma ensamblado es de aproximadamente 455 Mb distribuido en 75 476 contigs. Además el N50 reportado es de 29 434 pb que corresponde al contig número 4 094, según el parámetro del L50.

### Composición del genoma en *Drosophila amaguana*

A partir del genoma en formato FASTA, el pipeline de EDTA se ejecutó por alrededor de 87 horas, y arrojó un archivo de salida en formato EDTA.TE.anno.sum que contenía el

resumen de la anotación estructural para las secuencias de DNA repetitivo. Estos resultados permitieron estimar la composición del genoma que se muestra en la **Figura 21**.

**Figura 21.** Composición del genoma de *Drosophila amaguana*.



*Nota.* Porcentajes estimados de los componentes que conforman el genoma de *D. amaguana* luego de la anotación estructural de ETs con la herramienta EDTA. **Elaborado por:** Coba, M.

### Elementos transponibles en el genoma de *Drosophila amaguana*

El archivo de anotación en formato EDTA.TEanno.gff3 resultante de la ejecución del pipeline EDTA contiene información de todos los elementos transponibles identificados en el todo genoma, obtenidos por los algoritmos que utiliza cada uno de los programas que lo conforman.

Las anotaciones que en él se observan son realizadas por la estructura del elemento o por homología, esta última es realizada automáticamente basándose en la ontología de

secuencias (SO), este recurso permite la descripción de características y atributos en las secuencias ensambladas por contigs en el genoma de *D. amaguana*.

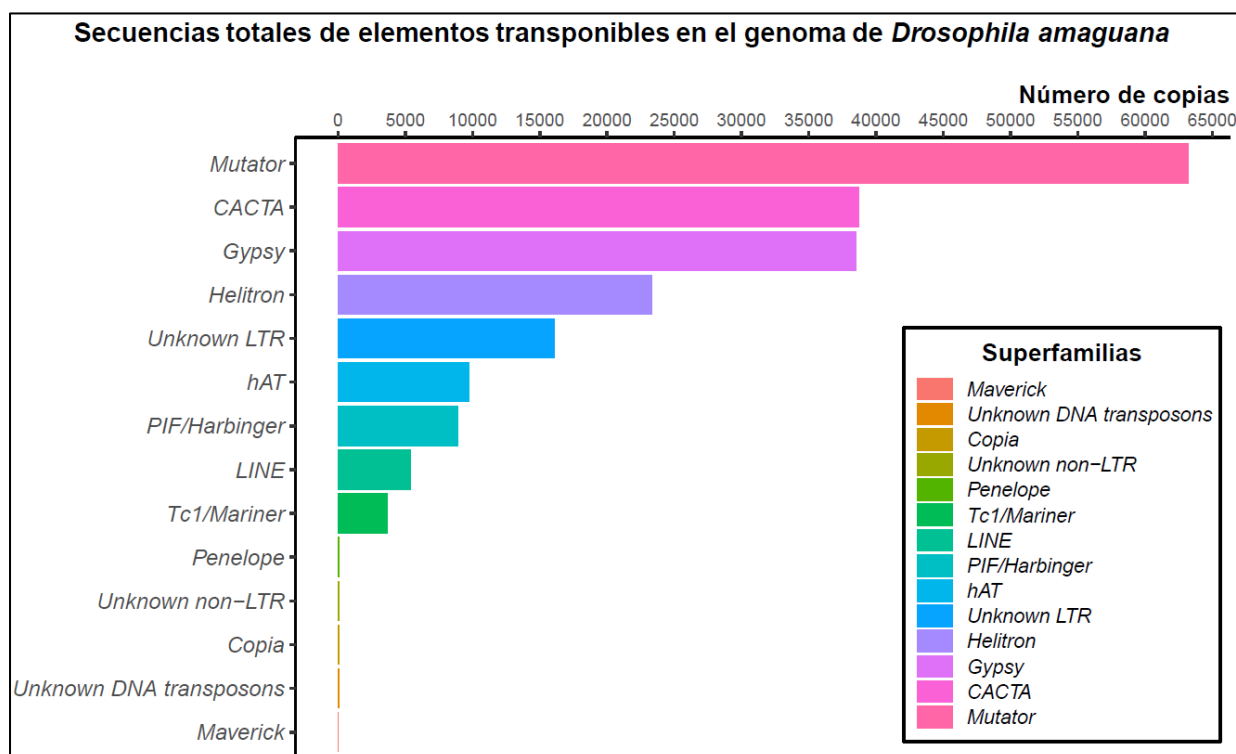
EDTA reconoció 454 916 558 pb en el genoma ensamblado de *D. amaguana*, del cual enmascaró 17 796 615 pb correspondiente a DNA altamente repetitivo y 53 967 332 pb correspondiente a DNA moderadamente repetitivo, donde se incluyen 35 769 226 pb para elementos transponibles clase II y 18 198 106 pb para retrotransposones.

**Tabla 2.** Elementos transponibles en el genoma de *Drosophila amaguana*.

<b>Superfamilia</b>	<b>Número de copias</b>
<i>Mutator</i>	63 185
<i>CACTA</i>	38 721
<i>Gypsy</i>	38 490
<i>Helitron</i>	23 316
<b>Unknown LTR</b>	16 091
<i>hAT</i>	9 718
<i>PIF/Harbinger</i>	8 906
<i>LINE</i>	5 374
<i>Tc1/Mariner</i>	3 674
<i>Penelope</i>	103
<b>Unknown non-LTR</b>	82
<b>Unknown DNA transposons</b>	43
<i>Copia</i>	73
<i>Maverick</i>	19

*Nota.* El número de copias por cada superfamilia se obtuvo a partir del archivo de anotación proporcionado como resultado del análisis con EDTA que considera todos los elementos estructuralmente intactos y fragmentados. Los *Unknown DNA transposons* es un nombre genérico que abarcan superfamilias del *PiggyBac* y *Sola2*. **Elaborado por:** Coba, M.

**Figura 22.** Totalidad de elementos transponibles en el genoma de *Drosophila amaguana*.



*Nota.* Superfamilias de los elementos transponibles identificados por EDTA en el genoma de *D. amaguana*, se muestra el número total de copias de estas secuencias que corresponderían a secuencias de elementos completos y elementos fragmentados. **Elaborado por:** Coba, M.

### Caracterización de elementos transponibles en *Drosophila amaguana*

El archivo de anotación en formato EDTA.intact.gff3 recopila información sobre los ETs estructuralmente intactos que fueron identificados en el genoma. A partir de él, se extrajo un registro para cada superfamilia con los datos respecto a la longitud de la secuencia en los elementos, nombre del elemento, nombre del contig en el cual se identificó, y, la clasificación realizada según la ontología de secuencias.

Para la caracterización, se identificó un elemento transponible clase II estructuralmente completo en todas las superfamilias del orden TIR como *Mutator*, *CACTA*, *hAT*, *PIF/Harbinger*, y, *Tc1/Mariner*. Adicionalmente de forma manual se reconoció y caracterizó un *Elemento P*, cuya superfamilia no fue identificada en el pipeline de EDTA.

Los elementos fueron considerados completos cuando se identificaron mediante un *dot plot* las secuencias de sus repeticiones terminales invertidas del tamaño correspondiente según cada superfamilia. Finalmente para cada elemento de estructura completa se realizó la anotación funcional de su secuencia.

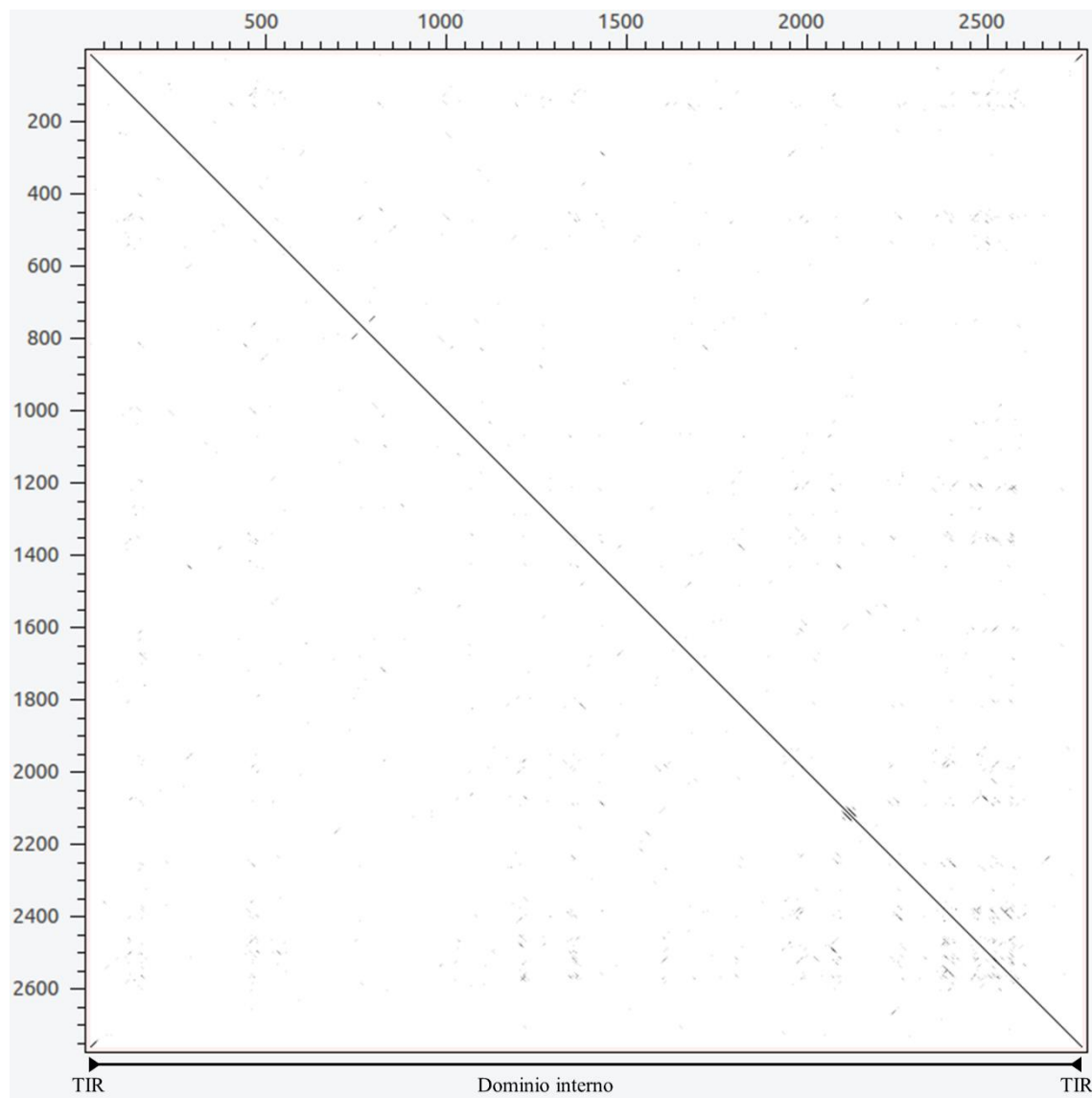
### ***Elemento P***

El resultado del alineamiento local del archivo de salida en formato EDTA.TElib.fa que contiene las secuencias no redundantes de elementos transponibles identificados por EDTA contra la base de datos creada con *elementos P* de referencia, no fue satisfactorio. Apenas se obtuvieron dos similitudes con un *elemento P* de *Drosophila ambigua* con un porcentaje de identidad entre 80 – 90 %, pero tan solo con una longitud de alineamiento de aproximadamente 150 pb, un indicativo de la falta de homología con este elemento completo.

Por otra parte, el alineamiento local entre el genoma ensamblado de *D. amaguana* y la base de datos creada con elementos de referencia mostró buenos resultados, indicando una similitud del 96.197 % entre el contig *jc7180002852566* y *elemento P* de *D. melanogaster* (GenBank Accession No. [X06779.1](#)) con una longitud de alineamiento de 2 051 pb.

El *elemento P* identificado manualmente en *D. amaguana* y caracterizado a través de un *dot plot* mostró una longitud de 2 772 pb flanqueado por TSDs de 8 pb (GCCAATAC), se visualizaron TIRs pequeños de 32 pb (CATGATGAAATAACATAAGGTGGTCTCGTCGG) y un dominio interno de 2 708 pb. Estructuralmente a este elemento se lo catalogó como un transposón intacto y completo.

**Figura 23.** *Dot plot* de un transposón de la superfamilia *elemento P* en *D. amaguana*.

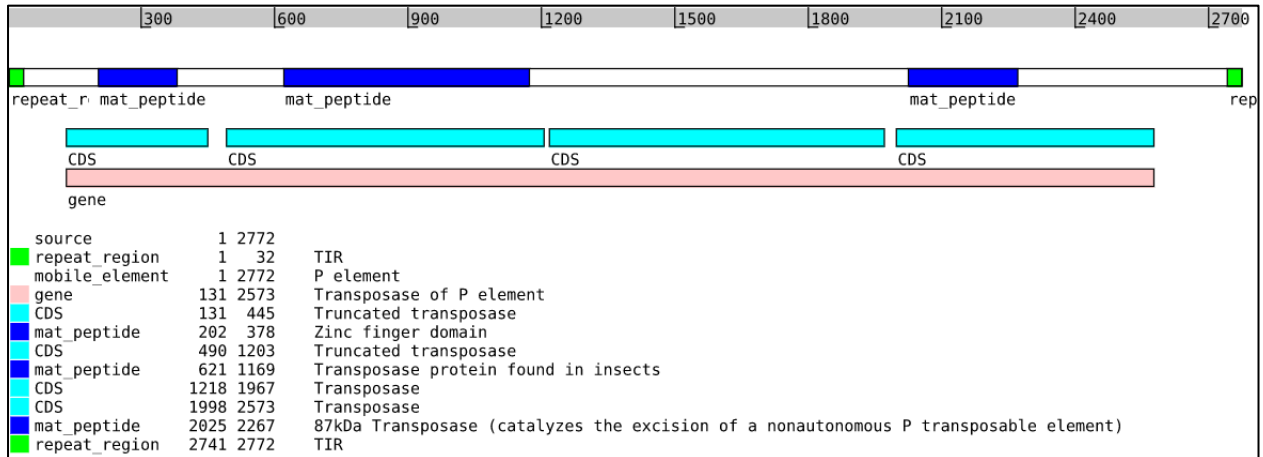


*Nota.* Alineamiento de la secuencia del *elemento P* de *D. amaguana* comprendido desde la posición 1 115 hasta la posición 3 886 del contig *jcf7180002852566* en su genoma ensamblado; las repeticiones terminales invertidas se muestran como pequeñas diagonales invertidas a los extremos del *dot plot*. La estructura gráfica del transposón se muestra debajo del *dot plot*, triángulos negros corresponden a TIR. **Elaborado por:** Coba, M.

La anotación funcional del *elemento P* mostró cuatro ORFs, cada uno de ellos se traduce en una transposasa de una longitud entre 200 a 400 AA. Adicionalmente, se identificó tres dominios proteicos conservados: el dominio *zinc finger* importante en la regulación de genes, un dominio de una transposasa característica de insectos, y, el dominio de una transposasa de 89 kDa asociado a un dominio THAP para catalizar la escisión de *elementos P* no autónomos.

Finalmente parece ser que los cuatro marcos de lectura corresponderían a un solo gen codificante para una transposasa, que posiblemente es un indicativo de la actividad de este transposón en *D. amaguana*.

**Figura 24.** Anotación funcional del *elemento P* en *D. amaguana*.



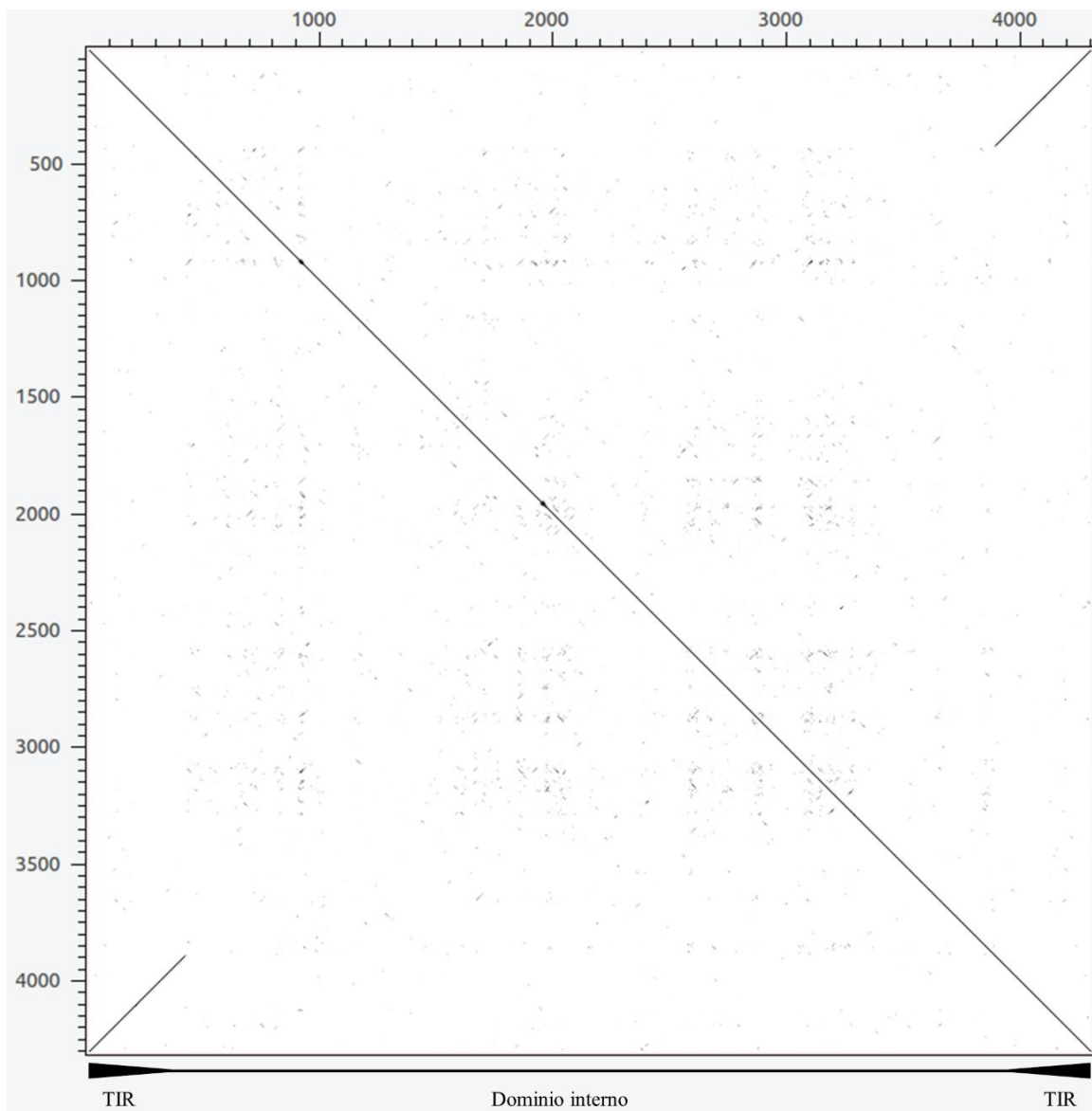
*Nota.* Información de las diferentes regiones que forman parte del *elemento P* identificado en el genoma de *D. amaguana*, donde se reportan cuatro regiones codificantes (CDS), dominios proteicos conservados (mat\_peptide), regiones repetitivas (repeat\_region) y un gen, donde para cada caso se reporta la posición que tienen en la secuencia. **Elaborado por:** Coba, M.

### **Mutator**

El reporte de EDTA mostró 632 secuencias intactas como parte de la superfamilia *Mutator*, sin embargo, tan solo 485 fueron clasificadas como tal propiamente, mientras que 147 secuencias fueron reportadas como MITEs, todas ellas con una longitud menor a 600 pb.

Se identificó un elemento completo con una longitud de 4 314 pb dentro del contig *jcf180002856425*. Estructuralmente se evidenció características propias de esta superfamilia en el elemento caracterizado. Se observó TIRs de 422 pb flanqueados por TSDs de 10 pb (AATTTAAATT), y un dominio interno de 3 470 pb.

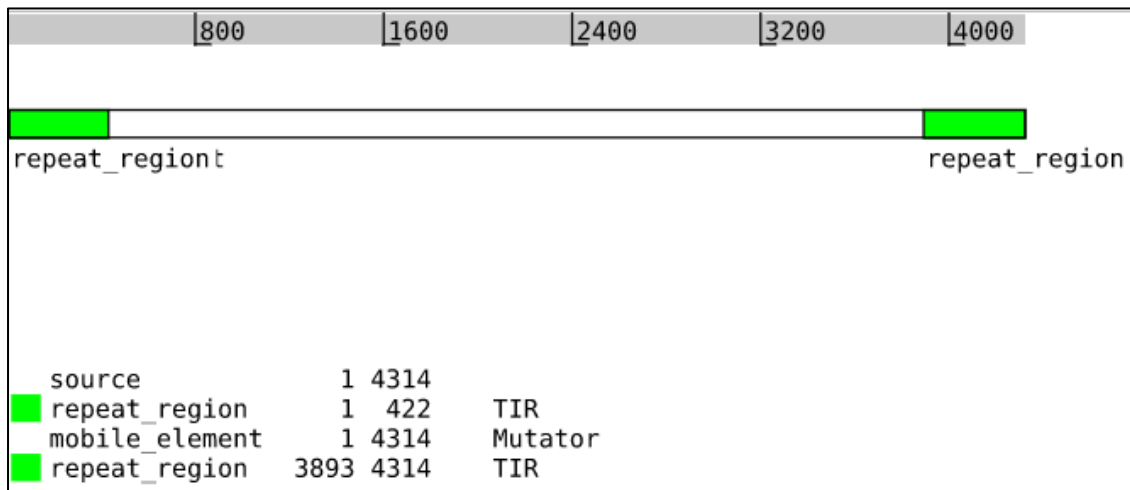
**Figura 25.** *Dot plot* de un transposón de la superfamilia *Mutator* en *D. amaguana*.



*Nota.* Alineamiento de la secuencia de un transposón *Mutator* en el genoma *D. amaguana* que está ubicado en contig *jcf180002856425* del genoma ensamblado. Se evidencian repeticiones terminales invertidas mayores a 300 pb que se muestran como diagonales invertidas a los extremos del *dot plot*. La estructura gráfica del transposón se muestra debajo del *dot plot*, triángulos negros corresponden a TIR. **Elaborado por:** Coba, M.

La anotación funcional en el elemento caracterizado no mostró secuencias codificantes a proteínas debido a que se observó codones *stop* consecutivos a lo largo del dominio interno en el transposón. Esta inactividad de la secuencia para poder transponerse autónomamente en el genoma está dada por la falta precisamente de una transposasa en el elemento transponible, que posiblemente está ligada a mutaciones que han generado múltiples codones de parada.

**Figura 26.** Anotación funcional del transposón *Mutator* en *D. amaguana*.



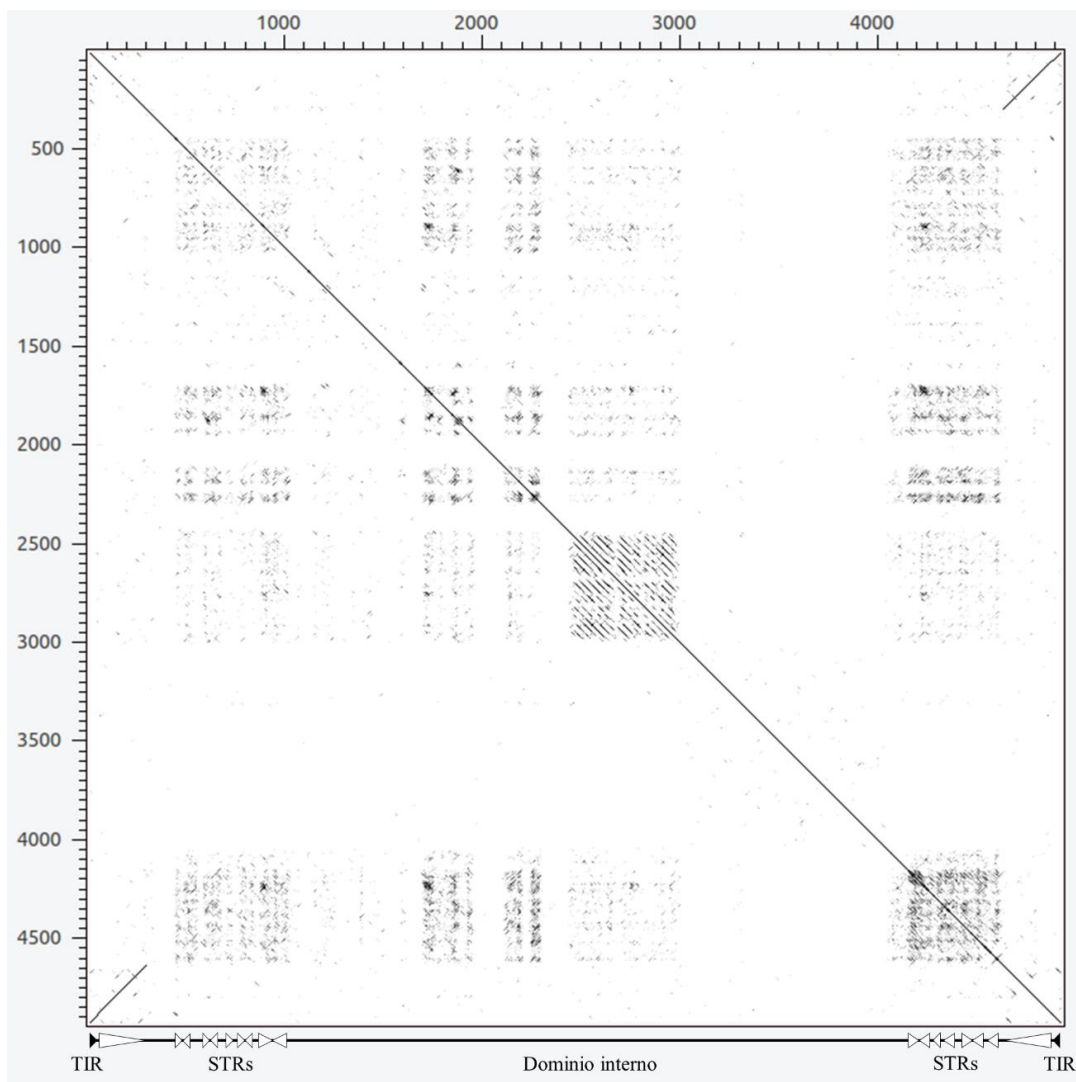
*Nota.* Anotación funcional en la secuencia del transposón de la superfamilia *Mutator* en el genoma de *D. amaguana*, donde se identifican TIRs largas (repeat\_region) y existe una carencia de secuencias codificantes. **Elaborado por:** Coba, M.

### **CACTA**

EDTA catalogó 345 secuencias intactas como parte de la superfamilia *CACTA*, de las cuales apenas 289 fueron propiamente parte de esta superfamilia, mientras que las 56 secuencias restantes se clasificaron bajo el orden de MITEs pertenecientes a esta misma superfamilia, sus secuencias presentaban una longitud por debajo de los 600 pb.

Dentro del contig *jcf7180002876874* se identificó un elemento de 4 940 pb que estructuralmente presenta aquellos elementos característicos para transposones *CACTA* reportados comúnmente en especies del reino vegetal. En este elemento se evidencia TIRs de 47 pb (CACTAATACTGAAACAAACTGTTGCTCTCTTTGTGAATCTCTCGA) acompañados de sub-repeticiones terminales (STRs). Las longitudes de todas estas STRs oscilaron variablemente entre al menos 256 hasta 571 pb aproximadamente. El elemento se encontró flanqueado por TSDs cortos de 2 pb (TA), y mostró un dominio interno de 4 324 pb.

**Figura 27.** *Dot plot* de un transposón de la superfamilia *CACTA* en *D. amaguana*.

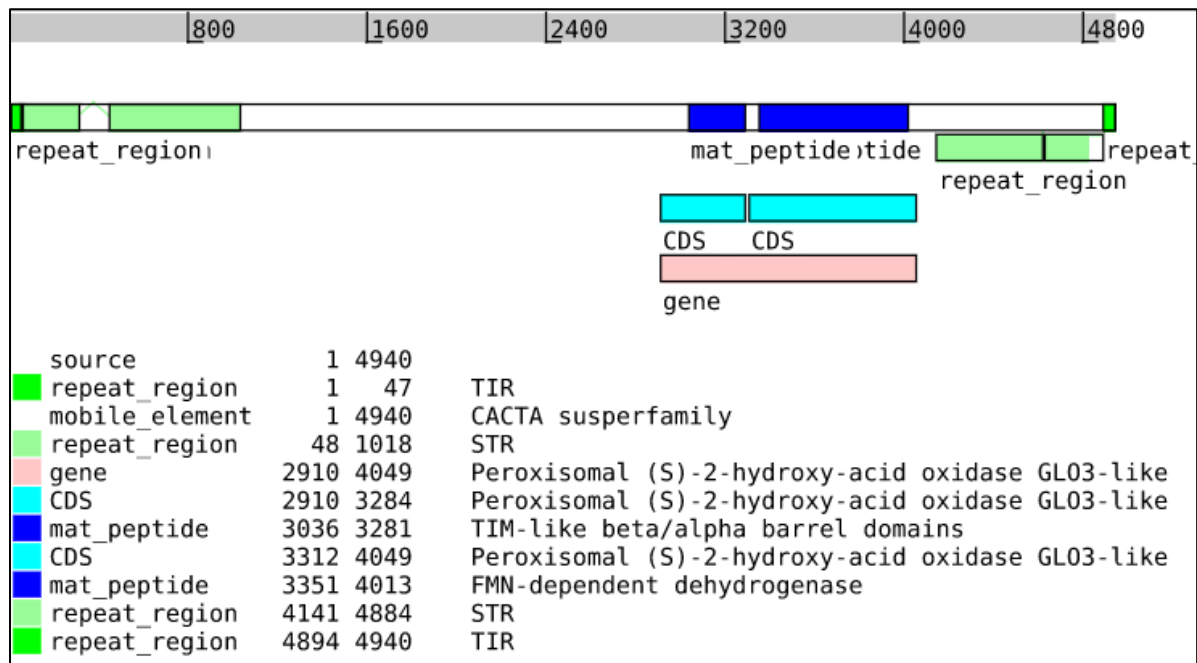


*Nota.* Alineamiento de la secuencia de un transposón *CACTA* de *D. amaguana* ubicado en contig *jcf7180002876874* del genoma ensamblado. Se evidencian repeticiones terminales invertidas que se muestran como pequeñas diagonales invertidas a los extremos del *dot plot*, y las repeticiones sub-terminales se muestran al finalizar cada TIR. La estructura gráfica del transposón se muestra debajo del *dot plot*, triángulos negros corresponden a TIR, y triángulos blancos corresponden a STRs. **Elaborado por:** Coba, M.

La anotación funcional de este transposón de DNA no mostró un ORF para la traducción a una transposasa que permita autonomía y funcionalidad en el elemento. Pero, por otra parte, se identificaron dos regiones codificantes hacia una misma proteína, la peroxisoma hidroxiácido oxidasa, aunque cada una con dominios distintos en cada CDS. Para el primero de ellos se identificó un dominio de barril de triosa-fosfato isomerasa (TIM) multifuncional

dentro de las vías metabólicas, mientras que en el siguiente CDS un segundo dominio fue correspondiente a flavoproteínas que contienen un flavín mononucleótido (FMN) y están implicadas en reacciones catalizadas por oxidorreductasas. Ambos CDS podrían corresponder a exones que provienen del mismo gen, como entonces lo hemos considerado.

**Figura 28.** Anotación funcional del transposón *CACTA* en *D. amaguana*.



*Nota.* Elementos anotados funcionalmente en el transposón *CACTA* identificado en el genoma de *D. amaguana*, se reportan regiones codificantes (CDS), dominios proteicos conservados (mat\_peptide), regiones repetitivas (repeat\_region) correspondientes a TIRs y STRs, y un gen. Para cada uno de ellos se detalla su posición en la secuencia. **Elaborado por:** Coba, M.

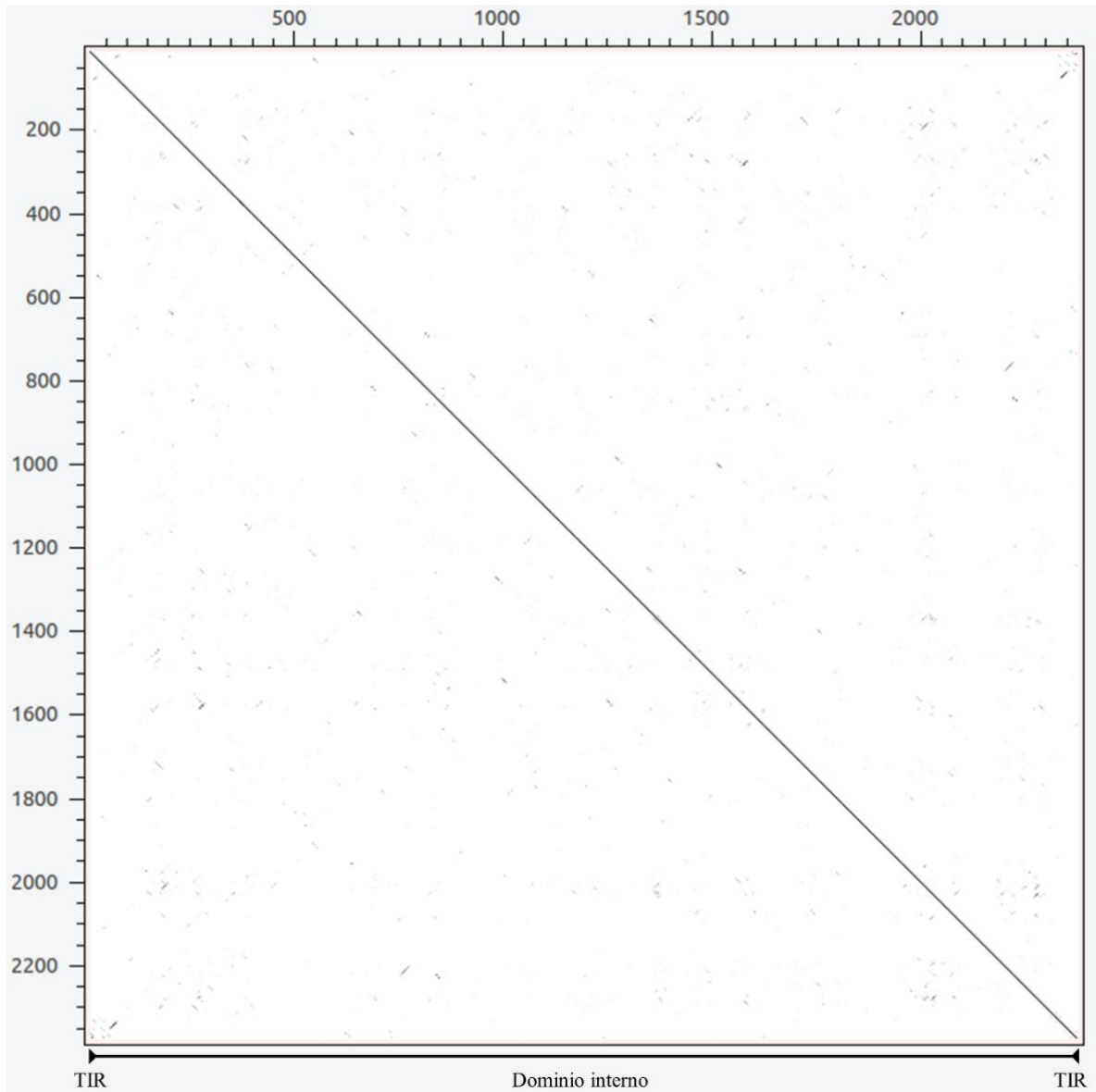
### *hAT*

La anotación para esta superfamilia mediante EDTA encontró 91 secuencias intactas de transposones de la superfamilia *hAT* y 38 secuencias con longitudes por debajo a 600 pb catalogadas como MITEs derivados de esta superfamilia.

Un elemento *hAT* completo estructuralmente fue identificado dentro del contig *jef18000288581* y su longitud fue de 2 385 pb. Se reportó TIRs de 21 pb

(TAGAGGTGGTAAACAATCGAT) y un dominio interno de 2 343 pb, todo ello comprendido entre TSDs de 8 pb (CCGTGCAC) característicos en esta superfamilia.

**Figura 29.** *Dot plot* de un transposón de la superfamilia *hAT* en *D. amaguana*.



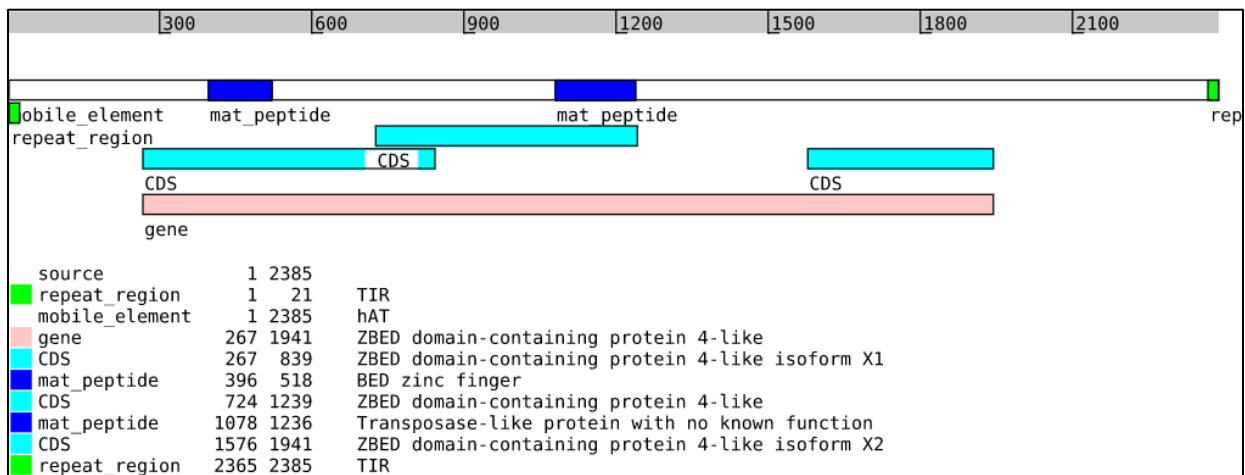
*Nota.* Alineamiento de la secuencia para el transposón *hAT* anotado en el contig *jcfl8000288581* del genoma de *D. amaguana*. Se evidencian repeticiones terminales invertidas que se muestran como pequeñas diagonales invertidas a los extremos del *dot plot*. La estructura gráfica del transposón se muestra debajo del *dot plot*, triángulos negros corresponden a TIR.

**Elaborado por:** Coba, M.

La anotación funcional en este elemento mostró tres CDS, todos ellos codificantes para un péptido similar a la proteína 4 que contenía un dominio ZBED. Se observó una

superposición entre dos CDS que requiere una revisión detallada, pues probablemente el CDS que tendría mayor posibilidad de encontrarse en el genoma sería el que se encuentra en la posición 724 – 1 239 del transposón, pues tiene un codón de inicio que puede corresponder al primer exón con el que se dé lugar a la traducción del mRNA maduro correspondiente. Pese a ello, hemos considerado estas tres secuencias codificantes para asumirlas como parte del mismo gen. Finalmente identificamos dos dominios proteicos importantes, el dominio de BED de dedos de zinc, mismo que actúa como regulador transcripcional; y, el dominio de una proteína de función desconocida pero asociada a una transposasa.

**Figura 30.** Anotación funcional del transposón *hAT* en *D. amaguana*.

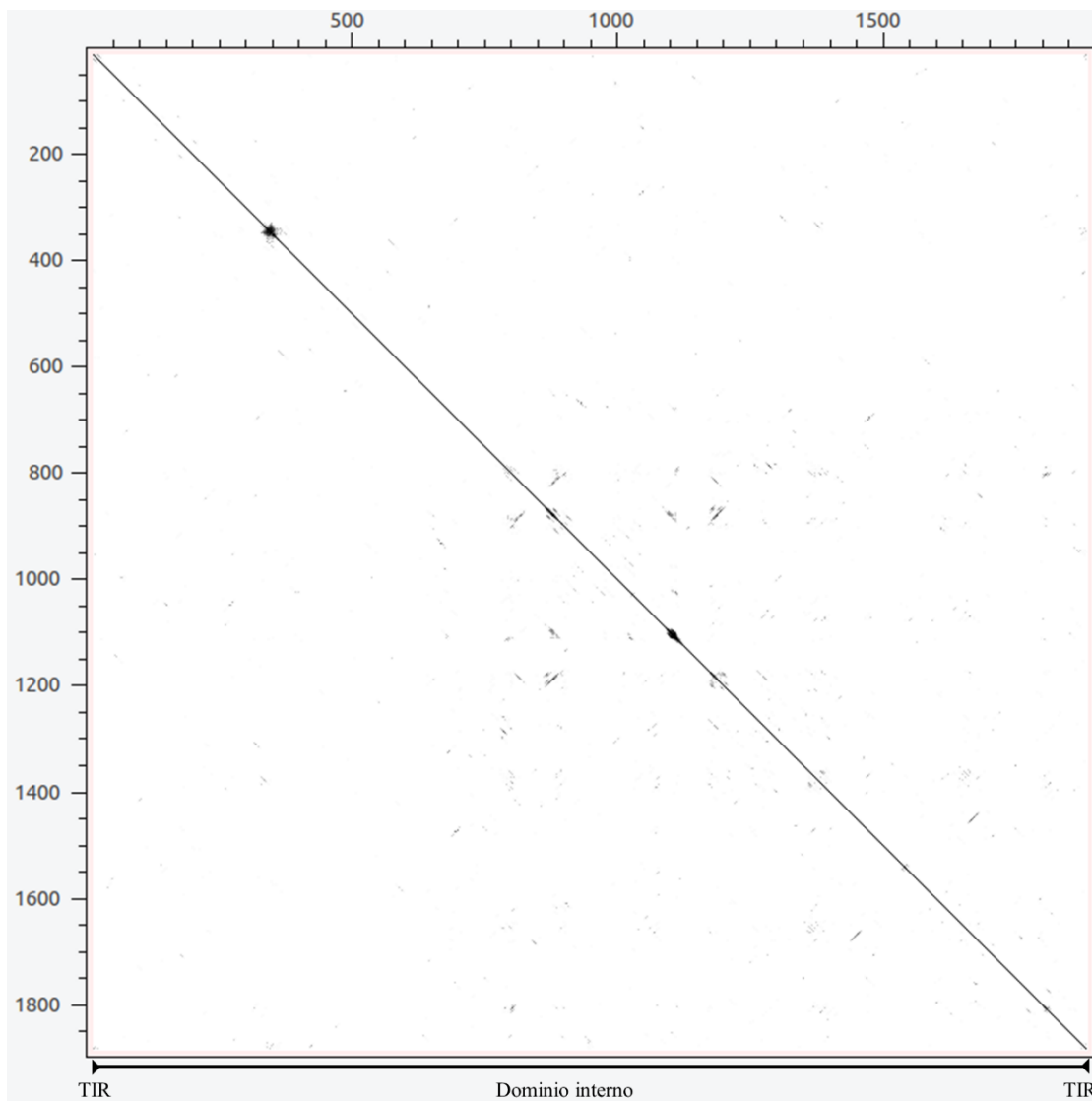


*Nota.* Elementos anotados funcionalmente en el transposón *hAT* identificado en el genoma de *D. amaguana*. Se detallan posiciones en la secuencias para regiones codificantes (CDS), dominios proteicos conservados (*mat\_peptide*), regiones repetitivas (*repeat\_region*) correspondientes a TIRs, y un hipotético gen. **Elaborado por:** Coba, M.

### ***PIF/Harbinger***

Para esta superfamilia, dentro del contig *jef180002881401* encontramos un transposón estructuralmente completo con un tamaño de 1 893 pb, en él se identificó 16 pb para sus dos TIRs (GAGCGGATGAGCGGCT), flanqueados dentro de 3pb (TTA) correspondientes a sus TSDs y un dominio interno de 1 861 pb.

**Figura 31.** *Dot plot* de un transposón de la superfamilia *PIF/Harbinger* en *D. amaguana*.

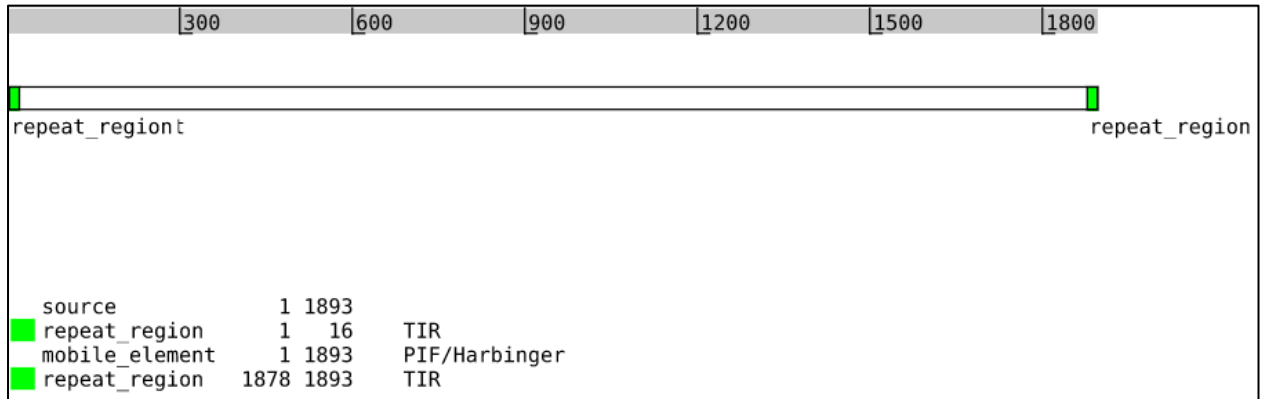


*Nota.* Alineamiento de la secuencia para el transposón *PIF/Harbinger* anotado en el contig *jef180002881401* del genoma de *D. amaguana*. Se evidencian repeticiones terminales invertidas que se muestran como pequeñas diagonales invertidas a los extremos del *dot plot*. La estructura gráfica del transposón se muestra debajo del *dot plot*, triángulos negros corresponden a TIR. **Elaborado por:** Coba, M.

El archivo de anotación final para ETs desde EDTA mostró 54 elementos *PIF/Harbinger* de un total de 65 secuencias intactas, pues apenas 11 elementos correspondieron a MITEs que se presentaron como secuencias de longitud menor a 600 pb.

Este elemento identificado como estructuralmente intacto, no es funcional, ya que no se observó secuencias codificantes para una transposasa, la proteína encargada de mover el elemento hacia otras regiones del genoma.

**Figura 32.** Anotación funcional del transposón *PIF/Harbinger* en *D. amaguana*.



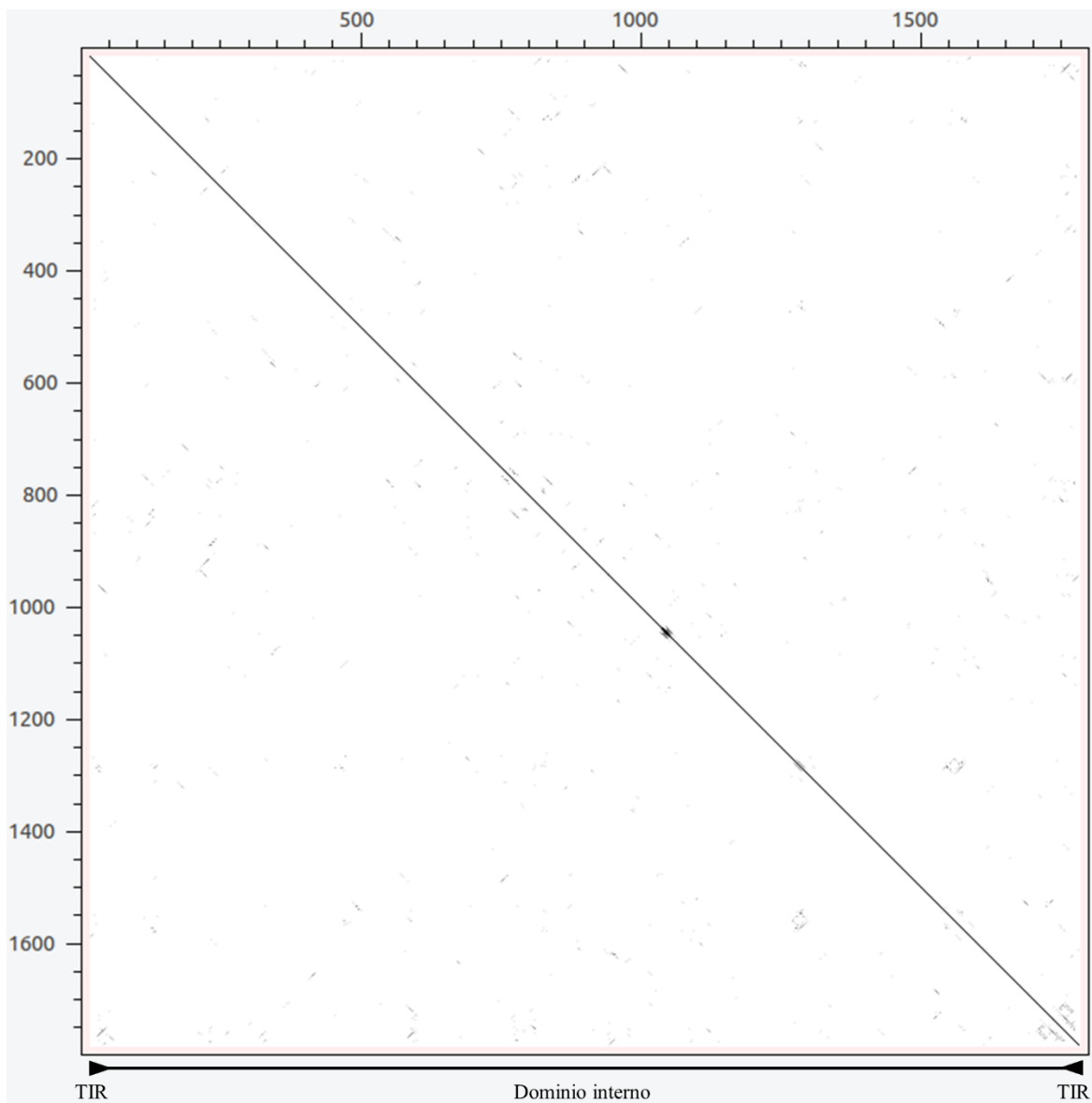
*Nota.* Identificación de regiones repetitivas (*repeat\_region*) correspondientes a TIRs en el transposón de la superfamilia *PIF/Harbinger* identificado en el genoma de *D. amaguana*, indicando las posiciones correspondientes en la secuencia. **Elaborado por:** Coba, M.

### ***Tc1/Mariner***

Esta superfamilia está representada por un elemento de 1 834 pb dentro del contig *jcf180002759449*, con una estructura completa que comprende TIRs de una longitud de 42 pb (AGTGAGACAGTAAAATATATATTTACACAGACTTAAAGTAAT) flanqueados por TSDs de 2 pb (TA) característicos en este grupo, dejando únicamente un dominio interno con un tamaño de 1 750 pb.

Aunque funcionalmente resultó ser un elemento no activo al carecer de al menos una secuencia codificante para una transposasa. Tampoco se observaron CDS para otras secuencias proteicas.

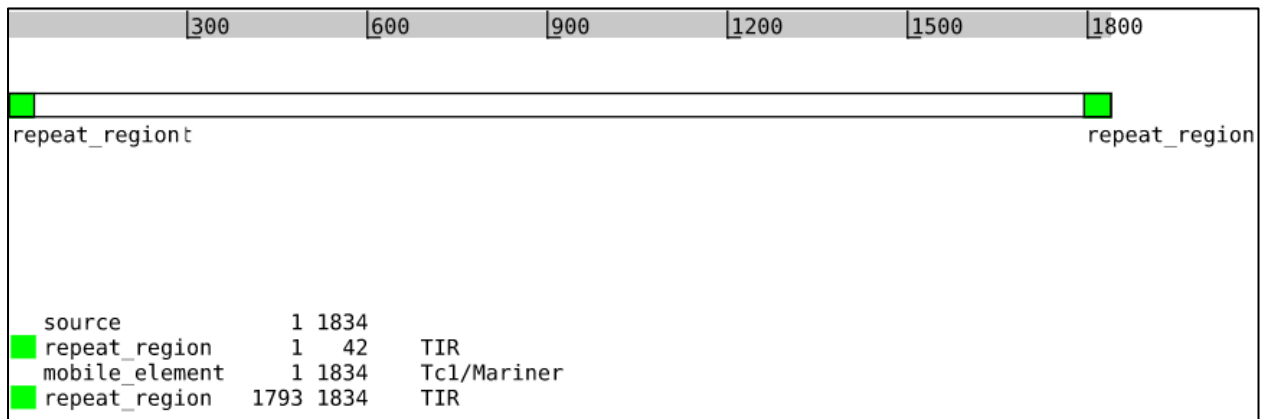
**Figura 33.** *Dot plot* de un transposón de la superfamilia *Tc1/Mariner* en *D. amaguana*.



*Nota.* Alineamiento de la secuencia para el transposón *Tc1/Mariner* anotado en el contig *jcf180002759449* del genoma de *D. amaguana*. Se evidencian repeticiones terminales invertidas que se muestran como pequeñas diagonales invertidas a los extremos del *dot plot*. La estructura gráfica del transposón se muestra debajo del *dot plot*, triángulos negros corresponden a TIR. **Elaborado por:** Coba, M.

EDTA identificó y anotó un total de 33 elementos intactos correspondientes a la superfamilia *Tc1/Mariner*, de los cuales 6 resultaron ser MITEs, catalogados así por ser secuencias menores a 600 pb. Esto dejó 27 de elementos intactos *Tc1/Mariner* reportados en el genoma de *D. amaguana*.

**Figura 34.** Anotación funcional del transposón *Tc1/Mariner* en *D. amaguana*.



*Nota.* Identificación de regiones repetitivas (repeat\_region) correspondientes a TIRs en el transposón de la superfamilia *Tc1/Mariner* identificado en el genoma de *D. amaguana*, indicando las posiciones correspondientes en la secuencia. **Elaborado por:** Coba, M.

### Contenido de elementos transponibles en otras especies de *Drosophila*

Utilizando reportes del contenido de ETs descritos en otras especies de *Drosophila* se han comparado esos valores con los obtenidos aquí en *D. amaguana* mediante EDTA.

Los elementos transponibles identificados en *D. melanogaster* se obtuvieron del estudio de Ou et al., (2019), quienes al desarrollar EDTA evaluaron su propia herramienta con varios genomas referenciales, entre ellos la versión r6.28 del genoma de *D. melanogaster*. La anotación de ETs lo hicieron con una librería curada de secuencias de RepBase (v.20170127).

**Tabla 3.** Porcentaje de ETs identificados en *D. melanogaster* y *D. amaguana*.

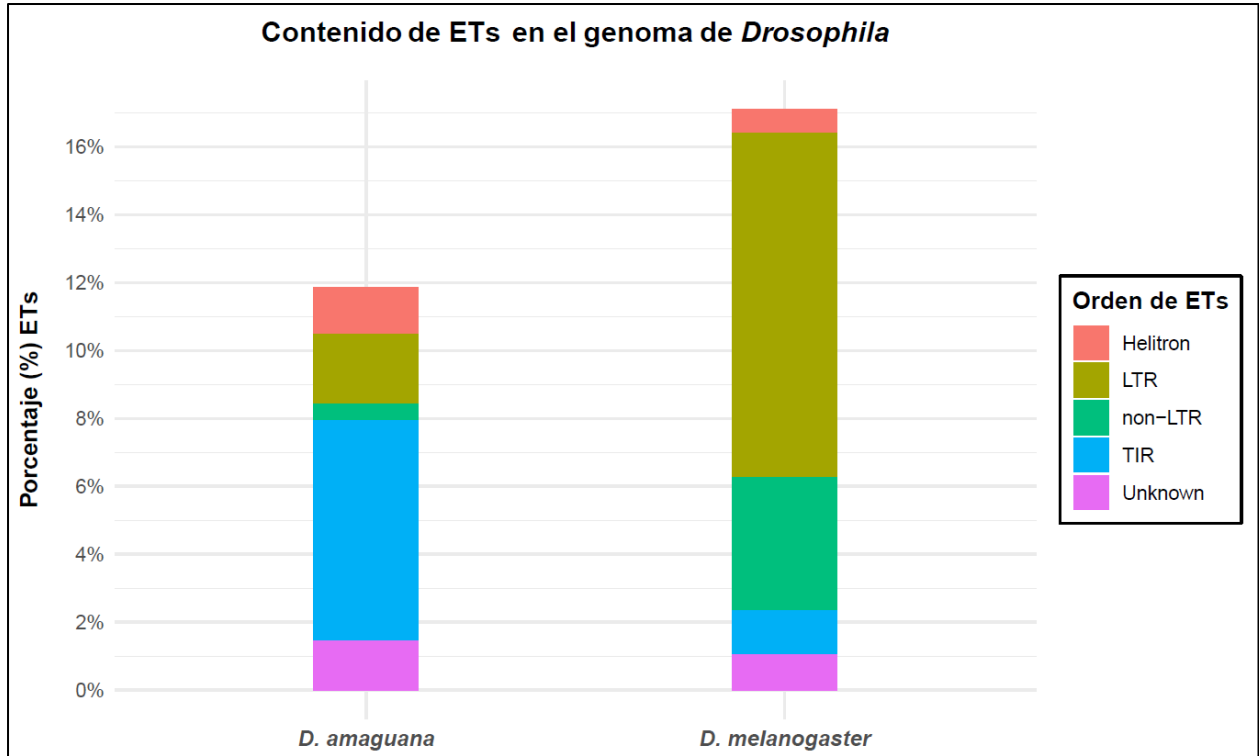
	LTR	non-LTR	TIR	Helitron	Unknown	Total
<i>D. melanogaster</i>	10.16 %	3.92 %	1.32 %	0.68 %	1.03 %	17.11 %
<i>D. amaguana</i>	2.06 %	0.49 %	6.5 %	1.37 %	1.45 %	11.87 %

*Nota.* Los porcentajes corresponden a la totalidad del contenido de ETs obtenidos mediante EDTA. Para *D. melanogaster*, Ou et al., (2019) utilizaron el genoma r6.28 con un tamaño estimado de 142.6 Mb. Para los elementos clasificados como Unknown se consideró a aquellos elementos que no encajaron en los órdenes propuestos. **Elaborado por:** Coba, M.

Como se observa comparativamente en la **Figura 35**, el contenido reportado en el genoma de *D. melanogaster* es mucho mayor en retrotransposones pero menor en los ETs clase

II, donde especialmente el contenido de TIR para *D. amaguana* se anotó en un 6.5 % de secuencias en todo el genoma.

**Figura 35.** Contenido de ETs anotados por EDTA en *Drosophila*.



*Nota.* Grafica de barras que muestra la abundancia relativa de cuatro órdenes de elementos transponibles anotados por EDTA en el genoma ensamblado de *D. amaguana* (454.9 Mb) y *D. melanogaster* (142.6 Mb). **Elaborado por:** Coba, M.

Por otra parte, también se analizó el porcentaje de elementos transponibles en *Drosophila* tomando los datos reportados por Mérel et al., 2020, quienes reconstruyeron copias de ETs combinando las herramientas RepeatMasker y OneCodeToFindThemAll, y utilizando secuencias consensos disponibles en RepBase para *Drosophila*.

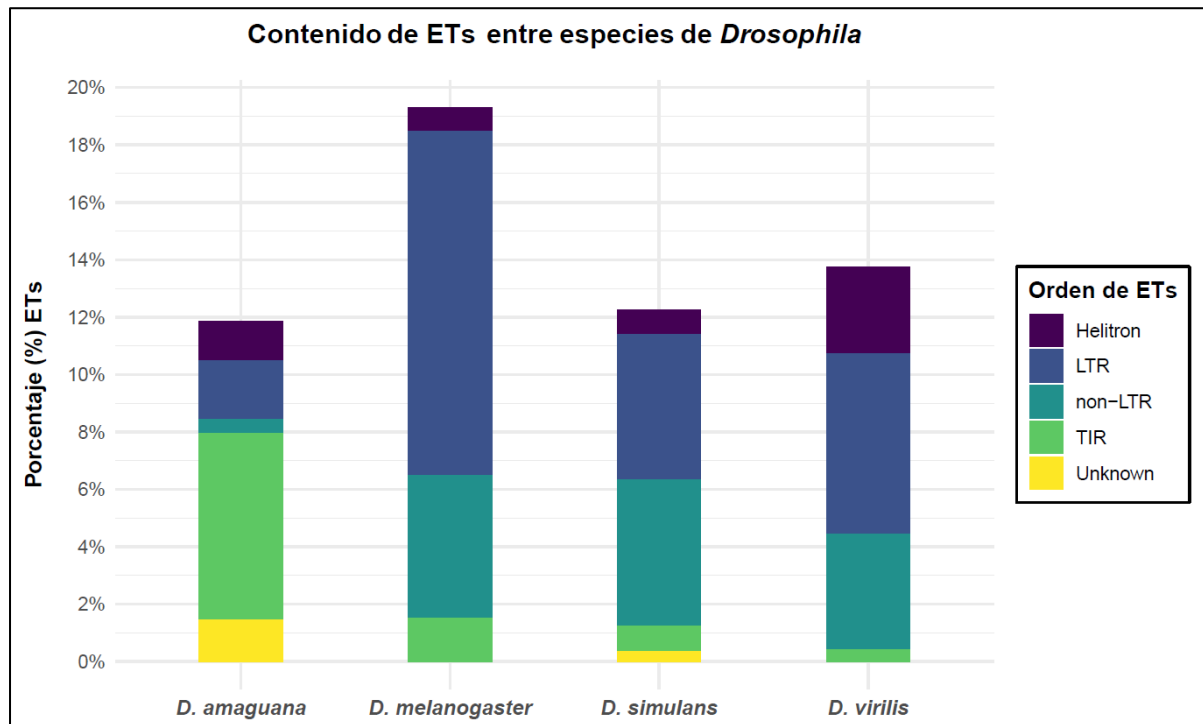
**Tabla 4.** Contenido de ETs identificados en especies de *Drosophila*.

	LTR	non-LTR	TIR	Helitron	Unknown	Total
<i>D. melanogaster</i>	12.0 %	5.0 %	1.5 %	0.8 %	N/A	19.3 %
<i>D. simulans</i>	5.1 %	5.1 %	0.9 %	0.8 %	0.34 %	12.24 %
<i>D. virilis</i>	6.3 %	4.05 %	0.4 %	3.0 %	N/A	14.95 %
<i>D. amaguana</i>	2.06 %	0.49 %	6.5 %	1.37 %	1.45 %	11.87 %

*Nota.* Los porcentajes corresponden a la ocupación de las secuencias para cada orden de ETs en el genoma ensamblado de tres especies de *Drosophila* realizado por Mérel et al., 2020, utilizando una biblioteca RepBase (v.20181026). **Elaborado por:** Coba, M.

En la **Figura 36** se visualiza que para especies de *D. melanogaster*, *D. simulans*, y *D. virilis* la variabilidad de ETs es evidente con el contenido que hemos encontrado en el genoma de *D. amaguana*. El contenido anotado de ETs en la subclase TIR en *D. amaguana* es el más abundante en comparación a las otras especies, donde principalmente su genoma está compuesto por elementos LTR y non-LTR, este último orden mostró ser el más bajo reportado *D. amaguana* con apenas 0.5 %.

**Figura 36.** Contenido de ETs identificado para especies de *Drosophila*.



*Nota.* Para cada especie de *Drosophila* se muestra el porcentaje de ETs identificados en los genomas ensamblados, mostrando su abundancia según su subclase. **Elaborado por:** Coba, M.

## Capítulo V: Análisis y Discusión de Resultados

El contenido de elementos transponibles en los genomas eucariotas y procariotas es muy variable entre cada especie, incluso pueden llegar a representar un porcentaje relevante de todo el material genético de los organismos. *Drosophila* se ha mantenido como un género que realiza aportes importantes sobre la evolución del genoma, se cree que apenas 1.6 % del total de especies descritas a nivel mundial de *Drosophila*, han sido estudiadas para comprender la organización del genoma (Gregory & Johnston, 2008). Particularmente en Ecuador en el último siglo se han descrito nuevas especies en nichos ecológicos importantes. La identificación de ETs en *D. amaguana*, una especie neotropical de *Drosophila* podría proporcionar información importante acerca del papel que estas secuencias desempeñan para permitir una radiación adaptativa en los grupos neotropicales de *Drosophila*.

Dentro de los principales resultados de calidad con QUAST se resume un genoma ensamblado con aproximadamente un tamaño de 456 Mb y un N50 de 29 kb. Si bien el tamaño del genoma estimado en *D. amaguana* corresponde a 632 Mb (dato no publicado), este difiere en cerca de 176 Mb con el genoma ensamblado; probablemente debido a que representa una parte del genoma con regiones altamente repetitivas que los algoritmos de los ensambladores no pueden resolver bien al trabajar con lecturas cortas. Sin embargo, el valor para el N50 nos indica una buena calidad en el ensamblaje, donde encontramos que al menos el 50 % del tamaño total del genoma está contenido en contigs con una longitud mayor o igual a 29 kb. Esto junto a los resultados de la temprana evaluación mediante BUSCO, donde se identificaron 2 764 genes completos (98.7 %), 20 genes fragmentados (0.7 %) y 15 genes faltantes (0.6 %) de un total de 2 799 genes, muestra que las métricas son similares a las reportadas por Rech et al., (2022), quienes obtuvieron genomas de alta calidad al secuenciar y ensamblar genomas *de novo* a partir de lecturas largas para 32 cepas de *D. melanogaster*, recolectadas en 11 localidades europeas y una norteamericana de cinco diferentes regiones climáticas.

*D. amaguana* perteneciente al grupo *mesophragmatica* mostró un contenido estimado de 15.78 % de DNA repetitivo en su genoma, del cual el 11.87 % corresponde a elementos transponibles, distribuidos entre transposones de DNA (7.87 %) y retrotransposones (4.0 %), mientras que 3.91 % correspondería a DNA altamente repetitivo. El poder comparar los resultados con otros de los que han sido reportados para otras especies del género, permitiría resolver interrogantes que involucran a estas secuencias genéticas móviles con el tamaño y la evolución de los genomas (Sessegolo et al., 2016).

Aunque si bien, hacer comparaciones directas del contenido de ETs entre especies siempre será arriesgado. Y es que precisamente la diversidad de software y metodologías a las que los investigadores tienen disponibilidad de uso, generan distintos resultados en la identificación de secuencias de ETs (Lerat, 2010). Así por ejemplo, en *D. willistoni* se ha estimado con dos diferentes programas bioinformáticos un contenido de ETs del 9 o del 16 % (Clark et al., 2007), o tal y como en *D. ananassae*, donde se obtuvo 12 o 20 % de ETs usando la misma herramienta pero con una librería de secuencias referenciales o mediante una anotación *de novo* (Clark et al., 2007).

En particular, el consorcio *Drosophila 12 genomes* halló en *D. melanogaster* un contenido de 20 % en elementos transponibles, del cual, en la región eucromática realmente se concentra entre 2 – 8 % de ETs, (McCullers & Steiniger, 2017), algo parecido al contenido de ETs reportado por (Kaminker et al., 2002), quienes en la secuencia eucromática para la misma especie identificaron 3.68 % de este DNA moderadamente repetitivo, aunque con conocimiento de estudios previos sobre que el contenido total de en el genoma de elemento móviles llegaba apenas al 9 % (Spradling & Rubin, 1981). Adicionalmente, se ha reportado resultados del contenido de ETs en otras especies del género como *D. busckii* con 4.65 %, *D. suzukii* con 30.80 %, *D. simulans* con ~ 7 %, *D. kikkawai* con ~ 18 %, *D. ananassae* con ~19 %, *D. virilis* con ~ 15 % (Sessegolo et al., 2016), algunas de ellas, neotropicales y exóticas, descritas y

colectadas al sur de Brasil (Cavasini et al., 2014), pero de poblaciones de las que aún no se tienen datos de la composición de su genoma.

Si bien el contenido de 11.87 % de ETs encontrado en el genoma *D. amaguana* es relativamente bajo comparado con el reportado con otras especies cuyos genomas tiene un tamaño de hasta 3 veces menor, como por ejemplo en *D. melanogaster* con un genoma estimado de 180 Mb donde se reporta 19.3 % de ETs (Mérel et al., 2020); hay evidencia acerca de la variación en el tamaño del genoma, asociada a la amplificación o eliminación de elementos transponibles a través de mecanismos en respuesta a factores de estrés ambientales en periodos particulares o entornos específicos (Canapa et al., 2015). Lo que, en parte, podría explicar el bajo porcentaje de secuencias de DNA repetitivo dentro de un genoma estimado de 632 Mb (dato no publicado), sin embargo se requiere una revisión y análisis más profundo.

Pero, ¿realmente se podría afirmar con seguridad que este es el contenido real de elementos transponibles en esta especie?, la respuesta a esta interrogante definitivamente es no, pues aquellos porcentajes de ETs en el genoma resultan ser apenas una aproximación, que si bien pueden acercarse a la realidad, depende en gran medida de la ensamblaje del genoma y de la región analizada (McCullers & Steiniger, 2017). Sin dejar también de lado la tecnología de secuenciación utilizada, la detección de más inserciones de ETs en genomas puede ser posible trabajando con secuenciación de tercera generación (Guio & González, 2019). Las tecnologías de esta generación permiten obtener lecturas con regiones importantes de repeticiones que no estén infrarrepresentadas y dificultan su ensamblaje (Sessegolo et al., 2016), un problema común que se presenta con lecturas cortas.

Con EDTA no fue posible identificar ningún *elemento P* en el genoma de *D. amaguana*. La distribución de esta superfamilia se ha estudiado con interés en el género *Drosophila*, aunque su presencia se ha extendido a otras especies de vertebrados (Hammer et al., 2005),

parece predominar en el orden Diptera (Silva & Kidwell, 2004). Por lo tanto, una verificación manual para identificar la presencia o ausencia de estos elementos en *D. amaguana* fue necesaria a partir de un elemento referencial.

Se encontró al menos un *elemento P* completo de 2.7 kb con cuatro CDS, cada uno de ellos codificante para una transposasa, dos de las cuales son proteínas truncadas. Considerando que la funcionalidad de un *elemento P* está dada por la codificación de una transposasa activa de 87 kDa, generada a partir de un mRNA transcrito por el *splicing* de cuatro exones que se han reportado en elementos P de longitud completa de hasta 2.9 kb (Majumdar & Rio, 2015). Y sin dejar de mencionar, que en algunos casos, el *splicing* alternativo puede dar lugar a un transcrito diferente de mRNA originado por la escisión de alguno de los exones traduciéndose en una proteína truncada de aproximadamente 66 kDa, y que puede presentar actividad represora transposicional, sobre todo cuando el exón escindido es aquel del extremo C-terminal (Ghanim et al., 2020). La transposasa del *elemento P* en *D. amaguana* demuestra falta de actividad autónoma, pues se identificó un codón *stop* en la mitad del primer CDS, que podría corresponder a un exón, truncando la traducción completa de la secuencia proteica.

Dentro del grupo *mesophragmatica* algunas especies neotropicales como *D. gaucha*, *D. gasici*, *D. brncici*, *D. pavani*, y *D. viracochi* (reportadas en Brasil, Colombia y Chile) no han mostrado presencia de algún *elemento P* homólogo al de *D. melanogaster* (da Silva Loreto et al., 1997; Germanos et al., 2006). Por lo tanto, la secuencia del *elemento P* descrito en *D. amaguana* permitiría rastrear el origen de esta superfamilia para conocer el mecanismo de transposición del elemento hacia esta especie; ya que la distribución de esta familia de ETs ha sido propuesta mayormente mediante eventos de transferencia horizontal (Silva & Margaret, 2000). Es decir, su aparición en nuevas especies neotropicales puede haber sido influenciada por factores geográficos, temporales y ecológicos entre la especie donante y receptora (Loreto et al., 2001).

El pipeline de EDTA ha identificado estructuralmente elementos de la superfamilia *Mutator* como la más abundante. La caracterización de un elemento completo en *D. amaguana* permitió identificar una secuencia de 4.3 kb con TIRs largos mayores a 100 pb, característico de estos transposones (K. Liu & Wessler, 2017). Su importancia dentro los genomas tiene que ver con la capacidad con la que estos transposones podrían portar fragmentos génicos del hospedador producto de inserciones en su dominio interno (Cerbin & Jiang, 2018), con este criterio se los ha catalogado como Pack-MULE (Jiang et al., 2004).

Los *Mutator* o actualmente conocidos como MULEs (*Mutator* like elements) son una fuente valiosa para comprender algunas interacciones con el huésped (Dupeyron et al., 2019), sin embargo, aún no han sido explorados por completo en más genomas de otras especies animales, a pesar de reportes de identificación en protozoos o insectos, como el del elemento *Muta1* en *Aedes aegypti* (K. Liu & Wessler, 2017), pues los estudios se han centrado mucho en hongos y algunas plantas (Dupeyron et al., 2019).

El elemento completo que reportamos en *D. amaguana* demostraría falta de actividad transposicional, pues carece de una transposasa funcional. Esta inactividad de transposones en esta superfamilia ya se ha observado en la gran parte de eucariotas, y, ha sido asociada a mutaciones, deleciones y/o mecanismos de regulación epigenética que silencian estos elementos (Lisch, 2009). Además a lo largo de su secuencia no se encontró marcos de lectura codificantes para alguna otra proteína, lo que excluye la posibilidad que el transposón sea un Pack-MULE. Aunque una vez más, la caracterización de este grupo es bastante raro en animales que incluso no se ha identificado en genomas de referencia (Tan et al., 2021).

Continuando el análisis, *CACTA* en *D. amaguana* mostró ser la segunda superfamilia de transposones de DNA con mayor abundancia en el genoma. Si bien estos transposones pueden estar presentes en los reinos vegetal, fungi y animal (Buchmann et al., 2014), por

ejemplo en insectos como *Mengenilla moldrzyki*, *Rhodnius prolixus*, *Bombyx mori* (Tang et al., 2015). Muchos reportes publicados se limitan a caracterización elementos *CACTA* en plantas, donde presentan gran abundancia en su genoma, como en el caso de algunas gramíneas que comprenden hasta 10 % de estos transposones (Fedoroff, 2013).

En *D. amaguana*, la caracterización de un elemento *CACTA* completo de 4.9 kb estructuralmente mostró componentes característicos de esta superfamilia tales como TIRs y STRs, como se ha reportado en genomas de la familia Triticeae (Wicker et al., 2003). Sin embargo, el dominio interno del transposón no mostró un ORF codificante para una transposasa que permita el movimiento del elemento a lo largo del genoma. Aunque dos CDS anotados codifican una peroxisoma (S)-2-hidroxiácido oxidasa, una proteína partícipe del metabolismo lipídico y de especies reactivas de oxígeno, cuya aparición puede correlacionarse con la expresión génica (Pridie et al., 2020), que a su vez estaría implicada por la asociación de elementos reguladores con los elementos transponibles *CACTA* (Wicker et al., 2003).

Y aunque reportes dentro de RepBase también muestran elementos *CACTA* para algunas especies de *Drosophila* de las cuales se tiene muy pocos estudios, quedará la interrogante a responder sobre si la identificación de esta superfamilia en *D. amaguana* es válida. Pues, hay que reconocer que existe un desafío para identificar elementos *CACTA* por la errónea catalogación de secuencias debido a la anidación que pudiese darse desde otros ETs (Belyayev et al., 2022), esto junto a la muy limitada conservación del motivo *CACTA* presente en los TIRs, dificulta el reconocimiento de elementos nuevos a través de repeticiones terminales invertidas de elementos ya conocidos (Wicker et al., 2003).

Todo esto supone en cierta medida tres escenarios, la falta de estudios profundos de esta superfamilia en otros organismos eucariotas, la inaccesibilidad a datos publicados o la

ausencia de elementos *CACTA* en sus genomas; temas que deben ser tratados para comprender interrogantes de ETs que no han podido estudiarse en un grupo de organismos más amplio.

En cuanto a la superfamilia *hAT*, que resultó ser la sexta más abundante en *D. amaguana*, dentro de *Drosophila* es una de las superfamilias de ETs más estudiadas, principalmente haciendo hincapié sobre el elemento *hobo*, uno de los más representativos de este grupo (N. R. Mota et al., 2010). Caracterizar estos elementos aporta información significativa para comprender sobre la distribución de elementos *hAT*, pues al tener una distribución muy amplia en eucariotas (Rubin et al., 2001), y al ser transposones que se han ido identificado por metodologías *in silico* todavía carecen de información biológica (Ladevèze et al., 2012).

Usualmente los transposones *hAT* presentan menos de 50 copias en todo el genoma del huésped (Ladevèze et al., 2012), algunas veces el número de copias se relaciona con la plasticidad cromosómica y genómica, esto impulsa cambios evolutivos para permitir que la especie ocupe un hábitat específico en una distribución geográfica (Bertocchi et al., 2022). Apenas se conoce que pocos de estos elementos se encontrarían en su forma activa (Atkinson, 2015), por ejemplo, en *D. willistoni* tres de cinco elementos *hAT* se han caracterizado como transposones activos (De Freitas Ortiz & Loreto, 2009), en *D. buzzatti* la identificación del elemento *BuT2* (Rossato et al., 2014), y, el elemento *hobo* que se ha identificado en *D. melanogaster* (Calvi et al., 1991) y en especies neotropicales como *D. simulans* o *D. kikkawai* (da Silva Loreto et al., 1997), el mismo que no ha sido encontrado en especies del grupo *mesophragmatica* (Germanos et al., 2006).

El elemento *hAT* completo que hemos caracterizado y anotado mostró una longitud de 2.3 kb con un CDS codificante para una proteína involucrada en procesos de regulación transcripcional (Ravasi et al., 2003), que muestra un dominio *BED-type zinc finger* originado

a partir de una familia de genes *ZBED* (*Zinc Finger Bed Domain*) (Hayward et al., 2013). Esto sugiere la presencia de un transposón domesticado *hAT* de la familia *Ac* en *D. amaguana*, ya que el origen de los genes *ZBED* se remota hacia la domesticación de transposones *hAT* en dos eventos independientes observados en ancestros primitivos de vertebrados (Hayward et al., 2013). Aunque todavía queda por estudiar el papel que desempeñaría esta proteína en el genoma de *D. amaguana*, pues la función de la triada de aminoácidos DDE observada en el dominio catalítico de genes *ZBED* domesticados va más allá de un mecanismo de cortar y pegar (Hayward et al., 2013).

Los resultados para la superfamilia *PIF/Harbinger*, los posicionan como la séptima superfamilia abundante que se identificó en *D. amaguana*. La mayor distribución de elementos *PIF* parece estar presente en genomas vegetales por sobre otras especies como vertebrados, insectos, gusanos y protistas, donde se ha visto una distribución más amplia de elementos *Harbinger* (Nouroz et al., 2016), tal como se ha encontrado en el genoma de algunos mosquitos (Kapitonov & Jurka, 2004).

Aquí, se logró caracterizar estructuralmente un transposón completo de esta superfamilia en *D. amaguana*, el mismo que mostró una longitud de 1.9 kb pero ninguna secuencia codificante en su dominio interno. Esto catalogó al elemento como inactivo al carecer de dos proteínas indispensables para su movimiento en el genoma, una transposasa y una proteína con un dominio MADF (myb/SANT-like domain in Adf-1), (Markova et al., 2022). Estudios basados en genomas de *Drosophila* reportan al menos cuatro clados de transposones *PIF-like* (*DPLT*) en especies como *D. willistoni*, *D. yakuba*, *D. pseudoobscura*, y *D. persimilis* (Casola et al., 2007), mismos que han dado origen a genes denominados *Drosophila PIF Like Genes* (*DPLG*) como parte de la domesticación de sus transposasas (Casola et al., 2007).

Esta información resulta ser la base para llevar a cabo más estudios que permitan comprender e identificar eventos de domesticación con estos transposones en *D. amaguana*, pues durante la evolución de los insectos es probable que hayan sucedido más de uno de estos eventos (Markova et al., 2022), quizá por la naturaleza inherente de las transposasas en esta superfamilia (Casola et al., 2007) que aún no han sido estudiados en su totalidad en nuevas especies de *Drosophila*.

A pesar de que muchos insectos son muy susceptibles a ser invadidos por transposones *Tc1/Mariner* (Jacobson et al., 1986), una de las superfamilias con mayor distribución en la naturaleza y en el reino eucariota (Y. Liu et al., 2021; Yuan & Wessler, 2011), los resultados para la superfamilia *Tc1/Mariner* en *D. amaguana* fueron de los menos abundantes del genoma, superando al de las superfamilias *Maverick*, *Copia*, y, *Penelope*, que mostraron un número de copias inferior a cien, lo que apenas representa el 0.01 % de secuencias en el genoma.

Sin embargo, la identificación y caracterización de estos transposones en esta especie andina genera un panorama inicial para comprender de mejor manera la evolución y diversidad de estos elementos transponibles, cuya información todavía sigue siendo un poco limitada (Dupeyron et al., 2020). Dentro del grupo *mesophragmatica* se ha identificado *in vitro* a través de PCR elementos *mariner* en especies como *D. gasici*, *D. brncici*, y, *D. viracochi* (Germanos et al., 2006).

Estructuralmente el elemento completo identificado presenta una longitud de 1.8 kb, pero internamente no mostró una región codificante para una transposasa, o algún otro gen. Esta inactividad actual del elemento analizado, podría ser consecuencia de eventos de transposición muy antiguos junto con la aparición de indels. A pesar de ello, durante el análisis de otras secuencias de las diferentes superfamilias que hemos abordado en este trabajo, se encontró una secuencia identificada dentro del grupo de los transposones *hAT* que contenía un

CDS codificante para una transposasa *Tc1-like*. Esta transposasa ha sido catalogada como parte del elemento *Bari-1*, una familia de transposones con similitud frente al transposón *Tc1* reportado por primera vez en *C. elegans* (Caizzi et al., 1993).

El hecho en cuestión abre una duda respecto a la correcta clasificación de las secuencias realizada por el pipeline de EDTA. El transposón de 2.1 kb que inicialmente fue clasificado como parte de la superfamilia *hAT* dentro del contig *jcf7180002877861* fue analizado en un *dot plot*, esto permitió identificar gráficamente TIRs de 36 pb flanqueados por TSDs muy cortos de 2 pb (TA), un motivo muy conservado y característico por excelencia en la superfamilia *Tc1/Mariner* (Wicker et al., 2007), pues para los elementos *hAT* se hubiese esperado TSDs de 8 pb. Sin embargo, el hallazgo sobre este transposón se convierte en un hecho interesante, ya que algunas familias de elementos transponibles como *Bari-1* y *Bari-3* han demostrado ser autónomamente funcionales en el género *Drosophila* (Palazzo et al., 2014), incluso en especies con genomas evolutivamente distantes (Palazzo et al., 2019) introduciendo novedades genómicas (Palazzo et al., 2022).

La identificación *de novo* de ETs activos en un genoma puede resultar todo un reto. Ya que la acumulación de mutaciones que generan codones *stop* prematuros en la transposasa de un ET (Ma et al., 2004) dificulta el reconocimiento de ETs activos, sobre todo cuando los elementos inactivos los superan en número (Z. Liu et al., 2021). Además, muchos elementos podrían estar silenciados (Atkinson, 2015) por la expresión de mecanismos represores como la metilación del DNA, modificación de la cromatina o interferencia de RNA (Fultz et al., 2015), sin embargo, estos hechos solo podrían determinarse a partir de datos transcriptómicos.

En *D. amaguana* la identificación *de novo* de elementos transponibles con repeticiones terminales invertidas (TIR), según la clasificación de Wicker et al., 2007, mostró resultados con ausencias de elementos activos en aquellas secuencias estructuralmente intactas que fueron

caracterizadas y anotadas manualmente. Este hecho tal vez podría asociarse a la idea planteada años atrás donde se hablaba respecto al fuerte sistema de defensa en los mosquitos, y su capacidad de silenciar efectivamente ETs exógenos (Scali et al., 2007). No obstante, encontrar la presencia de elementos activos no siempre debería limitarse a un solo ejemplar, pues podría existir la posibilidad de encontrarlos en otros individuos o poblaciones de la misma especie (Rossato et al., 2014).

Por otra parte, se observaron ciertas inconsistencias de los resultados finales arrojados por el pipeline, como por ejemplo el reporte de elementos transponibles que hasta ahora han sido identificados predominantemente solo en plantas y no en otras especies. Así mismo, con EDTA no fue posible identificar la presencia de un *elemento P* que por su importancia dentro del género de *Drosophila*, tuvo que ser buscado manualmente y fue encontrado estructuralmente intacto. Incluso se evidenció en un caso particular la asignación de una superfamilia a un transposón que posiblemente corresponde a otra superfamilia, como ya se lo mencionó en párrafos anteriores.

Y aunque si bien se han obtenido los primeros resultados respecto al genoma de esta especie, se deben considerar estos detalles para profundizar en los estudios y evaluar pipelines alternativos que de manera más precisa permitan la identificación y anotación de elementos transponibles, de tal modo que se aporte con datos verdaderos sobre ETs en especies neotropicales del grupo *mesophragmatica* que habitan los bosques andinos.

## Capítulo VI: Conclusiones y Recomendaciones

### Conclusiones

- Se identificó a través de una anotación estructural *de novo* utilizando la herramienta EDTA, siete superfamilias de elementos transponibles clase II en el genoma ensamblado de *Drosophila amaguana*, las mismas que fueron *Mutator*, *CACTA*, *Helitron*, *hAT*, *PIF/Harbinger*, *Tc1/Mariner*, y, *Maverick*. Mientras que una superfamilia adicional correspondiente a *elementos P* fue anotada por su estructura de forma manual en uno de los contigs del genoma.
- Se caracterizó un elemento transponible completo de clase II de la subclase TIR perteneciente a cada superfamilia identificados en el genoma de *Drosophila amaguana*. Estructuralmente la caracterización se ejecutó con DOTTER a través de un *dot plot* para la identificación de TIRs de tamaños específicos para cada superfamilia, mientras que la caracterización funcional se llevó a cabo mediante anotación de los componentes del transposón en su secuencia completa. Así, reportamos seis ETs de la subclase TIR con estructura completa pero sin capacidad de transposición en el genoma.
- Se determinó el porcentaje de elementos transponibles clase II identificados *de novo* en el genoma de *Drosophila amaguana*, el cual fue 7.87% de transposones de DNA en un total de 11.87 % de elementos móviles. El orden TIR fue el más abundante con 6.50 % de secuencias y que estuvo contribuido por 3.29 % de *Mutator*, 1.98 % de *CACTA*, 0.52 % de *hAT*, 0.46 % de *PIF/Harbinger*, 0.24 % de *Tc1/Mariner*, y una contribución de hasta el 0.01 % entre *Maverick*, un *elemento P*, y superfamilias desconocidas. Por otra parte, elementos *Helitron* representaron el 3.91 % del total del contenido de transposones.

- Se comparó el contenido de elementos transponibles clase II entre los valores que identificamos para *Drosophila amaguana* con aquellos obtenidos en reportes de otros estudios en el genoma de *Drosophila melanogaster*. Se observó que el contenido reportado entre 2 – 2.3 % de ETs clase II en el genoma referencial ensamblado de *D. melanogaster* con un tamaño estimado de 142.6 Mb, es evidentemente hasta 4 veces menor al contenido identificado en el presente estudio para *D. amaguana*, cuyo tamaño del genoma ensamblado fue 455.5 Mb, donde se encontró 7.87 % de transposones de DNA, con una predominancia en la subclase TIR.

## Recomendaciones

- Se recomienda evaluar y aplicar otras herramientas bioinformáticas que permitan la identificación elementos transponibles mediante una metodología *de novo*, y también considerando el uso de librerías curadas de secuencias homólogas de ETs en otras especies para los datos del genoma ensamblado de *D. amaguana*, y así, de esta manera, determinar el programa con mayor sensibilidad y especificidad para la anotación estructural de elementos móviles en esta especie.
- Realizar un análisis completo para cuantificar el número de copias exactas de cada secuencia en el genoma del organismo.
- Es importante abrir la invitación a la academia para poder solicitar, gestionar, y conseguir acceso mediante licencias académicas a bases de datos como RepBase, ya que cuentan con secuencias únicas de DNA repetitivo para un gran número de especies eucariotas que no son publicadas en otros repositorios o bases de datos, y por lo tanto, existe una limitante de acceso a información importante que puede requerirse en cualquier etapa de la investigación.
- Generar nuevos análisis de secuencias de ETs en *D. amaguana* partiendo de datos obtenidos con NGS de 3<sup>ra</sup> generación para su transcriptoma o nuevamente su genoma, y así, evitar subrepresentaciones de las secuencias repetitivas en el genoma ensamblado al utilizar solo lecturas obtenidas de la secuenciación de 2<sup>da</sup> generación.
- Finalmente, se debería continuar complementando otros estudios *in silico* e *in vitro* que permitan ampliar el conocimiento sobre otras especies andinas de *Drosophila*, pues los resultados presentados como parte de esta investigación, constituyen la primera información a nivel genético sobre una especie neotropical de *Drosophila* distribuida en el Ecuador.

## Referencias Bibliográficas

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., ... Craig Venter, J. (2000). The Genome Sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185–2195.  
<https://doi.org/10.1126/SCIENCE.287.5461.2185>
- Almojil, D., Bourgeois, Y., Falis, M., Hariyani, I., Wilcox, J., & Boissinot, S. (2021). The structural, functional and evolutionary impact of transposable elements in Eukaryotes. *Genes*, 12(6). <https://doi.org/10.3390/genes12060918>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410.  
[https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Aminetzach, Y. T., Macpherson, J. M., & Petrov, D. A. (2005). Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science (New York, N.Y.)*, 309(5735), 764–767. <https://doi.org/10.1126/SCIENCE.1112699>
- Animal Research. (n.d.). *Drosophila melanogaster*. Understanding Animal Research. Retrieved September 9, 2022, from <https://www.animalresearch.info/es/el-diseno-de-la-investigacion/animales-de-investigacion/drosophila-melanogaster/>
- Arkhipova, I. R. (2006). Distribution and Phylogeny of Penelope-Like Elements in Eukaryotes. *Systematic Biology*, 55(6), 875–885.  
<https://doi.org/10.1080/10635150601077683>
- Arkhipova, I. R., Yushenova, I. A., & Rodriguez, F. (2013). Endonuclease-containing

Penelope retrotransposons in the bdelloid rotifer *Adineta vaga* exhibit unusual structural features and play a role in expansion of host gene families. *Mobile DNA*, 4(1), 1–12.

<https://doi.org/10.1186/1759-8753-4-19/FIGURES/5>

Aroh, O., & Halanych, K. M. (2021). Genome-wide characterization of LTR retrotransposons in the non-model deep-sea annelid *Lamellibrachia luymesii*. *BMC Genomics*, 22(1).

<https://doi.org/10.1186/S12864-021-07749-1>

Atkinson, P. W. (2015). hAT Transposable Elements . *Microbiology Spectrum*, 3(4).

<https://doi.org/10.1128/MICROBIOLSPEC.MDNA3-0054-2014/ASSET/E936D398-F1C0-4C86-B468-C58A9111DC01/ASSETS/GRAPHIC/MDNA3-0054-2014-FIG6B.GIF>

Banho, C. A., Oliveira, D. S., Haudry, A., Fablet, M., Vieira, C., & Carareto, C. M. A.

(2021). *Transposable Element Expression and Regulation Profile in Gonads of Interspecific Hybrids of Drosophila arizonae and Drosophila mojavensis wrighti*.

<https://doi.org/10.3390/cells10123574>

Barrón, M. G., Fiston-Lavier, A. S., Petrov, D. A., & González, J. (2014). Population genomics of transposable elements in *Drosophila*. *Annual Review of Genetics*,

48(September), 561–581. <https://doi.org/10.1146/annurev-genet-120213-092359>

Beck, C. R., Garcia-Perez, J. L., Badge, R. M., & Moran, J. V. (2011). LINE-1 Elements in Structural Variation and Disease. *Annual Review of Genomics and Human Genetics*, 12,

187. <https://doi.org/10.1146/ANNUREV-GENOM-082509-141802>

Belyayev, A., Josefiová, J., Jandová, M., Kalendar, R., Mahelka, V., Mandák, B., & Krak, K.

(2022). The structural diversity of CACTA transposons in genomes of *Chenopodium* (Amaranthaceae, Caryophyllales) species: specific traits and comparison with the similar elements of angiosperms. *Mobile DNA*, 13(1), 1–16. <https://doi.org/10.1186/S13100->

- Bertocchi, N. Á., Oliveira, T. D. de, Deprá, M., Goñi, B., Valente, V. L. S., Bertocchi, N. Á., Oliveira, T. D. de, Deprá, M., Goñi, B., & Valente, V. L. S. (2022). Interpopulation variation of transposable elements of the hAT superfamily in *Drosophila willistoni* (Diptera: Drosophilidae): in-situ approach. *Genetics and Molecular Biology*, *45*(2), 20210287. <https://doi.org/10.1590/1678-4685-GMB-2021-0287>
- Biémont, C., & Vieira, C. (2006). Junk DNA as an evolutionary force. *Nature* *2006* *443*:7111, *443*(7111), 521–524. <https://doi.org/10.1038/443521a>
- Biémont, C., Vieira, C., Borie, N., & Lepetit, D. (1999). Transposable elements and genome evolution: The case of *Drosophila simulans*. *Genetica*, *107*(1–3), 113–120. <https://doi.org/10.1023/a:1003937603230>
- Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology* *2018* *19*:1, *19*(1), 1–12. <https://doi.org/10.1186/S13059-018-1577-Z>
- Brillet, B., Bigot, Y., & Augé-Gouillou, C. (2007). Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica*, *130*(2), 105–120. <https://doi.org/10.1007/S10709-006-0025-2>
- Britten, R. J., & Kohne, D. E. (1968). Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science (New York, N.Y.)*, *161*(3841), 529–540. <https://doi.org/10.1126/SCIENCE.161.3841.529>

- Brncic, D. (1957). A Comparative Study of Chromosomal Variation in Species of the Mesophragmatica Group of *Drosophila*. *Genetics*, 42(6), 798.  
<https://doi.org/10.1093/GENETICS/42.6.798>
- Brncic, D., & Koref, S. (1957). The Mesophragmatica Group of Species of *Drosophila*. *Evolution*, 11(3), 300–310. <https://doi.org/10.1111/J.1558-5646.1957.TB02899.X>
- Buchmann, J. P., Löytynoja, A., Wicker, T., & Schulman, A. H. (2014). Analysis of CACTA transposases reveals intron loss as major factor influencing their exon/intron structure in monocotyledonous and eudicotyledonous hosts. *Mobile DNA*, 5(1), 1–15.  
<https://doi.org/10.1186/1759-8753-5-24/FIGURES/3>
- Caizzi, R., Caggese, C., & Pimpinelli, S. (1993). Bari-1, a New Transposon-like Family in *Drosophila Melanogaster* with a Unique Heterochromatic Organization. *Genetics*, 133(2), 335. <https://doi.org/10.1093/GENETICS/133.2.335>
- Calvi, B. R., Hong, T. J., Findley, S. D., & Gelbart, W. M. (1991). Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: hobo, activator, and Tam3. *Cell*, 66(3), 465–471. [https://doi.org/10.1016/0092-8674\(81\)90010-6](https://doi.org/10.1016/0092-8674(81)90010-6)
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10.  
<https://doi.org/10.1186/1471-2105-10-421>
- Canapa, A., Barucca, M., Biscotti, M. A., Forconi, M., & Olmo, E. (2015). Transposons, Genome Size, and Evolutionary Insights in Animals. *Cytogenetic and Genome Research*, 147(4), 217–239. <https://doi.org/10.1159/000444429>
- Carver, T., Harris, S. R., Berriman, M., Parkhill, J., & McQuillan, J. A. (2012). Artemis: an

integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*, 28(4), 464–469.

<https://doi.org/10.1093/BIOINFORMATICS/BTR703>

Casacuberta, E., & González, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular Ecology*, 22(6), 1503–1517.

<https://doi.org/10.1111/mec.12170>

Casola, C., Lawing, A. M., Betrán, E., & Feschotte, C. (2007). PIF-like Transposons are Common in *Drosophila* and Have Been Repeatedly Domesticated to Generate New Host Genes. *Molecular Biology and Evolution*, 24(8), 1872–1888.

<https://doi.org/10.1093/MOLBEV/MSM116>

Cavasini, R., Buschini, M. L. T., Machado, L. P. B., & Mateus, R. P. (2014). Comparison of *Drosophilidae* (Diptera) assemblages from two highland *Araucaria* Forest fragments, with and without environmental conservation policies. *Brazilian Journal of Biology*, 74(4), 761–768. <https://doi.org/10.1590/1519-6984.00113>

Celniker, S. E., & Rubin, G. M. (2003). The *Drosophila Melanogaster* Genome. *Annual Review of Genomics and Human Genetics*, 4, 89–117.

<https://doi.org/10.1146/ANNUREV.GENOM.4.070802.110323>

Cerbin, S., & Jiang, N. (2018). Duplication of host genes by transposable elements. *Current Opinion in Genetics & Development*, 49, 63–69.

<https://doi.org/10.1016/J.GDE.2018.03.005>

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>

- Chen, S., & Li, X. (2008). Molecular characterization of the first intact Transib transposon from *Helicoverpa zea*. *Gene*, *408*(1–2), 51–63.  
<https://doi.org/10.1016/J.GENE.2007.10.015>
- Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics* *2016 18:2*, *18*(2), 71–86.  
<https://doi.org/10.1038/nrg.2016.139>
- Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., Iyer, V. N., Pollard, D. A., Sackton, T. B., Larracuente, A. M., Singh, N. D., Abad, J. P., Abt, D. N., Adryan, B., Aguade, M., Akashi, H., ... MacCallum, I. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, *450*(8), 203–218. <https://doi.org/10.1038/nature06341>
- Colonna Romano, N., & Fanti, L. (2022). Transposable Elements: Major Players in Shaping Genomic and Evolutionary Patterns. *Cells* *2022, Vol. 11, Page 1048*, *11*(6), 1048.  
<https://doi.org/10.3390/CELLS11061048>
- Cooley, L., Kelley, R., & Spradling, A. (1988). Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science (New York, N.Y.)*, *239*(4844), 1121–1128.  
<https://doi.org/10.1126/SCIENCE.2830671>
- Craig Venter, J., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., ... Zhu, X. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351. [https://doi.org/10.1126/SCIENCE.1058040/SUPPL\\_FILE/C10\\_SCIENCE.PDF](https://doi.org/10.1126/SCIENCE.1058040/SUPPL_FILE/C10_SCIENCE.PDF)
- da Silva Loreto, E. L. S., Basso da Silva, L., Zaha, A., & Da Valente, V. L. S. (1997). Distribution of transposable elements in neotropical species of *Drosophila*. *Genetica*,

101(3), 153–165. <https://doi.org/10.1023/a:1018381104700>

Dajoz, R. (2000). *Entomología Forestal. Los insectos y el bosque. Papel y diversidad de los insectos en el medio forestal*. Ediciones Mundi-Prensa.

De Freitas Ortiz, M., & Loreto, E. L. S. (2009). Characterization of new hAT transposable elements in 12 *Drosophila* genomes. *Genetica*, 135(1), 67–75.  
<https://doi.org/10.1007/S10709-008-9259-5>

Deininger, P. (2011). Alu elements: Know the SINEs. *Genome Biology*, 12(12), 1–12.  
<https://doi.org/10.1186/GB-2011-12-12-236/FIGURES/5>

Deloger, M., Cavalli, F. M. G., Lerat, E., Biémont, C., Sagot, M. F., & Vieira, C. (2009). Identification of expressed transposable element insertions in the sequenced genome of *Drosophila melanogaster*. *Gene*, 439(1–2), 55–62.  
<https://doi.org/10.1016/J.GENE.2009.03.015>

Drysdale, R. (2008). FlyBase : a database for the *Drosophila* research community. *Methods in Molecular Biology (Clifton, N.J.)*, 420, 45–59. [https://doi.org/10.1007/978-1-59745-583-1\\_3](https://doi.org/10.1007/978-1-59745-583-1_3)

Duncan, L., Bouckaert, K., Yeh, F., & Kirk, D. L. (2002). kangaroo, a mobile element from *Volvox carteri*, is a member of a newly recognized third class of retrotransposons. *Genetics*, 162(4), 1617. <https://doi.org/10.1093/GENETICS/162.4.1617>

Dunin, S., Feder, M., & Bujnicki, J. M. (2006). Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, 7(1), 1–19. <https://doi.org/10.1186/1471-2164-7-98/FIGURES/8>

Dupeyron, M., Baril, T., Bass, C., & Hayward, A. (2020). Phylogenetic analysis of the Tc1/mariner superfamily reveals the unexplored diversity of pogo-like elements. *Mobile*

- DNA*, 11(1), 1–14. <https://doi.org/10.1186/S13100-020-00212-0/TABLES/1>
- Dupeyron, M., Singh, K. S., Bass, C., & Hayward, A. (2019). Evolution of Mutator transposable elements across eukaryotic diversity. *Mobile DNA*, 10(1). <https://doi.org/10.1186/S13100-019-0153-8>
- Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, 9(1), 1–14. <https://doi.org/10.1186/1471-2105-9-18/TABLES/5>
- Esnault, C., Maestre, J., & Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nature Genetics* 24:4, 24(4), 363–367. <https://doi.org/10.1038/74184>
- Fablet, M., Salces-Ortiz, J., Menezes, B. F., Roy, M., & Vieira, C. (2017). Self and nonself from a genomic perspective: Transposable elements. *Evolutionary Biology: Self/Nonsel Evolution, Species and Complex Traits Evolution, Methods and Concepts*, 111–128. [https://doi.org/10.1007/978-3-319-61569-1\\_6/FIGURES/4](https://doi.org/10.1007/978-3-319-61569-1_6/FIGURES/4)
- Fedoroff, N. V. (2013). Molecular genetics and epigenetics of CACTA elements. *Methods in Molecular Biology (Clifton, N.J.)*, 1057, 177–192. [https://doi.org/10.1007/978-1-62703-568-2\\_13](https://doi.org/10.1007/978-1-62703-568-2_13)
- Fernández, M. A., Farr, C. L., Kaguni, L. S., & Garesse, R. (2007). *Drosophila melanogaster* as a Model System to Study Mitochondrial Biology. *Methods in Molecular Biology (Clifton, N.J.)*, 372, 33. [https://doi.org/10.1007/978-1-59745-365-3\\_3](https://doi.org/10.1007/978-1-59745-365-3_3)
- Feschotte, C. (2004). Merlin, a New Superfamily of DNA Transposons Identified in Diverse Animal Genomes and Related to Bacterial IS1016 Insertion Sequences. *Molecular Biology and Evolution*, 21(9), 1769–1780. <https://doi.org/10.1093/MOLBEV/MSH188>

- Feschotte, C., & Mouchès, C. (2000). Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. *Molecular Biology and Evolution*, *17*(5), 730–737.  
<https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026351>
- Feschotte, C., & Pritham, E. J. (2007). DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics*, *41*, 331–368.  
<https://doi.org/10.1146/ANNUREV.GENET.40.110405.090448>
- Figuro, M. (2017). *Filogenia molecular de especies ecuatorianas del grupo Drosophila mesophragmatica (Diptera, Drosophilidae)* [Pontificia Universidad Católica del Ecuador]. [http://repositorio.puce.edu.ec/bitstream/handle/22000/13207/Tesis M.L. Figuro 2017.pdf?sequence=1&isAllowed=y](http://repositorio.puce.edu.ec/bitstream/handle/22000/13207/Tesis%20M.L.%20Figuro%202017.pdf?sequence=1&isAllowed=y)
- Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends in Genetics : TIG*, *5*(4), 103–107. [https://doi.org/10.1016/0168-9525\(89\)90039-5](https://doi.org/10.1016/0168-9525(89)90039-5)
- Finnegan, D. J. (2012). Retrotransposons. *Current Biology : CB*, *22*(11).  
<https://doi.org/10.1016/J.CUB.2012.04.025>
- Fugmann, S. D. (2001). RAG1 and RAG2 in V(D)J recombination and transposition. *Immunologic Research*, *23*(1), 23–39. <https://doi.org/10.1385/IR:23:1:23>
- Fultz, D., Choudury, S. G., & Slotkin, R. K. (2015). Silencing of active transposable elements in plants. *Current Opinion in Plant Biology*, *27*, 67–76.  
<https://doi.org/10.1016/J.PBI.2015.05.027>
- García, M. P. (2012). What makes transposable elements move in the *Drosophila* genome? *Heredity*, *108*(5), 461–468. <https://doi.org/10.1038/hdy.2011.89>
- Gelbart, W. M., Crosby, M., Matthews, B., Rindone, W. P., Chillemi, J., Twombly, S. R.,

- Emmert, D., Bayraktaroglu, L., Ashburner, M., Drysdale, R. A., Whitfield, E. J., Millburn, G. H., De Grey, A., Kaufman, T., Matthews, K., Gilbert, D., Strelets, V., Tolstoshev, C., Rubin, G. M., ... Benos, T. (1998). FlyBase: A Drosophila database. *Nucleic Acids Research*, *26*(1), 85–88. <https://doi.org/10.1093/NAR/26.1.85>
- Gemmell, N. J. (2021). Repetitive DNA: genomic dark matter matters. *Nature Reviews Genetics* *2021 22:6*, *22*(6), 342–342. <https://doi.org/10.1038/s41576-021-00354-8>
- Germanos, E., Mota, N. R., & Loreto, E. L. S. (2006). Transposable elements from the mesophragmatica group of Drosophila. *Genetics and Molecular Biology*, *29*(4), 741–746. <https://doi.org/10.1590/S1415-47572006000400026>
- Ghanim, G. E., Rio, D. C., & Teixeira, F. K. (2020). Mechanism and regulation of P element transposition. *Open Biology*, *10*(12). <https://doi.org/10.1098/RSOB.200244>
- Goodwin, T. J. D., Butler, M. I., & Poulter, R. T. M. (2003). Cryptons: A group of tyrosine-recombinase- encoding DNA transposons from pathogenic fungi. *Microbiology*, *149*(11), 3099–3109. <https://doi.org/10.1099/MIC.0.26529-0/CITE/REFWORKS>
- Goodwin, T. J. D., & Poulter, R. T. M. (2001). The DIRS1 Group of Retrotransposons. *Molecular Biology and Evolution*, *18*(11), 2067–2082. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A003748>
- Grabundzija, I., Hickman, A. B., & Dyda, F. (2018). Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. *Nature Communications* *2018 9:1*, *9*(1), 1–14. <https://doi.org/10.1038/s41467-018-03688-w>
- Gregory, T. R., & Johnston, J. S. (2008). Genome size diversity in the family Drosophilidae. *Heredity* *2008 101:3*, *101*(3), 228–238. <https://doi.org/10.1038/hdy.2008.49>
- Grzebelus, D., Lasota, S., Gambin, T., Kucherov, G., & Gambin, A. (2007). Diversity and

- structure of PIF/Harbinger-like elements in the genome of *Medicago truncatula*. *BMC Genomics*, 8(1), 1–14. <https://doi.org/10.1186/1471-2164-8-409/FIGURES/7>
- Guio, L., & González, J. (2019). New Insights on the Evolution of Genome Content: Population Dynamics of Transposable Elements in Flies and Humans. *Methods in Molecular Biology (Clifton, N.J.)*, 1910, 505–530. [https://doi.org/10.1007/978-1-4939-9074-0\\_16](https://doi.org/10.1007/978-1-4939-9074-0_16)
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, 29(8), 1072–1075. <https://doi.org/10.1093/BIOINFORMATICS/BTT086>
- Haapa, S., Wahlberg, N., & Savilahti, H. (2014). Phylogenetic analysis of Maverick/Polinton giant transposons across organisms. *Molecular Phylogenetics and Evolution*, 78(1), 271–274. <https://doi.org/10.1016/J.YMPEV.2014.05.024>
- Hammer, S. E., Strehl, S., & Hagemann, S. (2005). Homologs of *Drosophila* P Transposons Were Mobile in Zebrafish but Have Been Domesticated in a Common Ancestor of Chicken and Human. *Molecular Biology and Evolution*, 22(4), 833–844. <https://doi.org/10.1093/MOLBEV/MSI068>
- Han, Y., Qin, S., & Wessler, S. R. (2013). Comparison of class 2 transposable elements at superfamily resolution reveals conserved and distinct features in cereal grass genomes. *BMC Genomics*, 14(1), 71. <https://doi.org/10.1186/1471-2164-14-71>
- Harnish, M., Link, N., & Yamamoto, S. (2021). *Drosophila* as a Model for Infectious Diseases. *International Journal of Molecular Sciences*, 22(5), 1–42. <https://doi.org/10.3390/IJMS22052724>
- Havecker, E. R., Gao, X., & Voytas, D. F. (2004). The diversity of LTR retrotransposons.

*Genome Biology*, 5(6), 1–6. <https://doi.org/10.1186/GB-2004-5-6-225/FIGURES/3>

Hayward, A., Ghazal, A., Andersson, G., Andersson, L., & Jern, P. (2013). ZBED Evolution:

Repeated Utilization of DNA Transposons as Regulators of Diverse Host Functions.

*PLOS ONE*, 8(3), e59940. <https://doi.org/10.1371/JOURNAL.PONE.0059940>

Hu, J., Zheng, Y., & Shang, X. (2018). MiteFinderII: A novel tool to identify miniature

inverted-repeat transposable elements hidden in eukaryotic genomes 06 Biological

Sciences 0604 Genetics. *BMC Medical Genomics*, 11(5), 51–59.

<https://doi.org/10.1186/S12920-018-0418-Y/TABLES/2>

Hu, K., Xu, K., Wen, J., Yi, B., Shen, J., Ma, C., Fu, T., Ouyang, Y., & Tu, J. (2019).

Helitron distribution in Brassicaceae and whole Genome Helitron density as a character

for distinguishing plant species. *BMC Bioinformatics*, 20(1), 1–20.

<https://doi.org/10.1186/S12859-019-2945-8/FIGURES/9>

Ivics, Z., Li, M. A., Mátés, L., Boeke, J. D., Nagy, A., Bradley, A., & Izsvák, Z. (2009).

Transposon-mediated Genome Manipulations in Vertebrates. *Nature Methods*, 6(6), 415.

<https://doi.org/10.1038/NMETH.1332>

Jacobson, J. W., Medhora, M. M., Hartlt, D. L., Simulans, D., Mauritiana, D., Sechel-Lia, D.,

Erecta, D., Orena, D., Yakuba, D., Teissieri, D., Woodruff, R. C., Lyman, R. F. D.,

Sechellia, D., Mauritiana W', D., & Melanogaster, D. (1986). Molecular structure of a

somatically unstable transposable element in *Drosophila*. *Proceedings of the National*

*Academy of Sciences*, 83(22), 8684–8688. <https://doi.org/10.1073/PNAS.83.22.8684>

Jennings, B. H. (2011). *Drosophila* – a versatile model in biology & medicine. *Materials*

*Today*, 14(5), 190–195. [https://doi.org/10.1016/S1369-7021\(11\)70113-4](https://doi.org/10.1016/S1369-7021(11)70113-4)

Jeong, H. H., Yalamanchili, H. K., Guo, C., Shulman, J. M., & Liu, Z. (2018). An ultra-fast

and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pacific Symposium on Biocomputing*, 0(212669), 168–179.

[https://doi.org/10.1142/9789813235533\\_0016](https://doi.org/10.1142/9789813235533_0016)

Jiang, N., Bao, Z., Zhang, X., Eddy, S. R., & Wessler, S. R. (2004). Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 2004 431:7008, 431(7008), 569–573. <https://doi.org/10.1038/nature02953>

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), 1–20.

Kanhayuwa, L., & Coutts, R. H. A. (2016). Short Interspersed Nuclear Element (SINE) Sequences in the Genome of the Human Pathogenic Fungus *Aspergillus fumigatus* Af293. *PLoS ONE*, 11(10). <https://doi.org/10.1371/JOURNAL.PONE.0163215>

Kapitonov, V., & Jurka, J. (2005). RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLOS Biology*, 3(6), e181. <https://doi.org/10.1371/JOURNAL.PBIO.0030181>

Kapitonov, V., & Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12), 4540–4545. [https://doi.org/10.1073/PNAS.0600833103/SUPPL\\_FILE/00833TABLE1.PDF](https://doi.org/10.1073/PNAS.0600833103/SUPPL_FILE/00833TABLE1.PDF)

Kapitonov, V. V., & Jurka, J. (2001). Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), 8714–8719. <https://doi.org/10.1073/PNAS.151269298/ASSET/2B569A99-E826-4C26-9D5E-06EA1A6B4026/ASSETS/GRAPHIC/PQ1512692006.JPEG>

- Kapitonov, V. V., & Jurka, J. (2003). Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(11), 6569–6574.  
<https://doi.org/10.1073/PNAS.0732024100>
- Kapitonov, V. V., & Jurka, J. (2004). Harbinger Transposons and an Ancient HARBI1 Gene Derived from a Transposase. *Https://Home.Liebertpub.Com/Dna*, *23*(5), 311–324.  
<https://doi.org/10.1089/104454904323090949>
- Kaufman, T. C. (2017). A Short History and Description of *Drosophila melanogaster* Classical Genetics: Chromosome Aberrations, Forward Genetic Screens, and the Nature of Mutations. *Genetics*, *206*(2), 665–689.  
<https://doi.org/10.1534/GENETICS.117.199950>
- Kazazian, H. H. (2011). Mobile DNA transposition in somatic cells. *BMC Biology*, *9*(1), 1–4.  
<https://doi.org/10.1186/1741-7007-9-62/FIGURES/1>
- Kidwell, M. (1983). Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(6), 1655. <https://doi.org/10.1073/PNAS.80.6.1655>
- Kidwell, M. G., & Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(15), 7704–7711.  
<https://doi.org/10.1073/PNAS.94.15.7704/ASSET/E0DE1AC7-E55A-4605-A570-BD91C51FC0CA/ASSETS/GRAPHIC/PQ1571507001.JPEG>
- Kojima, K. K. (2019). Structural and sequence diversity of eukaryotic transposable elements. *Genes & Genetic Systems*, *94*(6), 233–252. <https://doi.org/10.1266/GGS.18-00024>

- Kojima, K. K., & Jurka, J. (2011). Crypton transposons: Identification of new diverse families and ancient domestication events. *Mobile DNA*, 2(1), 1–17.  
<https://doi.org/10.1186/1759-8753-2-12/FIGURES/7>
- Kornberg, T. B., & Krasnow, M. A. (2000). The *Drosophila* genome sequence: implications for biology and medicine. *Science (New York, N.Y.)*, 287(5461), 2218–2220.  
<https://doi.org/10.1126/SCIENCE.287.5461.2218>
- Kramerov, D. A., & Vassetzky, N. S. (2011). Origin and evolution of SINEs in eukaryotic genomes. *Heredity* 2011 107:6, 107(6), 487–495. <https://doi.org/10.1038/hdy.2011.43>
- Kuo, Y. T., Ishii, T., Fuchs, J., Hsieh, W. H., Houben, A., & Lin, Y. R. (2021). The Evolutionary Dynamics of Repetitive DNA and Its Impact on the Genome Diversification in the Genus *Sorghum*. *Frontiers in Plant Science*, 12, 1735.  
<https://doi.org/10.3389/FPLS.2021.729734/BIBTEX>
- Ladevèze, V., Chaminade, N., Lemeunier, F., Periquet, G., & Aulard, S. (2012). General survey of hAT transposon superfamily with highlight on hobo element in *Drosophila*. *Genetica* 2012 140:7, 140(7), 375–392. <https://doi.org/10.1007/S10709-012-9687-0>
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., ... Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921.  
<https://doi.org/10.1038/35057062>
- Lerat, E. (2010). Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity*, 104(6), 520–533.  
<https://doi.org/10.1038/HDY.2009.165>

- Lerat, E., Casacuberta, J., Chaparro, C., & Vieira, C. (2019). On the importance to acknowledge transposable elements in epigenomic analyses. *Genes*, *10*(4), 1–18. <https://doi.org/10.3390/genes10040258>
- Li, X., Burnight, E. R., Cooney, A. L., Malani, N., Brady, T., Sander, J. D., Staber, J., Wheelan, S. J., Joung, J. K., McCray, P. B., Bushman, F. D., Sinn, P. L., & Craig, N. L. (2013). PiggyBac transposase tools for genome engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(25), E2279–E2287. [https://doi.org/10.1073/PNAS.1305987110/SUPPL\\_FILE/KEITHJOUNGFENGZHANGPODCAST.MP3](https://doi.org/10.1073/PNAS.1305987110/SUPPL_FILE/KEITHJOUNGFENGZHANGPODCAST.MP3)
- Lisch, D. (2002). Mutator transposons. *Trends in Plant Science*, *7*(11), 498–504. [https://doi.org/10.1016/S1360-1385\(02\)02347-6](https://doi.org/10.1016/S1360-1385(02)02347-6)
- Lisch, D. (2009). Epigenetic Regulation of Transposable Elements in Plants. <https://doi.org/10.1146/Annurev.Arplant.59.032607.092744>, *60*, 43–66. <https://doi.org/10.1146/ANNUREV.ARPLANT.59.032607.092744>
- Lisch, D. (2015). Mutator and MULE Transposons. *Microbiology Spectrum*, *3*(2), 1–24. <https://doi.org/10.1128/MICROBIOLSPEC.MDNA3-0032-2014/ASSET/2B6DEB71-DFAC-46DE-A942-E826E516DE7F/ASSETS/GRAPHIC/MDNA3-0032-2014-FIG8.GIF>
- Liu, K., & Wessler, S. R. (2017). Transposition of Mutator–like transposable elements (MULEs) resembles hAT and Transib elements and V(D)J recombination. *Nucleic Acids Research*, *45*(11), 6644–6655. <https://doi.org/10.1093/NAR/GKX357>
- Liu, Y., Zong, W., Diaby, M., Lin, Z., Wang, S., Gao, B., Song, C., & Ji, T. (2021). Diversity and evolution of pogo and tc1/mariner transposons in the apoidea genomes. *Biology*, *10*(9), 940. <https://doi.org/10.3390/BIOLOGY10090940/S1>

- Liu, Z., Zhao, H., Yan, Y., Wei, M. X., Zheng, Y. C., Yue, E. K., Alam, M. S., Smartt, K. O., Duan, M. H., & Xu, J. H. (2021). Extensively Current Activity of Transposable Elements in Natural Rice Accessions Revealed by Singleton Insertions. *Frontiers in Plant Science*, *12*, 2161. <https://doi.org/10.3389/FPLS.2021.745526/BIBTEX>
- Lopes, A. L. K., Kriegová, E., Lukeš, J., Krieger, M. A., & Ludwig, A. (2021). Distribution of Merlin in eukaryotes and first report of DNA transposons in kinetoplastid protists. *PLoS ONE*, *16*(5). <https://doi.org/10.1371/JOURNAL.PONE.0251133>
- López, I., & Garrido, M. A. (2012). The Repetitive DNA Content of Eukaryotic Genomes. *Genome Dynamics*, *7*, 1–28. <https://doi.org/10.1159/000337118>
- Loreto, E. L. D. S., Valente, V. L. D. S., Zaha, A., Silva, J. C., & Kidwell, M. G. (2001). *Drosophila mediopunctata* P Elements: A New Example of Horizontal Transfer. *Journal of Heredity*, *92*(5), 375–381. <https://doi.org/10.1093/JHERED/92.5.375>
- Ma, J., Devos, K. M., & Bennetzen, J. L. (2004). Analyses of LTR-Retrotransposon Structures Reveal Recent and Rapid Genomic DNA Loss in Rice. *Genome Research*, *14*(5), 860–869. <https://doi.org/10.1101/GR.1466204>
- Macko, A., Machaj, G., & Grzebelus, D. (2021). A global landscape of miniature inverted-repeat transposable elements in the carrot genome. *Genes*, *12*(6), 859. <https://doi.org/10.3390/GENES12060859/S1>
- Majumdar, S., & Rio, D. C. (2015). P transposable elements in *Drosophila* and other eukaryotic organisms. *Microbiology Spectrum*, *3*(2), 1–25. <https://doi.org/10.1128/MICROBIOLSPEC.MDNA3-0004-2014>
- Makalowski, W., Gotea, V., Pande, A., & Makalowska, I. (2019). Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics. In

*Evolutionary Genomics Statistical and Computational Methods, Methods in Biology Molecular* (2nd ed., Vol. 1910, pp. 177–207).

Malamy, M. H., Fiandt, M., & Szybalski, W. (1972). Electron microscopy of polar insertions in the lac operon of *Escherichia coli*. *Molecular and General Genetics MGG 1972* 119:3, 119(3), 207–222. <https://doi.org/10.1007/BF00333859>

Malicki, M., Spaller, T., Winckler, T., & Hammann, C. (2020). DIRS retrotransposons amplify via linear, single-stranded cDNA intermediates. *Nucleic Acids Research*, 48(8), 4230–4243. <https://doi.org/10.1093/NAR/GKAA160>

Marchler-Bauer, A., & Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Research*, 32(Web Server issue). <https://doi.org/10.1093/NAR/GKH454>

Markova, D. N., Ruma, F. B., Casola, C., Mirsalehi, A., & Betrán, E. (2022). Recurrent co-domestication of PIF/Harbinger transposable element proteins in insects. *Mobile DNA*, 13(1), 1–15. <https://doi.org/10.1186/S13100-022-00282-2/FIGURES/7>

Markow, T. A. (2015). The secret lives of *Drosophila* flies. *ELife*, 4(JUNE). <https://doi.org/10.7554/ELIFE.06793>

Martinez, A. (2008). *Drosophila melanogaster* and the development of biology in the 20th century. *Methods in Molecular Biology (Clifton, N.J.)*, 420, 1–25. [https://doi.org/10.1007/978-1-59745-583-1\\_1/COVER](https://doi.org/10.1007/978-1-59745-583-1_1/COVER)

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, 36(6), 344–355. <https://doi.org/10.1073/PNAS.36.6.344>

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science*, 226(4676), 792–801. <https://doi.org/10.1126/SCIENCE.15739260/ASSET/A3643FE5->

38B1-457C-A50B-DC2C01D1BBAE/ASSETS/SCIENCE.15739260.FP.PNG

- McCullers, T. J., & Steiniger, M. (2017). Transposable elements in *Drosophila*. *Mobile Genetic Elements*, 7(3), 1–18. <https://doi.org/10.1080/2159256x.2017.1318201>
- Mérel, V., Boulesteix, M., Fablet, M., & Vieira, C. (2020). Transposable elements in *Drosophila*. *Mobile DNA*, 11(1), 1–20. <https://doi.org/10.1186/s13100-020-00213-z>
- Mérel, V., Gibert, P., Buch, I., Rodriguez Rada, V., Estoup, A., Gautier, M., Fablet, M., Boulesteix, M., & Vieira, C. (2021). The Worldwide Invasion of *Drosophila suzukii* Is Accompanied by a Large Increase of Transposable Element Load and a Small Number of Putatively Adaptive Insertions. *Molecular Biology and Evolution*, 38(10), 4252–4267. <https://doi.org/10.1093/MOLBEV/MSAB155>
- Mota, N. R., Ludwig, A., Valente, V. L. D. S., & Loreto, E. L. S. (2010). harrow: new *Drosophila* hAT transposons involved in horizontal transfer. *Insect Molecular Biology*, 19(2), 217–228. <https://doi.org/10.1111/J.1365-2583.2009.00977.X>
- Mota, Nina R., Robe, L. J., Valente, V. L. S., Budnik, M., & Loreto, E. L. S. (2008). Phylogeny of the *Drosophila mesophragmatica* group (Diptera, Drosophilidae): an example of Andean evolution. *Zoological Science*, 25(5), 526–532. <https://doi.org/10.2108/ZSJ.25.526>
- Muñoz, M., & Garcia, J. (2010). DNA Transposons: Nature and Applications in Genomics. *Current Genomics*, 11(2), 115–128. <https://doi.org/10.2174/138920210790886871>
- Muszewska, A., Steczkiewicz, K., & Ginalski, K. (2013). DIRS and Ngaro Retrotransposons in Fungi. *PLOS ONE*, 8(9), e76319. <https://doi.org/10.1371/JOURNAL.PONE.0076319>
- Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H. J., Remington, K. A., Anson, E. L.,

- Bolanos, R. A., Chou, H. H., Jordan, C. M., Halpern, A. L., Lonardi, S., Beasley, E. M., Brandon, R. C., Chen, L., ... Venter, J. C. (2000). A whole-genome assembly of *Drosophila*. *Science*, *287*(5461), 2196–2204.  
[https://doi.org/10.1126/SCIENCE.287.5461.2196/SUPPL\\_FILE/1049666S3\\_THUMB.GIF](https://doi.org/10.1126/SCIENCE.287.5461.2196/SUPPL_FILE/1049666S3_THUMB.GIF)
- Nesmelova, I. V., & Hackett, P. B. (2010). DDE Transposases: Structural Similarity and Diversity. *Advanced Drug Delivery Reviews*, *62*(12), 1187.  
<https://doi.org/10.1016/J.ADDR.2010.06.006>
- Neumann, P., Novák, P., Hošťáková, N., & MacAs, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA*, *10*(1), 1–17.  
<https://doi.org/10.1186/S13100-018-0144-1/FIGURES/4>
- Nikitin, A. G., & Woodruff, R. C. (1995). Somatic movement of the mariner transposable element and lifespan of *Drosophila* species. *Mutation Research*, *338*(1–6), 43–49.  
[https://doi.org/10.1016/0921-8734\(95\)00010-4](https://doi.org/10.1016/0921-8734(95)00010-4)
- Niño, J. G., & Katzourakis, A. (2021). Phylogenomics of the Maverick Virus-Like Mobile Genetic Elements of Vertebrates. *Molecular Biology and Evolution*, *38*(5), 1731–1743.  
<https://doi.org/10.1093/MOLBEV/MSAA291>
- Nouroz, F., Noreen, S., & Heslop-Harrison, J. (2016). *Characterization and Diversity of Novel Pif/Harbinger Dna Transposons in Brassica Genomes*.
- Ohshima, K., & Okada, N. (2005). SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. *Cytogenetic and Genome Research*, *110*(1–4), 475–490.  
<https://doi.org/10.1159/000084981>

- Ou, S., & Jiang, N. (2018). LTR\_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiology*, *176*(2), 1410–1422. <https://doi.org/10.1104/PP.17.01310>
- Ou, S., & Jiang, N. (2019). LTR\_FINDER\_parallel: Parallelization of LTR\_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mobile DNA*, *10*(1), 1–3. <https://doi.org/10.1186/S13100-019-0193-0/TABLES/1>
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, *20*(1), 1–18. <https://doi.org/10.1186/S13059-019-1905-Y/FIGURES/6>
- Palacios-Gimenez, O. M., Koelman, J., Palmada-Flores, M., Bradford, T. M., Jones, K. K., Cooper, S. J. B., Kawakami, T., & Suh, A. (2020). Comparative analysis of morabine grasshopper genomes reveals highly abundant transposable elements and rapidly proliferating satellite DNA repeats. *BMC Biology*, *18*(1), 1–21. <https://doi.org/10.1186/S12915-020-00925-X/TABLES/4>
- Palazzo, A., Caizzi, R., Moschetti, R., & Marsano, R. M. (2022). What Have We Learned in 30 Years of Investigations on Bari Transposons? *Cells 2022, Vol. 11, Page 583*, *11*(3), 583. <https://doi.org/10.3390/CELLS11030583>
- Palazzo, A., Lorusso, P., Miskey, C., Walisko, O., Gerbino, A., Marobbio, C. M. T., Ivics, Z., & Marsano, R. M. (2019). Transcriptionally promiscuous “blurry” promoters in Tc1/mariner transposons allow transcription in distantly related genomes. *Mobile DNA*, *10*(1), 1–11. <https://doi.org/10.1186/S13100-019-0155-6/TABLES/2>
- Palazzo, A., Moschetti, R., Caizzi, R., & Marsano, R. M. (2014). The *Drosophila mojavensis*

- Bari3 transposon: Distribution and functional characterization. *Mobile DNA*, 5(1), 1–14.  
<https://doi.org/10.1186/1759-8753-5-21/FIGURES/6>
- Pathak, D., & Ali, S. (2012). Repetitive DNA: A Tool to Explore Animal Genomes/Transcriptomes. *Functional Genomics*. <https://doi.org/10.5772/48259>
- Perveen, F. K. (2018). Introduction to Drosophila. *Drosophila Melanogaster - Model for Recent Advances in Genetics and Therapeutics*. <https://doi.org/10.5772/67731>
- Piégu, B., Bire, S., Arensburger, P., & Bigot, Y. (2015). A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity. *Molecular Phylogenetics and Evolution*, 86, 90–109.  
<https://doi.org/10.1016/J.YMPEV.2015.03.009>
- Plasterk, R. H. A., Izsvák, Z., & Ivics, Z. (1999). Resident aliens the Tc1/mariner superfamily of transposable elements. *Trends in Genetics*, 15(8), 326–332.  
[https://doi.org/10.1016/S0168-9525\(99\)01777-1](https://doi.org/10.1016/S0168-9525(99)01777-1)
- Poulter, R. T. M., & Butler, M. I. (2015). Tyrosine Recombinase Retrotransposons and Transposons. *Microbiology Spectrum*, 3(2).  
<https://doi.org/10.1128/MICROBIOLSPEC.MDNA3-0036-2014/ASSET/2BF79017-CA76-4918-A0CE-FEEC07AB0B40/ASSETS/GRAPHIC/MDNA3-0036-2014-FIG4.GIF>
- Pray, L. (2008). Transposons: The Jumping Genes. *Nature Education*, 1(1), 204.
- Pridie, C., Ueda, K., & Simmonds, A. J. (2020). Rosy Beginnings: Studying Peroxisomes in Drosophila. *Frontiers in Cell and Developmental Biology*, 8.  
<https://doi.org/10.3389/FCELL.2020.00835>
- Pritham, E. J., Putliwala, T., & Feschotte, C. (2007). Mavericks, a novel class of giant

- transposable elements widespread in eukaryotes and related to DNA viruses. *Gene*, 390(1–2), 3–17. <https://doi.org/10.1016/J.GENE.2006.08.008>
- Qi, X., & Sandmeyer, S. (2012). Nonhomologous Recombination: Retrotransposons. *Encyclopedia of Biological Chemistry: Second Edition*, 4, 283–291. <https://doi.org/10.1016/B978-0-12-378630-2.00508-9>
- Ramos, E., & Rafael, V. (2017). Cinco especies nuevas del género *Drosophila* (Diptera, Drosophilidae) en la provincia de Napo, Ecuador. *Iheringia. Série Zoologia*, 107, 1–12. <https://doi.org/10.1590/1678-4766E2017022>
- Ravasi, T., Huber, T., Zavolan, M., Forrest, A., Gaasterland, T., Grimmond, S., Arakawa, T., Carninci, P., Kawai, J., & Hayashizaki, Y. (2003). Systematic Characterization of the Zinc-Finger-Containing Proteins in the Mouse Transcriptome. *Genome Research*, 13(6b), 1430. <https://doi.org/10.1101/GR.949803>
- Rech, G. E., Radio, S., Guirao-Rico, S., Aguilera, L., Horvath, V., Green, L., Lindstadt, H., Jamilloux, V., Quesneville, H., & González, J. (2022). Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nature Communications* 2022 13:1, 13(1), 1–16. <https://doi.org/10.1038/s41467-022-29518-8>
- Rice, P., Longden, L., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics : TIG*, 16(6), 276–277. [https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2)
- Richardson, S. R., Doucet, A. J., Kopera, H. C., Moldovan, J. B., Garcia-Pérez, J. L., & Moran, J. V. (2015). The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiology Spectrum*, 3(2). <https://doi.org/10.1128/MICROBIOLSPEC.MDNA3-0061-2014>

- Romero, D. A. (2018). *Análisis morfométrico del edeago y alas en poblaciones naturales de Drosophila amaguana y D. neoamaguana provenientes de Río Guango y Laguna de Papallacta provincia de Napo, Ecuador* [Pontificia Universidad Católica del Ecuador].  
<http://repositorio.puce.edu.ec/handle/22000/14690>
- Rossato, D. O., Ludwig, A., Deprá, M., Loreto, E. L. S., Ruiz, A., & Valente, V. L. S. (2014). BuT2 Is a Member of the Third Major Group of hAT Transposons and Is Involved in Horizontal Transfer Events in the Genus *Drosophila*. *Genome Biology and Evolution*, 6(2), 352–365. <https://doi.org/10.1093/GBE/EVU017>
- Rubin, E., Lithwick, G., & Levy, A. A. (2001). Structure and Evolution of the hAT Transposon Superfamily. *Genetics*, 158(3), 949–957.  
<https://doi.org/10.1093/GENETICS/158.3.949>
- Saleh, A., Macia, A., & Muotri, A. R. (2019). Transposable elements, inflammation, and neurological disease. *Frontiers in Neurology*, 10(AUG), 894.  
<https://doi.org/10.3389/FNEUR.2019.00894/XML/NLM>
- Sassi, A. K., Herédia, F., Loreto, É. L. da S., Valente, V. L. da S., & Rohde, C. (2005). Transposable elements P and gypsy in natural populations of *Drosophila willistoni*. *Genetics and Molecular Biology*, 28(4), 734–739. <https://doi.org/10.1590/S1415-47572005000500013>
- Scali, C., Nolan, T., Sharakhov, I., Sharakhova, M., Crisanti, A., & Catteruccia, F. (2007). Post-integration behavior of a Minos transposon in the malaria mosquito *Anopheles stephensi*. *Molecular Genetics and Genomics*, 278(5), 575–584.  
<https://doi.org/10.1007/S00438-007-0274-5/METRICS>
- 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).

<https://doi.org/10.1098/RSBL.2016.0407>

Sha, M., Lee, X., Li, X. ping, Veldman, G. M., Finnerty, H., Racie, L., LaVallie, E., Tang, X. Y., Edouard, P., Howes, S., Keith, J. C., & McCoy, J. M. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*, *403*(6771), 785–789. <https://doi.org/10.1038/35001608>

Shen, D., Song, C., Miskey, C., Chan, S., Guan, Z., Sang, Y., Wang, Y., Chen, C., Wang, X., Müller, F., Ivics, Z., & Gao, B. (2021). A native, highly active Tc1/mariner transposon from zebrafish (ZB) offers an efficient genetic manipulation tool for vertebrates. *Nucleic Acids Research*, *49*(4), 2126–2140. <https://doi.org/10.1093/NAR/GKAB045>

Shi, J., & Liang, C. (2019). Generic Repeat Finder: A High-Sensitivity Tool for Genome-Wide De Novo Repeat Detection. *Plant Physiology*, *180*(4), 1803. <https://doi.org/10.1104/PP.19.00386>

Silva, J. C., & Kidwell, M. G. (2004). Evolution of P Elements in Natural Populations of *Drosophila willistoni* and *D. sturtevantii*. *Genetics*, *168*(3), 1323. <https://doi.org/10.1534/GENETICS.103.025775>

Silva, J. C., & Margaret, G. K. (2000). Horizontal Transfer and Selection in the Evolution of P Elements. *Molecular Biology and Evolution*, *17*(10), 1542–1557. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026253>

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. <https://doi.org/10.1093/BIOINFORMATICS/BTV351>

Simmons, M. J. (2013). P Elements. *Brenner's Encyclopedia of Genetics: Second Edition*,

258–260. <https://doi.org/10.1016/B978-0-12-374984-0.01112-8>

Sinzelle, L., Kapitonov, V. V., Grzela, D. P., Jursch, T., Jurka, J., Izsvák, Z., & Ivics, Z. (2008). Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4715–4720. [https://doi.org/10.1073/PNAS.0707746105/SUPPL\\_FILE/07746FIG15.PDF](https://doi.org/10.1073/PNAS.0707746105/SUPPL_FILE/07746FIG15.PDF)

Sonnhammer, E. L. L., & Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, *167*(1–2), GC1. [https://doi.org/10.1016/0378-1119\(95\)00714-8](https://doi.org/10.1016/0378-1119(95)00714-8)

Spradling, A. C., & Rubin, G. M. (1981). Drosophila genome organization: conserved and dynamic aspects. *Annual Review of Genetics*, *15*, 219–264. <https://doi.org/10.1146/ANNUREV.GE.15.120181.001251>

Su, W., Gu, X., & Peterson, T. (2019). TIR-Learner, a New Ensemble Method for TIR Transposable Element Annotation, Provides Evidence for Abundant New Transposable Elements in the Maize Genome. *Molecular Plant*, *12*(3), 447–460. <https://doi.org/10.1016/J.MOLP.2019.02.008>

Tan, S., Ma, H., Wang, J., Wang, M., Wang, M., Yin, H., Zhang, Y., Zhang, X., Shen, J., Wang, D., Banes, G. L., Zhang, Z., Wu, J., Huang, X., Chen, H., Ge, S., Chen, C. L., & Zhang, Y. E. (2021). DNA transposons mediate duplications via transposition-independent and -dependent mechanisms in metazoans. *Nature Communications* *2021* *12:1*, *12*(1), 1–14. <https://doi.org/10.1038/s41467-021-24585-9>

Tang, Z., Zhang, H. H., Huang, K., Zhang, X. G., Han, M. J., & Zhang, Z. (2015). Repeated horizontal transfers of four DNA transposons in invertebrates and bats. *Mobile DNA*, *6*(1), 1–10. <https://doi.org/10.1186/S13100-014-0033-1/FIGURES/5>

- Thomas, J., & Pritham, E. J. (2015). Helitrons , the Eukaryotic Rolling-circle Transposable Elements . *Microbiology Spectrum*, 3(4).  
<https://doi.org/10.1128/MICROBIOLSPEC.MDNA3-0049-2014/ASSET/95A8D437-BC8F-4B31-92C7-3631304977F3/ASSETS/GRAPHIC/MDNA3-0049-2014-FIG13.GIF>
- Thomas, M. C., Macias, F., Alonso, C., & López, M. C. (2010). The biology and evolution of transposable elements in parasites. *Trends in Parasitology*, 26(7), 350–362.  
<https://doi.org/10.1016/J.PT.2010.04.001>
- Tiedeman, Z., & Signor, S. (2021). The Transposable Elements of the *Drosophila serrata* Reference Panel. *Genome Biology and Evolution*, 13(9), 1–12.  
<https://doi.org/10.1093/gbe/evab100>
- Tolwinski, N. S. (2017). Introduction: *Drosophila*—A Model System for Developmental Biology. *Journal of Developmental Biology*, 5(3). <https://doi.org/10.3390/JDB5030009>
- Vela, D., & Rafael, V. (2000). *Drosophila* distribution in Ecuador. *Drosophila Information Service*, 83, 85–88.
- Vela, D., & Rafael, V. (2003). Estudios sobre la diversidad del genero *Drosophila* (Diptera, Drosophilidae) en el Bosque Pasochoa de la provincia de Pichincha - Ecuador. *Rev. Pontif. Univ. Catol. Ecuador*, 71, 117–127.
- Vela, D., & Rafael, V. (2004). Three new andean species of *Drosophila* (Diptera, Drosophilidae) of the mesophragmatica group. *Iheringia. Série Zoologia*, 94(3), 295–299. <https://doi.org/10.1590/S0073-47212004000300012>
- Venkatesh, & Nandini, B. (2020). Miniature inverted-repeat transposable elements (MITEs), derived insertional polymorphism as a tool of marker systems for molecular plant

breeding. *Molecular Biology Reports* 2020 47:4, 47(4), 3155–3167.

<https://doi.org/10.1007/S11033-020-05365-Y>

- Vieira, C., Fablet, M., Lerat, E., Boulesteix, M., Rebollo, R., Burlet, N., Akkouche, A., Hubert, B., Mortada, H., & Biéumont, C. (2012). A comparative analysis of the amounts and dynamics of transposable elements in natural populations of *Drosophila melanogaster* and *Drosophila simulans*. *Journal of Environmental Radioactivity*, 113, 83–86. <https://doi.org/10.1016/J.JENVRAD.2012.04.001>
- Wang, W., & Kirkness, E. F. (2005). Short interspersed elements (SINEs) are a major source of canine genomic diversity. *Genome Research*, 15(12), 1798. <https://doi.org/10.1101/GR.3765505>
- Wicker, T., Guyot, R., Yahiaoui, N., & Keller, B. (2003). CACTA Transposons in Triticeae. A Diverse Family of High-Copy Repetitive Elements. *Plant Physiology*, 132(1), 52. <https://doi.org/10.1104/PP.102.015743>
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973–982. <https://doi.org/10.1038/nrg2165>
- Woo, A. C., Gaia, M., Guglielmini, J., Cunha, V. Da, & Forterre, P. (2019). Evolution of the PRD1-adenovirus lineage: a viral tree of life incongruent with the cellular universal tree of life. *BioRxiv*, 741942. <https://doi.org/10.1101/741942>
- Xiong, W., He, L., Lai, J., Dooner, H. K., & Du, C. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(28), 10263–10268. [https://doi.org/10.1073/PNAS.1410068111/SUPPL\\_FILE/PNAS.1410068111.SAPP.PD](https://doi.org/10.1073/PNAS.1410068111/SUPPL_FILE/PNAS.1410068111.SAPP.PD)

## F

- Xu, Z., & Wang, H. (2007). LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, *35*(Web Server issue).  
<https://doi.org/10.1093/NAR/GKM286>
- Yamaguchi, M., & Yoshida, H. (2018). *Drosophila* as a model organism. *Advances in Experimental Medicine and Biology*, *1076*, 1–10. [https://doi.org/10.1007/978-981-13-0529-0\\_1/TABLES/1](https://doi.org/10.1007/978-981-13-0529-0_1/TABLES/1)
- Yan, Y., Buckler-White, A., Wollenberg, K., & Kozak, C. A. (2009). Origin, antiviral function and evidence for positive selection of the gammaretrovirus restriction gene Fv1 in the genus *Mus*. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3259–3263. <https://doi.org/10.1073/PNAS.0900181106>
- Yuan, Y. W., & Wessler, S. R. (2011). The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7884–7889.  
[https://doi.org/10.1073/PNAS.1104208108/SUPPL\\_FILE/SD01.TXT](https://doi.org/10.1073/PNAS.1104208108/SUPPL_FILE/SD01.TXT)
- Zattera, M. L., Gazolla, C. B., Soares, A. de A., Gazoni, T., Pollet, N., Recco-Pimentel, S. M., & Bruschi, D. P. (2020). Evolutionary Dynamics of the Repetitive DNA in the Karyotypes of *Pipa carvalhoi* and *Xenopus tropicalis* (Anura, Pipidae). *Frontiers in Genetics*, *11*, 637. <https://doi.org/10.3389/FGENE.2020.00637/BIBTEX>
- Zhang, L., Yan, L., Jiang, J., Wang, Y., Jiang, Y., Yan, T., & Cao, Y. (2014). The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast *Candida albicans*. *Virulence*, *5*(6), 655. <https://doi.org/10.4161/VIRU.32180>
- Zhang, R. G., Li, G. Y., Wang, X. L., Dainat, J., Wang, Z. X., Ou, S., & Ma, Y. (2022).

TEsorter: An accurate and fast method to classify LTR-retrotransposons in plant genomes. *Horticulture Research*, 9. <https://doi.org/10.1093/HR/UHAC017>

Zhao, S., Jiang, E., Chen, S., Gu, Y., Shangguan, A. J., Lv, T., Luo, L., & Yu, Z. (2016).

PiggyBac transposon vectors: the tools of the human gene encoding. *Translational Lung Cancer Research*, 5(1), 120. <https://doi.org/10.3978/J.ISSN.2218-6751.2016.01.05>

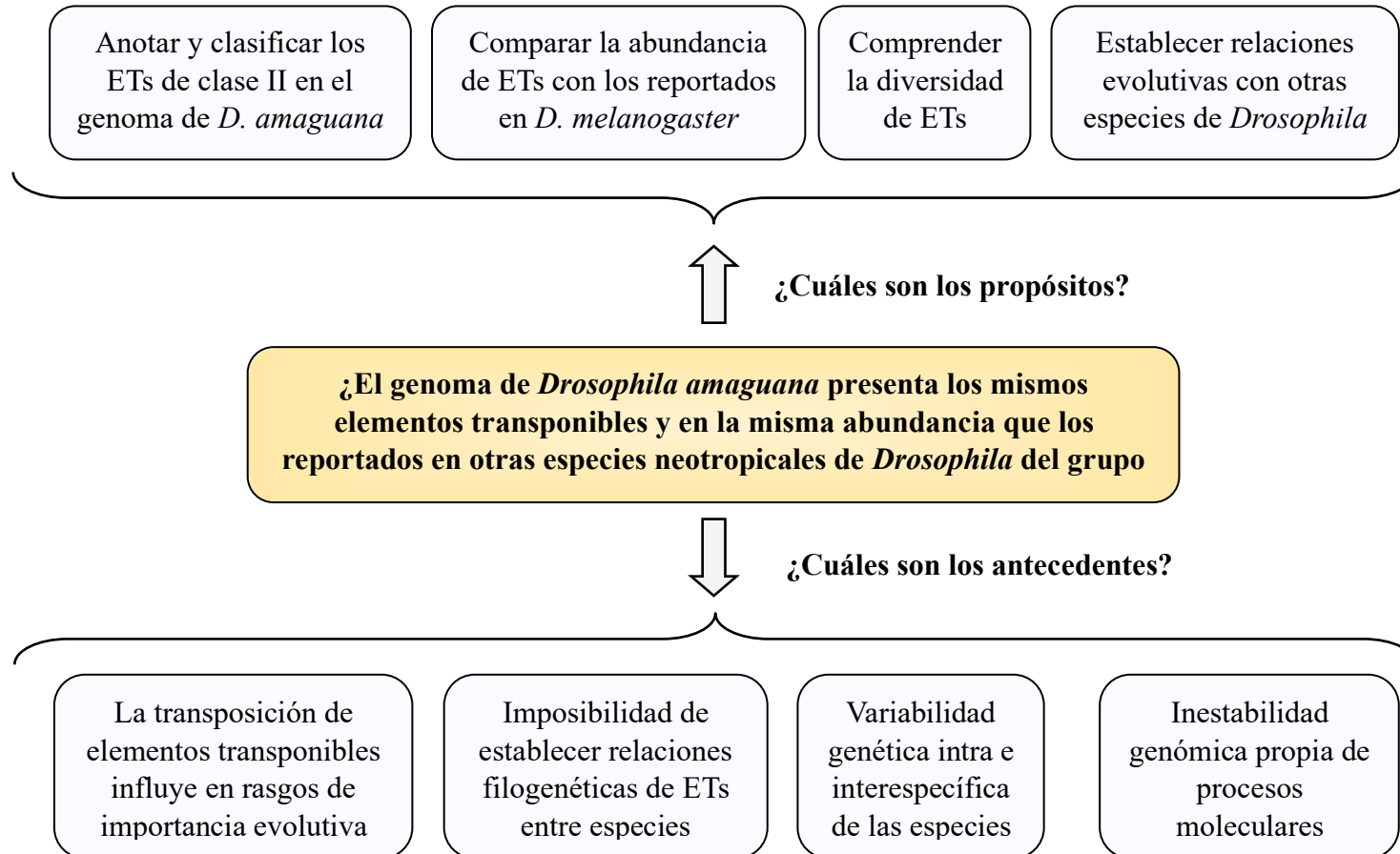
Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The

MaSuRCA genome assembler. *Bioinformatics*, 29(21), 2669–2677.

<https://doi.org/10.1093/BIOINFORMATICS/BTT476>

## Anexos

### Anexo A. Árbol de Problemas



Anexo B. Anotación del elemento P en *D. amaguana* en formato GenBank.

```

BASE COUNT      956 a    475 c    554 g    787 t
ORIGIN
source          1..2772
/organism="Drosophila amaguana"
/country="Ecuador"
/mol_type="genomic DNA"
repeat_region  1..32
/note="TIR"
/colour=3
mobile_element 1..2772
/note="P element"
gene           131..2573
/product="Transposase"
/colour=12
/note="Transposase of P element"
CDS           131..445
/note="Truncated transposase"
/colour=5
mat_peptide   202..378
/note="Zinc finger domain"
/colour=4
CDS           490..1203
/note="Truncated transposase"
/colour=5
mat_peptide   621..1169
/note="Transposase protein found in insects"
/colour=4
CDS           1218..1967
/note="Transposase"
/colour=5
CDS           1998..2573
/note="Transposase"
/colour=5
mat_peptide   2025..2267
/note="87kDa Transposase (catalyzes the excision of a
nonautonomous P transposable element)"
/colour=4
repeat_region 2741..2772
/note="TIR"
/colour=3
1 catgatgaaa taacataagg tggctctgtc ggaagccgaa gcttaccgaa gtatacactt
61 aacttcagtg cacgtttgct tgttgagagg aaagcttggtg tgcggacgaa tttttttttt
121 aaaacattaa cccttacgtg gactaaaaaa aatgaaatat tgcaaatgtg gctgcaaaagc
181 tgtgactgga gtgaaattaa ttcacgtgcc gaagtgtgctt attaagagaa aattgtggga
241 gcagagcctg ggggtgcagcc ttggtagaaa ctcccaaatt tgtgatacc accttaatga
301 ttcgcagtgg aagtcctcac ctgcaaaagg tcagacattt aaaaggaggc gacttaacgc
361 agatgccgta cctagtaaag tgatagagcc tgaaccagaa aagataaaa aaggctatgc
421 cagtgaggag acacaaaacag agtaagtttg aatagtaaaa aaaaatcatt tatgtaaaca
481 ataaagtgc tgtgcgttag gtcctgttca ttgtttaatg aaaataagag cttgagggaa
541 aaaattcgtc ctttggagta cgaatgcgt cgttagagc agcagctaa ggagctcaa
601 aagttggagg agtctctac caacatcttc acggaaacgc agatacggat actgaaagat
661 ggtggacaaa gagcgcagtt caattccgaa gacatttcta cggctatttg tctccacacc
721 gcaggccctc gagcgtataa ccatctgtac aaaaaaggat ttccttacc cagtctgagc
781 actttgtaca gatggttatc agatgtggac ataaaaacag gatgtttgga tgttgcata
841 gacctaatgg acagtgatgg agttgatgac gccgacaagc tttgcgtact gcctttcgac
901 gagatgaagg tcgctgctgc cttcgagtat gacagctctg ctgatattgt ttacgagcca
961 agcgactatg tccaactggc tattgttcgt ggtctaaaaa aatcgtggaa gcagccagtt
1021 tttttcgatt ttaataccc aatggaccgc gatactcta acaatatatt aagcaaaactg
1081 cataggaaa gatatttagt agttgctatt gtatccgatt taggtaccgg aaaccaaag
1141 ctatggacag agctcggtat atcagaatgt aagtttcgta tatgccttaa aaaaattaga
1201 taatccttta aatttaattt ttagcaaaaa cctggtttag ccatcctgca gatgaccact
1261 taaagatttt cgttttttcg gatacgtcac attttaattaa gttagtcctg aaccactatg
1321 tggattccgg attaacaata aatggaaaaa aattaacaaa aaaaaaatt caggaggcac
1381 tcatctttg taacaagtc gatctgtcta tctctttaa aattaatgaa aatcacatta
1441 atgttcgatc gctcgcaaaa cagaaggtta aattggctac ccaactggtt tcgaatacca
1501 ctgctagctc catcagagc tgctattcat tggggtatga cattgaaaat gccaccgaaa
1561 ctgaggactt cttcaaatg atgaatgatt ggttcgacat ttttaattct aaattgtcca
1621 catccaattg cattgagtc tcgcaacctt atggcaagca gttggatata cagaatgaca
1681 ttttgaacag aatgtcggaa attatgcaaa caggaatttt ggataaaccc aaaaggctcc
1741 catttcaaaa agggatcatt gtgaataatg ctctccctga tggcttgtat aatatttgc
1801 aagaaaactt cagtatgcaa tacatattaa caagcctct caaccaagac attgtggagc
1861 atttttttgg cagcatcgca tcgagagggtg gacaattcga ccatccgact ccaactgagc
1921 tgaagtatag gttaaagaaa tatataatag gtatgacaaa aatgtaattt catgatttat
1981 aattgtttaa tgtttagcta tatgtttcag gaaagtcca gttgaaatc gaagaggaaa
2041 actctgaaaag ctggcttaat ttagacttaa gttctaaaaa aaaggaaaat aagactaaag
2101 atgatgagcc tgtcgtgat gagcctgtcg atgatgagct gagcataata agcaatatag
2161 atgtcaccca aatggatgag ttaacggagg atgcatgga atatatcgc ggctatgta
2221 ttaaaaaatt gagaatcagt gacaaaagtaa aagaaaattt gacattaca tacgtcgac
2281 aggtgtctca cggcggactt attaaaccgt ccgaaaaatt tcaagagaag ctgaaaagagc
2341 tagaatgat tttttcgag tatacaata ataataatt tgaattaca aataatgtaa
2401 aggaaaaatt aatattagca gcccaaacg tcgatgtcga taaactagta aaatctttt
2461 attttaaaat tagaataata tttagaatta agtacttaaa caaaaaatt gcaattaaaa
2521 atcaaaaaa aaagttaatt ggaaactcca aattattaaa aataaaactt taaaaataat
2581 ttcgtctaata taatattatg agttaaattca aaccccagc acatgctaag ggtaataca
2641 caatcatatc gctttctcac tcagactcaa tatgacactc agtagtctac ttcctttcac
2701 tcgcacttat tgcaagcata cgtaaagtgt atgtctcttg ccgacgggac caccttatgt
2761 tatttcatca tg

```

//

Anexo C. Anotación del transposón *Mutator* en *D. amaguana* en formato GenBank.

```

source          1..4314
                /organism="Drosophila amaguana"
                /country="Ecuador"
                /mol_type="genomic DNA"
repeat_region   1..422
                /note="TIR"
                /colour=3
mobile_element   1..4314
                /note="Mutator"
repeat_region   3893..4314
                /note="TIR"
                /colour=3

BASE COUNT      1395 a    745 c    724 g    1450 t
ORIGIN
1  gtaatttgagt tcaagaatcc agttgcta atgattttat gctagaatgc ctcgtggatc
61 acatttgctg gacaaagaac aaggctgat cacagggttg gctgaagggt gaaagaatat
121 tccatataatt gctgctaaag ttgataggca tagaaacaca gtggccaatt tcctcagaaa
181 tccagaaaaa tacagaaaaa ataagccggc cggccgacct accgacattg ataatcgctg
241 taagaggcaa attcggcgaa tgacggtgga agacagcatg tcatgtaccg aaatccaacg
301 tcagctacag ctcaatgtag acgggtgaac cagattttac aggagaacga cacggttcaa
361 tatgcttccc gcgaggctgt acccaggcta ttaaaacgcc acataaacgc cgcctagca
421 ttatctgaac tttttgttt tcgaaaattt cgttgtataa acatagtatc aacatagttt
481 gcttacaact atggttctat tgttttctga ctctcacatt ctttccatct ttcattgtgt
541 tctggtcatt tacaacaaa taataccgat atttgtttgc taaccgggt tcaatttaaa
601 aagtacaaaa cacaaatcgt agtaagaatt gtattatgtt ttgcaatggt gataatgtaa
661 aagacaatgt caaaaacata atactagtat gtaattttgc ttacgatatc taaaaaac
721 aaaaagtttt caatacgaag tgcagctata tgtgtactat ttatatatca ggaactggca
781 atactaatgt caatagctaa aatctggaaa caataataaa gttttgtatt tttaagtata
841 aaagtccaac gttgccattg gctttgttta cagtgaagta agaaaaatcaa ttggcccaa
901 aactgaaaat atcaaaaaaa aaaaaaaaaa taaaaaacct ttcggtggat aatttttcac
961 agtttttgac gtgaattatc tgaaccagaa tgtgtagatg aacttagttt tttagcgtg
1021 aattcaatcg tggctgtagt tcgggcgcca acgaattccg tgaaggctgt ctaaatcggt
1081 tgtgtgccca caaacatcg gtgcagtgcg tgaactgata atgtaagatc gtcagtgaac
1141 aaaagtttga tgtctttggt gtaaagaaat gctttaaaac ttcgaccgtg gtgcttcaa
1201 agacgtgtat taatcagct gatagctgac gaaacatgga tatataaata cgagccggaa
1261 accaaaccgt gtggatgccc taagaccggc caaatctaac aaaaattagt ctccccatg
1321 actgatgatt cattctgcat atagttacaa ttggcacct gatgacatca tcttgttcc
1381 gaatatcaaa ggtaaactgc ctggtcaaca atttacgagc gccgaaaaag catttgaagc
1441 gttctacaac catgtttgtg aggtttatca agaggaatgg aaaaactgct atgacaaaatg
1501 gattaaacgt attgaaaagt gtattgataa ccaagggtg gattgataa acaatggggg
1561 cattcagcaa gtaattttga ctaattgatt acaacatttg tcattcaaca tgaattttgt
1621 aacactcaaa atgtcaaaat ttggtatcgg tacaacaaaa aaataaattg aagtgtgata

```

```

1681 cgtaatacca atttatttat atgaagaaag tttgaaatac ttaggtggct actttaataa
1741 aaagctataa aggctgatct ttataaaaata gtataatgct attcagcagg tactgattac
1801 tgatacatca caaggaaata agggattact gataaagcga ttttaatttt aaaatgtttg
1861 taaatatttt ctattctcga aacataaata gtactttagc taccaatcga aagaaaaata
1921 aaagtatgaa agacttatgt ttagttaa ttagctta ttaatttaatt tattaatttt
1981 tattgaacac ttaagcataa tttcttgttt gtttttaaat cttaacata ttaaatgtat
2041 atttaacaa tcaaaaaact aagtaggta actttcaaaa caaatagtc agtaatgatc
2101 aagtaaacac agttttaacg ctgataaagc ctttaacaca agaaatcga aaatgactt
2161 tgttcatctg tttgttctt ggcttatgca atgaatcata gctcgattc gcaatctatt
2221 tttattgctg tttgaaatg ctccttttc ctcgattata gcaaacatat tctctattg
2281 cgtgctatcc ctttctttt gctcagcgtt gcaactctg cgccttcgc ttgttcctg
2341 catatatttt gtaccgcaa gtaaaaaata gcatatgtag ccttaacctt atttttgtt
2401 ctgattcatc gttgttctg ttgtttttg ttggtcagtc tacatcgcca caaatgactt
2461 tgtcgtaca attacaaagt catctacata gttcctctg aagtaattac cctgtttta
2521 acaacctctg gaagatgtt ttcctaaaa tagccacca agactttta ttgtaaaaca
2581 tatattttta tctgaaaaa atttaaaaa tcgaatgta ataaatttag ttgttttct
2641 agctactctt ttattattg caagtcttat attaaaaaa ttttaacc tgtaacctg
2701 aaatgtaaat atacatcac taaattgca gaattaaatt tgggtgctta aataaaattg
2761 caaacagact ttacttacac agatttatat aataaatgct gctaatggct gcataaacct
2821 aatcttttga aatttccaa tgttggattt tataatgttt ttatacaatt tcaatattaa
2881 atttatatta tattttatta gtacaacttg gcgctgaag tgatcga aaaataatg
2941 aagaataagg aagttaaac atgcgacaca aaaccgttag gctcgtcatt tcaatgctt
3001 gtattaagat atctatctat gcgtgtattg tacctcagtg aaataaatgc catttttga
3061 agaaaatcat tgaaaaacta tattacaatt tttgttttt tcaaatatt tagtatacat
3121 gtactgctt agaaaataa ttttttttt tattatctc aagaaacta aaatattaa
3181 ttaacaaagt ttatttcata tgaaaaaagt atatgtgcaa aatgataaat gtgaagaaa
3241 aagaagtagt ttgttctt atccaataat attgttatt taatagttaa tatgtttcc
3301 cttattaaca agcatttgac tagttcgact tgccattgaa tttgacaagt tttgagagt
3361 tgaaaagtc atttcatcc atctttgat tagtgcctt ttaaggtcct ttacagtatc
3421 aaactggcga ttattttgt agacttgcg tgacaatatt ccccaagat ctcaatcgg
3481 gcttaggtcc ggactgatc caggccattc caaaataggt atttctctat ccgcaacaa
3541 atttttggt ttccgagcag tatgaattg cgcaattgct tgtccataag tccatgtatc
3601 tccgtatagg tctccagca actcaataag tattgtgtt aacaattcta tataaattc
3661 tgcatttgt ctggtgggta taaaacagag ttagacttt cctgcatc caaatgcggc
3721 ccaaatcata aaggatcct caccgtggt gcgctatga catgaattac gtggttgcct
3781 tacatctcgc caaactttt ggaacagtc tggccatct aagttaaact tctttttgtc
3841 ggaaaaaata acatttttg actcttctga ccaaatgta ttttttccc caaatgctag
3901 gcgggcttt atgtggcgt ttaatagcct gggtacagc tcgagggaag catattgaac
3961 cgtgtcttc tcctgtaaaa tctggttcac ccgtctacat tagctgttag ctgacttgg
4021 atttcggtac atgacatgct gcttccacc gccattcgc gaatttgcct cttacagca
4081 ttatcaatgt cgtaggtct gccggcggc ttattttcc tgaattttc tggatttctg
4141 aggaattttg ccactgtgt tctatgcta tcaactttag cagcaataa tgaatattc
4201 tttccacct cagccaactc tgtgatcag ccttgttct tgtcgttag atgacttca
4261 cgaggcttc taacataaaa tcacattagc aactgtattc ttgaactcaa atac

```

//

## Anexo D. Anotación del transposón *CACTA* en *D. amaguana* en formato GenBank.

```

source          1..4940
                /organism="Drosophila amaguana"
                /country="Ecuador"
                /mol_type="genomic DNA"
repeat_region  1..47
                /note="TIR"
                /colour=3
mobile_element 1..4940
                /note="CACTA superfamily"
repeat_region  join(48..303,447..1018)
                /note="STR"
                /colour=8
gene           2910..4049
                /product="Peroxisomal (S)-2-hydroxy-acid oxidase
                GLO3-like"
                /colour=12
                /note="Peroxisomal (S)-2-hydroxy-acid oxidase GLO3-like"
CDS           2910..3284
                /note="Peroxisomal (S)-2-hydroxy-acid oxidase GLO3-like"
                /colour=5
mat_peptide   3036..3281
                /note="TIM-like beta/alpha barrel domains"
                /colour=4
CDS           3312..4049
                /note="Peroxisomal (S)-2-hydroxy-acid oxidase GLO3-like"
                /colour=5
mat_peptide   3351..4013
                /note="FMN-dependent dehydrogenase"
                /colour=4
repeat_region join(4141..4612,4629..4884)
                /note="STR"
                /colour=8
repeat_region 4894..4940
                /note="TIR"
                /colour=3

BASE COUNT    1588 a    860 c    807 g    1685 t
ORIGIN
1  cactaataca ctgaacaaa ctgttgcctt ctttgtgaat ctctcgattt cacgaagttc
61  aaagcagctc aacagctgct tccatatac gagctcacac gcatagcctt agctgagcgt
121  aaaactllga ctgtgttggg gaactctatg aactcaccia tacagt1aaa gacaaaagacg
181  ctctctttac agaaataggc aaagagcata acatgtgagt tagtattgga aattgtatga
241  cacttactaa cactaacaaa ccaacagctg ctctctttcc atttcaattc tctgtttatg
301  ttttgcctca gcttcttgaa ctccaatcaa tgttttttgc atcagctcct gttcgtgggc
361  taattgaaca ggttttctgag aattggccea aataacgcca ggggccattg agattttggc
421  agctgagcgt gtctcagtga aatttgtgaa attaataaaa tttatatggt gaattcttga
481  agaataaac gaataatata tttatacatt tatttaacaa attaaaacta atallaaata
541  tcaacttaat acaaaat1ta aagagtattg tttcctt1ga ctattacgaa tattacatat
601  atataataat gtgaagatgt atgtaaatat tgtaataata atataatgtaa tgcacatatat
661  ttatataatg ttttaaatg tgaaatgccc tttctgaatt tgtagatttt gttcattttt
721  aaaagctaaa aataatgata tcataggaac aatacacaat gaatgaggtt ttatatatg
781  ataaaattta atatttcttt gattt1t1ta atactaaagt gttat1aata ttatacatat
841  acattttgaa ggctttacct attcgtt1ta agtgaatatt atatatatat catatagata
901  ctatatattt tgaataaaat agtattattt cataactcaat tgtatttaat atttataattt
961  attttcagca aatatgtttt agctataaaa taattaaata tttcatttca aatagt1ttt
1021  ttattatctg aaccatttga aaatgggatt ttgggt1aaa atttt1t1t1gga aatgtat1t1ga
1081  acaggcagaa ggaggcgtgg cagacctcat aagctc1ataa actaca1aaa tggagagctg
1141  aaattgtaaa catatattcc ataatacctt acgcagttca agttiatctt tttttctc1g
1201  ataactt1ac taattttctt ttatcgataa ctatcga1aat ttgga1ttt1g ttttgacag
1261  ataactt1gaa aaccaagagc gctaaacaca ccaattt1gg caggtatatt cccaat1tga
1321  gtacacaaaa tatgacatct ttcgattttg gatacctaac ctcaactccc ccaaaaaataa

```

Anexo E. Anotación del transposón *hAT* en *D. amaguana* en formato GenBank.

```

BASE COUNT      808 a   416 c   476 g   685 t
ORIGIN
    1 tagagggtgt aaacaatcga tggcgctatc gacactatcg atggggccgc actcccgata
    61 gtggtcgata gtgccataga taagttgtgt ttacacagca ctgaaattgc gcctttttca
source          1..2385
                /organism="Drosophila amaguana"
                /country="Ecuador"
                /mol_type="genomic DNA"
    121 tttgtgacgc cgcttgatgt gttgatttgc ttaacaaatt atattcccac tattttaaac
    181 aattaaatgt aaaaaaatgg atcgctttat tcacaaaggt aagtaacata acacacagtt
    241 gaagcagttt taatgcagtt aattaaattt tggttttttt tttaaaagga caacgtcgaa
    301 actcgaagga aatatgcgat tgcaatcctt cagagtctga aaaaaatttt gcgtcgaaaa
repeat_region   1..21
                /note="TIR"
                /colour=3
    361 agccaaacc  aaaattgaaa aagaaaagta aaatttcgga tgtttggaat tatttcaaag
    421 gatcggacga caaaaaattc gctaagtgtt tgaattgcgg aaaggaatat aaaaactagc
    481 ggaatacttc gaatttgcga gatcacctca aaaggtggtt cgcaaattta aaggatgatg
    541 gaactgacgc cattattgat gcaggtgctt ctcttggggg cgacaacata cctctacaa
mobile_element  1..2385
                /note="hAT"
gene           267..1941
                /product="ZBED domain-containing protein 4-like"
                /colour=12
                /note="ZBED domain-containing protein 4-like"
    601 gcagcagtg  tcggtcaagc atgcgatcaa tggattccta tattaanaag gctgtactat
    661 atgattccaa ctcagaaagg aaaaaatgta ttgataaatc tttagcggag atgattgcta
    721 tagatgtcca accctataga gaatgcgaga ttatacaat  actgcaaaat tttggattgc
    781 agatacacgc tgccaagtcg cgggcattta aaatatgata tggtaagca  attatttaag
    841 gaaacttcag caaaactttt agtaaatgtt gaaagtgttt cgtgtattgc gattacgtgt
    901 gacctatgga cgtcgcgtgc aaatacaagt ttctaatag  tgactggatc ctttgtcac
CDS           267..839
                /note="ZBED domain-containing protein 4-like isoform X1"
                /colour=5
    961 gacaatagcc tgaaaacagc atccttacca acgagaaagt tattggatgt aactaccat
    1021 tcggctcaaa acgttgcaaa cactttacac gatgtactaa ctttttgtaa tttgtggaa
    1081 aaaaacagtct gtgtagtcc  ggataatgct agttcgatgg tcaaaagctt gcaattgtgt
mat_peptide    396..518
                /note="BED zinc finger"
                /colour=4
    1141 aagatacga  acattcctgt cttgctcat  acactaaatt tagtagatgg actgcaattt
    1201 gaagatgact caactatcgc ttttaattcag aagtgctaga cgatcgtaa  gttttttta
    1261 aaaaaagttc gattgccaac gagaagttaa aattgtcca  agaaaacttt gcatattcgt
    1321 tgctgcagga aacaccaacc cgatggaata gtttttatta tatgattgaa tattgaaat
CDS           724..1239
                /note="ZBED domain-containing protein 4-like"
                /colour=5
    1381 tgacaacaca tgaggaatc  gcagcagtat tattggccac agtaaatgca ccgcaatcgt
    1441 ttaatgcagt ggaaataaat gtcttaaaga tatggaaaaa atacttgcgt tatttttgaa
    1501 agcctccgag aaagtttccg gagggaaata cgttacgatt tctctaataa ctccattaac
    1561 gtatggactt tttaaaaaat taacaatatt tcttccagat tgaaaaactc aagtggggaa
    1621 aacatgggta ctgttttgct ggaggcggct gcaaagcgtt tggcgccata gaaaaaacga
    1681 acagttacga gaatggcgac cattttggac ccacgattca aaaaaattgg gttccaacaa
    1741 aatacaaatg cggacaagc  tgcaatttgc tttgaagacg agttggtgac tataagtaat
    1801 aaggaaaatc gagatagctc taatgaggga gtcaccaaaa ttgatgggga ctcatcttc
    1861 aactttttgg accatcaagc cagctgctcg catgggaacg cgctatcgga cgcaattatt
    1921 gcgaaaagga gagggcaata gtacctcaat atatgaatcc tttgttatgg atacaagtaa
    1981 ataaaaacat ttaactatcg atattcaaaa atttatttat ttcaacatta atttttatt
    2041 caggttaacc aggcggttt  tccttcatt  aaaacaataa tgttcaataa cctcagtata
    2101 ccggccacat ctgtggagtc tgagcggata tttagcaagg cagggcaaat cgtctcagac
    2161 cgcagaacaa gattaaaaga ggaaaatgct aatgttgaac caaaactttt gcattattca
    2221 attaaagtcc agcaaaaata aggaacatgt tgaacatcag agcaataaaa aattcatttt
    2281 taaatactaa ttacagttgt ttcaaatgg  gcactatcga ggcactatcg atgccatttt
    2341 cgtccactat cgatggggca cactatcgat ggttttccac ctcta
//

```

Anexo F. Anotación del ET *PIF/Harbinger* en *D. amaguana* en formato GenBank.

```

source          1..1893
                /organism="Drosophila amaguana"
                /country="Ecuador"
                /mol_type="genomic DNA"
repeat_region  1..16
                /note="TIR"
                /colour=3
mobile_element 1..1893
                /note="PIF/Harbinger"
repeat_region  1878..1893
                /note="TIR"
                /colour=3
    
```

```

BASE COUNT      535 a      385 c      400 g      573 t
ORIGIN
    
```

```

1 gagcggatga gggcgtgagc ggaggaacag ctgaagatgt tcaacttca gccatccagc
61 tgtctgctct ttgcatagat ccgccaatgc tggacgctct ttggaacgtt aacggattgg
121 aaatgcccat acggcaatgc ctacatgtgt ttgcttccct gtgggaaacg ttgacatgac
181 aatgattaat aagaccctgt gtcatgtcga atatgacgga actcatttaa aaacggatct
241 ggagacggag atcagcctca caaacgaaat ggagcatgag ctatacatag tatgtcagag
301 gggggatgga agcagtgtaa taataatgag ggagagatag agagagagag ggagagagag
361 ctgcacagag cgaggtaaag acagcggcgt tgtcttcgac cggtttcttc gaagagcgcg
421 aggggggagg cgattaacag agtttgggtg gctccctagc ccagctgagg tgccaggcat
481 cagcagatat ttcagtactc aacgaactct tggccgcaac ggacgtttgg cgctctcctt
541 gctgcatggg caaacttggg gagcgcaacg tgggacgagc gacgaaacgc ttgtagttag
601 agtgaatgta aacgtgtcgc gcatgcgtcc gtcgacttca attgttgctg ttgcccgttc
661 ggttcgggtc ggcacgtctc tcaccgttat tataatttgt gtgctgctgc ttttatctat
721 gcaaatgtag aagccaagtg atgtgctggc aaattgttga aattgtttgt cgaaaattgc
781 aaataggaaa aggaaaaaca agaaaagaaa aaccagcaac tccaccgctc gcaatgtttt
841 taaatagtct ccatgttggg tgtggctgct gttgttggg cttctttttt fgatttttgg
901 ttgttgcttt ttgggtcgac gtttcgttcg gtttgccatc tcttttgctc acacgcaaat
961 tcgtcgtctt cggttcggtc gtcttttcgc ctctctgctt tgttcatttt tagcgtcgtc
1021 gtcttcttgt cttttgctta tgtacgtttt cgttcatgaa tataatttat attcgtgttt
1081 atttctgctg ctgctgctgc ttcttcttct tcttcttctt tcaaaaacca cctctccaca
1141 atcatcccac tctccgcctg ttcaatacaa caacaaaaat aaccaacaac tacaacaaac
1201 acaattcaca aaatttagtt gtcgcttgcg caattttagc gcttaataaa attattatta
1261 gcaaaaccaa aataaattcc aagtacgaaa attaaatata gcggaaagtgc aaaaataaac
1321 gttaagtgtg tgcaaaatca taagtgtcga agataattta aaaaaataaa taaggaaata
1381 taatgagaaa aataaatgaa atgaaagtcg caagcaacga ccagcttata aaccaagtct
1441 ttgaatatgt ctcttagaga ttgcataca aattattata aactagttag ctttaagacc
1501 cgatatattc ccaaggcctc aagagcatac ttacatacac acatatgtac gtatatgact
1561 taagagagag aagcgaacag cttgaactaa ggacgaacaa taatatgttg ttccttattc
1621 gttgctattc ttaagaata aagtaaaatt taaaatcagt aaaaaacata ttcaatgttt
1681 ttttttctc cacaagtctt ttaattggca ataattactc ttggaggcgc ttcgggcatc
1741 aagtgcccaa aaattattca acctttgtct gcttggcttc caaaaccgca taattttttt
1801 ctttcttctc ttgcatcca ttataattat ggctttcaac tctgccacgc tttcttgcat
1861 ttctattgct ccagctctca cgctcttccg ctc
    
```

//

**Anexo G.** Anotación del transposón *Tc1/Mariner* en *D. amaguana* en formato GenBank.

```

source          1..1834
                /organism="Drosophila amaguana"
                /country="Ecuador"
                /mol_type="genomic DNA"
repeat_region   1..42
                /note="TIR"
                /colour=3
mobile_element  1..1834
                /note="Tc1/Mariner"
repeat_region   1793..1834
                /note="TIR"
                /colour=3
BASE COUNT      618 a    293 c    324 g    599 t
ORIGIN
    1 agtgagacag taaatatat atttacacag acttaaagta attttgcttt aaaaggtggt
   61 cattgtcctc atgatatgca agcttttaaa agaggaaatt aacagactat tgttatgcat
  121 tagtaaatTTT aaaagtatat ataatcgagt gggaaagaag ctatataca gctaaacttc
  181 acgtttattg cgtctgacat ttcttgaggt gttcataact taaattccat gtatgtaata
  241 cagggcattt aaatgacagt caaccatta aacactcgcc aaataaacca gtcttgttat
  301 agttataata agcttaaaaa catacaaac aagtaaaaaa gttgtagttg agcgtaatcg
  361 actacgtaat actcgtctac agtttaacga cttaaaagag tttctcaggc aaattaacta
  421 aaagattaga cctcagtaaa tatgaacact tactttgatg tagatataca cttgatataa
  481 atagtaagtt tcaagtctct atttatagtc aaaccaatat actcttttca ctctgtgagc
  541 gagcataaaa acagagcttt caagaatcag catgaaagga tttgcatcga gtaaatatgt
  601 ataagtcttg cagatgaatg caaatatgat tgaatcattt aaataaattg tctgcctgca
  661 gctggagggt gtgacaaatg gatcagagcc cgaggatagg gaaatcgccg ccattaacaa
  721 atatttttat tatacccaaa taacaatttc acgtctgctg cgtccgcgaa ccgaatttat
  781 ttacgtcatt tatttgccaa tcatttattt tgcattaatt aaatatgtgg tttgttttat
  841 acaaaaaccg ctaaatattt atgtttttt gtgcataaat ttaacagtgg atgagaaaaa
  901 tgcattttta agctgtaata caaacatgaa aatccaatga aaacattcac ggattttcct
  961 ttttgaagaa gtgaacaaaa ctgaaatgaa atggaaaaag cctttgagtg attttgattt
 1021 tttttcttct gtgggtgctg catgtttttg acaaattgac tttggaagca acaataataa
 1081 caacagcaac aacagcagca aatggctctc aagtactata ttcatattga tttcaaacga
 1141 aactcaacat gcaaatgaca gatttatttt tcatgctgtc gtttcgctg tcaatcaact
 1201 tggccaaggt gttgattgat gtctagactc cttagcgcag ctggttgagc tcgtatgcc
 1261 cagccaggcg attttccca actctaaagc tgagcagcaa aaacatttca tatacataca
 1321 tatacatata cgtattatac gttttttgac agctacgcat tcagttttt tcattcggtt
 1381 ctggttccat atttggattt cttagatgtt ggggtgcaatg cccgccaaat agagcgcctt
 1441 gactacggct gggcacactt gattgaattt ggtatgggga aacgagtgga gctactaagc
 1501 tgggcattag caatgcttag aaatgcaatg acgcatgtag cgtgccatgt tgaataggat
 1561 tatattacaa tattcaagta cctatgatg tacatattat gtacatatgt atgtttgaaa
 1621 ttatactcat atgtctgtca tccagcaaat gaaatgaaat gaaatgggcg tcagtcaaaa
 1681 ctttttcgtg aatcgacttg agaaagtagt caaaagcact ttataagcat gttcaagaat
 1741 ttgatgaatt cgattttgat taaagtaact taatctattt taaataaaat tgattgaatt
 1801 ttttttaaat tgatgtcatt ttagtgtct cact

```

//