



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

## LICENCIATURA EN ECOLOGÍA

Escuela Nacional de Estudios Superiores,  
Unidad Morelia

**Análisis de genómica comparativa del género *Tapirus*:  
secuenciación y ensamblaje del mitogenoma de un  
ejemplar silvestre de tapir centroamericano (*Tapirus  
bairdii*)**

### TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADA EN ECOLOGÍA

P R E S E N T A

GLORIA NOEMI CHASSIN SOLÍS

DIRECTOR DE TESIS: DR. LUIS EDUARDO SERVÍN GARCIDUEÑAS

MORELIA, MICHOACÁN

AGOSTO, 2024



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

## LICENCIATURA EN ECOLOGÍA

Escuela Nacional de Estudios Superiores,  
Unidad Morelia

**Análisis de genómica comparativa del género *Tapirus*:  
secuenciación y ensamblaje del mitogenoma de un  
ejemplar silvestre de tapir centroamericano (*Tapirus  
bairdii*)**

### TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADA EN ECOLOGÍA

P R E S E N T A

GLORIA NOEMI CHASSIN SOLÍS

DIRECTOR DE TESIS: DR. LUIS EDUARDO SERVÍN GARCIDUEÑAS

MORELIA, MICHOACÁN

AGOSTO, 2024



ESCUELA  
NACIONAL  
DE ESTUDIOS  
SUPERIORES  
UNIDAD MORELIA

10  
años  
(2011-2021)

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO ESCUELA  
NACIONAL DE ESTUDIOS SUPERIORES UNIDAD MORELIA  
SECRETARÍA GENERAL  
SERVICIOS ESCOLARES

**MTRA. IVONNE RAMÍREZ WENCE**

DIRECTORA

DIRECCIÓN GENERAL DE ADMINISTRACIÓN ESCOLAR

**PRESENTE**

Por medio de la presente me permito informar a usted que en la **sesión ordinaria 10 del Comité Académico de la Licenciatura en Ecología** de la Escuela Nacional de Estudios Superiores (ENES), Unidad Morelia, celebrada el día **25 de septiembre de 2024**, se acordó poner a su consideración el siguiente jurado para la presentación del Trabajo Profesional de la alumna **Gloria Noemí Chassin Solís** de la Licenciatura en **Ecología**, con número de cuenta **418128930**, con el trabajo titulado: **"Análisis de genómica comparativa de género *Tapirus*: secuenciación y ensamblaje del motigenoma de un ejemplar silvestre de tapir Centroamericano (*Tapirus bairdii*)"**, bajo la dirección como tutor del **Dr. Luis Eduardo Servín Garcidueñas**.

El jurado queda integrado de la siguiente manera:

<b>Presidente:</b>	M. en C. Violeta Patiño Conde
<b>Vocal:</b>	Dr. Rodolfo Pérez Rodríguez
<b>Secretario:</b>	Dr. Luis Eduardo Servín Garcidueñas
<b>Suplente:</b>	Dra. Yunuen Tapia Torres
<b>Suplente:</b>	Dr. Rafael Aguilar Romero

Sin otro particular, quedo de usted.

Atentamente  
"POR MI RAZA HABLARÁ EL ESPÍRITU"  
Morelia, Michoacán a 24 de junio de 2024.

**DRA. ANA YESICA MARTÍNEZ VILLALBA**  
SECRETARIA GENERAL

---

**CAMPUS MORELIA**

Antigua Carretera a Pátzcuaro N° 8701, Col. Ex Hacienda de San José de la Huerta  
58190, Morelia, Michoacán, México. Tel: (443)689.3500 y (55)5623.7300. Extensión Red UNAM: 80614  
[www.enesmorelia.unam.mx](http://www.enesmorelia.unam.mx)

## **AGRADECIMIENTOS INSTITUCIONALES**

Agradezco a mi asesor de tesis, el Dr. Luis Eduardo Servín Garcidueñas, quien hizo posible la realización de este proyecto, por guiarme y apoyarme durante la elaboración de esta tesis.

Agradezco a los miembros del jurado, a la M. en C. Violeta Patiño Conde, Dra. Yunuen Tapia Torres, Dr. Rodolfo Pérez Rodríguez, Dr. Rafael Aguilar Romero y el Dr. Luis Eduardo Servín Garcidueñas por todo su apoyo, por mantener siempre la mejor disposición y sus valiosas observaciones.

Agradezco a la Licenciatura en Ecología, a todas y todos mis profesores que participaron activamente en mi formación como ecóloga y sobre todo por despertar en sus alumnxs la curiosidad y el interés en generar ciencia.

De igual manera agradezco a la UNAM, al Laboratorio de Microbiómica, y a todo el personal de la ENES Morelia por brindarme la oportunidad de estudiar la Licenciatura en Ecología en un espacio de aprendizaje y desarrollo.

## **AGRADECIMIENTOS PERSONALES**

Agradezco al Dr. Rafael Reyna Hurtado y sus colaboradores de ECOSUR, Campeche por la enorme disposición de compartir sus muestras y su valioso tiempo en campo. Al M. en C. Roberto Marín Paredes por acompañarme y aconsejarme constantemente durante la realización de la parte bioinformática del proyecto y al Dr. Luis Eduardo Servín Garcidueñas por proporcionarme un espacio para llevar a cabo este trabajo sobre el tapir en México.

Agradezco a mi madre por ser mi modelo a seguir como mujer en la ciencia y por siempre mostrarse maravillada ante la hermosa naturaleza. Su perseverancia y pasión por la biología lograron que despertara una enorme curiosidad en mí.

Agradezco a mi padre por el apoyo incondicional y por inculcarme desde siempre a cuidar mi capacidad de asombro, la cual me orilló a realizar este proyecto con un mamífero tan icónico.

Agradezco a todas mis maestras de la licenciatura en ecología quienes diariamente durante sus clases nos demostraron sus excelentes capacidades como científicas lo que generó un enorme sentimiento de admiración y respeto entre sus estudiantes.

Agradezco a mis amigas de licenciatura quienes fueron mi círculo de apoyo en todos los aspectos de mi vida durante la carrera y con las cuales aprendí a enamorarme de la ecología.

<b>I. RESUMEN</b> .....	<b>7</b>
<b>II. ABSTRACT</b> .....	<b>8</b>
<b>III. INTRODUCCIÓN</b> .....	<b>9</b>
Historia general de la familia Tapiridae.....	9
Estado de conservación del género <i>Tapirus</i> y datos genómicos disponibles.....	9
Situación actual e importancia del tapir centroamericano <i>Tapirus bairdii</i> .....	12
Estudios moleculares sobre el tapir centroamericano.....	14
Colecta de un ejemplar silvestre de tapir centroamericano.....	15
Muestra de ADN de un ejemplar silvestre de tapir centroamericano.....	17
Importancia de obtener un genoma.....	19
Análisis de genomas mitocondriales.....	21
Relevancia del proyecto.....	23
<b>IV. HIPÓTESIS</b> .....	<b>24</b>
<b>V. OBJETIVO GENERAL</b> .....	<b>25</b>
<b>VI. OBJETIVOS ESPECÍFICOS</b> .....	<b>25</b>
<b>VII. MATERIALES Y MÉTODOS</b> .....	<b>25</b>
Sitio de estudio.....	25
Obtención de muestra.....	26
Extracción y secuenciación de ADN.....	26
Evaluación del control de calidad.....	27
Preprocesamiento de lecturas de secuenciación.....	28
Mapeo del genoma mitocondrial de <i>Tapirus bairdii</i> .....	29
Ensamble del genoma.....	29
Selección de ensamble.....	30
Eliminación duplicados.....	30
Búsqueda de secuencia de la región control.....	31
Anotación.....	31
Cálculo de distancia genoma a genoma.....	32
Análisis filogenómicos.....	32
<b>VIII. RESULTADOS</b> .....	<b>33</b>
Secuenciación de ADN, evaluación y filtrado de calidad.....	33
Ensamble y selección de valores.....	37
Genoma mitocondrial de <i>Tapirus bairdii</i> .....	40
Porcentaje de identidad.....	42
Anotación del genoma mitocondrial de <i>Tapirus bairdii</i> .....	43
Alineamiento y filogenómica.....	46
<b>IX. DISCUSIÓN</b> .....	<b>48</b>
Filtros de calidad.....	49
Genoma mitocondrial.....	51
Filogenómica de <i>Tapirus bairdii</i> .....	52

Sugerencias y perspectivas a futuro.....	54
Sugerencias en campo.....	54
Sugerencias en el procesamiento de datos.....	55
Perspectivas a futuro.....	55
<b>X. CONCLUSIÓN.....</b>	<b>57</b>
<b>XI. REFERENCIAS.....</b>	<b>58</b>
<b>XII. ANEXOS.....</b>	<b>66</b>
Alineamientos.....	66
Líneas de comando.....	83

## I. RESUMEN

En las últimas décadas, la preocupación por el futuro de los ecosistemas a nivel mundial ha alcanzado un nivel nunca antes visto. Entre las principales preocupaciones está el incremento del número de especies que se vean orilladas a la extinción por causas no naturales como los conflictos entre humanos y vida silvestre. El tapir centroamericano, *Tapirus bairdii*, es el mamífero terrestre más grande del Neotrópico y es considerado un ingeniero del ecosistema ya que mantiene una dieta herbívora minuciosa. Se ha observado que suelen desplazarse y defecar con más frecuencia en las áreas degradadas de los bosques, por lo que su presencia es fundamental para la dispersión de semillas y la regeneración de los bosques tropicales. Esta especie, al igual que las demás integrantes del género *Tapirus*, forma parte de la lista roja de la Unión Internacional para la Conservación de la Naturaleza como especie en peligro. Fácilmente podemos reconocer que aún se cuenta con pocos conjuntos de datos genómicos disponibles en las bases de datos a nivel internacional para el género *Tapirus*. Los datos genómicos de especies silvestres son útiles para conocer la diversidad genética de las poblaciones. En este trabajo se secuenció y ensambló el genoma mitocondrial de un individuo silvestre sujeto a monitoreo en la Reserva de la Biósfera de Calakmul, México. El ADN se obtuvo a partir de los folículos de una muestra de pelaje que se encontraba almacenada en etanol durante 11 años a temperatura ambiental. El análisis de los datos de secuenciación mostró daños en la secuencia de ADN, principalmente atribuidos a procesos de depurinación y desaminación. El uso de técnicas bioinformáticas permitió sobrepasar las limitaciones técnicas y se logró recuperar un mitogenoma con un tamaño de 16,567pb y un contenido de GC de 38.4, A(34.15% 5658), C(25.77% 4269), G(12.63% 2092), T(27.45% 4548). La anotación del mitogenoma permitió la identificación de las secuencias correspondientes a los 13 genes codificantes para proteínas, 22 tRNAs y 2 RNA ribosomales. Los análisis de genómica comparativa, tales como alineamientos de secuencias, métricos de identidad y análisis de contenido genético, mostraron que los mitogenomas disponibles para el género *Tapirus* se encuentran altamente conservados. La secuenciación de próxima generación por *whole genome shotgun* ha permitido rebasar barreras que anteriormente dificultan la obtención de datos genómicos como la secuenciación de muestras con material genético degradado.

## II. ABSTRACT

In recent decades, concern about the future of ecosystems worldwide has reached a level never seen before. Among the main concerns is the increase in the number of species that are pushed to extinction due to unnatural causes such as conflicts between humans and wildlife. The Central American tapir, *Tapirus bairdii*, is the largest land mammal in the Neotropics and is considered an ecosystem engineer as it maintains a meticulous herbivorous diet. It has been observed that they tend to move and defecate more frequently in degraded areas of forests, so their presence is essential for seed dispersal and the regeneration of tropical forests. This species, like the other members of the *Tapirus* genus, is part of the red list of the International Union for Conservation of Nature as an endangered species. We can easily recognize that there are still few genomic data sets available in international databases for the genus *Tapirus*. Genomic data from wild species are useful to understand the genetic diversity of populations. In this work, the mitochondrial genome of a wild individual subject to monitoring in the Calakmul Biosphere Reserve, Mexico, was sequenced and assembled. DNA was obtained from the follicles of a fur sample that was stored in ethanol for 11 years at room temperature. Analysis of the sequencing data showed damage to the DNA sequence, mainly attributed to depurination and deamination processes. The use of bioinformatics techniques made it possible to overcome technical limitations and a mitogenome with a size of 16,567bp and a GC content of 38.4, A(34.15% 5658), C(25.77% 4269), G(12.63% 2092), T(27.45% 4548) was recovered. The annotation of the mitogenome allowed the identification of the sequences corresponding to the 13 protein-coding genes, 22 tRNAs and 2 ribosomal RNAs. Comparative genomics analyses, such as sequence alignments, identity metrics, and gene content analyses, showed that the available mitogenomes for the genus *Tapirus* are highly conserved. Next-generation sequencing by whole genome shotgun has made it possible to overcome barriers that previously made it difficult to obtain genomic data, such as the sequencing of samples with degraded genetic material.

### III. INTRODUCCIÓN

#### Historia general de la familia Tapiridae

Dentro del orden Perissodactyla Owen, 1848, del griego *perissós* (impar) y *dáktylos* (dedo), se encuentra la familia Tapiridae Gray, 1821, la cual está compuesta por un solo género: *Tapirus* Brünnich, 1772 (Hulbert, 1995). El género *Tapirus* actual comprende cinco especies: *Tapirus bairdii*, *Tapirus terrestris*, *Tapirus pinchaque*, *Tapirus indicus*, y la quinta y más recientemente propuesta *T. kabomani* (Cozzuol et al., 2013). Sobre esta última especie, es necesario ampliar sus estudios para validar su identidad como nueva especie. Todas las especies son Neotropicales, a excepción de *T. indicus*, que se distribuye en Asia, principalmente en la península de Malasia (Fog, 2015; Ruiz-García et al., 2016)

Durante el Oligoceno tardío hace ~25Ma a Plioceno temprano hace ~1.8Ma, la familia de los tapires tuvo una distribución mundial y conformaron un grupo extremadamente diverso. En el norte del continente americano, la recuperación de fósiles de géneros extintos (ej. *Nexuotapirus*, *Miotapirus*, *Tapiravus* y *Protapirus*) ha permitido que la familia Tapiridae sea considerada biogeográficamente importante.

Se propone que hace ~3Ma la región del Istmo de Panamá cerró completamente, lo que conectó el norte y el sur del continente americano. Ahora, en América, las especies actuales son las sobrevivientes del Gran Intercambio Biótico Americano (GABI). Al igual que los grandes mamíferos terrestres del continente, los tapires migraron a Sudamérica escapando del congelamiento del hemisferio norte. Se ha propuesto que los tapires se resguardaron de la última extinción del Pleistoceno cruzando por el Istmo de Panamá. (Futuyma, 1979; Coates y Stallard, 2013; Tedford et al., 2004; Ashley et al., 1996; Holanda y Ferrero, 2013; Cozzuol et al., 2013).

#### Estado de conservación del género *Tapirus* y datos genómicos disponibles

*T. terrestris* está considerada como especie en estado vulnerable de acuerdo a la Lista Roja de la Unión Internacional para la Conservación de la Naturaleza. El resto de las especies que conforman el género se clasifican actualmente como especies en peligro (IUCN, 2021). Esto remarca la importancia de realizar investigación con enfoques multidisciplinarios sobre la situación actual de los tapires con la finalidad de enriquecer los programas de conservación actualmente implementados. En México contamos con el “Programa para la conservación de la

especie, Tapir Centroamericano (*Tapirus bairdii*)” por la SEMARNAT, 2009” (SEMARNAT, 2009).

El uso de técnicas de secuenciación de ADN y de bioinformática permiten el acceso al análisis de la diversidad genómica de las especies. Actualmente se cuenta con un número limitado de datos genómicos disponibles a nivel internacional para el género. La plataforma del *National Center for Biotechnology Information* (NCBI) tiene disponibles los siguientes genomas (**Tabla 1**).

**Tabla 1.** Genomas de especies del género *Tapirus* disponibles en NCBI. Fecha de consulta: Mayo, 2024.

<b>Especie</b>	<b>Genoma</b>	<b>Número de acceso</b>	<b>Fecha de liberación</b>	<b>Información adicional</b>
<i>T. indicus</i>	Nuclear	<b>GCA_03187870 5.1</b>	Septiembre, 2023	
<i>T. indicus</i>	Nuclear	<b>GCA_03187865 5.1</b>	Septiembre, 2023	
<i>T. indicus</i>	Nuclear	<b>GCA_00402490 5.1</b>	Enero , 2019	
<i>T. indicus</i>	Mitocondrial	<b>CM063009.1</b>	Agosto, 2023	
<i>T. indicus</i>	Mitocondrial	<b>KJ417809.1</b>	Febrero, 2014	
<i>T. indicus</i>	Mitocondrial	<b>NC_023838</b>	Febrero, 2014	Este registro ha sido curado por el personal del NCBI. La secuencia de referencia es idéntica a <b>KJ417809</b> .
<i>T. indicus</i>	Mitocondrial	<b>KJ417810.1</b>	Febrero,2014	
<i>T. indicus</i>	Mitocondrial	<b>KJ417808.1</b>	Febrero,2014	
<i>T. indicus</i>	Mitocondrial	<b>NC_023838.1</b>	Marzo, 2014	Este registro ha sido curado por el personal del NCBI. La secuencia de referencia es idéntica a <b>KJ417809</b> .
<i>T. terrestris</i>	Nuclear	<b>GCA_00402502 5.1</b>	Enero, 2019	
<i>T. terrestris</i>	Nuclear	<b>GCA_02853325 5.1</b>	Febrero 2023	

<i>T. terrestris</i>	Mitocondrial	<b>MW257204.1</b>	Noviembre, 2020	
<i>T. terrestris</i>	Mitocondrial	<b>NC_053962.1</b>	Marzo, 2021	Este registro ha sido curado por el personal del NCBI. La secuencia de referencia es idéntica a <b>MW257204</b> .
<i>T. pinchque</i>	Mitochondrial	<b>OQ420428.1</b>	Febrero, 2023	
<i>T. bairdii</i>	Mitochondrial	<b>OM935749.1</b>	Diciembre, 2022	
<i>T. bairdii</i>	Mitochondrial	<b>NC_063943.1</b>	Marzo, 2022	Este registro ha sido curado por el personal del NCBI. La secuencia de referencia es idéntica a <b>OM935749</b> .

### **Situación actual e importancia del tapir centroamericano *Tapirus bairdii***

El tapir centroamericano, también conocido como danta, anteburro o tapir de Bairdi, es el mamífero terrestre más grande del Neotrópico. Los individuos adultos pueden alcanzar una longitud de dos metros y su peso puede variar entre 150 y 300 kilogramos (Ceballos et al., 2020).

Al ser un mamífero de talla grande, presenta requerimientos espaciales extensos, mantiene una distribución restringida a los tipos de vegetación nativos, y presenta densidades poblacionales naturalmente bajas. Prefieren bosques bien conservados con disponibilidad de comida y agua donde puedan consumir de 12 a 15 kg de biomasa por día (Foerster y Vaughan, 2015). Estas características de historia de vida hacen que el tapir centroamericano sea altamente sensible a la fragmentación de su hábitat (Muench, 2006).

En México la distribución verificada del tapir centroamericano incluye los estados de: Campeche (Reservas de Calakmul, Balamkú, Balam-kin, Laguna de Términos y Los Petenes); Chiapas (Reservas Lacantún, Montes Azules, El Ocote, El Triunfo, La Sepultura y algunos ejidos de la Sierra Madre y la Selva Lacandon); Oaxaca (la región de Chimalapas, el Río Verde, la Sierra Mixe y la Sierra Norte); Quintana Roo (Reserva SianKa'an y ejidos grandes del centro y sur del estado); y Veracruz (la cuenca alta del río Uxpanapa) (Naranjo, 2018).

A pesar de que las poblaciones de esta especie son naturalmente bajas, en las últimas tres décadas estas se han reducido posiblemente hasta el 50% (Castellanos et al., 2008). En el 2014, *T. bairdii* fue integrado a la lista roja de la IUCN como especie En Peligro según los criterios A2abcd+3bcde (García et al., 2016). La pérdida de hábitat representa el factor de mayor riesgo para *T. bairdii* en México, ocasionada principalmente por incendios forestales, grandes obras de infraestructura (presas hidroeléctricas y autopistas), expansión urbana, transmisión de enfermedades por animales domésticos, contaminación de cuerpos de agua, y efectos del cambio climático global (Naranjo, 2009; García et al., 2016; Monette, 2019).

El tapir centroamericano presenta múltiples características que lo convierten en un ingeniero de los ecosistemas. Se consideran a especies ingenieras de ecosistemas, a aquellas que modulan directa o indirectamente la disponibilidad de recursos para otras especies, al hacerlo, modifican, mantienen y crean nuevos hábitats (Jones et al., 1994).

Al igual que otras especies pertenecientes al orden Perissodactyla, *T. bairdii* presenta diferencias anatómicas (menos volumen en el colon proximal) y metabólicas (metabolismo lento, baja frecuencia de defecación e ingesta de alimentos, y mayor retención de líquidos y partículas) en su tracto intestinal. Debido a esto, ellos mantienen una selección minuciosa de alimentos (hojas, frutos, cortezas y ramas de especies leñosas y no leñosas) a lo largo de todo el año. También se ha observado que suelen desplazarse y defecar con más frecuencia en las áreas degradadas de los bosques, por lo que su presencia es fundamental para la dispersión de semillas y en la regeneración de los bosques tropicales (Yanez-Montalvo et al., 2021; Polucci et al., 2018; Monette, 2019).

Un estudio realizado en la península de Yucatán por Carrillo-Reyna y colaboradores (2015), prueba que *T. bairdii* es la única especie conocida que puede dispersar con éxito semillas grandes de especies arbóreas como el chicozapote (*Manilkara zapota*). Se ha observado que los bosques se tornan extremadamente diferentes cuando este animal desaparece de un área

determinada, y se sabe que las poblaciones a lo largo de su área de distribución se están extinguiendo rápidamente. Es posible que si no se brinda más atención a la especie y no se maneja esta situación en el futuro sólo será posible encontrar pequeñas poblaciones aisladas sin conexión entre ellas, posteriormente las poblaciones de tapir se tornarán inviables (O'Farrill et al. 2014; Martins, 2014; Carrillo-Reyna et al., 2015).

A pesar de ser una especie clave para los bosques sudamericanos, el tapir centroamericano no ha sido estudiado a profundidad. En México se han realizado escasos estudios, únicamente de las poblaciones que se encuentran en la Reserva Montes Azules - Selva Lacandona, Reserva el Triunfo y La Sepulta -Sierra Madre de Chiapas, Los Chimpalas y Calakmul. Investigaciones empleando técnicas de fototrampeo y telemetría han logrado describir el ámbito hogareño de individuos, así como los cambios en su comportamiento. Se ha demostrado que en zonas aledañas a los ejidos, los tapires suelen ser más activos a comparación de aquellos que se encuentran dentro de una reserva. Este cambio en su conducta podría ser explicado por la presencia humana, sin embargo hay poca información del tapir centroamericano en México que ayude a resolver este tipo de preguntas (Naranjo et al., 2015; Sánchez-Pinzón et al., 2019).

### **Estudios moleculares sobre el tapir centroamericano**

*Tapirus bairdii* cuenta con 40 pares de cromosomas nucleares, de los cuales el último par corresponde a los cromosomas sexuales (Houck et al., 2000).

Anteriormente se han realizado algunos trabajos con enfoques moleculares del tapir centroamericano. En Panamá y Costa Rica se colectaron muestras de sangre y pelo (n=33) y se eligieron nueve microsatélites previamente establecidos para *T. bairdii* y *T. Terrestris* y con los datos obtenidos se pudieron realizar análisis de variabilidad y estructura de las poblaciones de *T. bairdii* de ambos países. Además se logró estimar que las poblaciones actuales derivan de una sola población panmíctica y posiblemente hayan sufrido un evento de cuello de botella hace 50-100 años. Actualmente las poblaciones en Panamá y Costa Rica presentan valores de heterocigosidad y diversidad alélica dentro de los más bajos reportados para mamíferos de talla grande (Norton y Ashley, 2004). Por lo tanto, es necesario generar e integrar más datos moleculares sobre la especie en toda su área de distribución incluyendo México.

Se ha logrado confirmar que *T. bairdii* se encuentra en la situación menos favorable ya que a pesar de que su distribución geográfica es amplia, se ha

reducido y fragmentado dramáticamente en los últimos dos siglos y se estima que actualmente quedan alrededor de 6.000 individuos en estado silvestre. (Ruiz-García et al., 2016; Ashley et al. 1996).

En la actualidad se cuenta para *T. bairdii* con 37 secuencias de genes completos mitocondriales; las cuales incluyen a; la NADH deshidrogenasa (subunidades 1-6); ATP sintasa (F0 subunidades 6 y 8); citocromo C oxidasa (subunidades 1-3); tRNAs (codificantes a 22 aminoácidos); las subunidades 12S y 18S rARN ribosomales. Además, como se observa en la **Tabla 1**, se encuentra disponible de un genoma mitocondrial completo de individuo en cautiverio en el zoológico de Xcaret, Quintana Roo, que fue publicado en 2022, durante la realización de esta tesis (Ennis et al., 2022).

### **Colecta de un ejemplar silvestre de tapir centroamericano**

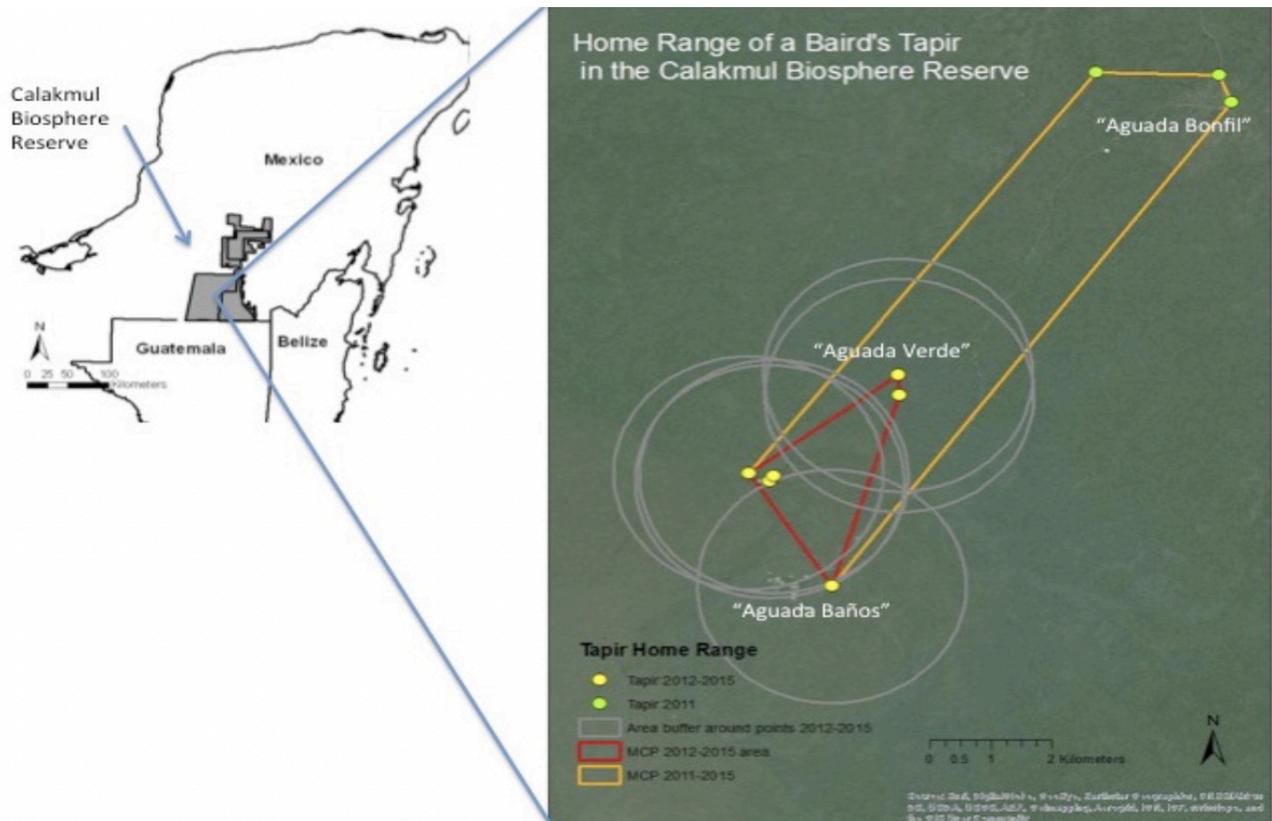
Como parte de un proyecto fundado por el Consejo Nacional de Ciencia y Tecnología (CONACyT), aunado con la Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT) se realizó la investigación titulada “Insights into the multiannual home range of a Baird’s tapir (*Tapirus bairdii*) in the Maya Forest”. Este fue el primer reporte del ámbito hogareño, capacidad de viaje y distancias del tapir en la Reserva de la Biósfera de Calakmul (Reyna-Hurtado et al., 2016).

Reyna-Hurtado y colaboradores lograron reconstruir la distribución multianual de un individuo monitoreado 20 días por telemetría y tiempo completo por cámaras-trampa durante cinco años (mayo 2011- septiembre 2015). Registraron un rango de distribución a 24-40 km<sup>2</sup>, mayor a los valores antes registrados para la especie en otros estudios (**Figura 1.**) (Reyna-Hurtado et al., 2016; Foster, 2001).

Con la finalidad de colocar un collar de telemetría (**Figura 2.**) se obtuvieron permisos de colecta otorgados por la Dirección General de Vida Silvestre número: SGPA/DGVS/02665/13 y con la autorización de las Autoridades de la Reserva de la Biosfera de Calakmul y del comité de ética de El Colegio de La Frontera Sur, ECOSUR.

En el 2015 se logró capturar un macho juvenil silvestre disparando un dardo tranquilizador de sistema de entrega remota (Daninject Inc. Austin, TX) con una combinación de butorfanol y clorhidrato de xilazina, así después de un par de horas pasa completamente el efecto de la anestesia y se le permite al individuo

continuar con su trayecto habitual (Reyna-Hurtado et al., 2016; Foester et al., 2000).



**Figura 1.** Imagen obtenida del artículo “*Insights into the multiannual home range of a Baird's tapir (Tapirus bairdii) in the Maya Forest*” por Reyna-Hurtado y colaboradores, 2016, que muestra registros de presencia de tapir centroamericano (mayo 2011 - septiembre 2015) en la Reserva de la Biósfera de Calakmul. Se muestran datos con polígono convexo mínimo de todas las localidades (mayo de 2011 a septiembre de 2015), de solo el área donde se dispersó el individuo (noviembre de 2011 a septiembre de 2015), y el polígono de rango de hogar estimado usando un buffer de la mitad de la distancia máxima recorrida de los lugares donde se dispersó (noviembre de 2011 a septiembre de 2015).



**Figura 2.** Imagen capturada en el 2011 durante la manipulación de “Emiliano”, el ejemplar de tapir silvestre muestreado por el equipo de trabajo de Reyna-Hurtado y colaboradores.

### **Muestra de ADN de un ejemplar silvestre de tapir centroamericano**

Durante la manipulación directa de “Emiliano” se lograron coleccionar algunos pelos con folículo y un par de centímetros de tejido de la piel. Las muestras fueron almacenadas en un tubo Falcon con alcohol al 70% durante una década (2011-2021) en condiciones de temperatura ambiental.

La muestra que se logró obtener en el 2011 fue almacenada en La Escuela del Sur, ECOSUR, y posteriormente fue donada al Laboratorio de Microbiómica de la Escuela Nacional de Estudios Superiores, unidad Morelia con la finalidad de realizar el presente trabajo de secuenciación del genoma mitocondrial de un individuo silvestre de *T. bairdii*.

Para obtener la secuenciación de un genoma, ensamblarlo y analizarlo se requiere seguir una serie de pasos en campo, de laboratorio y computacionales utilizando herramientas bioinformáticas. El primer paso es conseguir al menos una

muestra del organismo *in situ* (se puede utilizar sangre, excretas, tejido, o pelos con folículo).

El tapir centroamericano es predominantemente nocturno y se encuentra en bajas densidades, lo que complica el trabajo en campo y el esfuerzo de muestreo suele ser extensivo para capturar un tapir silvestre (Yanez-Montalvo et al.,2021). Pueden pasar meses e inclusive años para capturar un ejemplar, además se debe planear con anticipación el material necesario para coleccionar una muestra que se analizará a escala molecular. Aquí resalta la importancia del método y del tiempo de almacenamiento de la muestra debido a que las características de la misma juegan un papel crucial en la conservación y extracción de ADN.

Es importante recordar que las células de organismos vivos mantienen mecanismos de reparación enzimática las cuales favorecen la conservación de la integridad del genoma. El ADN sufre continuamente agresiones químicas, y al morir, estos mecanismos de regulación desaparecen. Cuando las nucleasas intracelulares ya no están secuestradas en la célula, es posible que estas puedan acceder al ADN y degradarlo, así como los microorganismos comúnmente presentes en los tejidos en descomposición. Factores físicos, químicos y biológicos pueden conducir a la pérdida de todo el ADN recuperable.

La degradación del material genético es prevenible cuando se emplean técnicas de conservación adecuadas como congelar o disecar el tejido inmediatamente después de la colecta (Dabney et al., 2013).

Cabe señalar que la depurinación es uno de los mecanismos mejor caracterizados de degradación del ADN, en donde el enlace N -glicosilo entre un azúcar y un residuo de adenina o guanina se divide, dando como resultado un sitio abásico (apurínico o apuriridimínico). Particularmente esto suele suceder en condiciones de almacenamiento con pH bajo. Las bases de nitrogenadas que configuran al ADN son susceptibles a la desaminación de citosina a uracilo. Estos mecanismos favorecen a la incorporación de A frente a la U y, a su vez, una T frente a la A, lo que adicionalmente resultan en sustituciones aparentes de C por T o de G por A. Las ADN polimerasas leen las bases erróneamente (Dabney et al., 2013).

Por este motivo, la depurinación y desaminación son algunos de los mecanismos principales que conducen a lesiones de codificación errónea en el ADN y deben ser consideradas al momento de trabajar con muestras de ADN antiguo (Dabney et al., 2013; Sunkari et al.,2023).

## Importancia de obtener un genoma

La introducción de datos moleculares a la sistemática filogenética causó una increíble mejora en la comprensión de la historia evolutiva de las especies (Delsuc et al., 2005).

Los avances en las tecnologías de secuenciación han hecho posible secuenciar genomas completos de organismos no modelo, incluidas especies de interés para la conservación, y a medida que la secuenciación por *Whole Genome Shotgun* (WGS) se ha vuelto más eficiente, la cantidad de organismos secuenciados por esta estrategia ha aumentado rápidamente a escala global (Seremi et al., 2019; Delsuc et al., 2005).

Esta riqueza de datos moleculares que era inimaginable hace sólo una década, está dando lugar a nuevas áreas de investigación como la filogenómica. La cual utiliza principios filogenéticos para dar sentido a la escala de los datos genómicos y así generar reconstrucciones evolutivas más precisas. (Seremi et al., 2019; Delsuc et al., 2005).

Al ser campo emergente, la filogenómica presenta errores sistemáticos como a los cuales anteriormente se enfrentó la filogenética. Con el análisis de datos y selección de métodos de inferencia adecuados, es posible disminuir drásticamente errores estocásticos o de muestreo (Delsuc et al., 2005).

Otro ejemplo en donde se integran herramientas genómicas con otros campos es en la genómica del paisaje. Esta rama emergente de la genética del paisaje tiene como objetivo identificar factores ambientales o paisajísticos que influyen en la diversidad genética adaptativa en especies silvestres utilizando el genoma. Con la genómica es posible realizar exploraciones con una gran cantidad de marcadores moleculares genotipados por individuo (Manel y Holderegger, 2013).

Reconocemos que el desarrollo de la genómica del paisaje ha sido en gran medida por la aparición de grandes conjuntos de datos genómicos y ambientales en genómica de poblaciones. Hasta ahora, la mayoría de los estudios genómicos del paisaje han utilizado herramientas de la genómicas de poblaciones, por ejemplo, la detección de loci atípicos en grandes conjuntos de bases. Esto es importante ya que la distribución alélica atípica puede ser relacionada directa o indirectamente con factores ambientales (Manel y Holderegger, 2013).

En la actualidad se discute si realmente es importante obtener esta gran cantidad de información para especies eucariotas y emplear recurso en crear nuevos análisis evolutivos previamente resueltos con marcadores moleculares conocidos (Philippe et al., 2011), sin embargo recordemos el debate entre la comunidad científica hace poco menos de 25 años con la aparición del Proyecto del Genoma Humano (HGP).

Antes se cuestionaba fuertemente si realmente valía la pena mapear las vastas regiones no codificantes del genoma. Ahora, se reconoce la importancia de recuperarlas ya que la mayoría de las secuencias funcionales del genoma no codifican proteínas y actúan como reguladoras de genes (Gates et al., 2021). Además, contar con una gran cantidad de datos posibilita corroborar análisis anteriores, especialmente resultados confusos o con poca significación estadística empleando conjuntos de datos más pequeños (Lockhart y Winzeler, 2000).

Además de secuenciar especies nuevas y completamente desconocidas, la comunidad científica también ha reconocido el valor de secuenciar los genomas completos de varios individuos como se realiza en la genómica de poblaciones (Hohenlohe et al., 2021).

Los datos de WGS de un número relativamente pequeño de individuos pueden proporcionar información a diferentes escalas de tiempo, por ejemplo, la historia demográfica y las relaciones filogenéticas entre poblaciones aisladas o determinar el nivel de endogamia (Hohenlohe et al., 2021).

También, se reconoce que los datos genómicos pueden permitir evaluar la estructura de la población desde la perspectiva de la conectividad, lo que es de suma importancia considerar en las acciones de conservación (Hohenlohe et al., 2021). Por ejemplo el estudio realizado por Robinson y colaboradores (2019) con una población de lobos (*Canis lupus*) en Isle Royale, Michigan, Estados Unidos.

En este caso se obtuvieron muestras de 11 individuos recolectados entre 1988 y 2012 para la secuenciación y el análisis del genoma completo (Robinson et al., 2019). Anteriormente la población sufrió un proceso de cuello de botella y en la actualidad se han detectado deficiencias en la adecuación biológica de los individuos. Se estimó que el tamaño de la población en la isla oscila entre tan solo 8 y 30 individuos.

Se detectaron múltiples regiones en los genomas con homocigosidad, posiblemente influenciadas por la historia demográfica de la población. Finalmente se propone llevar a cabo un flujo genético periódico asistido por humanos con la

finalidad de disminuir la endogamia y promover la presencia largo plazo de esta población reintroducida en Isle Royale (Robbinson et al., 2019)

El campo de la genómica comparativa cuantifica y describe las coincidencias entre genomas completos, y gracias a la nueva secuenciación e integración de genomas, es posible comparar homologías entre especies y poblaciones (Hardison, 2003). Por ejemplo, el proyecto *Zoonomia Project* es un proyecto que pretende comparar los genomas de cientos de mamíferos ya secuenciados con el propósito de identificar regiones conservadas en todas las especies posibles (*Zoonomia Consortium*, 2020).

El proyecto confirma a nivel genómico que probablemente aquellas posiciones que se han mantenido iguales por aproximadamente 100 millones de años de tiempo evolutivo en la mayoría de las especies de mamíferos, cumplen alguna función importante; por lo que si se presentan modificaciones en estas regiones podrían tener un efecto negativo en los individuos afectando su salud. Actualmente el proyecto *Zoonomia* da prioridad a la obtención de genomas no disponibles, aún no se ha integrado el genoma de *T. bairdii* (*Zoonomia Consortium*, 2020).

Por otro lado, cuando concluye el proceso de análisis de un genoma, existe la posibilidad de compartirlo a la comunidad científica a través de las bases de datos públicas. Esto abre la posibilidad de realizar análisis de genómica comparativa incluyendo el archivo obtenido y archivos de otros genomas disponibles.

Algunas de las principales fuentes de datos de secuencias de nucleótidos son aquellas involucradas con la International Nucleotide Sequence Database Collaboration INSDC-DDBJ, European Nucleotide Archive ENA, European Molecular Biology Laboratory- European Bioinformatics Institute EMBL-EBI y el *National Center for Biotechnology Information NCBI-GenBank*. Entre ellas existe un intercambio de datos nuevos una vez cada 24 horas, principalmente lo que mantiene la buena disponibilidad de información pública actualizada (Lipman y Pearson, 1985; INSDC, 2023).

### **Análisis de genomas mitocondriales**

Las mitocondrias desempeñan un papel central en el metabolismo, la apoptosis, las enfermedades y el envejecimiento de los organismos. Son el sitio donde se lleva a cabo la fosforilación oxidativa, proceso bioquímico esencial para una

variedad de funciones en el que sintetiza ATP y una molécula de oxígeno resulta como aceptor terminal de electrones.

Una de las técnicas más utilizadas en sistemática molecular durante la última década ha sido el análisis del ADN mitocondrial. Actualmente la información del genoma mitocondrial se ha convertido en la fuente de marcadores moleculares más empleada para los estudios de diversidad molecular en animales (Nabholz et al., 2008).

Los genomas mitocondriales o mitogenomas, en animales, se transmiten de forma uniparental por lo que no hay recombinación del material genómico. En la gran mayoría de los casos es por vía materna, ya que la hembra es la que hereda sus mitocondrias al cigoto.

Se estima que este evoluciona entre cinco y diez veces más rápidamente que el ADN nuclear y mantiene una tasa general de sustitución de nucleótidos del 1% por millón de años. La tasa de sustitución depende de la región considerada ya que se han identificado regiones de evolución rápida y lenta dentro de un mismo genoma. Además, también se ha identificado que las tasas de mutación son variables entre linajes animales (Castellana, et al., 2011; Pesole et al., 1999; Nabholz et al., 2008).

Entre los metazoos (animales multicelulares), el genoma mitocondrial es casi siempre una molécula circular cerrada (a excepción de ciertas clases de cnidarios con mitogenomas lineales) que mantiene un tamaño de 16.5kb aproximadamente. Regularmente contiene 37 genes, dos para ARNr, 13 para proteínas y 22 para ARNt. Se han encontrado genomas mitocondriales mucho más grandes, sin embargo, estos casos ocurren por procesos de duplicaciones de genomas o regiones del ADN mitocondrial (Jeffrey, 1999).

En el grupo de los Metazoos, la región control (RC) de la mitocondria forma gran parte de la región no codificante del genoma y en ella se localiza el inicio mitocondrial de replicación y transcripción. Además de la subunidad 1 de la citocromo oxidasa y el gen ribosomal 18S rRNA), la RC se ha utilizado para estimar relaciones filogenéticas entre especies estrechamente relacionadas (Gates et al., 2021; Mithiaux, 2016).

Para el género *Tapirus*, se han realizado análisis genéticos empleando marcadores moleculares como el citocromo C oxidasa (Ashley et al., 1996) y citocromo b oxidasa (Ruiz-García et al., 2014), y genómicos en donde se utilizaron múltiples marcadores, dos ARNr y 13 genes codificadores de proteínas (Ruiz-García et al., 2016).

## Relevancia del proyecto

Se estima que la precipitación en los últimos 50 años ha disminuido un 16% en la región del Gran Calakmul en la Península de Yucatán, lo que ha afectado la conectividad de los recursos hidrológicos causando una modificación negativa en la estructura espacial y conectividad de mamíferos como el tapir de bairdii, pecarí de labios blancos y jaguar (O'Farril et al., 2014).

También se ha observado que a medida que la disponibilidad de agua se reduce, los tapires se ven obligados a buscar nuevas fuentes para hidratarse, estos mamíferos suelen consumir de 10-15L de agua por día, así que cuando en las aguadas de la reserva no hay suficiente recurso, los tapires suelen acercarse a los bebederos de ganado cerca de los poblados modificando el rango de distribución de los mismos. Se ha observado que como medida para evitar contacto con humanos, los tapires han cambiado su comportamiento a hábitos cada vez más nocturnos a diferencia de lo anteriormente registrado en vida silvestre (Pérez- Flores et al., 2021).

Por otro lado, un estudio realizado por Benítez y colaboradores en el 2019 muestra que dentro de las estimaciones preliminares del Fondo Nacional de Fomento al Turismo (FONATUR), quienes coordinan el proyecto ferroviario del *Tren Maya*, se espera que en la próxima década, esta construcción propiciará un incremento de 7,500% con respecto al número de visitantes actual (40,000 personas/año) en la zona arqueológica de Calakmul, Campeche. Además la infraestructura del tren generará impactos negativos sobre la biodiversidad de la región, cuya magnitud dependerá del alcance de los efectos de barrera (bloqueo de movimiento y atropellamiento) y de borde (ruido), causados por la carretera 186 y la vía férrea (Benítez et al., 2019).

La cacería ilegal no es un problema grave a comparación de la pérdida de hábitat. Anteriormente se han realizado encuestas a los habitantes nativos aledañas al hábitat natural de los tapires y se reportó que la caza de tapir persiste principalmente como prevención ya que *T. bairdii* llega a forrajear en los cultivos de uso personal de las familias (maíz, calabaza y frijol) y como respuesta, los campesinos los cazan ilegalmente para evitar que invadan sus cultivos nuevamente. Esta práctica ilegal permanece a pesar de que se reconoce que los daños ocasionados por la presencia de tapires no son significativos en la pérdida de cultivos (Reyna-Hurtado y Tanner 2007; Reyna-Hurtado et al., 2021).

Las proyecciones a futuro para el tapir centroamericano en México con desalentadores y la información sigue siendo limitada. En este trabajo se pretende secuenciar el genoma de “Emiliano”, un individuo silvestre monitoreado por Reyna-Hurtado y colaboradores (2016) en su hábitat natural. La intención es recuperar y analizar el genoma mitocondrial para la obtención de marcadores moleculares. Emplear herramientas bioinformáticas será útil para recuperar el genoma mitocondrial a partir de los datos de secuenciación sobrepasando las limitaciones técnicas de degradación de la muestra.

Además del mitogenoma de *T. bairdii* secuenciado y ensamblado en este proyecto, como se mencionó anteriormente, existe otro disponible en las bases de datos de libre acceso para la misma especie, el material genético secuenciado provenía de un tapir en cautiverio del Parque Xcaret, Quintana Roo, México. Debido a que la muestra fue obtenida de un tapir en cautiverio sería importante compararlo con el genoma de uno silvestre. Se ha observado que los individuos en cautiverio suelen presentar divergencias genéticas y fenotípicas (McPhee y Carlstead, 2010).

Esta escasa pero valiosa información genómica de distintas especies facilita la reconstrucción de la historia evolutiva del género, así como el incremento del acervo genómico disponible en las bases de datos de libre acceso. Además, de acuerdo con Ennys y colaboradores, 2022, el análisis de la anotación del mitogenoma completo de *T. bairdii* contribuirán a la mejor comprensión de su diversidad genómica y estructura de la población, posteriormente ayudará a la conservación y protección de sus poblaciones con el fin de conservar a este icónico megamamífero neotropical en peligro de extinción (Ennys et al., 2022; Gene; NCBI).

#### **IV. HIPÓTESIS**

A partir de la secuenciación del ADN total de una muestra por *Whole Genome Shotgun* (WGS) es posible recuperar el ensamble del genoma mitocondrial de un individuo silvestre de *T. bairdii*.

Los estudios basados en mitogenomas permiten obtener reconstrucciones filogenéticas con más resolución que los estudios basados en marcadores morfológicos y moleculares de genes concatenados utilizados tradicionalmente.

La filogenómica mitocondrial resuelve mejor las relaciones filogenéticas del género *Tapirus* y permite conocer mejor las características de los genomas de tapires y especies afines.

## **V. OBJETIVO GENERAL**

Realizar la secuenciación del genoma mitocondrial de un ejemplar silvestre de *Tapirus bairdii* y utilizar herramientas de genómica comparativa para el análisis de los mitogenomas del género *Tapirus* disponibles en las bases de datos.

## **VI. OBJETIVOS ESPECÍFICOS**

- Obtener una muestra de tejido de un ejemplar silvestre de *Tapirus bairdii* con la finalidad de realizar una secuenciación masiva de ADN.
- Ensamblar y anotar el genoma mitocondrial del tapir *Tapirus bairdii* a partir de los datos de secuenciación.
- Llevar a cabo análisis de genómica comparativa del genoma mitocondrial de *Tapirus bairdii* contra los demás genomas mitocondriales disponibles para el género *Tapirus* en el NCBI.
- Realizar un árbol filogenómico y analizar la historia evolutiva de las especies actuales del género *Tapirus*.

## **VII. MATERIALES Y MÉTODOS**

### **Sitio de estudio**

La región del Gran Calakmul es la segunda región de bosque tropical más grande y biológicamente importante de América, por lo que el 90% de esta se encuentra protegida (Pérez-Flores et al., 2021)

La muestra analizada fue recolectada en el área de La Reserva de la Biosfera Calakmul (RBC), la cual tiene una extensión de 7238.5 km<sup>2</sup>. Esta se ubica en el municipio de Calakmul, Campeche. El estado de Campeche forma parte de la Provincia Biótica Yucateca, cuya región cuenta con características geológicas

homogéneas sin demasiados relieves. Se estima que la zona emergió hace ~3 Ma y se ha convertido en un gran corredor biológico para las comunidades silvestres.

Presenta un clima cálido subhúmedo (Aw) de acuerdo al sistema de clasificación de Köppen modificado por García, con lluvias en verano (Ruiz, 2010; Gallina-Tessaro, 2019). La vegetación está compuesta por selvas medianas subcaducifolias, selvas bajas subcaducifolias, caducifolias, selvas bajas inundables, selva alta subperennifolia, y la vegetación dominante es la selva mediana subperennifolia. Este bosque tropical semi-perenne se ha mantenido prácticamente intacto desde hace 1100 años que fue abandonado por los Mayas (Reyna-Hurtado et al., 2016; Carillo-Reyna et al., 2015, Pennington y Sarukhán, 2005).

### **Obtención de muestra**

En marzo del 2021 con la finalidad de realizar la secuenciación del mitogenoma del tapir, nos contactamos con el equipo de trabajo de Reyna-Hurtado y colaboradores.

Como se mencionó anteriormente, en su investigación pudieron obtener material biológico de un individuo silvestre en el 2011, con el cual se pudo llevar a cabo la obtención del genoma mitocondrial. Las muestras fueron enviadas desde Campeche a Michoacán, en donde fueron transferidas a un tubo Eppendorf con agua estéril para así almacenarlas y enviarlas a secuenciar.

Adicional al proyecto actual, en abril del 2023 se logró capturar un nuevo individuo de tapir macho del cual únicamente se colectaron pelos con folículo e inmediatamente se transfirieron a un tubo Falcon con alcohol al 70%. Nuevamente se solicitó a la compañía MacroGen en Corea del Sur el servicio de secuenciación, para esta nueva muestra únicamente fue posible obtener el gen codificante para proteínas de la Subunidad 2 de la NADH Deshidrogenasa.

### **Extracción y secuenciación de ADN**

Debido al tiempo de almacenamiento (2011-2021) y el tipo de almacenamiento de la primera muestra se optó por solicitar un servicio de extracción y secuenciación de ADN a MacroGen, Inc. Seúl, Corea del Sur.

La obtención de ADN se realizó con un kit de extracción TruSeq Nano DNA, siguiendo el protocolo TruSeq Nano DNA *Sample Preparation Guide*, Part # 15041110 Rev. D.

La empresa cuenta con protocolos rigurosos para la construcción de librerías, en donde primero se purifica el material genético de la muestra y posteriormente se realiza un control de calidad en donde las muestras calificadas pasan a la construcción de la librería. La librería de secuenciación se prepara mediante la fragmentación aleatoria de la muestra de ADN, seguida de la ligadura del adaptador 5' y 3'. Alternativamente, la "tagmentación" combina las reacciones de fragmentación y ligadura en un solo paso y luego los fragmentos ligados al adaptador se amplifican por PCR y se purifican en gel.

Posteriormente se carga en una celda de flujo en donde los fragmentos se capturan por oligonucleótidos unidos a la superficie de la celda, los cuales complementan a los adaptadores de la librería. Luego, cada fragmento se amplifica en grupos clonales a través de una amplificación de puente. Cuando se completa la generación de los grupos, las plantillas están listas para la secuenciación.

Para esta muestra se utilizó la tecnología de secuenciación de nueva generación de Illumina NovaSeq con un método patentado basado en terminadores reversibles que detecta bases individuales a medida que se incorporan a las cadenas de plantilla de ADN previamente preparadas. Finalmente se ejecuta una secuenciación base por base de alta precisión que prácticamente elimina los errores específicos del contexto de la secuencia, incluso dentro de regiones de secuencias repetitivas y homopolímeros.

Como resultado al paso previo obtuvimos dos archivos en formato fastq cada uno con los resultados de la secuenciación ("TAP\_1.fastq" para las lecturas forward y "TAP\_2.fastq" para las lecturas *reverse*) en donde se muestra la información genética nuclear, mitocondrial e incluso información de microorganismos contaminantes que pudieron haberse desarrollado en el medio que contenía las muestras de foliculo y tejido.

## **Evaluación del control de calidad**

El control de calidad es un primer paso esencial en el análisis de datos de secuenciación, fue necesario evaluar si el contenido de las librerías permite

cumplir con las expectativas del experimento y si los datos son adecuados para el análisis posterior. A pesar de que la mayoría de los secuenciadores generan un informe de control de calidad (que generalmente solo se enfoca detectar problemas generados por el propio secuenciador) es necesario también realizar un informe de control de calidad que pueda detectar problemas del secuenciador, así como los del material de la librería inicial.

Para esto se realizó un control de calidad que consiste en una serie de pruebas; recuento de la frecuencia relativa de los nucleótidos en cada posición de un conjunto de lecturas para detectar posibles desviaciones de las frecuencias esperadas, el resumen de la distribución de las puntuaciones de calidad de Phred para identificar las posiciones base con baja calidad global (lo que sugiere una degeneración en el proceso de secuenciación), y medir la frecuencia de adaptadores de secuenciación y contaminantes que no se espera que sean ADN biológico de la muestra (University of Missouri; de Sena Brandine y Smith, 2019) La puntuación de calidad Phred de 20 significa una precisión del 99 % y las lecturas con una puntuación de calidad superior a 20 se aceptan generalmente como lecturas de buena calidad (Illumina).

FastQC: La calidad de las lecturas se comprobó utilizando FastQC (Andrews, 2010), una herramienta que tiene como objetivo proporcionar una forma sencilla de realizar algunas comprobaciones de control de calidad como los valores de calidad, contenido GC, longitud y distribución de las lecturas, etc. en los datos de secuencia sin procesar que provienen de procesos de secuenciación de alto rendimiento.

### **Preprocesamiento de lecturas de secuenciación**

El preprocesamiento de los archivos de secuencias consiste en el filtrado y recorte de aquellas lecturas que hayan obtenido valores bajos en el paso anterior. La abundancia y la longitud de lectura de alta calidad son los factores principales que diferencian las lecturas correctas de las erróneas producidas por los instrumentos Illumina. Se ha demostrado que un sesgo en la eliminación de estos errores puede entorpecer los resultados (Bokulich et al., 2013), y con esto los análisis ecológicos posteriores.

Trimmomatic: El filtrado y recorte de lecturas se realizó con Trimmomatic (Bolger et al., 2014) la cual es una herramienta de preprocesamiento más flexible y eficiente que maneja correctamente datos paired-end posteriores a una

secuenciación WGS. Se utilizaron calidades de recorte de 15 y 30 con la finalidad de comparar si un filtrado más estricto lograba reconstruir de mejor manera el genoma mitocondrial y eliminar errores.

### **Mapeo del genoma mitocondrial de *Tapirus bairdii***

Para obtener únicamente en material genético mitocondrial del tapir se realizaron mapeos mediante el programa Bowtie2 (Langmead y Salzberg 2012).

El proceso del mapeo de un genoma consiste en comparar todas las lecturas ingresadas entre sí y calcular cuáles se superponen para determinar el punto de origen probable de la lectura con respecto a un genoma de referencia. En este caso se utilizó la secuencia del mitogenoma de la especie hermana de *T. bairdii*, *T. terrestris*, cuyos datos se encuentran disponibles públicamente en la base de datos de NCBI con número de referencia NC\_053962.1 (Nucleotide). Posteriormente en el 2022 con la publicación del genoma de un individuo en cautiverio en NCBI con número de referencia OM935749.1 (Ennis et al., 2022) se ejecutó un nuevo mapeo empleando este genoma.

En el mapeo se utiliza una técnica de indexación basada en oligómeros cortos que ayudan a identificar a las lecturas y reconocer a las que probablemente se superpongan, con esto es más fácil generar una lista de ubicaciones de alineación candidatas (Langmead, 2012), esta se extraen en un nuevo archivo para luego realizar un alineamiento en cualquier par de lecturas que tengan una interacción significativa (Pop, 2004).

Bowtie 2: Este proceso se realizó empleando la herramienta de análisis de secuencias Bowtie2 (Langmead y Salzberg 2012). Bowtie2, que es un algoritmo de alineamiento muy eficiente basado en la transformación de Burrows-Wheeler (Langmead, 2012).

### **Ensamble del genoma**

Después del paso de superposición se debe realizar el diseño, aquí se posicionaron las lecturas con precisión entre sí, produciendo una alineación múltiple de todas las lecturas. Esta alineación múltiple se utiliza para producir el consenso, es decir, la secuencia de ADN final. Algunos algoritmos también incluyen un paso adicional en el que los *contigs* se unen utilizando la información de enlace de las lecturas de extremos emparejados, creando estructuras más

grandes conocidas como scaffolds (Pop,2004), sin embargo, debido al tamaño del mitogenoma y la selección de programa para el ensamble no fue necesario el scaffolding.

SPAdes: El ensamble se realizó con el algoritmo de ensamblaje SPAdes (Bankevich et al., 2012) que es una herramienta de ensamblaje basada en la construcción de gráficos de Brujin. Es una herramienta reconocida como una de las mejores para lecturas Illumina, mantiene buena calidad de resultados, precisos y métodos de corrección de errores (González de la Fuente, 2018).

Por defecto, el programa ensambla utilizando valores de k-meros de longitudes 21, 33 y 55, además elige el ensamblaje con la mejor puntuación N50. Debido a que nuestras lecturas tenían una longitud de 151 bases, para evitar que el máximo valor predeterminado fuera muy pequeño se ejecutó SPAdes con un rango ampliado de k-meros (Anexos). Se realizaron tres ensayos con distintos valores con la finalidad de compararlos y seleccionar el más adecuado.

### **Selección de ensamble**

Para seleccionar los valores de k-meros que se utilizarán en el ensamble, los cuales se utilizarán para el resto de los análisis genómicos se realizó una evaluación de cada set de valores (**ANEXOS**)

Quast: Se utilizó la herramienta Quast (Gurevich et al., 2013) que evalúa la calidad de los ensamblajes y selecciona el mejor ensamblaje de todos los proporcionados.

### **Eliminación duplicados**

Es posible que dentro del archivo de nuestra secuencia ya ensamblada podamos encontrar regiones duplicadas, para eliminar estas secuencias tendremos que visualizarlas; existen programas que identifican y eliminan duplicados automáticamente, sin embargo, debido al tamaño de nuestro archivo optamos por realizarlo manualmente.

BLAST: Con la herramienta BLAST (Madden, 2013) logramos visualizar los nodos obtenidos y compararlos contra el mitogenoma ya disponible de *T. bairdii*. De esta

manera es fácil identificar aquellas regiones que no necesitaremos en el archivo final.

Con la finalidad de mejorar el manejo de la secuencia obtenida corregimos la dirección de esta misma, para esto se identifica la dirección de nuestro genoma, es decir, observar si es una sola cadena con dirección 5' 3' (que se verá representado con un +/+) o tenemos fragmentos inversos (representado con +/-).

Reverse Complement: Con la herramienta Reverse Complement (Stothard, 2000) facilitamos la corrección de la dirección de nuestra secuencia.

### **Búsqueda de secuencia de la región control**

Posterior a los pasos previos, comparando nuestro resultado con los datos de mitogenomas disponibles en NCBI (Nucleotide), se observó que la región correspondiente al D-loop, lo que conocemos como la región control o no codificante del genoma mitocondrial no se había recuperado en su totalidad. Anteriormente Ennis y colaboradores, 2022 lograron reconstruir la RC con una longitud de 1,247pb, sin embargo, nuestras herramientas solo lograron recuperar con el método de mapeo la mitad de la longitud esperada.

Debido a esto, se realizó nuevamente un ensamble con SPAdes incluyendo la secuencia *paired-end* de *T.bairdii* del *Sequence Read Archive* SRA de NCBI “SRR17086167” y posteriormente se volvió a mapear únicamente la secuencia correspondiente a la RC contra nuestros datos de WGS. Finalmente se logró reconstruir el mitogenoma completo incluyendo las secuencias no codificantes de proteínas.

### **Anotación**

La anotación del genoma mitocondrial del tapir *T. bairdi* se realizó mediante el programa Mitos Web Server (Bernt, 2013). Recuperamos el reverso complementario de la secuencia para obtener una cadena con dirección +. La visualización de los genomas se realizó con MAUVE (Darling et al., 2004)

Como resultado del paso anterior se generó un archivo final del mitogenoma completo de *T.bairdii*, con esto es posible realizar la anotación. El objetivo

principal de la anotación a nivel de proteína es asignar una función a los productos del genoma (Stein, 2001).

MITOS: Se utilizó Mito Web Server (Bernt, 2013) el cual es un programa que realiza anotaciones de genes codificadores de proteínas, ARNr y ARNt. de mitogenomas.

BLAST: Adicionalmente se realizaron alineaciones pareadas de los 13 genes codificantes para proteínas en BLAST (Madden, 2013) del genoma en cuestión y la anotación del genoma de *T. bairdii* (Número de referencia: OM935749.1) con la finalidad de comprobar y completar la identidad de los genes mitocondriales obtenidos previamente en MITOS (Bernt, 2013).

Proksee: Se utilizó la plataforma de Proksee (Grant et al., 2023), el cual es un sistema experto para el ensamblaje, anotación y visualización de genomas.

### **Cálculo de distancia genoma a genoma**

Se obtuvieron distancias de DDH (Hibridación ADN-ADN) empleando dos métodos. El primero incluye sólo aquellos pares de segmentos de alta puntuación (identidades/HSP), se realiza este recorte de valores con el propósito de que este proporcione resultados más precisos (Auch, et al., 2010). El segundo método incluye los caracteres totales de las secuencias (identidades/ longitud total).

GGDC: *Genome-to-Genome Distance Calculator* (Meier-Kolthoff et al., 2022) el programa infiere distancias de genoma a genoma entre pares de genomas secuenciados total o parcialmente, es un estimador digital y altamente confiable de la relación de los genomas. En la actualidad sigue siendo una de las herramientas en línea más populares para el cálculo de valores de DDH y es empleado para la delimitación de (sub)especies in silico en especies procarióticas. Debido a las características del ADN mitocondrial esta herramienta fue empleada.

### **Análisis filogenómicos**

Para realizar el análisis filogenómico de los datos de mitogenomas disponibles del género *Tapirus* se incluyeron tres grupos externos, *Rhinoceros unicornis* (Número de referencia: X97336.1) perteneciente a la familia Rhinocerotidae Gray, 1821, *Equus caballus* (Número de referencia: X79547.1), y *Equus zebra* (Número de referencia: JX312724.1), ambos pertenecientes a la familia Equidae Gray, 1821.

Todos nuestros grupos externos forman parte del orden Perissodactyla Owen, 1848. También, se incluyeron todos los datos de genomas mitocondriales completos disponibles en el NCBI (Nucleotide). Finalmente para el análisis filogenómico se excluyeron aquellas secuencias idénticas entre sí (**Tabla 1**).

Se realizó un árbol filogenético de máxima verosimilitud. También, se evaluó el modelo de sustitución que se ajustara a todo el mitogenoma de *T. bairdii*.

Con la finalidad de corroborar la topología del árbol filogenómico obtenido, se realizaron 2 árboles filogenéticos adicionales. Se concatenaron los 13 genes codificantes de proteínas para el primer árbol filogenético, mientras que en el segundo, únicamente se concatenaron ambos ARNr y COX1.

ORPA: Se utilizó el programa ORPA (Bi et al., 2023) que realiza una reconstrucción filogenómica específica de organelos celulares con comandos amigables incluso para usuarios no bioinformáticos. En nuestro caso utilizamos únicamente el alineamiento obtenido por el mismo programa ya que este emplea Gblocks (Castresana, 2000) que elimina posiciones mal alineadas y regiones divergentes.

PhyML: Con la finalidad de reconstruir los árboles filogenómico y filogenéticos, se utilizó PhyML (Guindon et al., 2010) el cual es un software de filogenias basado en el principio de máxima verosimilitud

## **VIII. RESULTADOS**

Se obtuvieron dos muestras, ambas utilizadas en el proyecto. Estas son las únicas que se han colectado de individuos silvestres con fines de análisis genómicos para la especie.

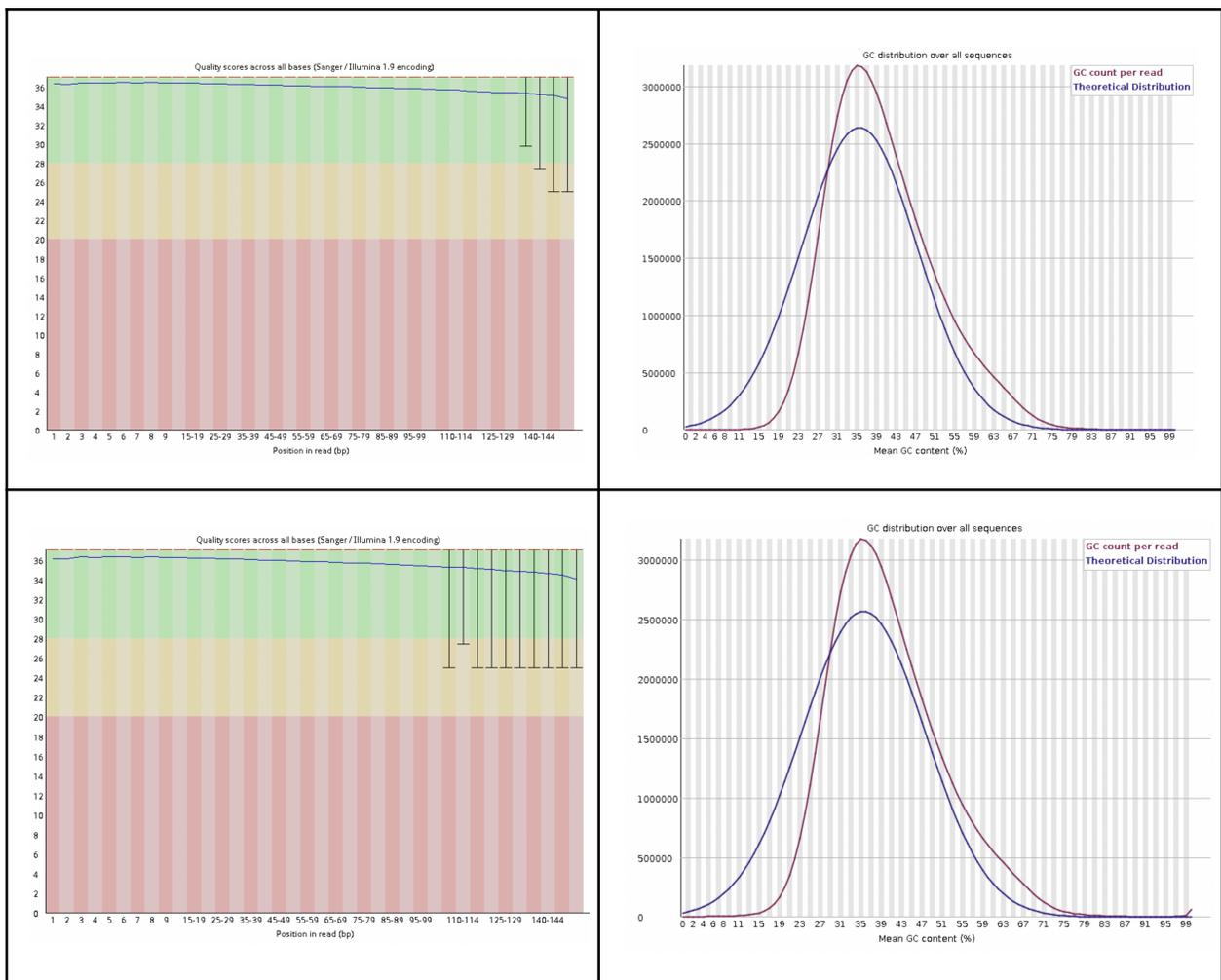
### **Secuenciación de ADN, evaluación y filtrado de calidad.**

A partir de los datos de secuenciación se obtuvieron 23.6 gigabases de lecturas pareadas con una calidad promedio de 36 (Phred) y un contenido de GC 39.75% (**Tabla 2.**)

**Tabla 2.** Datos crudos obtenidos de la secuenciación con Illumina.

<b>Muestra</b>	<b>Número de bases secuenciadas (pb)</b>	<b>Número total de lecturas (reverse y forward)</b>	<b>CG (%)</b>	<b>AT/%)</b>	<b>Q20(%)</b>	<b>Q30(%)</b>
<i>Tapirus bairdii</i>	23,607,169,672	156,338,872	39.75	60.25	97.33	92.5

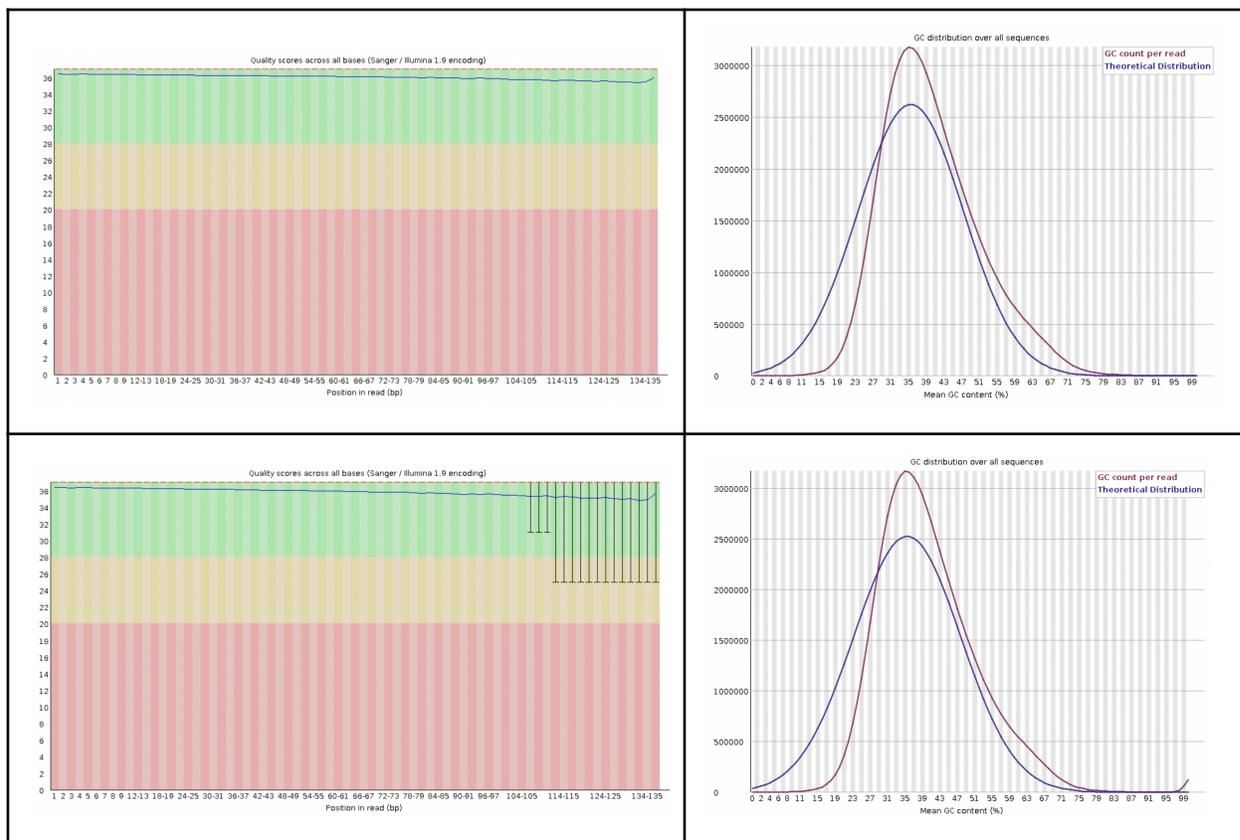
Se generaron dos reportes de calidad FastQC (Andrews, 2010) , uno para el archivo TAP\_1 y TAP\_2 (**Figura 3**)



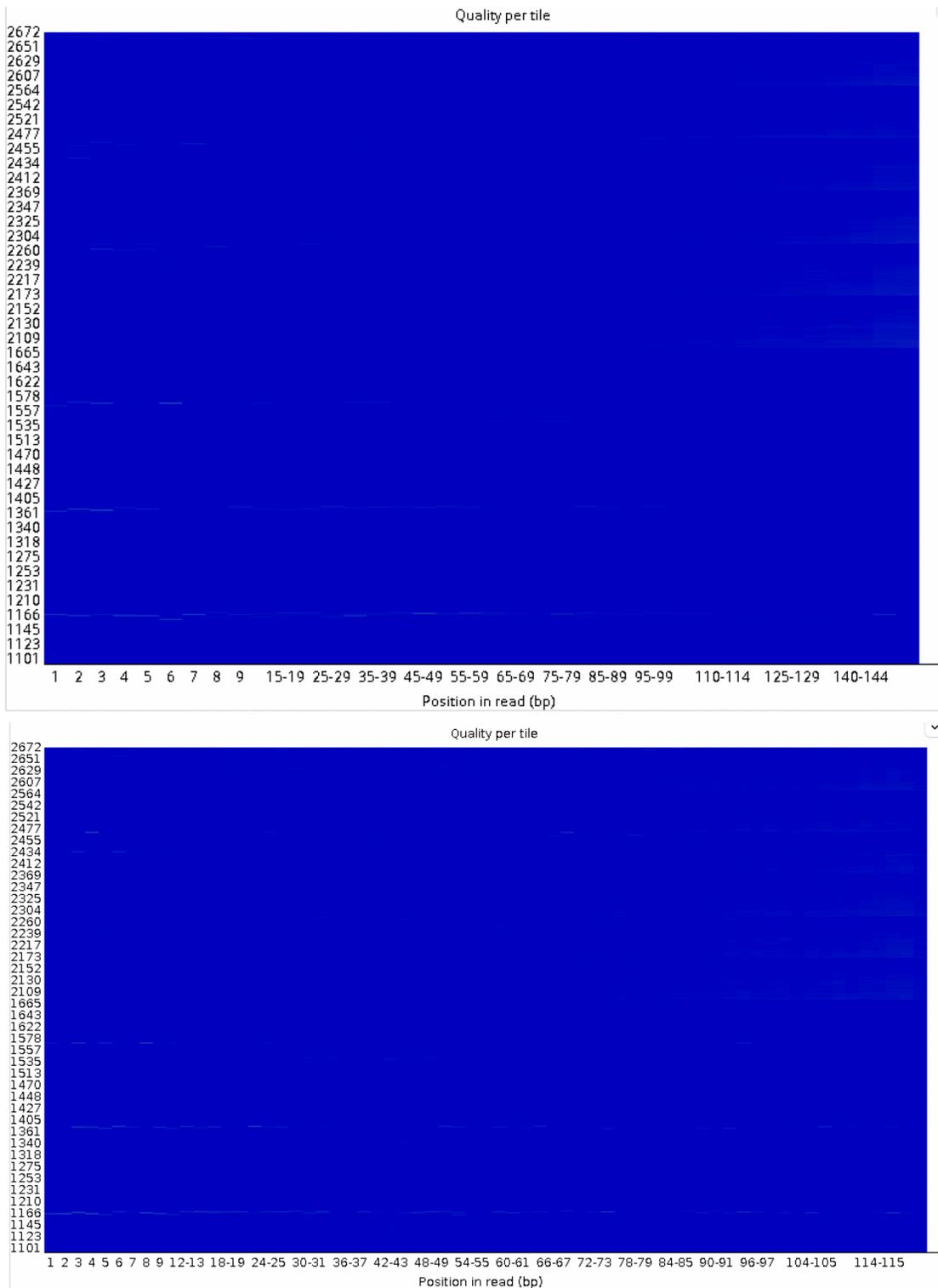
**Figura 3.** Gráficos de puntaje de calidad y contenido de GC a lo largo de todas las bases generados por FastQC (Andrews, 2010) para arriba TAP\_1 y abajo TAP\_2. Los gráficos de calidad muestran el rango de valores de calidad en cada ciclo realizado por el secuenciador. El eje X y el eje Y son, respectivamente, la posición de la lectura por pares de bases y la puntuación de valores de de Phred.

Posterior al filtrado de nuestras secuencias pudimos obtener dos archivos derivados de los archivos crudos llamados TAP\_1\_Tp y TAP\_2\_Tp, posteriormente evaluamos la condición general de las lecturas con gráficos de calidad y contenido de GC con (**Figura 4**) y la calidad de secuencia por mosaico de la secuencia forward (**Figura 5**) con FastQC, (Andrews, 2010).

Para ambas cadenas se obtuvo una secuencia de 76,870,878 lecturas pareadas con un tamaño de 21-136 pb y se identificaron 0 secuencias marcadas de baja calidad. Las lecturas filtradas tienen un contenido de GC de 39% (**Tabla 2**).



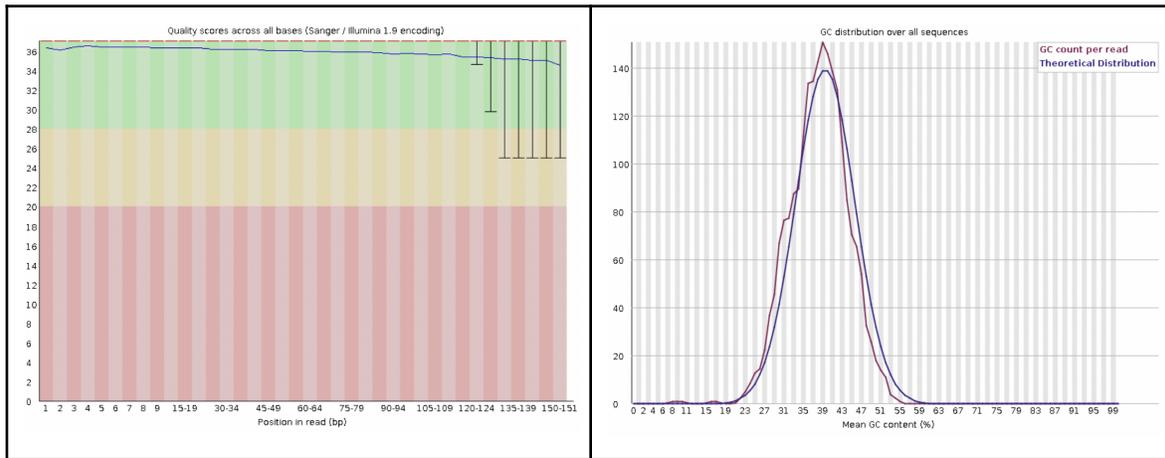
**Figura 4.** Gráficos de puntaje de calidad y contenido de GC a lo largo de todas las bases generado por FastQC (Andrews, 2010) para arriba TAP\_1\_Tp y abajo TAP\_2\_Tp.



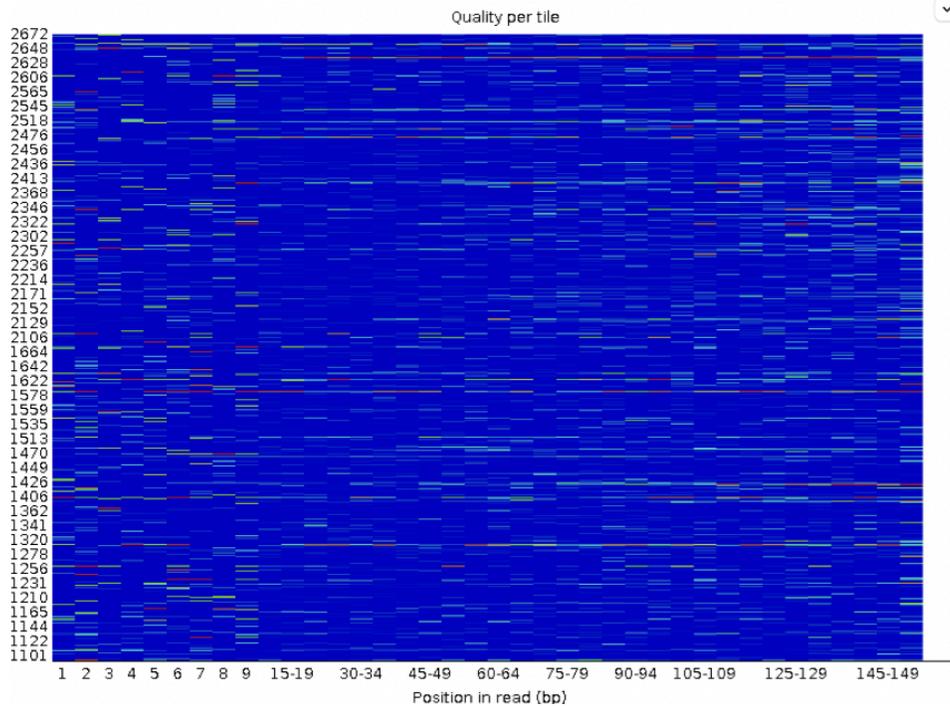
**Figura 5.** Gráfico de calidad por secuencia de mosaicos generados por FastQC (Andrews, 2010) para la secuencia forward con valores de recorte de arriba 15 y abajo 30.

## Ensamble y selección de valores

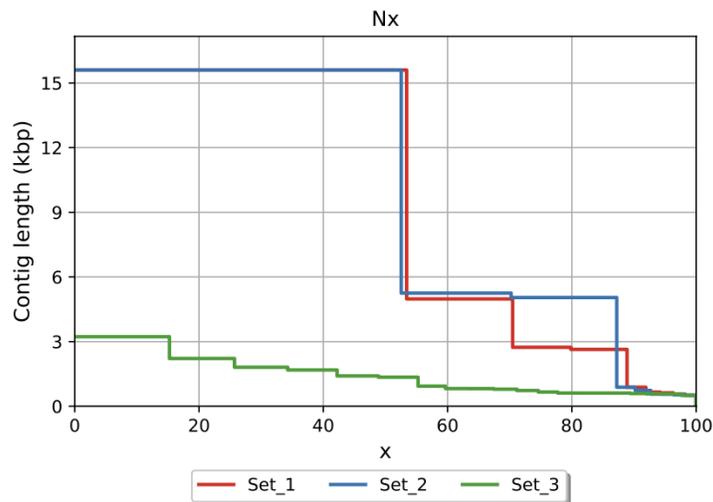
Se obtuvieron ensamblajes con el programa SPAdes (Bankevich et al., 2012) de todos los archivos mapeados con Bowtie2 (Langmead y Salzberg 2012), únicamente el primer mapeo ejecutado que se repitió tres veces con distintos valores fue analizado con Quast (Gurevich et al., 2013). El segundo y tercero mostraron consistencia en los resultados utilizando los valores de k-meros predeterminados por el programa (**Tabla 3**). Se obtuvieron gráficos de puntaje de calidad y contenido de GC (**Figura 6**) y un gráfico de calidad por secuencia de mosaicos (**Figura 7**) con FastQC (Andrews, 2010).



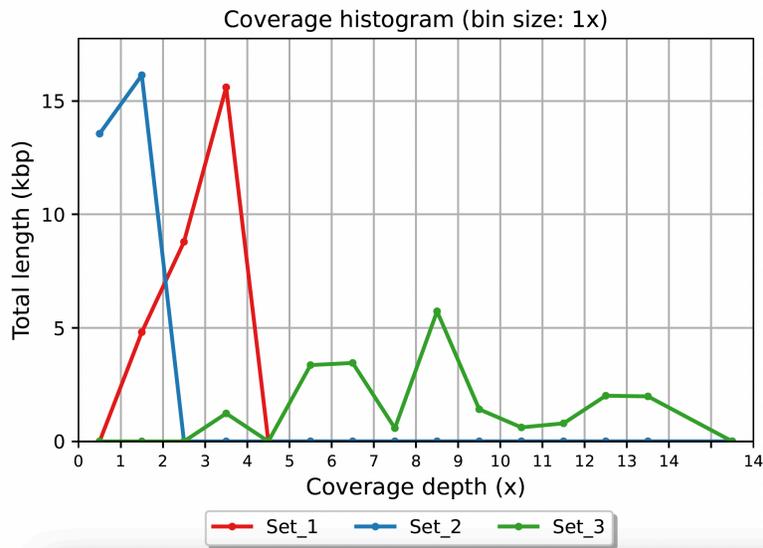
**Figura 6.** Gráficos de puntajes de calidad y contenido de GC generados por FastQC (Andrews, 2010) para la secuencia *forward* y *reverse* concatenadas posterior al mapeo con Bowtie2 (Langmead y Salzberg 2012).



**Figura 7.** Gráficos de calidad por secuencia de mosaicos generados por FastQC (Andrews, 2010) para la secuencia *forward* y *reverse* concatenadas posterior al mapeo con Bowtie2 (Langmead y Salzberg 2012).



**Figura 8.** Gráfico de longitud de contigs por set de valores de k-meros obtenido con Quast (Gurevich et al., 2013). Cada color representa cada una de las 3 configuraciones aplicadas para el ensamblaje del genoma.



**Figura 9.** Gráfico de profundidad de cobertura por set de valores de k-meros obtenido con Quast (Gurevich et al., 2013). Cada color representa cada una de las 3 configuraciones aplicadas para el ensamble del genoma.

**Tabla 3.** Resultados de análisis de ensambles con Quast (Gurevich et al., 2013) para el primer set de datos de k meros. En azul se indica el ensamble de mejor calidad y en rojo el más bajo.

Statistics without reference	Set_1	Set_2	Set_3
# contigs	9	9	20
# contigs (>= 0 bp)	45	38	82
# contigs (>= 1000 bp)	4	3	6
# contigs (>= 5000 bp)	1	3	0
# contigs (>= 10000 bp)	1	1	0
# contigs (>= 25000 bp)	0	0	0
# contigs (>= 50000 bp)	0	0	0
Largest contig	15 601	15 601	3228
Total length	29 205	29 692	21 176
Total length (>= 0 bp)	40 547	39 564	35 871
Total length (>= 1000 bp)	25 958	25 899	11 700
Total length (>= 5000 bp)	15 601	25 899	0
Total length (>= 10000 bp)	15 601	15 601	0
Total length (>= 25000 bp)	0	0	0
Total length (>= 50000 bp)	0	0	0
N50	15 601	15 601	1350
N90	884	884	592
auN	9753.4	10 068	1505
L50	1	1	6
L90	5	4	17
GC (%)	38.55	38.6	38.51
<b>Mismatches</b>			
# N's per 100 kbp	0	0	0
# N's	0	0	0

En este caso, para reconstruir la secuencia no codificante RC se tuvieron que realizar al menos tres mapeos, cada uno más específico que el anterior. El ensamble por SPAdes de los datos de *T. bairdii* disponibles en el SRA demostró ser la mejor opción para reconstruir el genoma de una manera más completa (16,567), la segunda mejor opción de ensamble fue aquella que se realizó con el genoma de referencia de la misma especie (15,521) y finalmente la menos completa fue el ensamble incluyendo como referencia a *T. terrestris* (15,855), la especie más genéticamente emparentada a nuestro modelo de estudio.

### Genoma mitocondrial de *Tapirus bairdii*

Se lograron obtener múltiples archivos con los cuales fue finalmente posible obtener el mitogenoma del ejemplar silvestre de *T. bairdii* utilizando solo el 0.7% de las lecturas totales obtenidas con el secuenciador NovaSeq de Illumina. El mitogenoma obtenido consiste de un solo *contig* con una longitud de 16567 bases, y un contenido de GC de 38.4% (**Tabla. 4**), A(34.15% 5658) | C(25.77% 4269) | G(12.63% 2092) | T(27.45% 4548).

**Tabla 4.** Tabla de número de secuencias, bases y longitud de secuencia para cada archivo obtenido durante la realización del proyecto. Se muestran resultados obtenidos antes, durante y post mapeo. Datos obtenidos con FastQC (Andrews, 2010).

FastQC	Número de secuencias	Número de bases	Longitud de secuencia	GC(%)
Datos crudos cadena <b>forward</b> (TAP_1)	78,169,436	11.8 Gbp	151 bp	39
Datos crudos cadena <b>reverse</b> (TAP_2)	78,169,436	11.8 Gbp	151 bp	39
Post trimming cadena <b>forward</b> recorte en 30	76,870,878	9.1 Gbp	6-121 bp	39
Post trimming cadena <b>reverse</b> recorte en 30	76,870,878	9.1 Gbp	6-121 bp	39

Post trimming cadena <b>forward</b> recorte en 15 (TAP_1_Tp)	76,870,878	10.3 Gbp	21-136 bp	39
Post trimming cadena <b>reverse</b> recorte en 15 (TAP_2_Tp)	76,870,878	10.3 Gbp	151 bp	39
Post mapeo cadena <b>forward</b> con <b>T. terrestris</b> como referencia	2,068	279 Kbp	23-136 bp	37
Post mapeo cadena <b>reverse</b> con <b>T. terrestris</b> como referencia	2,068	278.5 Kbp	22-136 bp	38
Post mapeo cadena <b>forward</b> con <b>T. bairdii</b> como referencia	2,121	286.5 Kbp	23-136 bp	37
Post mapeo cadena <b>reverse</b> con <b>T. bairdii</b> como referencia	2121	285.4 Kbp	23-136 bp	38
Post ensamble con SPAdes, mapeo con <b>T.terrestris</b>	38	40,183 bp	121-15,861 bp	38.7
Post ensamble con SPAdes, mapeo con <b>T. bairdii</b>	52	43,421 bp	101-12,247 bp	38.5
Primera selección de <i>contigs</i> o nodos post ensamble	3	17,157 bp	542, 750 y 15,861 bp	38.4
Ensamble final	1	16,567 bp	16,567 bp	38.4

## Porcentaje de identidad

El resultado de las **Tablas 5 y 6** muestran los porcentajes de identidad con valores de DDH para los dos genotipos mitocondriales actuales de *T. bairdii* en México.

**Tabla 5.** Valores de DDH (Hibridación ADN-ADN) empleando la fórmula identidades/HSP (pares de segmentos de alta puntuación), obtenido con GGDC (Meier-Kolthoff et al., 2022)

	<i>T. indicus</i> (CM063 009.1)	<i>T. indicus</i> (CM063 009.1)	<i>T. indicus</i> (CM063 009.1)	<i>T. indicus</i> (CM063 09.1)	<i>T. terrestris</i> (MW2572 04.1)	<i>T. terrestris</i> (NC_0539 62.1)	<i>T. bairdii</i> (NC_063 943.1)	<i>T. bairdii</i> (OM9357 49)	<i>T. bairdii</i> (OQ4204 28.1)	<i>T. bairdii</i> Calakmul
<i>T. bairdii</i> (OM935749 .1)	36.5	36.9	36.9	37	86.6	86.6	100	100	88.8	48.2
<i>T. bairdii</i> , Calakmul	37.9	38.6	38.6	38.6	47.5	47.5	48.2	48.2	48.2	100

**Tabla 6.** Valores de DDH (Hibridación ADN-ADN) empleando la fórmula identidades/longitud total, obtenido con GGDC (Meier-Kolthoff et al., 2022)

	<i>T. indicus</i> (CM063 009.1)	<i>T. terrestris</i> (MW2572 04.1)	<i>T. terrestris</i> (NC_053 962.1)	<i>T. bairdii</i> (NC_063 943.1)	<i>T. bairdii</i> (OM9357 49)	<i>T. bairdii</i> (OQ4204 28.1)	<i>T. bairdii</i> Calakmul			
<i>T. bairdii</i> (OM935749 .1)	89.6	89.8	89.8	89.8	99.7	99.7	100	100	99.8	94.6
<i>T. bairdii</i> , Calakmul	87.3	87.7	87.7	87.7	94.2	94.2	94.6	94.6	94.5	100

## Anotación del genoma mitocondrial de *Tapirus bairdii*

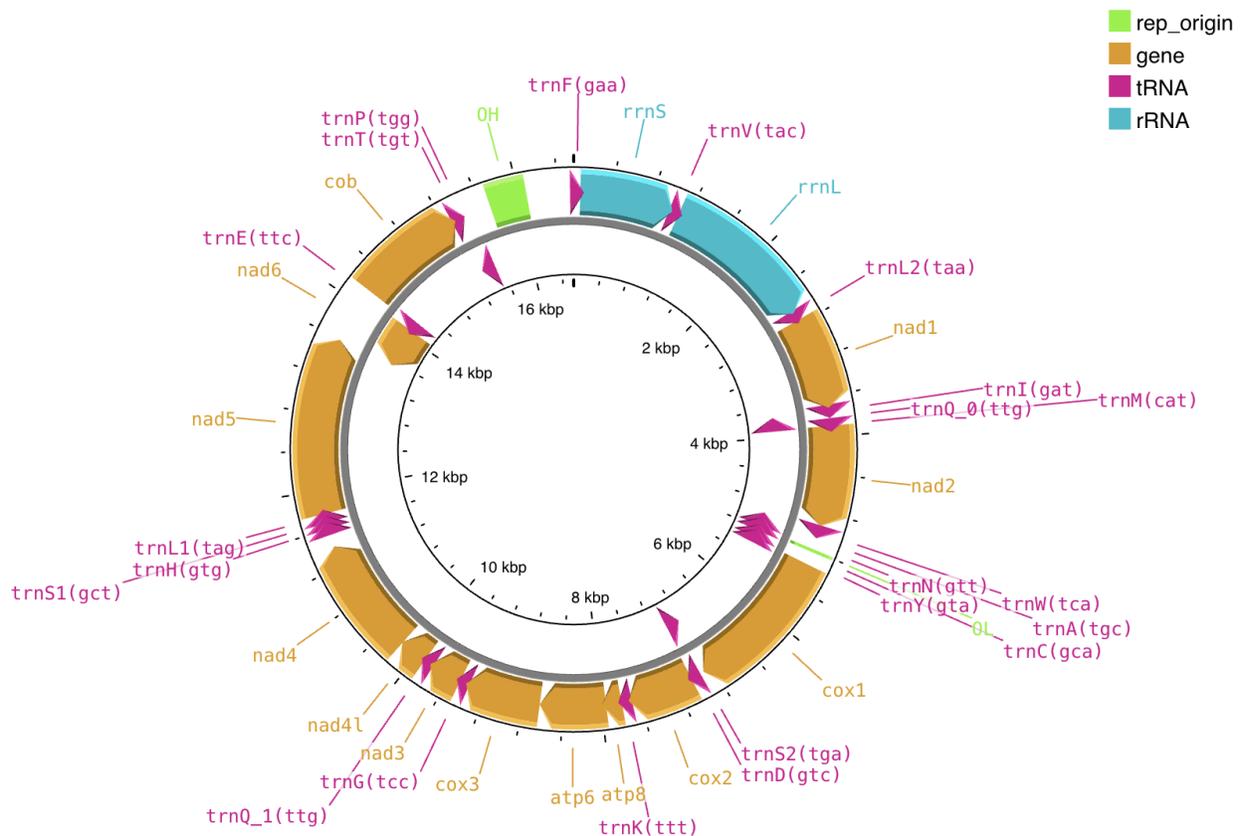
Se lograron identificar las secuencias que corresponden a los 13 genes codificantes de proteínas, 22 ARN de transferencia y 2 ARN ribosomales del mitogenoma completo de *T. bairdii* (Tabla 7 y Figura 11)

**Tabla 7.** Resultados de la anotación de mitogenoma de *T. bairdii* con MITOS (Bernt, 2013).

Gen	Identificador	Tipo	Posición	Dirección
RNA de transferencia trnF(gaa)	trnF(gaa)	ARNt	1-68	+
ARN ribosomal 12s	rrnS	ARNr	69-1036	+
RNA de transferencia trnV(tac)	trnV(tac)	ARNt	1036-1102	+
ARN ribosomal 18s	rrnL	ARNr	1103-2681	+
RNA de transferencia trnL2(taa)	trnL2(taa)	ARNt	2682-2756	+
Subunidad 1 de NADH deshidrogenasa	nad1	Codificante a proteína	2758-3714	+
RNA de transferencia trnQ_0(ttg)	trnQ_0(ttg)	ARNt	3782-3854	-
RNA de transferencia trnM(cat)	trnM(cat)	ARNt	3857-3925	+
Subunidad 2 de NADH deshidrogenasa	nad2	Codificante a proteína	3925-4979	+
RNA de transferencia trnW(tca)	trnW(tca)	ARNt	4968-5036	+

RNA de transferencia trnA(tgc)	trnA(tgc)	ARNt	5042-5110	-
RNA de transferencia trnN(gtt)	trnN(gtt)	ARNt	5112-5184	-
RNA de transferencia trnC(gca)	trnC(gca)	ARNt	5218-5283	-
RNA de transferencia trnY(gta)	trnY(gta)	ARNt	5284-5349	-
Subunidad 1 del citocromo oxidasa	cox1	Codificante a proteína	5350-6895	+
RNA de transferencia trnS2(tga)	trnS2(tga)	ARNt	6893-6961	-
RNA de transferencia trnD(gtc)	trnD(gtc)	ARNt	6970-7036	+
Subunidad 2 del citocromo oxidasa	cox2	Codificante a proteína	7036-7720	+
RNA de transferencia trnK(ttt)	trnK(ttt)	ARNt	7724-7790	+
Subunidad 8 del ATP sintasa	atp8	Codificante a proteína	7791-7995	+
Subunidad 6 del ATP sintasa	atp6	Codificante a proteína	7952-8633	+
Subunidad 3 del citocromo oxidasa	cox3	Codificante a proteína	8632-9416	+
RNA de transferencia trnG(tcc)	trnG(tcc)	ARNt	9417-9484	+
Subunidad 3 de NADH deshidrogenasa	nad3	Codificante a proteína	9485-9831	+
RNA de transferencia trnQ_1(ttg)	trnQ_1(ttg)	ARNt	9833-9900	+

Subunidad 4l de NADH deshidrogenasa	nad4l	Codificante a proteína	9900-10197	+
Subunidad 4 de NADH deshidrogenasa	nad4	Codificante a proteína	10190-11568	+
RNA de transferencia trnH(gtg)	trnH(gtg)	ARNt	11569-11637	+
RNA de transferencia trnS1(gct)	trnS1(gct)	ARNt	11638-11696	+
Subunidad 5 de NADH deshidrogenasa	nad5	Codificante a proteína	11766-13587	+
Subunidad 6 de NADH deshidrogenasa	nad6	Codificante a proteína	13589-14095	-
RNA de transferencia trnE(ttc)	trnE(ttc)	ARNt	14099-14167	-
Citocromo b	cob	Codificante a proteína	14172-15312	+
RNA de transferencia trnT(tgt)	trnT(tgt)	ARNt	15313-15380	+
RNA de transferencia trnP(tgg)	trnP(tgg)	ARNt	15381-15446	-

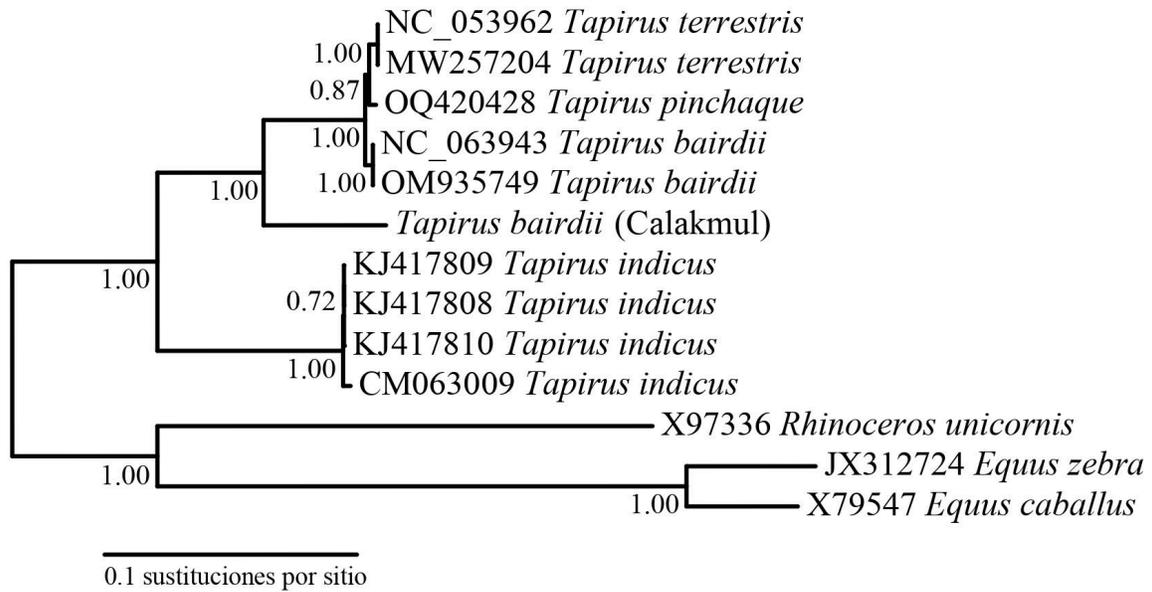


**Figura 11.** Mapa del mitogenoma recuperado en este trabajo para *T. bairdii*, obtenido con Proksee (Grant et al., 2023)

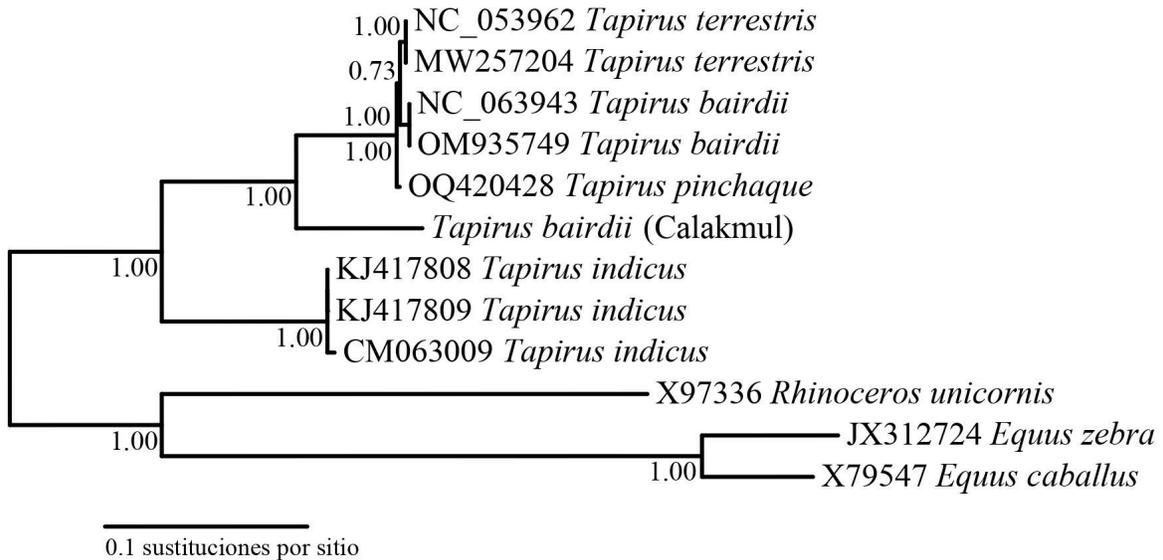
## Alineamiento y filogenómica

Se obtuvo la reconstrucción del árbol filogenómico empleando los genomas mitocondriales completos utilizando el método de máxima verosimilitud (**Figura 12**) Adicionalmente, se realizaron pruebas filogenéticas empleando los 13 genes codificantes de proteínas (**Figura 13**), ambos ARNr y el COX1 concatenados (**Figura 14**).

El programa ORPA (Bi et al., 2023) seleccionó como mejor modelo de sustitución de acuerdo al criterio de inferencia bayesiana (BIC) el modelo TIM2+F+G4 para el filogenoma y la filogenia de los 13 genes codificantes concatenados. En cuanto al análisis filogenético empleando ambos ARNr y el COX1 se seleccionó el modelo TIM2+F+R2 de acuerdo al criterio de inferencia bayesiana (BIC) .

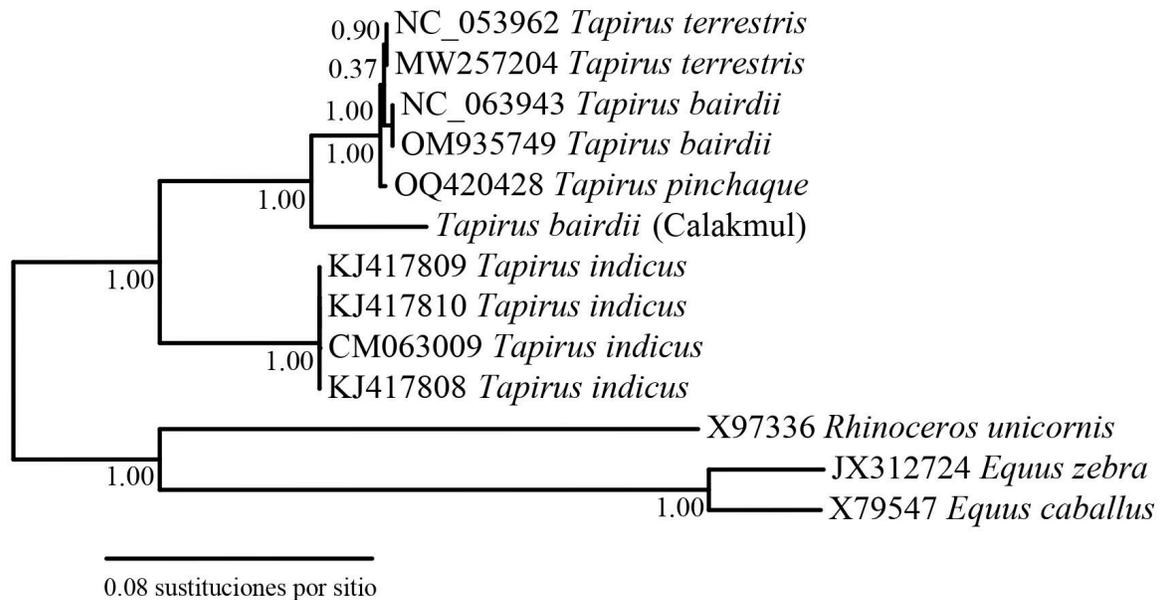


**Figura 12.** Análisis filogenómico de los genomas mitocondriales del género *Tapirus*. El alineamiento incluye secuencias de 13 organismos y tiene una longitud de 14,880 caracteres. El árbol se construyó mediante el método de máxima verosimilitud. Realizado con ORPA (Bi et al., 2023) y PhyML (Guindon et al., 2010).



**Figura 13.** Análisis filogenético de los 13 genes codificantes de proteínas pertenecientes al genoma mitocondrial del género *Tapirus*. El alineamiento incluye secuencias de 13 organismos y tiene una longitud de 8,231 caracteres. El árbol se

construyó mediante el método de máxima verosimilitud. Realizado con ORPA (Bi et al., 2023) y PhyML (Guindon et al., 2010).



**Figura 14.** Análisis filogenético de los dos genes de ARN ribosomales (mtARNs) y la Subunidad 1 del citocromo c oxidasa (Cox1) concatenados del género *Tapirus*. El alineamiento incluye secuencias de 13 organismos y tiene una longitud de 4,128 caracteres. El árbol se construyó mediante el método de máxima verosimilitud. Realizado con ORPA (Bi et al., 2023) y PhyML (Guindon et al., 2010).

## IX. DISCUSIÓN

El resultado de la secuenciación, ensamblaje y anotación de un genoma es prácticamente el mapa genético completo de un individuo (genes heredados). Con él podemos llegar a inferir comportamientos y rasgos fenotípicos, sin embargo, para tener una mejor comprensión de los procesos biológicos es necesario emplear principios del campo de la genómica comparada, la cual propone que múltiples características comunes de dos o más organismos están codificadas dentro del ADN que se conserva entre las especies (Hardison, 2003).

Se espera que las secuencias de ADN codificantes para proteínas, los ARN responsables de funciones y aquellas regiones no codificantes reguladoras que se conservaron del último ancestro común, también se conserven en secuencias genómicas contemporáneas. Generar un alineamiento de genomas es la parte

fundamental de la genómica comparativa, además el poder identificar y representar estas diferencias con distancias filogenéticas permite obtener información valiosa sobre la evolución de diferentes taxa (Hardison, 2003).

Es importante reconocer que la recuperación del genoma mitocondrial de un individuo en cautiverio por Ennis y colaboradores en el 2022 (Ennis et al., 2022) facilitó en gran medida el proceso de mapeo y ensamblaje en nuestro proyecto, ya que antes de la aparición de este nuevo mitogenoma de *T. bairdii*, nuestra secuencia de referencia inicial era la *T. terrestris*, la segunda especie con mayor parentesco genético y mitogenoma disponible en las bases de datos públicas .

Posteriormente en el 2023 se liberó en el NCBI el mitogenoma completo de la especie hermana de *T. bairdii*, *T. pinchaque* por Gutierrez, Ortega y Baeza (Nucleotide), la cual fue integrada a los análisis filogenómicos del proyecto.

Esta situación demostró claramente que la disponibilidad de secuencias genómicas es de suma importancia en los estudios de genómica comparativa. En nuestro caso, la aparición de nuevos genomas facilitó el mapeo del genoma mitocondrial.

### **Filtros de calidad**

Los filtros de calidad aparentemente son buenos antes del mapeo con bowtie. Como se observa en la sección de resultados, en la **Figura 3**, la parte superior que corresponde al archivo de TAP\_1 con dirección forward, en su mayoría, tiene un margen de error de bases menor a 1 en 1000, excepto por la región a partir de la posición 140 a la 151 en donde incrementa el error a 1 en 100. En la parte inferior se observa que el archivo de TAP\_2 con dirección reverse también tiene un margen de error de base menor a 1 en 1000, a partir de la posición 110 a la 151 incrementa el error a 1 en 100.

Estos valores de calidad cambian en ambas cadenas después de ejecutar el filtrado con Trimmomatic se observa que el archivo TAP\_1\_Tp mantiene en su totalidad un margen de error de 1 en 1000 bases, mientras que para TAP\_2\_Tp el rango con margen de error de 1 en 100 se reduce 4 posiciones, de la 114-151 (**Figura 4**).

Posterior al mapeo y ensamblaje de los archivos filtrados se obtuvieron 10 *contigs* menos que los que no se filtraron y no se recortaron (38 y 48 *contigs* respectivamente), por lo que el resto del proyecto se realizó con el archivo de los datos filtrados. Adicionalmente se realizaron dos mapeos más para recuperar

aquellas regiones de la secuencia no detectadas en el primer mapeo (principalmente de la RC) siguiendo la misma línea de comando.

En la **Figura 5** donde se comparan los resultados del quality per tile (gráfico que permite observar los puntajes de calidad de cada mosaico en todas sus bases para ver si hubo una pérdida de calidad asociada con solo una parte de la celda de flujo. Los colores más cálidos (rojo) indican que las lecturas tienen peores cualidades para esa posición en comparación a las demás) utilizando valores de recorte de 15 y 30 respectivamente, no se observa diferencia en la tonalidad de las celadas. Sin embargo, después del mapeo con bowtie en la **Figura 6**, se aprecia claramente una disminución en las calidades de esta misma lo cual no fue detectado en el gráfico de calidad de la **Figura 7**.

Esto puede ser atribuido a la degradación del ADN molecular identificado durante la secuenciación con el secuenciador NovaSeq de Illumina. Esto hace sentido debido a la condición de almacenamiento de la muestra.

El resultado del Quast (Gurevich et al., 2013) indicó que la ampliación del rango de valores de k-meros representado en el Set\_2 del primer mapeo mejoró el desempeño de este mismo ya que reconstruyó 38 contigs (**Figura 8 y 9**) de los cuales, tres de ellos recuperaron las primeras 17,157pb obtenidas inicialmente (**Cuadro 3**).

Adicionalmente, se realizó una extracción de ADN en laboratorio de la muestra del 2023 y se mandó secuenciar como manera de corroborar la diferencia entre la condición del material genético de ambos individuos, esto como manera de identificar la diferencia entre una muestra que presentó características de ADN antiguo y otra fresca. Debido al tiempo y recursos, en febrero del 2024 se logró obtener únicamente la subunidad 2 de la NADH deshidrogenasa del segundo individuo muestreado en abril de 2023 que corresponde a la obtenida por métodos bioinformáticos (**Anexos**). A pesar de que se presentan pocas diferencias entre ambas muestras de los dos individuos silvestres para este gen, es apresurado llegar a alguna conclusión previa sin reiterar la necesidad de más datos moleculares en estudios futuros.

El análisis filogenómico del género *tapirus* ubicó a la muestra obtenida en el 2011 de *T.bairdii* como una especie externa al clado de los tapires neotropicales. Este resultado inusual puede ser resultado de un proceso de degradación del material genético, específicamente depurinación y desaminación del ADN debido al periodo y condiciones de almacenamiento como se explica en la introducción.

Es importante reconocer que esta degradación no fue detectada en la evaluación de control de calidad como los gráficos de puntaje o gráficos de calidad por secuencia de mosaico ya que aparentemente son calidades adecuadas. Sin embargo, los gráficos de contenido de GC muestran anomalías que corresponden a las características de ADN antiguo (Dabney et al., 2013) evaluadas posterior al ensamble.

## **Genoma mitocondrial**

En la actualidad, se reconoce que varias regiones mitocondriales no recuperan de manera consistente la topología encontrada en análisis filogenéticos del mitogenoma. Incluso utilizando modelos mejorados y eliminando regiones “filogenéticamente ruidosas”, como homoplasias, no se eliminan estas incongruencias. Por esto, se propone que las regiones individuales deben ser utilizadas con extrema precaución (Jeffrey, 1999).

Gracias a las características de los genomas mitocondriales se sabe que es posible analizar a mayor nivel de detalle aspectos evolutivos como la evolución de los reordenamientos genéticos, la regulación genética, los sistemas de procesamiento de mensajes, los mecanismos de replicación, entre otros. Los datos de mitogenomas normalmente suelen ser manejables a un nivel de detalle y comprensión que actualmente no es tan común para los sistemas nucleares (Jeffrey, 1999).

Gracias a la fácil obtención y manipulación de los genomas mitocondriales, sería pertinente implementar su uso. De igual manera, la integración de múltiples regiones mitocondriales concatenadas incrementarían el acervo de información para realizar con mayor facilidad análisis filogenómicos y filogenéticos. También proponemos que es necesario un mayor desarrollo de programas y modelos para maximizar la información filogenómica de los datos mitocondriales en el futuro. De este modo, en futuras investigaciones cuando éstas se combinen con datos nucleares, se podrán obtener mejores estimaciones de los árboles de diversas especies (Meiklejohn et al., 2014).

En nuestro estudio se logró reconstruir la RC del mitogenoma, la cual resultó ser la región con más variación del mitogenoma, al igual que como es típicamente descrita en otros tapires y mamíferos en general ( Muangkram et al., 2016).

Es importante mencionar que gracias a la disponibilidad de programas bioinformáticos fue posible realizar esta búsqueda de regiones mitocondriales a nivel computacional en vez de emplear protocolos de laboratorio ya sea con kits o métodos de extracción específicos para el aislamiento de organelos como la

mitocondria, sin mencionar que estos suelen ser laboriosos y requieren mucho tiempo (Hartwig et al., 2009).

A medida que incrementa el número de usuarios de programas diseñados para proceso de secuenciación y ensamblaje de genomas, estos se ven obligados a volverse más amigables para aquellos científicos sin formación bioinformática como tal. Un ejemplo claro es el programa ORPA (Bi et al., 2023), que específicamente fue diseñado con la finalidad de realizar análisis filogenómicos de material genético mitocondrial con líneas de comando sencillas e intuitivas para la mayoría de los usuarios (Bii et al., 2023).

### **Filogenómica de *Tapirus bairdii***

En la literatura científica se pueden reconocer proyectos en los que se han empleado marcadores moleculares mitocondriales con la finalidad de reconstruir árboles filogenéticos para el género *Tapirus*. Uno de los primeros proyectos fue el realizado por Ashley y colaboradores, 1996, donde utilizaron la secuencia del marcador citocromo-c oxidasa para reconstruir un árbol filogenético por el método de máxima parsimonia. En este trabajo se logró integrar a *T. terrestris* y *T. pinchaque* en un mismo clado hermano a *T. bairdii* para el grupo de los tapires neotropicales (Ashley et al. 1996). Posteriormente, Cozzuol y colaboradores, 2013, emplearon el marcador del citocromo-b oxidasa para reconstruir un árbol empleando el método de inferencia bayesiana. En dicho trabajo se mantiene la topología propuesta en 1996, con *T. bairdii* excluido del clado de *T. terrestris* y *T. pinchaque* (Cozzuol et al., 2013). En el 2016, Ruiz-García y colaboradores lograron reconstruir dos árboles filogenéticos de máxima verosimilitud. El primero empleó las secuencias concatenadas de citocromo-b oxidasa, citocromo oxidasa 1 y 2. El segundo árbol se realizó empleando el método de inferencia bayesiana para la secuencia de 15 marcadores moleculares del genoma mitocondrial concatenados. Ambos árboles resultaron concordantes con los trabajos previos descritos en la literatura. El único árbol en el que se incluyó una muestra de *T. bairdii* mexicano fue el que concatenando las secuencias de 15 genes mitocondriales. En este, el *T. bairdii* de Yucatán se muestra alejado del resto de los de su especie, lo que concuerda con su distribución geográfica (Ruiz-García et al., 2016).

Por otro lado, las topologías de estos árboles corresponden a las previamente propuestas con métodos de reconstrucción morfológica como la de Ferrero y Noriega, 2007 en donde las muestras de *T. bairdii* se agrupan en un solo clado más emparentado con *T. pinchaque* y posteriormente a *T. terrestris* obviando

especies del género extintas. Estos resultados descritos de igual manera concuerdan con la distribución y barreras geográficas entre las especies Neotropicales del género *Tapirus*. (Ruiz-García et al., 2007; Ferrero y Noriega, 2007).

Es importante mencionar que en estudios moleculares previos, la cantidad de datos recolectados para *T. bairdii* son reducidos a comparación de las demás especies muestreadas del mismo género (Ashley et al. 1996; Ruiz-García et al., 2007; Ferrero y Noriega, 2007; Ruiz-García et al., 2016).

Recientemente, el trabajo de Ennis y colaboradores, 2022 consistió en la reconstrucción del primer genoma mitocondrial completo de un *T. bairdii*. En este proyecto se presenta el árbol completo del orden Perissodactyla, empleando el método de máxima verosimilitud para los 13 genes mitocondriales codificantes de proteínas, sin embargo, las únicas 2 muestras de tapires neotropicales (*T. terrestris* y *T. bairdii*) se ubican en un solo clado alejados de los tapires asiáticos *T. terrestris* (Ennis et al., 2022).

A diferencia de lo anteriormente descrito por Ennis y colaboradores, 2022, nuestro estudio con enfoque genómico presenta una topología concordante con el resto de los árboles que han incluido a *T. bairdii* a pesar de que se realizaron pruebas filogenéticas empleando los 13 genes codificantes de proteínas y posteriormente las secuencias concatenadas de los marcadores 12 y 18S ARN ribosomal y el citocromo oxidasa 1 concatenados (**Figuras 12, 13 y 14**).

La consistencia de la reconstrucción topológica de los árboles, con distintos marcadores moleculares, distintos filtros de calidad, y empleando distintos programas, nos indica claramente la enorme necesidad de incrementar los esfuerzos de muestreo, técnicas de conservación de muestras, desarrollo de herramientas bioinformáticas con la finalidad de resolver de mejor manera la historia evolutiva y situación actual de las poblaciones de *T. bairdii* en la RBC, y en el resto de México. La diferencia entre el genoma mitocondrial propuesto por Ennis et al., 2022 y el nuestro podrá ser corroborando secuenciando un nuevo individuo y nuevamente realizando análisis de genómica comparativa.

Por otro lado, proponemos que es importante tomar en cuenta la historia de vida y ecología de los individuos de los cuales se obtienen muestras genómicas. En este caso tenemos información clara de los hábitos de distribución del individuo muestreado durante poco más de cuatro años (Nov 2011- Sep 2015).

Consideramos que sería importante conocer la historia del individuo en cautiverio previamente secuenciado, es decir, tener la información necesaria para poder

responder preguntas como ¿cuántos años lleva ese individuo en cautiverio?, ¿cuántas generaciones antes de él fueron extraídas de su hábitat natural?, ¿existe recombinación genética artificial?, ¿de qué población específica fue extraído ese individuo?, etcétera. Esto con la finalidad de comparar las diferencias genómicas entre dos tapires con hábitos e historias de vida distintas.

### **Sugerencias y perspectivas a futuro**

La muestra utilizada para la secuenciación en este estudio fue una muestra almacenada en condiciones no óptimas durante una década, aún con esta limitante fue posible reconstruir el genoma mitocondrial completo de un tapir centroamericano. Debido a la creciente presión selectiva ejercida hacia los tapires en la actualidad, sale a relucir la importancia de las colecciones científicas como un reservorio de individuos biológicos, y en ciertos casos como el nuestro, material genético.

Como se menciona en la introducción, las proyecciones a futuro para esta especie neotropical son desalentadoras. Sin embargo, recuperar información de ejemplares de campo, en cautiverio y colecciones brindarán información adecuada para reconstruir la historia evolutiva del género. Recordemos que las colecciones científicas tienen como característica el concentrar archivos del conocimiento en un espacio determinado. Esto con el fin de conservarlos y mantenerlos disponibles en el largo plazo, preservando una parte de la biodiversidad que existe y existió (Cristín y Perrilliat, 2011).

Debido a esto, sugerimos implementar el acervo y la disponibilidad de las colecciones científicas de mamíferos ungulados en México. Nuestro proyecto es ejemplo de que una muestra almacenada por varios años en condiciones no óptimas puede ser la base para una investigación con enfoque genómico, lo que será útil al momento de utilizar otro tipo de muestra perteneciente a colecciones de animales en futuras investigaciones.

### **Sugerencias en campo**

No se registraron parámetros cardiopulmonares (frecuencia cardíaca, frecuencia respiratoria, temperatura corporal, electrocardiograma, porcentaje de saturación de oxígeno y presión arterial indirecta) ni análisis de gases en sangre arterial para corroborar la salud de los individuos durante la captura y manipulación del individuo silvestre, a diferencia del protocolo sugerido en Foerster et al., 2000,

recomendamos que la interacción con los organismos sea con la mayor cantidad de medidas de seguridad posible (Foerster et al., 2000).

Cabe recalcar que la captura del individuo no fue con la finalidad de obtener material biológico para secuenciar. Debido a esto, el método y tiempo de almacenamiento no fue el más adecuado para facilitar la extracción de material genético. Proponemos que en futuras investigaciones se anticipe con más tiempo y a más detalle el protocolo y material que se utilizará para llevar a cabo la colecta en campo.

### **Sugerencias en el procesamiento de datos**

Es importante mencionar que inicialmente, el proceso bioinformático se llevó a cabo considerando la integridad del material genético de la muestra obtenida. Posterior a la realización de los mapeos, ensamblajes, y análisis comparativos se lograron identificar claramente cambios de C por T o de G por A, o bien, depurinaciones y desaminaciones a lo largo de la secuencia comparando el mitogenoma liberado por Ennis et al., 2022 (**Anexos**).

Debido a esto proponemos conocer de manera detallada las condiciones de temperatura, humedad, pH, catalizadores, entre otros factores del medio que conduzcan a la degradación del ADN con la finalidad de considerar el impacto de estos sobre la secuencia en cuestión (Sunkari et al., 2023).

### **Perspectivas a futuro**

Se reconoce que la genética tiene una larga historia como herramienta en la conservación de la vida silvestre, sin embargo, investigaciones de análisis moleculares quedan excluidas de planes de manejo y conservación. Esto se debe principalmente a la falta de comprensión de conceptos genéticos, la financiación limitada, la poca disponibilidad de experiencia y la comunicación limitada entre científicos y el resto de la comunidad. En consecuencia, el flujo de información genética y evolutiva entre investigadores y administradores de conservación se ve obstaculizado (Seremi et al., 2019; Taft et al., 2020).

Es necesario destacar la importancia de conservar y fomentar el incremento de los bancos de recursos genéticos y genómicos para la conservación animal. Esto como herramienta útil en planes y propuestas de manejo para promover el mantenimiento de la diversidad genética en especies amenazadas (Holt y Pickard, 1999).

Un punto clave dentro de las perspectivas de este proyecto es el reconocer la importancia de considerar las características y protocolo de colecta de la muestra. En nuestro caso, se utilizó una muestra que presentó características de ADN antiguo como las mencionadas por Dabney et al., 2013. A pesar de que sí fue posible reconstruir el genoma mitocondrial completo del tapir centroamericano, proponemos sistematizar herramientas y métodos de conservación de muestras para evitar degradación de ADN en futuros proyectos .

Por otro lado, se calcula que las tasas de extinción de especies a escala global han sido entre 1000 y 10,000 veces mayor a las tasas observadas a lo largo de la mayor parte de la historia geológica, y claramente algunos taxa son más vulnerables que otros (Purvis et al., 2000).

A pesar de el conocimiento actual sobre la crisis en la pérdida de diversidad biológica a escala global, se propone que el número de especies extintas en esta “incipiente extinción masiva de antropoceno” es menor al de los grandes eventos de extinción recuperados por registros fósiles, por lo que se considera que aún hay tiempo de actuar y tomar las medidas correspondientes para evitar una pérdida de biodiversidad aún mayor (Marshall, 2023).

Las oportunidades de investigación posteriores a la obtención de un genoma son ilimitadas. A pesar de que difícilmente se puede comprender el estado genético de un grupo o población a partir de los datos de un solo individuo, la metodología empleada y los resultados de este proyecto abre camino a la comparación y planteamiento de nuevas técnicas que se podrán extrapolar y seguir explorando sobre el género *Tapirus*.

Se propone que empleando el estudio de los genomas será posible en un futuro contestar preguntas que involucren la relación de aspectos fisiológicos, mecanismos moleculares, la distribución o historia de vida empleando la variación genómica observada. También se podrá comprender qué factores moleculares coevolucionan a través del tiempo y cómo pudieron ocurrir modificaciones en los patrones evolutivos de los mecanismos de expresión genética en la especie y el género (Jeffrey, 1999).

Las características del material genómico mitocondrial permiten la fácil manipulación y análisis de datos, el tamaño de la secuencia y el creciente número de herramientas bioinformáticas hacen que este sea un buen marcador molecular al que cada vez más usuarios pueden acceder.

Otra parte importante es reconocer que actualmente, debido a los conflictos entre humanos y vida silvestre son la causa principal de la reducción y extinción de las poblaciones de especies en todo Mesoamérica, se han dedicado recursos para fomentar el manejo y protección de las especies. Sin embargo, estos están mayormente dirigidos a especies carismáticas, mientras que la actitud y percepción de los locales hacia las especies no carismáticas ha recibido poca atención (Pérez-Flores et al., 2021). La disponibilidad de información genómica del género *Tapirus* es limitada, por lo que es necesario dirigir atención a este grupo e invitar a la comunidad científica a involucrarse en estudios que liberen datos genómicos.

La secuenciación de próxima generación por WGS ha permitido rebasar barreras que anteriormente dificultaban la obtención de datos genómicos como la secuenciación de muestras de ADN antiguas, esto ha ampliado sustancialmente el alcance del análisis metagenómico de muestras derivadas del medio ambiente y de organismos en la actualidad (Mardis, 2008). En este trabajo reafirmamos la posibilidad de obtener datos genómicos de calidad a partir de una muestra almacenada durante al menos 11 años gracias a los avances bioinformáticos.

Finalmente, reiteramos la importancia de fomentar el interés en la obtención de datos genómicos con la finalidad de facilitar el acceso a información en futuras investigaciones. Gracias a la creciente facilidad de acceso a herramientas genómicas, es importante sugerir su implementación y de este modo obtener una mayor cantidad de datos de calidad promoviendo la investigación interdisciplinaria con enfoques moleculares.

## **X. CONCLUSIÓN**

A partir de la secuenciación por whole genome shotgun (WGS) fue posible obtener datos de secuenciación y posteriormente ensamblar el genoma mitocondrial de un individuo de tapir centroamericano silvestre a partir de una muestra almacenada durante poco más de una década en etanol a temperatura ambiente.

La muestra analizada presentó características de degradación de ADN posiblemente por procesos de desaminación y depurinación. A pesar de esto, fue posible reconstruir el genoma mitocondrial completo de un individuo silvestre monitoreado durante cinco años.

No fue posible realizar reconstrucciones filogenéticas con más resolución que los estudios basados en marcadores morfológicos y moleculares de genes

concatenados utilizados anteriormente. Los árboles generados mantuvieron la misma topología previamente descrita por distintos autores.

La filogenómica mitocondrial y la genómica comparativa pueden resolver mejor las relaciones filogenéticas del género *Tapirus* y permiten conocer mejor las características de los genomas de tapires y especies afines, sin embargo es necesario incrementar el acervo de datos genómicos del género *Tapirus* para obtener información novedosa empleando herramientas genómicas.

## XI. REFERENCIAS

- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
- Ashley, M. V., Norman, J. E., Stross, L. (1996). Phylogenetic analysis of the perissodactylan family Tapiridae using mitochondrial cytochrome c oxidase (COII) sequences. *Journal of Mammalian Evolution*, 3(4), 315-326.
- Auch, A. F., von Jan, M., Klenk, H. P., y Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in genomic sciences*, 2, 117-134.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Pribelski, A. D., Pyskin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., y Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology*, 19(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., y Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, 19(5), 455-477.
- Benítez, J. A., Pozo-Montuy, G., Alexander, S. M., Vargas-Contreras, J. A., Escalona-Segura, G., Sánchez-Acuña, M., y Prieto-Díaz, S. (2021). Impacto de la Vía Férrea y del Crecimiento Turístico Asociado al Tren Maya; medidas de mitigación y cambios al diseño para las reservas de Calakmul y Balam-kú. Impacto de las vías de comunicación sobre la fauna silvestre en áreas protegidas. Estudio de caso para el suresta de Mexico, 576-634.

- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., y Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution*, 69(2), 313-319.
- Bi, G., Luan, X., y Yan, J. (2023). ORPA: a fast and efficient phylogenetic analysis method for constructing genome-wide alignments of organelle genomes. *Journal of Genetics and Genomics*.
- Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., y Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods*, 10(1), 57-59.
- Bolger, A. M., Lohse, M., y Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
- Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic acids research*, 27(8), 1767-1780.
- Carrillo-Reyna, N., Reyna-Hurtado, R., y Schmook, B. (2015). Abundancia relativa y selección de hábitat de *Tapirus bairdii* en las reservas de Calakmul y Balam Kú, Campeche, México. *Revista Mexicana de biodiversidad*, 86(1), 202-207.
- Castellana, S., Vicario, S., y Saccone, C. (2011). Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. *Genome biology and evolution*, 3, 1067-1079.
- Castellanos, A., Foerster, C., Lizcano, D., Naranjo, E., Cruz-Aldan, E., Lira-Torres, I. et al. (2008). *Tapirus bairdii*. In IUCN Red list of threatened species. Version 2012.2. Recuperado el 15 abril 2023. Disponible en <http://www.iucnredlist.org>.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17, 540-552
- Ceballos, G., L. Mendoza G. O'Farrill. (2020). El tapir en México distribución, ecología y conservación. Universidad Nacional Autónoma de México, Ciudad de México.
- Coates, A. G., y Stallard, R. F. (2013). How old is the Isthmus of Panama?. *Bulletin of Marine Science*, 89(4), 801-813.
- Corradine, L. F (2022). Dantas: la historia de una hermana perdida. *Revista Pesquisa Javeriana*. Recuperado el 10 de enero de 2023. Disponible en <https://www.javeriana.edu.co/pesquisa/dantas-la-historia-de-una-hermana-perdida/>
- Cozzuol, M. A., Clozato, C. L., Holanda, E. C., Rodrigues, F. H., Nienow, S., De Thoisy, B., Santos, F. R. (2013). A new species of tapir from the Amazon. *Journal of Mammalogy*, 94(6), 1331-1345.

- Cristín, A., y Perrilliat, M. D. C. (2011). Las colecciones científicas y la protección del patrimonio paleontológico. *Boletín de la Sociedad Geológica Mexicana*, 63(3), 421-427.
- Dabney, J., Meyer, M., y Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harbor perspectives in biology*, 5(7), a012567.
- Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, 14(7), 1394-1403.
- de Sena Brandine, G., y Smith, A. D. (2019). Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Research*, 8.
- Delsuc, F., Brinkmann, H., y Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5), 361-375.
- Ferrero, B. S., y Noriega, J. I. (2007). A new upper Pleistocene tapir from Argentina: remarks on the phylogenetics and diversification of Neotropical Tapiridae. *Journal of Vertebrate Paleontology*, 27(2), 504-511.
- Foerster CR y Vaughan C (2015) Diet and foraging behavior of a female Baird's tapir (*Tapirus bairdii*) in a Costa Rican lowland rainforest. *Cuad Inv UNED* 7:259–267
- Foerster, C., Myers, M., y Villalobos, J. (2001, November). Results of a five-year telemetry study of Bairds tapir in Costa Rica. In *Proceedings of the First International Tapir Symposium*. San Jose, Costa Rica (p. 9).
- Foerster, S. H., Bailey, J. E., Aguilar, R., Loria, D. L., y Foerster, C. R. (2000). Butorphanol/xylazine/ketamine immobilization of free-ranging Baird's tapirs in Costa Rica. *Journal of Wildlife Diseases*, 36(2), 335-341.
- Foerster, S. H., Bailey, J. E., Aguilar, R., Loria, D. L., y Foerster, C. R. (2000). Butorphanol/xylazine/ketamine immobilization of free-ranging Baird's tapirs in Costa Rica. *Journal of Wildlife Diseases*, 36(2), 335-341.
- Futuyma, D. J. (1979). *Evolutionary biology*. Sunderland, Mass., Sinauer Associates.
- Gallina-Tessaro (2019). Ecology and conservation of tropical ungulates in Latin America (pp. 371-393). Springer International Publishing. 89-93.
- García, M., Jordan, C., O'Farril, G., Poot, C., Meyer, N., Estrada, N., Leonardo, R., Naranjo, E., Simons, Á., Herrera, A. ., Urgilés, C., Schank, C., Boshoff, L. y Ruiz-Galeano, M. 2016. *Tapirus bairdii* . Lista Roja de Especies Amenazadas de la UICN 2016. Consultado el 20 de octubre de 2023 .
- García, M., Jordan, C.A., O'Farril, G., Poot, C., Meyer, N., Estrada, N., Ruiz-Galeano, M. (2016). *Tapirus bairdii*. M. *Tapirus bairdii*. The IUCN Red List of Threatened Species.
- Gates, A. J., Gysi, D. M., Kellis, M., & Barabási, A. L. (2021). A wealth of discovery built on the Human Genome Project—by the numbers. *Nature*, 590(7845), 212-215.

- Genomics Education Programme. *Glossary: Genomics NHS choices*. Disponible en: <https://www.genomicseducation.hee.nhs.uk/about-us/> (Consultado: 15 de septiembre del 2023).
- González de la Fuente, S. (2018). *Ensamblaje de novo y anotación génica del genoma de Leishmania major mediante secuenciación masiva*.
- Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C, Graham M, Van Domselaar G, y Stothard P. (2023). Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*. Consultado el 2 de febrero de 2024. Disponible en <https://proksee.ca/>.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., y Gascuel, O. (2010). *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0*. *Systematic biology*, 59(3), 307-321.
- Gurevich, A., Saveliev, V., Vyahhi, N., y Tesler, G. (2013). *QUAST: quality assessment tool for genome assemblies*. *Bioinformatics*, 29(8), 1072-1075.
- Hardison, R. C. (2003). *Comparative genomics*. *PLoS biology*, 1(2), e58.
- Hartwig, S., Feckler, C., Lehr, S., Wallbrecht, K., Wolgast, H., Müller-Wieland, D., y Kotzka, J. (2009). *A critical comparison between two classical and a kit-based method for mitochondria isolation*. *Proteomics*, 9(11), 3209-3214.
- Hohenlohe, P. A., Funk, W. C., y Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1), 62-82.
- Holanda, E. C., Ferrero, B. S. (2013). *Reappraisal of the genus Tapirus (Perissodactyla, Tapiridae): systematics and phylogenetic affinities of the South American tapirs*. *Journal of Mammalian Evolution*, 20(1), 33-44.
- Holt, W. V., y Pickard, A. R. (1999). *Role of reproductive technologies and genetic resource banks in animal conservation*. *Reviews of reproduction*, 4, 143-150.
- Houck, M. L., Kingswood, S. C., y Kumamoto, A. T. (2000). Comparative cytogenetics of tapirs, genus *Tapirus* (Perissodactyla, Tapiridae). *Cytogenetics and cell genetics*, 89(1-2), 110-115.
- Hulbert RC. (1995). *The giant tapir, Tapirus haysii, from Leisey Shell Pit 1A and other Florida Irvingtonian localities*. *Bull Fla Mus Nat Hist* 37: 515-551.
- Illumina. Sequencing quality scores. Recuperado el 15 de septiembre del 2023. Disponible en <https://www.illumina.com/science/technology/next-generation-sequencing/pl-an-experiments/quality-scores.html>
- INSDC (2023) International Nucleotide Sequence Database Collaboration. Disponible: <https://www.insdc.org/> (Consultado el 15 de septiembre de 2023).
- IUCN. The IUCN Red List of Threatened Species. Versión 2021-1. Disponible en:

<https://www.iucnredlist.org/search?query=tapirus&searchType=species>.  
Consultado el 18 de 08 del 2022.

- Jones, C. G., Lawton, J. H., y Shachak, M. (1994). Organisms as ecosystem engineers. *Oikos*, 373-386.
- Kapli, P., Yang, Z., y Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, 21(7), 428-444.
- Langmead, B., Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359.
- Lipman, D. y Pearson, W. (1985). Rapid and sensitive protein similarity searches. *Science*, 227(4693), 1435-1441.
- Lockhart, D. J., y Winzler, E. A. (2000). Genomics, gene expression and DNA arrays. *Nature*, 405(6788), 827-836.
- Madden, T. (2013). The BLAST sequence analysis tool. In *The NCBI Handbook* [Internet]. 2nd edition. National Center for Biotechnology Information (US).
- Meiklejohn, K. A., Danielson, M. J., Faircloth, B. C., Glenn, T. C., Braun, E. L., y Kimball, R. T. (2014). Incongruence among different mitochondrial regions: a case study using complete mitogenomes. *Molecular phylogenetics and evolution*, 78, 314-323.
- Manel, S., y Holderegger, R. (2013). Ten years of landscape genetics. *Trends in ecology & evolution*, 28(10), 614-621.
- Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, 9, 387-402.
- Marshall, C. R. (2023). Forty years later: The status of the “Big Five” mass extinctions. *Cambridge Prisms: Extinction*, 1, e5.
- Martins, A., (2014). El Tapir, El Desconocido Jardinero De Los Bosques Sudamericanos. BBC Mundo Bbc.com.Disponible en: [https://www.bbc.com/mundo/noticias/2011/12/111215\\_tapir\\_extraordinario\\_am](https://www.bbc.com/mundo/noticias/2011/12/111215_tapir_extraordinario_am). Consultado el 8 de febrero 2021.
- McPhee, M. E., Carlstead, K. (2010). The importance of maintaining natural behaviors in captive mammals. *Wild mammals in captivity: Principles and techniques for zoo management*. 2, 303-313.
- Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, y Marshall D. (2013). Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform.* 14:193-202.
- Mithiaux Cosin, V. E. 2016. Causas de las diferencias en las tasas de sustitución de nucleótidos en genes codificantes para proteínas mitocondriales de mamíferos. *Universitat de les Illes Balears*.
- Monette, V. D. (2019). Ecological Factors Associated with Habitat Use of Baird’s Tapirs (*Tapirus bairdii*). *Electronic Theses and Dissertations*. 1645.
- Muangkram, Y., Wajjwalku, W., Kaolim, N., Buddhakosai, W., Kamolnorrath, S., Siriaroonrat, B., y Sanannu, S. (2016). The complete

mitochondrial genome of the Asian tapirs (*Tapirus indicus*): the only extant Tapiridae species in the Old World. *Mitochondrial DNA Part A*, 27(1), 413-415.

- Muench, C. (2006). Corredores de vegetación y conectividad de hábitat para el tapir (*Tapirus bairdii*) en la selva Lacandona, Chiapas. Universidad Nacional Autónoma de México, México.
- Nabholz, B., Glémin, S., y Galtier, N. (2008). Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Molecular biology and evolution*, 25(1), 120-130.
- Naranjo, E. J. (2009). Ecology and conservation of Baird's tapir in Mexico. *Tropical Conservation Science*, 2(2), 140-158
- Naranjo, E. J., Amador-Alcalá, S. A., Falconi-Briones, F. A., y Reyna-Hurtado, R. A. (2015). Distribución, abundancia y amenazas a las poblaciones de tapir centroamericano (*Tapirus bairdii*) y pecarí de labios blancos (*Tayassu pecari*) en México. *Therya*, 6(1), 227-249.
- Naranjo, E.J. (2018). Baird's tapir ecology and conservation in Mexico revisited. *Tropical Conservation Science* 11:1-4
- National Center for Biotechnology Information (NCBI). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 1988. Disponible en: <https://www.ncbi.nlm.nih.gov/>. Consultado 13 de agosto del 2021.
- Norton, J. E., Ashley, M. V. (2004, May). Genetic variability and population structure among wild Baird's tapirs. In *Animal Conservation forum* (vol. 7, no. 2, pp. 211-220). Cambridge University Press.
- Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. Disponible es <https://www.ncbi.nlm.nih.gov/nucleotide/?term=tapirus>
- O'Farrill G, Gauthier-Schampaert K, Rayfield B, Bodin Ö, Calmé S, Sengupta R, y Gonzalez A (2014) The potential connectivity of waterhole networks and the effectiveness of a protected area under various drought scenarios. *PLoS ONE* 9(5): e95049.
- Paolucci, L. N., Pereira, R. L., Rattis, L., Silvério, D. V., Marques, N. C., Macedo, M. N., y Brando, P. M. (2019). Lowland tapirs facilitate seed dispersal in degraded Amazonian forests. *Biotropica*, 51(2), 245-252.
- Park, S. T., y Kim, J. (2016). Trends in next-generation sequencing and a new era for whole genome sequencing. *International neurology journal*, 20(Suppl 2), S76.
- Pennington, T. D., Sarukhán, J. (2005). Árboles tropicales de México: manual para la identificación de las principales especies. UNAM.
- Pérez-Flores, J., Mardero, S., López-Cen, A., y Contreras-Moreno, F. M. (2021). Human-wildlife conflicts and drought in the greater Calakmul

Region, Mexico: implications for tapir conservation. *Neotropical Biology and Conservation*, 16(4), 539-563

- Pesole, G., Gissi, C., De Chirico, A., y Saccone, C. (1999). Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of molecular evolution*, 48, 427-434.
- Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., y Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS biology*, 9(3), e1000602.
- Pop, M., Phillippy, A., Delcher, A. L., y Salzberg, S. L. (2004). Comparative genome assembly. *Briefings in bioinformatics*, 5(3), 237-248.
- Purvis, A., Jones, K. E., y Mace, G. M. (2000). Extinction. *BioEssays*, 22(12), 1123-1133.
- Reyna-Hurtado, R., y Tanner, W. (2007). Ungulate relative abundance in hunted and non-hunted sites in Calakmul Forest (Southern Mexico). *Biodiversity Conservation*, 16, 743-756. <https://doi.org/10.1007/s10531-005-6198-7>
- Reyna-Hurtado, R., Molina-Rosales, D., Naranjo-Piñera, E., y Serrano-Mac-Gregor, I. (2021). Baird's tapir: predicting patterns of crop damage surrounding the Calakmul Biosphere Reserve, Campeche, Mexico. *Revista Mexicana de Biodiversidad*, 92, 1-13.
- Reyna-Hurtado, R., Sanvicente-López, M., Pérez-Flores, J., Carrillo-Reyna, N., Calmé, S. (2016). Insights into the multiannual home range of a Baird's tapir (*Tapirus bairdii*) in the Maya Forest. *Therya*, 7(2), 271276.
- Robinson, J. A., Räikkönen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., y Wayne, R. K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances*, 5(5), eaau0757.
- Ruiz-García, M., Castellanos, A., Bernal, L. A., Pinedo-Castro, M., Kaston, F., Shostell, J. M. (2016). Mitogenomics of the mountain tapir (*Tapirus pinchaque*, Tapiridae, Perissodactyla, Mammalia) in Colombia and Ecuador: phylogeography and insights into the origin and systematics of the South American tapirs. *Mammalian Biology*, 81(2), 163-175.
- Ruiz-García, M., Vásquez, C., Pinedo-Castro, M. O., Sandoval, S., Castellanos, A., Kaston, F., Shostell, J. (2012). Phylogeography of the Mountain Tapir (*Tapirus pinchaque*) and the Central American Tapir (*Tapirus bairdii*) and the origins of the three Latin-American tapirs by means of mtCyt-B sequences. *Current topics in phylogenetics and phylogeography of terrestrial and aquatic systems*. InTech, Rijeka, 83116.
- Ruiz, R. (2010). Estimación y actualización al 2009 de la Tasa de Transformación del hábitat de las áreas naturales protegidas SINAP I y SINAP II del FANP: Reserva de la Biosfera Calakmul. Fondo mexicano

para la conservación de la naturaleza, A.C, Comisión Nacional de las Áreas Naturales Protegidas.

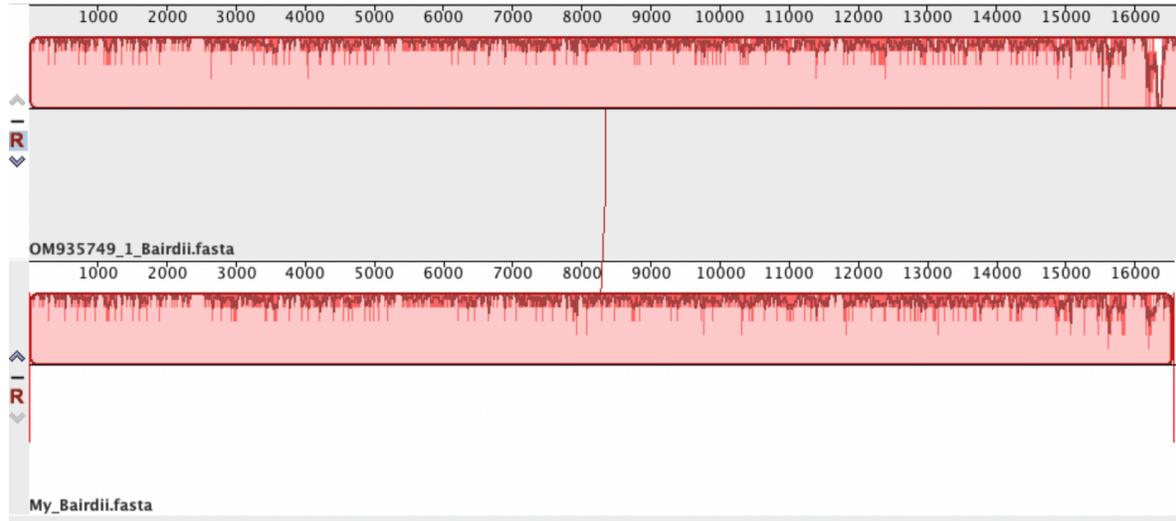
- Sanchez-Pinzon, K., Reyna-Hurtado, R., Meyer, N. F. (2019). Moon light and the activity patterns of Baird's tapir in the Calakmul region, Southern México. *Therya*, 11(1), 137.
- Saremi, N. F., Supple, M. A., Byrne, A., Cahill, J. A., Coutinho, L. L., Dalén, L., y Shapiro, B. (2019). Puma genomes from North and South America provide insights into the genomic consequences of inbreeding. *Nature communications*, 10(1), 4769.
- SEMARNAT, (2009). PROGRAMA DE ACCIÓN PARA LA CONSERVACIÓN DE LA ESPECIE Tapir Centroamericano (*Tapirus bairdii*).
- Smith D. R. (2016). The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs?. *Briefings in functional genomics*, 15(1), 47–54.
- Stein, L. (2001). Genome annotation: from sequence to biology. *Nature reviews genetics*, 2(7), 493-503.
- Stothard P (2000) The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28:1102-1104.
- Sunkari, Y. K., Nguyen, T. L., Siripuram, V. K., y Flajolet, M. (2023). Impact of organic chemistry conditions on DNA durability in the context of DNA-encoded library technology. *Iscience*, 26(9).
- Taft, H. R., McCoskey, D. N., Miller, J. M., Pearson, S. K., Coleman, M. A., Fletcher, N. K., y Barbosa, S. (2020). Research–management partnerships: An opportunity to integrate genetics in conservation actions. *Conservation Science and Practice*, 2(9), e218.
- Tedford, R., Albright, L., Barnosky, A., Ferrusquia-Villafranca, I., Hunt, R., Storer, J., Swisher, C., Voorhies, M., Webb, S. y Whistler, D. (2004). 6. Mammalian Biochronology of the Arikareean Through Hemphillian Interval (Late Oligocene Through Early Pliocene Epochs). In M. Woodburne (Ed.), *Late Cretaceous and Cenozoic Mammals of North America: Biostratigraphy and Geochronology* (pp. 169-231). New York Chichester, West Sussex: Columbia University Press. <https://doi.org/10.7312/wood13040-008>
- University of Missouri. FastQC 1. Introduction 1.1 What is FastQC - DNA Core. Missouri University. Disponible en: [https://dnacore.missouri.edu/PDF/FastQC\\_Manual.pdf](https://dnacore.missouri.edu/PDF/FastQC_Manual.pdf) (Accessed: November 10, 2022).
- Yanez-Montalvo, A., Gaona, O., Águila, B., Arias-Dominguez, N., Falcón, L. I., y Perez-Flores, J. (2021). *Tapirus bairdii*-associated Fecal microbiome from a critical conservation area: Calakmul, México. *Current Microbiology*, 78(7), 2648-2659.

- Zoonomia Consortium. (2020) A comparative genomics multitool for scientific discovery and conservation. *Nature* 587, 240–245.

## XII. ANEXOS

### Alineamientos

Alineamiento pareado con MAUVE (Darling et al., 2004) de *T. bairdii* (OM935749.1) y *T. bairdii* de Calakmul.



Alineamiento pareado con BLAST (Madden, 2013) de *T. bairdii* (OM935749.1) y *T. bairdii* de Calakmul.

Score	Expect	Identities	Gaps	Strand
23398 bits(12670)	0.0	15100/16303(93%)	47/16303(0%)	Plus/Plus
Query 1		GTTAATGTAGCTTAATCCAAAGCGAGGCACTGAAGATGCCAGATGAGTATTACTAACTC		60
Sbjct 1		GTTAATGTAGCTTAATCCAAAGCGAGGCACTGAAGATGCCAGATGAGTATTACTAACTC		60
Query 61		CATAAACACACAGGTTTGGTCCTAGCCTTTATATTAATTGTTAATAAAAATTACACATGCA		120
Sbjct 61		CATAAACACATAGGTTTGGTCCTAGCCTTTATATTAATTGTTAATAAAAATTACACATGCA		120
Query 121		AGTATCCGCACCCAGTGAAAATGCCCTCCAAATCACATTAACGATCAAAGGAGCAGGT		180
Sbjct 121		AGTATCCGCACCCAGTGAAAATGCCCTCCAAATCACACTAACGATCAAAGGAGCAGGT		180
Query 181		ATCAAGCACGCTAAATTAGCAGCTTACAACACCTTGCTTAACCACACCCCCACGGGAAAC		240
Sbjct 181		ATCAAGCACGCTAAATTAGCAGCTCAGCACCTTGCTTAACCACACCCCCACGGGAAAC		240
Query 241		AGCAGTGATAAAAAATTAAGCCATAAACGAAAGTTCGACTAAGTTATATTAATCAGGGTT		300
Sbjct 241		AGCAGTGATAAAAAATTAAGCCATAAACGAAAGTTCGACTAAGTTATATTAAGTTAGGGTT		300
Query 301		GGTAAATTTTCGTGCCAGCCACCGCGGTCATACGATTAACCCAAATTAATATACTCCGGCG		360
Sbjct 301		GGTAAATTTTCGTGCCAGCCACCGCGGTCATACGATTAACCCAAATTAATATACTCCGGCG		360
Query 361		TAAAGAGTGTCAAAGatatacatataataaagttaaaaaactaactaagctgtaaaaaag		420
Sbjct 361		TAAAGAGTGTCAAAGATACATAC--ACAATAAAGTTAAAAACCAACTAAGCTGTA AAAAG		418
Query 421		ccacagcgaaaaaataaaataaactacgaaaagTGACTTTAACACCTCTGACTACACGACAG		480
Sbjct 419		CCACAGC-AAAAATAAAATAAGCTACGAAAGTGACTTTAACATCTCTGACTACACGACAG		477
Query 481		CTAAGATCCAAACTGGGATTAGATACCCCACTATGCTTAGCCATAAACTAAAATAATTTA		540
Sbjct 478		CTAAGACCCAAACTGGGATTAGATACCCCACTATGCTTAGTCATAAACCAAGATAA-TTA		536
Query 541		TCCAACAAAATTATTGCGCCAGAGTACTACTAGCAATAGCTTAAAACCTCAAAGGACTTGGC		600
Sbjct 537		TCCAACAAAATTATTGCGCCAGAGTACTACTAGCAATAGCTTAAAACCTCAAAGGACTTGGC		596
Query 601		GGTGCTTTATATCCCTCTAGAGGAGCCTGTTCCATAATCGATAAACCCCGATAAACCTTA		660
Sbjct 597		GGTGCTTTATACCCCTCTAGAGGAGCCTGTTCCGTAATCGATAAACCCCGATAAACCTTA		656
Query 661		CCATCCCTTGCCAACACAGCCTATATACCGCCATCTTCAGCAAACCTAAAAAGGACTCA		720
Sbjct 657		CCATCCCTTGCCCAATACAGCCTATATACCGCCATCTTCAGCAAACCTAAAAAGGAACCA		716
Query 721		AAGTAAGCACAAATACAAGACATAAAAAAGTTAGGTCAAGGTGTAGCCTATGGGATGGGG		780
Sbjct 717		AAGTAAGCACAAATATAAGACATAAAAAAGTTAGGTCAAGGTGTAGCCTATGGGATGGAG		776
Query 781		AGAAATGGGCTACATTTTCTACTCAAGAACAACATAACTCATTACACGAAAGTTTTTATG		840
Sbjct 777		AGAAATGGGCTACATTTTCTACCAAGAACAACACAACCTCGTCACACGAAAGTTTTTATG		836
Query 841		AAACTGAAAACCTAAAGGAGGATTTAGCAGTAAATTAAGAATAGAGAGCTTAATTGAACTA		900
Sbjct 837		AAATTA AAAACCTAAAGGAGGATTTAGTAGTAAATTAAGAGTAGAGAGCTTAATTGAACTA		896
Query 901		GGCCATGAAGCACGCACACACCCGCCGTACCCTCCTTAAATATTACATACTACAACCTTA		960
Sbjct 897		GGCCATGAAGCACGCACACACCCGCCGTACCCTCCTTAAATATTACATACTATAACCTTA		956
Query 961		ACATA-TAACAATTGTTAAACATATGAAAGGAGACAAGTCGTAACAAGGTAAGCGTACTG		1019
Sbjct 957		ACATATTAATAACTGTCAAACATATGAAAGGAGACAAGTCGTAACAAGGTAAGCGTACTG		1016
Query 1020		GAAAGTGTGCTTGGACAACCAAAGTGTAGCTTAAACAAAGCATCTAGTTTACACCTAGAA		1079
Sbjct 1017		GAAAGTGTGCTTGGACAACCAAAGTGTAGCTTAAACAAAGCATCTAGTTTACACCTAGAA		1076
Query 1080		GATTTACACAAAAATGACCACTTTGAACTAAAGCTAGCCCAAACGTCATTCAACTCAACT		1139
Sbjct 1077		GATTTACAGTAAAATGACCACTTTGAACCAAAGCTAGCCCAAACACCATCCAACCTCAACT		1136
Query 1140		ACCCCAA AAAATTTTCAAATAAAAACATTACCCACCACCATTA AAAAGTATAGGAGATAGAA		1199
Sbjct 1137		ACCC--AAGACCTTCAAATAAAAACATTACCTAACACCATTA AAAAGTATAGGAGATAGAA		1195
Query 1200		ATTTTAATCGGCGCTATAGAGAAAGTACCGCAAGGGAACGATGAAAGAAATTA AAAAGT		1259
Sbjct 1196		ATTTTAATAGGCGCTATAGAGAAAGTACCGCAAGGGAATGATGAAAGAAATTTTA AAAAGT		1255

Query	1260	ACCAAACAGCAAAGCTTACCCCTTCTACCTTTTGCATAATGATTTAACTAGAATAATTTA	1319
Sbjct	1256	ACTGAACAGCAAAGCTTACCCCTTCTACCTTTTGCATAATGATTTAACTAGAATAATTTA	1315
Query	1320	GCAAAGAGAACTTAAGCTAAATACCCGAAACCAGACGAGCTACCTATGAACAGTCCCAA	1379
Sbjct	1316	GCAAAGAGAACTTAAGCTAAACATCCCGAAACCAGACGAGCTACCTATGAACAGTCCCAA	1375
Query	1380	AGAACAAACTCATCTATGTTGCAAAATAGTGAGAAGATTTATAGGTAGAGGTGAAAAGCC	1439
Sbjct	1376	AGAACAAACTCATCTATGTTGCAAAATAGTGAGAAGATTTATAGGTAGAGGTGAAAAGCC	1435
Query	1440	AAACGAGCCTGGTGATAGCTGGTTATCCAGAACAGAATATTAGTTCAAATTTAAATTTAC	1499
Sbjct	1436	AAACGAGCCTGGTGATAGCTGGTTATCCAGAATAGAATATTAGTTCAAATTTAAATTTAC	1495
Query	1500	CTAACAAAAGCACCAACTCTAATGTAATTTAAATTTTAACTAAAGAGGTACAGCTC	1559
Sbjct	1496	CTAACAGAAACCCATAACTCTAACGTAAATTTAAATTTTAACTAAAGAGGTACAGCTC	1555
Query	1560	TTTAGACACAGGATACAACCTTCTTAGAGAGTAAAGACACAAACAAACCATAGTTGGCC	1619
Sbjct	1556	TTTAGACACAGGATACAACCTTCTTAGAGAGTAAAGAACGAATAAACCATAGTTGGCC	1615
Query	1620	TAAAAGCAGCCATCAATTAAGAAAGCGTTTAAAGCTCAACAACACAACCATCTTAATCCCA	1679
Sbjct	1616	TAAAAGCAGCCATCAATTGAGAAAGCGTTCAAGCTCAACAACATAAACCATCTTAATCCCA	1675
Query	1680	ACAATAACCAACAACCTCTAATATAGCACTGGATTAATCTATTAATATATAGAAGCAAT	1739
Sbjct	1676	ACAACAACCTAAATAACTCTAATATAATACTGGATTAATCTATTAATATATAGAAGCAAT	1735
Query	1740	AATGTTAATATAAGTAACAAGAAATATTTCTCCCGCATGAGCTTATATCAGAACGGATA	1799
Sbjct	1736	AATGTTAATATAAGTAACAAGAAATATTTCTCCCGCATGAGCTTATATCAGAACGGATA	1795
Query	1800	ACCACTGATAGTTAAACAACAAAATAAACAACCAACAATTAACAATTAATAAAT	1859
Sbjct	1796	ACCACTGATAGTTAAACAACAAAATAAACAACCAACAATTAACAATTAATAAAT	1855
Query	1860	AACTGTTAATCCAACACAGGCATGCTATACTCCAAGGAAAGATTAAGAAAGTAAAGG	1919
Sbjct	1856	AACTGTTAATCCAACACAGGCATGCTATACTCCAAGGAAAGATTAAGAAAGTAAAGG	1915
Query	1920	AACTCGGCAAAACAAACCCCGCCTGTTTACCAAAAACATCACCTCTAGCATTACTAATA	1979
Sbjct	1916	AACTCGGCAAAATACAAACCCCGCCTGTTTACCAAAAACATCACCTCTAGCATTACTAATA	1975
Query	1980	TTAGAGGCACTGCCTGCCAGTGACATCTGTTTAAACGGCCGCGGTATCCTAACCGTGCA	2039
Sbjct	1976	TTAGAGGCACTGCCTGCCAGTGACATCTGTTTAAACGGCCGCGGTATCCTAACCGTGCA	2035
Query	2040	AAGGTAGCATAATCACTTGTTCTCTAAATAAGGACTTGTATGAATGGCCACACGAGGGTT	2099
Sbjct	2036	AAGGTAGCATAATCACTTGTTCTCTAAATAAGGACTTGTATGAACGGCCACACGAGGGTT	2095
Query	2100	TAACTGTCTCTACTCTAATCAGTGAAATTTGACCTTCCCGTGAAGAGGCGGGAATAACA	2159
Sbjct	2096	TAACTGTCTCTACTCTAATCAGTGAAATTTGACCTTCCCGTGAAGAGGCGGGAATAACA	2155
Query	2160	AAATAAGACGAGAAGACCTATGGAGCTTCAATTAATCAATTAACAAGTATAAACTTC	2219
Sbjct	2156	AAATAAGACGAGAAGACCTATGGAGCTTCAATTAATCAATTAACAAGTATAAACTTC	2215
Query	2220	CAACCTACCAGGCAT-AAAAAATTTAACTGAATTCGACAATTTTCGGTTGGGGTGACCTC	2278
Sbjct	2216	AAACCTACCAGGTATAAAAAAATTTAACTGAATTCGACAATTTTCGGTTGGGGTGACCTC	2275
Query	2279	GGAGAACaaaaaaCCTCCGAGTGATTAAGCTTAGACTAACAGTCAAAGCACTACATC	2338
Sbjct	2276	GGAGAACAAAATAACCTCCGAGCGATTAAGCTCAGACTAACAGTCAAAGCATTACATC	2335
Query	2339	ACTTATTGATCCAAATTATTGATCAACGGAACAAGTTACCCTAGGGATAACAGCGCAATC	2398
Sbjct	2336	ACTTATTGATCCAAATTATTGATCAACGGAACAAGTTACCCTAGGGATAACAGCGCAATC	2395
Query	2399	CTATTCTAGAGTCCATATCGACAATAGGGTTTACGACCTCGATGTTGGATCAGGACATCC	2458
Sbjct	2396	CTATTCTAGAGTCCATATCGACAATAGGGTTTACGACCTCGATGTTGGATCAGGACATCC	2455
Query	2459	CAATGGTGCAACCGCTATTAATGGTTCGTTTGTTCACGATTAAGTCTACGTGATCTG	2518
Sbjct	2456	CAATGGTGCAACCGCTATTAATGGTTCGTTTGTTCACGATTAAGTCTACGTGATCTG	2515
Query	2519	AGTTCAGACCGGAGCAATCCAGGTCGGTTTCTATCTATTTCTATTTTCTCCAGTACGAA	2578
Sbjct	2516	AGTTCAGACCGGAGCAATCCAGGTCGGTTTCTATCTATTTCTATTTTCTCCAGTACGAA	2575

Query	3899	AAAATGTTGGATCATACCCTTCCCGTACTAATAAATCCTCTAATCTCCATAACCATCATA	3958
Sbjct	3896	AAAATGTTGGATCATACCCTTCCCGTACTAATAAACCTCTAATCTTCACAACCATTATA	3955
Query	3959	ACAACTGTTATTCTAGGAACCATAATTGTAATAACAAGCTCCCAGTACTAATAATCTGA	4018
Sbjct	3956	ACAACGTTATCCTAGGAACCATAATTGTAATAACAAGCTCTCACTGACTAATAATCTGA	4015
Query	4019	ATTGGTTTCGAAATAAACCTTCTAGCTATCATCCCAATCCTAATaaaaaaTTCAACCCA	4078
Sbjct	4016	ATCGGCTTCGAAATAAATCTCTTAGCCATCATCCCAATCCTAATAAAAAATTTAACCCA	4075
Query	4079	CGGACAATAGAAGCATCCACCAAATACTTTCTAACACAAGCCACCGCATCTATACTACTC	4138
Sbjct	4076	CGGACAATAGAAGCATCCACCAAATACTTTTAAACAAGCCACCGCATCCATACTACTC	4135
Query	4139	ATAATAGCAATCATTATTAACCTCATATACTCAGGCCAATGAGCTATTACAAAAATTTTT	4198
Sbjct	4136	ATAATGGCAATTATTGTTAACCTCATATACTCAGGCCAATGAACTATTACGAAAAATTTTT	4195
Query	4199	AATCCAATAGCGTCCGTTATTATAACAACAGCTCAATTATAAAACTTGGATTATCCCC	4258
Sbjct	4196	AATCCAGCAGCATCCATCATATAACAACAGCCCTAATCATAAACTTGGATTATCTCCC	4255
Query	4259	TTTCACTTCTGAGTACCCGAAGTAACACAAGGTATCTCACTTACTTCAGGCTTAATTCTA	4318
Sbjct	4256	TTCCACTTCTGAGTACCCGAAGTAACACAAGGCATCTCGTTACCTCAGGCTTAATTCTA	4315
Query	4319	CTCACGTGACAAAAATTAGCCCCAATATCAATCTTATACCAAATTTACCATCCATCAAC	4378
Sbjct	4316	CTCACATGACAAAAATTAGCCCCAATATCAATCTTATACCAAATTTACCATCCATCAAT	4375
Query	4379	CCAAACATACTACTAACCGCGCTATTTTATCCATTATAGTAGGAGGCTGAGGAGGCCTA	4438
Sbjct	4376	CCAAACATACTACTAACTGCTGCCATATATCCGTTATAGTAGGAGGCTGAGGAGGCCTA	4435
Query	4439	AACCAAACACTCAACTACGAAAAATCATAGCATATTCATCAATTGCCACATAGGATGAATA	4498
Sbjct	4436	AATCAAACACTCAACTACGAAAAATCATAGCATACTCATCAATTGCTCACATAGGATGAATA	4495
Query	4499	ACAGCCATCCTAATTTACAACCCAAACAATAACAATACTAAACATACTAATCTATATCATA	4558
Sbjct	4496	ACAGCTATCCTAATTTACAACCCAAACAATAACAATACTAAACATACTAATTTACATCATA	4555
Query	4559	ACAACACTTACCATATTCATGATATTCATAAATCACTCCTCAACCACAACCTTATCACTC	4618
Sbjct	4556	ACAACACTTACCATATTTATAATATTCATAAACCACCTCCTCAACCACAACCTTGTCACTC	4615
Query	4619	TCAAATACATGAAATAAAAAATCCCCTAATTACCACACTTATCCTAACACCCTACTATCT	4678
Sbjct	4616	TCAAACACGTGAAATAAAGACCCCCTAATCACTACACTCATCTTAACAACCCTACTGTCC	4675
Query	4679	CTAGGTGGCCTCCCACCCTATCAGGATTCATACCCAAATGACTAATCATCCAAGAACTT	4738
Sbjct	4676	CTAGGTGGCCTCCCACCATTATCGGGATTCATACCCAAATGACTAATCATCCAAGAACTT	4735
Query	4739	ACAAAAAATAATAGCATTATTCTACCAACATTAATAGCCATTACAGCACTACTCAACCTA	4798
Sbjct	4736	ACAAAAACAACAGCATTATTCTACCAACATTAATAGCTATTACAGCATTACTCAACCTG	4795
Query	4799	TATTTCTACATACGACTGGCCTACTCCACATCACTAACAAATATTTCCATcaacaataac	4858
Sbjct	4796	TATTTCTACATACGACTGGCCTACTCCACATCACTAACAAATATTTCCATCAACAAACAAT	4855
Query	4859	ataaaaaataaatgacagtttgaaaacacaaaaaaataatCCTCCTACCAATACTAATC	4918
Sbjct	4856	ACAAAAATAAATGACAAATTCGAAAACCTCAAAACAAATAATCCTCCTACCAACACTAATC	4915
Query	4919	ACTATATCTACACTTCTCTTACCCTAACACCAATAATTTCCATCCTAGAATAGGAATTT	4978
Sbjct	4916	ACCATATCTACGCTCCTCCTACCCTAACACCAATAATTTCCATCCTAGAATAGGAATTT	4975
Query	4979	AGGTTACATAGACCAAGAGCCTTCAAAGCTCTAAGCAAGAGaaaaaaTCACTTAATTCC	5038
Sbjct	4976	AGGTTACGACAGACCAAGAGCCTTCAAAGCTCTAAGCAAGAG-AAAAAATCACTTAATTCC	5034
Query	5039	TGCGTACTAAGGGCTGCAAGATTATACCTTACATCAATTGAACGCAAAATCAAACTTTA	5098
Sbjct	5035	TGCATACTAAGGGCTGCAAGACTACACCTTACATCAATTGAACGCAAAATCAAACTTTA	5094
Query	5099	ATTAAGCTAAGCCCTCACTAGATTGACGGGCTATTATCCCAGGAAATTTTAGTTAACAGC	5158
Sbjct	5095	ATTAAGCTAAGCCCTCACTAGATTGACGGGCTATTATCCCAGGAAATTTTAGTTAACAGC	5154
Query	5159	TAAATACCCTAAACAACCTGGCTTCAATCTACTTCTCCCGCCGCTa-aaaaaaaGGCGG	5217
Sbjct	5155	TAAACACCCTAAACAACCTGGCTTCAATCTACTTCTCCCGCCTCCTGGAAAAAAGGCGG	5214

Query	5218	GAGAAGCCCCGGCAGTATTGAAGCTGCTTCTTCGAATTTGCAATTC AATGTGAAATTCAC	5277
Sbjct	5215	GAGAAGCCCCGGCAGTATTGAAGCTGCTTCTTTGAATTTGCAATTC AATGTGAAATTCAC	5274
Query	5278	CACGAGGCTTGGCAAAAAGAGGGCTTTGACCTCTGTCTTTAGATTTACAGTCTAATGCTT	5337
Sbjct	5275	CACAAGGCTTGGTAAAAAGA-GGCTTTGACCTCTGTCTTTAGATTTACAGTCTAATGCTT	5333
Query	5338	ACTCAGCCATTTTACCTATGTTTCATCAACCGCTGATTATTTTCAACAAACCACAAAGATA	5397
Sbjct	5334	ACTCAGCCATTTTACCTATGTTTCATCAACCGCTGATTATTTTCAACAAACCACAAAGATA	5393
Query	5398	TTGGTACCCTATACTTGCTATTTGGCGCCTGAGCTGGAATAGTAGGCACAGCTCTAAGTC	5457
Sbjct	5394	TTGGTACCCTATACTTACTATTTGGCGCCTGAGCTGGAATGGTAGGCACAGCTCTAAGTC	5453
Query	5458	TCTTAATCCGTGCCGAATTAGGTCAACCAGGAACCTTATTAGGAGATGACCAAATCTATA	5517
Sbjct	5454	TCTTAATTCGTGCCGAATTAGGTCAACCAGGAACCTTACTAGGAGATGACCAAATCTATA	5513
Query	5518	ATGTAGTGGTAACCGCTCATGCATTTGTAATAATTTTCTTCATAGTTATGCCAATCATAA	5577
Sbjct	5514	ACGTAGTGGTAACCGCTCATGCATTTGTAATAATTTTCTTTATAGTTATACCAATCATAA	5573
Query	5578	TCGGAGGATTTGGGAACGTATTAGTCCATTAATGATTGGAGCACCCGACATAGCATTTC	5637
Sbjct	5574	TTGGAGGGTTTGGGAACGTATTAGTCCATTAATGATTGGAGCACCCGACATAGCGTTTC	5633
Query	5638	CACGAATAAACAAACATAAGCTTCTGATTACTTCCCCATCCTTTCTGCTCCTACTAGCAT	5697
Sbjct	5634	CACGAATAAACAAATATAAGCTTCTGATTACTTCCCCATCCTTCTGCTTCTACTAGCAT	5693
Query	5698	CTTCAATAATTGAAGCTGGTGCAGGTACAGGCTGAACGTCTATCCACCCCTAGCCGGTA	5757
Sbjct	5694	CCTCAATAATTGAAGCCGGTGCAGGTACAGGTTGAACGTCTATCCACCCCTAGCCGGTA	5753
Query	5758	ACTTAGCACATGCAGGAGCTTCCGTTGACTTAACCATTTTCTCCCTCCACCTTGCAGGTG	5817
Sbjct	5754	ACTTAGCACACGCAGGAGCTTCCGTTGACTTAACCATTTTCTCCCTCCATCTTGCAGGTG	5813
Query	5818	TATCCTCAATTCTAGGCGCTATTAACCTTATTACTACAATTATTAATAAAAACCACCAG	5877
Sbjct	5814	TATCTTCAATTCTAGGCGCCATTAACCTTATCACTACAATTATTAACATAAAAACCACCAG	5873
Query	5878	CTATATCACAGTACCAAACCTTTATTTGTTTGATCAGTCCTAATTACAGCAGTGTTAC	5937
Sbjct	5874	CTATATCACAAATACCAAACCTCTATTTGTTTGATCAGTCCTAATTACAGCAGTATTAC	5933
Query	5938	TATTACTAGCACTTCCAGTTCTAGCAGCAGGGATTACCATACTACTAACAGACCGTAACC	5997
Sbjct	5934	TACTACTAGCACTCCAGTTCTAGCAGCAGGAATTACCATGCTACTCACAGACCGCAACC	5993
Query	5998	TAAATACTACTTTCTTTGACCCTGCAGGAGGAGGAGATCCTATTCTATATCAACATCTAT	6057
Sbjct	5994	TAAATACTACTTTCTTTGATCCTGCAGGAGGAGGAGATCCTATTCTATATCAACATCTAT	6053
Query	6058	TTTGATTCTTTGGACATCCTGAAGTTTATATCCTCATTTTACCAGGCTTTGGAATAATCT	6117
Sbjct	6054	TCTGATTCTTTGGACACCCTGAAGTTTATATCCTCATTTTACCAGGTTTTGGAATAATCT	6113
Query	6118	CACATATTGTCACATACTATT CAGGTAAAAAAGAACCATTTGGCTATATAGGAATAGTAT	6177
Sbjct	6114	CTCACATCGTCACATACTATT CAGGTAAAAAAGAACCATTTGGCTACATAGGAATAGTAT	6173
Query	6178	GAGCTATAATATCCATCGGCTTCTTAGGCTTTATCGTATGAGCTCACCATATGTTTACAG	6237
Sbjct	6174	GAGCTATAATATCCATCGGCTTCTTAGGTTTATCGTATGAGCTCACCATATGTTTACGG	6233
Query	6238	TAGGTATAGATGTTGATACACGAGCATACTT CACATCAGCTACTATAATTATTGCTATCC	6297
Sbjct	6234	TAGGCATAGATGTTGACACACGAGCATACTT CACATCAGCCACTATAATCATTGCTATCC	6293
Query	6298	CCACTGGTGTAAAAGTGTTTAGTTGACTAGCGACCCTCCACGGAGGGAACATCAAATGAT	6357
Sbjct	6294	CCACTGGTGTAAAAGTATTTAGTTGACTAGCTACCCTCCATGGAGGAAATATCAAGTGAT	6353
Query	6358	CCCCTGCCCTATTATGAGCCTTAGGCTTTAAttttttATTACAGTAGGAGGCTGACAG	6417
Sbjct	6354	CCCCTGCTCTACTATGAGCCCTAGGCTTTATTTTTTTATTTACAGTAGGGGGCTAACAG	6413
Query	6418	GAATCGTACTAGCCAATTCATCCCTAGATATCGTACTCCACGACACATACTATGTGGTAG	6477
Sbjct	6414	GGATCGTATTAGCCAATTCATCCTTAGACATTGTA CTCCACGACACGTA CTATGTGGTAG	6473
Query	6478	CTCACTTCCACTATGTTTTATCCATAGGAGCAGTTTTCGCTATTATAGGAGGATTTGTTC	6537
Sbjct	6474	CTCACTTCCACTATGTTTTATCCATAGGAGCAGTTTTCGCTATTATAGGAGGATTTGTTC	6533

Query	6538	ACTGATTTCCATTATTCTCAGGATATACACTCAATCAAACCTGAGCAAAAATTCACTTCA	6597
Sbjct	6534	ATTGATTTCCATTATTCTCAGGATATACACTTAACCAAACCTGAGCAAAGATTCACTTCA	6593
Query	6598	CAATCATATTTGTAGGTGTAATAATAACATTTCTCCCTCAACATTTTCTTGGCTTAGCAG	6657
Sbjct	6594	CAATCATATTTGTAGGTGTAATAATAACATTTCTCCCTCAACACTTTCTCGGCTTGTGAG	6653
Query	6658	GAATGCCACGACGTTATTCTGACTACCCAGATGCATATACAACATGAAACACTATTTTCAT	6717
Sbjct	6654	GTATACCACGACGCTATTCCGACTACCCAGATGCATATACAACATGAAACACTATTTTCAT	6713
Query	6718	CTATAGGATCCTTCATCTCACTCACAGCAGTTATACTAATAGCATTTATAGTATGAGAAG	6777
Sbjct	6714	CAATAGGATCCTTCATCTCACTCACAGCAGTTATACTAATAGCATTTATAGTATGAGAAG	6773
Query	6778	CATTCGCATCCAAACGAGAAGTACTAGCAGTAGAACTAACTACAACCAACCTCGAATGAC	6837
Sbjct	6774	CATTTGCATCCAAACGAGAAGTACTAGCAGTAGAACTAACTACAACCAACCTCGAATGAC	6833
Query	6838	TAAATGGGTGCCCTCCACCTTACCACACATTTGAAGAACCTGTTTATATTAACCTTAAAT	6897
Sbjct	6834	TAAATGGGTGCCCTCCACCTTATCACACATTTGAAGAACCTGTTTACGTCAATCTTAAAT	6893
Query	6898	AAGAAAGGAAGGAATCGAACCCCTCTGACTGGTTTTCAAGCCAATATCATAACCATTATG	6957
Sbjct	6894	AAGAAAGGAAGGAATCGAACCCCTCTGACTGGTTTTCAAGCCAACATCATAACCATTATG	6953
Query	6958	TCTTTCTC-AAAATATGAGATATTAGTAAAACCTTACATAACTTTGTCAAAGTTAAATTAT	7016
Sbjct	6954	TCTTTCTCAAATAATGAGATATTAGTAAAATTTACATAACTTTGTCAAAGTTAAATTAT	7013
Query	7017	AGGTTAAAACCTTATATATCTCCATGGCATATCCCTTCCAAC TAGGATTTCAAGATGCAA	7076
Sbjct	7014	AGGTTAAAACCTTATATATCTCCATGGCGTACCCCTTCCAAC TAGGATTTCAAGATGCAA	7073
Query	7077	CATCTCCTATTATAGAAGAACTTCTCCATTTCCACGACCATACACTAATAATCGTATTCC	7136
Sbjct	7074	CATCTCCATTATAGAAGAACTTCTCCACTTCCATGACCATACACTAATAATTGTATTCC	7133
Query	7137	TAATTAGCTCTTTAGTATTATATATTTTTCATTAATACTTACAACCAAACCTAACCCACA	7196
Sbjct	7134	TGATCAGCTCCTTAGTATTATATATTTTCTACTAATACTTACAACCTAAATTAACCTATA	7193
Query	7197	CAAGTACTATAGATGCTCAAGAAGTAGAAACCATTTGAACTATCCTACCAGCCATCATT	7256
Sbjct	7194	CAAGTACCATAGATGCCAAGAAGTAGAAACCATTTGAACTATCCTACCAGCCATCATT	7253
Query	7257	TAATCTTAATTGCCCTACCATCTCTACGAATCTTTATATAATAGACGAAATTAACAACC	7316
Sbjct	7254	TAATCTTAATCGCCCTACCATCCCTACGAATCTTTACATAATAGACGAAATCAACAACC	7313
Query	7317	CTTCTTTAACCGTCAAACCTATAGGACACCAATGATACTGAAGCTATGAATATACAGATT	7376
Sbjct	7314	CTTCTTTAACTGTCAAACCTATAGGCCATCAGTGATACTGAAGCTATGAGTATACAGATT	7373
Query	7377	ATGAAGACCTAAACTTTGATTCTACATAGTACCTACATCAGATCTAAAACCCGGAGAGT	7436
Sbjct	7374	ATGAAGACCTGAACTTCGACTCCTACATGGTACCTACATCGGATCTAAAACCCGGAGAAC	7433
Query	7437	TACGACTACTTGAAGTTGATAACCGAGTAGTTTTACCATAGAAATAACAATTCGAATAC	7496
Sbjct	7434	TACGATTACTTGAAGTTGACAACCGAGTAGTATTACCTATAGAAATAACGATTCGAATAC	7493
Query	7497	TCATTTTCATCCGAAGATGTCCTGCACTCATGAGCCGTACCTTCCCTTAGGCCTAAAACTG	7556
Sbjct	7494	TCATTTTCATCCGAAGATGTCCTGCACTCATGAGCTGTACCTTCCCTTAGGCCTAAAACTG	7553
Query	7557	ACGCTATTCCAGGACGCCTAAACCAAACGACTTTAGTAGCTATACGACCCGGATTATACT	7616
Sbjct	7554	ATGCTATCCAGGACGCCTAAATCAAACAACCTTAGTAGCTATACGACCCAGGACTATATT	7613
Query	7617	ACGGACAATGCTCAGAAATCTGCGGATCAAACCACAGCTTCATACCTATTGTTCTTGAAC	7676
Sbjct	7614	ACGGACAATGCTCAGAAATCTGCGGATCAAACCATAGCTTTATACCTATCGTTCTTGAAC	7673
Query	7677	TGGTTCCATTAAGTACCTTTGAAAAATGATCTACATCAATACTTTAAAGTCATTAAGAAG	7736
Sbjct	7674	TAGTCCCATTAAGCACCTTTGAAAAATGATCTACATCAATACTTTAAAGTCATTAAGAAG	7733
Query	7737	CTATATAGCATTAAACCTTTTAAAGTTAAAGACTGAGGGTCTAAGTCCCTCCTTAATGATAT	7796
Sbjct	7734	CTATATAGCATTAAACCTTTTAAAGTTAAAGACTGAGGGTCTAAGTCCCTCCTTAATGATAT	7793
Query	7797	GCCACAACCTAGACACATCAACATGATTTATTAATATCACATCAATAATTATCACACTATT	7856
Sbjct	7794	GCCACAACCTAGACACATCAACATGATTTATTAATATTACATCAATAATTATTACATTATT	7853

Query	7857	TATTATATTCCTCAACTAAAAATTTCAAACACCTGTACCCCTATGACCCAAAACCTAAAAAC	7916
Sbjct	7854	TATCATATTCCTCAACTAAAAATTTGAAACACCTATACCCCTATGACCCAAAACCTAAAAAC	7913
Query	7917	AACTAAAGTGACAAAATATACCACCCCTTGAGAATCAAATGAACGAAAATCTATTTCGCT	7976
Sbjct	7914	AACTAAATAACAGGACATATACCCCTTGAGAATCAAATGAACGAAAATCTATTTCGCT	7973
Query	7977	TCTTTTCGCTACCCCAACAATAATGGGCTACCTATTGTTATCCTAATCATCATATTCCCA	8036
Sbjct	7974	TCTTTTCGCTACCCCAACAATAATGGGCTACCCATTGTCATCCTAATCATCATATTCCCA	8033
Query	8037	AGTATTCTATTTCCCTCATCCAACCGATTAATTAACAACCGTCTAATCTCTATCCAACAA	8096
Sbjct	8034	ACTATTCTATTTCCCTCGTCCAACCGATTAATCAGTAACCGCTAATCTCTATCCAACAA	8093
Query	8097	TGACTAGTTCAACTTACATCAAAACAAATAATGTCCATCCATAATCACAAGGGACAACA	8156
Sbjct	8094	TGACTAATTCAACTTACATCAAAACAAATGATGTCCATTATAATCACAAGGGACAATA	8153
Query	8157	TGAACACTAATACTTATATCACTTATTTTATTTCATTGGCTCTACTAATCTTCTAGGCTTA	8216
Sbjct	8154	TGAACACTAATGCTCATATCATTATTTTATTTCATTGGCTCCACTAATCTACTAGGCTTA	8213
Query	8217	CTACCACACTCATTACACCTACCACACAACCTATCAATGAATCTAGGCATAGCTATTCCA	8276
Sbjct	8214	CTACCACACTCATTACACCTACTACACAACCTATCAATGAATCTAGGCATGGCTATTCCA	8273
Query	8277	CTATGAGCAGGCACAGTATTATAGGCTTTTCGTCAAAAACAAAAGCATCCTTAGCCCAT	8336
Sbjct	8274	CTATGAGCAGGCACAGTGTTCGTAGGCTTCGTCAAAAACAAAAGCATCCTTAGCCCAT	8333
Query	8337	TTTCTTCCCAAGGAACTCCCATCTTCTAATTCCTAATCACTAGTTATCATTGAAACCATT	8396
Sbjct	8334	TTTCTTCCCAAGGAACTCCTATCTTCTAATTCCTAATCACTAGTTATCATTGAAACTATT	8393
Query	8397	AGCCTATTTATTCAACCAATAGCCTTAGCTGTACGACTTACAGCTAATATCACTGCTGGA	8456
Sbjct	8394	AGCCTATTTATCCAGCCAATAGCTTAGCTGTACGACTTACAGCTAATATCACTGCTGGA	8453
Query	8457	CATCTCCTGATGCACCTAATTGGAGGAGCAACACTGGCTCTAATAAATATTAGTCCCTACC	8516
Sbjct	8454	CATCTCCTAATACACCTAATTGGAGGAGCAACACTAGCTCTAATAAATATTAGTCCCTACT	8513
Query	8517	ACAGCCCTCATTACATTATCATTCTCATTCTTACTCACTATCCTTGAATTTGCAGTAGCC	8576
Sbjct	8514	ACAGCCCTCATTACATTATTCTTATCTTACTCACCATCCTTGAATTTGCAGTAGCC	8573
Query	8577	TTAATTCAGCTTACGTATTTACTCTCTTAGTAAGCCTTTATCTACACGACAACACTTAA	8636
Sbjct	8574	TTAATTCAGCTTACGTATTTACCCTCTTAGTAAGCCTTTACTACACGACAACACTTAA	8633
Query	8637	TGGCCCAACCAACTCACTCATAACCACATAGTTAACCCAGCCCATGACCACTTACAGGAG	8696
Sbjct	8634	TGGCCCAACCAACCCTCATAACCACATAGTTAACCCAGCCCTGACCACTTACAGGAG	8693
Query	8697	CTCTATCAGCTCTCCTTATAACTTCAAGCTTAGCTATATGATTCACCTTCAACTCAATAT	8756
Sbjct	8694	CTTTATCAGCCCTCCTTATAACATCGGGCTTAGCCATATGATTCATTTTAACTCAATAC	8753
Query	8757	TACTACTATATTTAGGATTAACAACCTAATTAATACTATATCAATGATGACGAGATG	8816
Sbjct	8754	TACTACTATACCTAGGACTAACAACCTAATTAATACTATATCAATGATGACGAGATG	8813
Query	8817	TTATCCGAGAAAGCACATTTCAAGGTCACCACACCAACCGTCCAAAAGGACTTCGAT	8876
Sbjct	8814	TTATTCGAGAAAGCACATTTCAAGGCCATCATAACCAACCGTCCAAAAGGACTTCGAT	8873
Query	8877	ACGGCATGATCTTATTCTGCTCAGAAGTATTTTCTTCTCAGGATCTTCTGAGCCT	8936
Sbjct	8874	ACGGCATGATCCTATTCTTATTCTCAGAAGTATTTTCTTCTCAGGATCTTCTGAGCCT	8933
Query	8937	TTTACCCTCAAGCTTAGCCCGACACCTGAACTAGGCGTTGCTGACCACCTACAGGCA	8996
Sbjct	8934	TTTACCCTCAAGCTTAGCCCAACACCTGAGCTAGGTGGTTGCTGACCACCTACAGGCA	8993
Query	8997	TCCATCCCTTAAATCCCCTAGAACTACCTCTTCTCAATACCTCAGTACTTCTAGCTTCTG	9056
Sbjct	8994	TCCACCCCTAAACCCCCTAGAACTACCCCTTCTCAATACCTCAGTACTTCTAGCTTCTG	9053
Query	9057	GAGTCTCCATCACTTGAGCCACCATAGCCTGATAGAAGGAGATCGCAAACATATATTAC	9116
Sbjct	9054	GAGTCTCCATCACCTGAGCCACCATAGCCTAATAGAAGGAGACCGCAAACATATATGAC	9113
Query	9117	AAGCCCTATTCATTACAATCTCTCTAGGAATCTACTTTACCTTACTTCAAGCCTCAGAA	9176
Sbjct	9114	AAGCCCTATTTATTACAATCTCCCTAGGAGTCTACTTTACCTTACTTCAAGCCTCAGAA	9173

Query	9177	ACTACGAAGCATCATTTACAATTTCTGATGGCGTATATGGTTCAACCTTCTTTGTAGCCA	9236
Sbjct	9174	ACTACGAAGCATCATTTACAATTTCTGATGGCGTATATGGCTCAACCTTCTTTGTAGCCA	9233
Query	9237	CAGGCTTCCACGGACTACACGTAATTATTGGTTCTACCTTTCTTATTGTATGCTTCTTAC	9296
Sbjct	9234	CAGGCTTCCATGGACTGCACGTAATCATTGGCTCTACCTTTCTTATTGTATGTTTCTTAC	9293
Query	9297	GCCAACTAAAGTTCACCTTTACATCTAATCACCACCTTTGGCTTTGAAGCAGCTGCCTGAT	9356
Sbjct	9294	GCCAACTAAAATTCACCTTTATATCCAACCACCACCTTTGGCTTCGAAGCAGCTGCCTGAT	9353
Query	9357	ACTGACATTTCTAGACGTAGTATGACTATTTCTTATACGTATCTATTTATTGATGAGGGT	9416
Sbjct	9354	ACTGACATTTCTAGACATAGTGTGATTATTTCTTATATGTATCTATTTATTGATGAGGGT	9413
Query	9417	CATATTCTCTTAGTACCAAATAGTACAATTGACTTCCAATCAATTAGTTTCGGTATAATC	9476
Sbjct	9414	CATATTCTTTTAGTATCAAATAGTACAATTGACTTCCAATCAATTAGTTTCGGTTAATC	9473
Query	9477	CGAAAGAGAATAATAAACTTAATATTCACCCTATTTGTCAACACAACATTAGCTTCACTT	9536
Sbjct	9474	CGAAAAAGAATAATAAACTTAATATTCACCCTACTTGTAAACACAACATTAGCTTCACTT	9533
Query	9537	CTTGATTAATTGCATTCTGATTACCTCAACTAAACATCTATACGAAAAAGCCACTCCC	9596
Sbjct	9534	CTTGACTAATCGCATTCTGATTACCCCAACTAAACATCTATACGAAAAAGCCACCCCC	9593
Query	9597	TATGAATGCGGATTTGATCCTATAGGATCAGCACGTCTGCCTTTTTCTATAAAATTTTTT	9656
Sbjct	9594	TATGAATGCGGATTTGATCCTATAGGATCAGCACGTTACCTTCTCCATAAAATTTTTT	9653
Query	9657	CTAGTGGCCATCACATTTCTCCTATTCGACCTAGAAATTGCACTCCTACTACCACTACCA	9716
Sbjct	9654	CTAGTAGCTATCACATTTCTTCTATTCGACCTAGAAATTGCACTCCTATTACCATTACCA	9713
Query	9717	TGAGCGTCTCAAACAACAAATCTTAAAACCATACTTACAATAGCGCTGACTCTTATCTCA	9776
Sbjct	9714	TGAGCATCCCAAACAACAGACCTTAAAACCATACTTACAATAGCACTAATCCTTATCTCA	9773
Query	9777	TTACTAGCCATTAGCCTAGCCTACGAATGAACTCAAAAAGGACTAGAATGAACTGAATAT	9836
Sbjct	9774	TTACTAGCTATTAGCCTAGCCTACGAGTGAATCAAAAAGGACTAGAATGGACCGAATAT	9833
Query	9837	GATAATTAGTTTAAACAAAAACAAGTGATTTGACTCACTAGACTATGATTTACTTCATAA	9896
Sbjct	9834	GGTAAATTAGTTTAAATAAAAAACAAGTGATTTGACTCACTAGACTATGACTTATCTCATAA	9893
Query	9897	TTATCAAATGTCCCTAGCCCATATCAATATCTTTCTAGCATTATAGTATCCCTTATAGG	9956
Sbjct	9894	TTATCAAATGTCCCTAGCCACATCAATATTTTTCTGGCATTATAGTATCTCTTATAGG	9953
Query	9957	ACTACTAATATATCGATCCCATCTAATATCCTCACTACTGTGCCTAGAAGGTATAATATT	10016
Sbjct	9954	ATTGCTAATATACCGATCCCATTAAATATCTTCACTACTATGCCTAGAAGGCATGATACT	10013
Query	10017	ATCACTATTTATTATAGCAACCATAATAATCCTAAACTCCCACCTTACACTAGCTAGTAT	10076
Sbjct	10014	ATCACTATTCATTATAGCAACCATAATAATCCTAAACTCCCACCTTACACTAGCCAGCAT	10073
Query	10077	AATACCCATTATCCTACTAGTATTCGAGCCTGTGAAGCAGCGCTAGGATTATCTCTCCT	10136
Sbjct	10074	AATACCCATTATCCTACTAGTATTCGAGCCTGTGAAGCAGCACTAGGATTATCTCTCCT	10133
Query	10137	AGTTATAGTATCCAACACCTATGGAATAGACTATGTACAAAACCTCAATCTACTCCAATG	10196
Sbjct	10134	AGTCATAGTATCCAATACCTATGGAATAGACTATGTACAAAACCTCAATCTACTTCAATG	10193
Query	10197	CTAAAAATCATCTTCCCTACTATAATACTAATTCCTCACTAACATGATTATCaaaaaaaGT	10256
Sbjct	10194	CTAAAAATCATCTTCCCACTATAATACTAATTCCTCACTAACATGATTATCAAAAAAAGC	10253
Query	10257	ATAATCTGAATTAATACTACAATATACAGCCTACTAATTAGTCTTATTGCTTACCCCTT	10316
Sbjct	10254	ATAATCTGAATTAACACTACAATATATAGCCTATTAATTAGCCTTATCACCCCTACCCCTT	10313
Query	10317	CTAAACCAACCTGACGATAATAGCCTCAATTTCTCACTACTATTTTTCTCCGATCCCTTA	10376
Sbjct	10314	CTAAATCAACCTAATGACAACAGCCTCAATTTCTCACTACTATTTTTCTCCGATTCCCTA	10373
Query	10377	TCAGCACCTCTCCTAGTACTAACAACATGACTGCTACCATTAATACTAATAGCTAGTCAA	10436
Sbjct	10374	TCAGCACCCCTCTTAGTACTAACAACATGATTACTACCACCTAATACTTATAGCCAGCCAA	10433
Query	10437	CACCACCTATCAAAAGAACCATTAACTCGaaaaaaCTATATATTACTATACTAGTCTTA	10496
Sbjct	10434	TACCACCTATCAAAAGAATCACTAACCAGAAAAAACTGTATATTACCATGCTAATCTTA	10493

Query	10497	CTCCAAATATTCTAATTATAACCTTCACTGCTACAGAACTAATCTCCTTTTACATTTTA	10556
Sbjct	10494	CTCCAAATATTCTAATCATAACCTTTACTGCTACAGAACTAATCTCCTTCTACATTCTA	10553
Query	10557	TTCGAAGCAACACTAGTACCAACACTTATCATCATTACCCGCTGAGGTAACCAACAGAA	10616
Sbjct	10554	TTCGAAGCAACACTAGTACCAACACTTATCATTATTACCCGCTGAGGTAACCAACAGAA	10613
Query	10617	CGACTAAATGCAGGCCTTTACTTTTCTATTCTATACGTTAGTAGGATCCCTCCCCTTCTA	10676
Sbjct	10614	CGACTAAATGCAGGCCTTTATTTTCTATTTTATACACTAGTAGGGTCACTCCCCTTCTA	10673
Query	10677	GTTGCATTAATCACTACCCAAAACTTTACAGGCTCGCTAAATCTACTATTAACCCAATAC	10736
Sbjct	10674	GTTGCATTAATCACTATCCAAAACTTTACAGGCTCACTAAACCTATTATTAACCCAATAC	10733
Query	10737	TGAACCCAACCACTACCCAACCTCTGATCTAGTATATTCCTATGAACAGCATGCATAATA	10796
Sbjct	10734	TGAACCCAACCACTACCCAACCTCTGATCCAGTATATTCCTATGAATAGCATGCATAATA	10793
Query	10797	GCATTCATGGTAAAAATACCACTATATGGCCTACATCTATGACTACCAAAAGCACATGTA	10856
Sbjct	10794	GCATTCATGGTAAAAATACCACTATACGGCCTAAACCTATGACTACCAAAAGCACATGTA	10853
Query	10857	GAAGCCCTATCGCTGGCTCTATAGTACTAGCAGCTATCCTACTAAAACCTAGGAGGCTAT	10916
Sbjct	10854	GAAGCTCCTATTGCTGGCTCCATAGTGTAGCAGCTATTCTACTAAAATAGGAGGCTAT	10913
Query	10917	GGAAATATTACGAATTACCACTATCCTAAATCCCTAACAAGCTACATAGCTTACCCTTT	10976
Sbjct	10914	GGAACTACTACGAATTACCACCATCCTAAACCCCTAACAAGCTATATAGCTTATCCATTC	10973
Query	10977	CTCATATTATCCTATGAGGAATAATTATAACAAGCTCCATTTGCCTCCGCCAACAGAT	11036
Sbjct	10974	CTCATACTATCCTTATGAGGGATGATTATAACAAGCTCTATCTGCCTTCTCCAACAGAT	11033
Query	11037	CTTAAATCTCTCATCGCTTACTCTTCCGTTAGCCACATGGCTCTAGTAATTGTAGCTATC	11096
Sbjct	11034	CTAAAAATCCCTCATTGCTTACTCCTCTGTAGTCACATAGCTCTAGTAATTGTAGCCATC	11093
Query	11097	CTCATCCAAACACCCTGAAGTTACATAGGAGCTACAGCCCTAATAATCGCACACGGTCTA	11156
Sbjct	11094	CTTATCCAAACACCCTGAAGTTATATAGGAGCTACAGCCCTAATAATTGCACACGGCCTA	11153
Query	11157	ACCTCATCAATATTATTCTGCCTAGCAAACCTCAAACCTATGAACGAATCCATAGTCGAACT	11216
Sbjct	11154	ACTTCATCAATACTATTCTGTCTAGCAAACCTCAAACCTATGAACGAATCCATAGTCGAACT	11213
Query	11217	ATAATCCTAGCCCGGGACTCCAAACACTCCTACCCTAATAGCAGCATGATGATTAATA	11276
Sbjct	11214	ATAATCCTAGCCCATGGACTCCAAACACTCCTACCCTTATAGCAGCATGATGATTAATA	11273
Query	11277	GCAAGCCTAACCTAATCTAGCTTTACCCCTACAATTAATCTGATCGGAGAATTATTTGTA	11336
Sbjct	11274	GCAAGCCTAACCAACCTAGCTTTACCCCTACAATTAACCTAATTGGAGAATATTTGTA	11333
Query	11337	GTTATAACATCATTCTCATGATCCAACATCACCATTATCCTAACAGGAACCAACATTATC	11396
Sbjct	11334	GTTATAACATCATTCTCATGATCCAATATCACTATTATCCTAATAGGAATTAATATCACC	11393
Query	11397	ATTACTGCCCTATATTCCTCTACATATTAATCACAACACAACGAGGCAAATATACATAT	11456
Sbjct	11394	ATTACCGCTTATATTCCTCTAGATACTAATCACAACACAACGAGGCAAATATACACAT	11453
Query	11457	CACATTAACAATGTAAGCCATCATTACACGAGAAAACGCACTCATAGCTCTACATTTA	11516
Sbjct	11454	CACATTAATAATGTAAGCCATCATTACACGAGAAAACGCACTCATGGCCCTACACCTA	11513
Query	11517	CTACCCCTTTTACTACTATCACTCAATCCTAAAATCATCCTAGGAGTCACATACTGTAGA	11576
Sbjct	11514	CTACCCCTCTTACTACTATCACTCAACCCCAAAAATCATCTAGGAATCACATACTGTAAA	11573
Query	11577	TATAGTTTAAACAAAACATTAGATTGTGAATCTAACATAGAAAACCTAAAACCTTCTTATT	11636
Sbjct	11574	TATAGTTTAAACAAAACATTAGATTGTGAATCTAACATAGAAAACCTAAAACCTTCTTATT	11633
Query	11637	TACCGAAAAAGTATGCAAGAACTGCTAATTCATGCTCCCATGTATAACAACATGGCTTTT	11696
Sbjct	11634	TACCGAAAAAGTATGCAAGAACTGCTAATTCATGCTCCCATGTATAACAACATGGCTTTT	11693
Query	11697	TCAACTTTTAAAGGATAAGAGTTATCCATTGGCTTAGGAACCAAAAAATTTGGTGCAACT	11756
Sbjct	11694	TCAACTTTTAAAGGATAAGAGTTATCCATTGGCTTAGGAACCAAAAAATTTGGTGCAACT	11753
Query	11757	CCAAATAAAAAGTAATAAACATATTTCTTCTTATACTAACTTCATTAATTATACTAAC	11816
Sbjct	11754	CCAAATAAAAAGTAATAAACATATTTCCCTCTTATATTAACCTCACTAATTATACTAAC	11813

Query	11817	ACTACCAATCATCATAACCGCCCTCAATTACCACAAAAATGAATTGTACCCCTTACCATGT	11876
Sbjct	11814	ACTACCAATCATCATAACTACTCTTAATTATCACAAAAATGATTATACCCCTTACTATGT	11873
Query	11877	AAAAAACATTATTTTCATATTCCTTCACTACTAGCCTCATCCCTACAATAATATTCATCCA	11936
Sbjct	11874	AAAAAACATTATTTTCATACTCCTTTATTACTAGTCTTATCCCCACAATAATATTTATCCA	11933
Query	11937	CTCCGGCCAAGAAACAATTATCTCAAACCTGACATTGAATAACCATTCAAACACTAAAAC	11996
Sbjct	11934	ATCTGGCCAAGAAACAATTATCTCAAACCTGACATTGAATAACTATTCAAACACTAAAAC	11993
Query	11997	ATCACTCAGCTTTAAACTAGATTATTTTTCAATAATCTTTGTACCAGTGGCTCTCTTCGT	12056
Sbjct	11994	ATCACTCAGTTTTAAACTAGATTACTTTTTCAATAATCTTCGTACCAGTAGCTCTCTTCGT	12053
Query	12057	CACATGATCTATCATAGAGTTCTCAATATGATATATACACTCAGACCCCTACATTACTCA	12116
Sbjct	12054	CACATGATCCATCATAGAGTTCTCAATATGATATATACACTCAGATCCCTACATCACTCA	12113
Query	12117	ATTCTTTAAGTATCTCCTCACATTCTTAATTACCATAATAATTCTAGTCACAGCAAACAA	12176
Sbjct	12114	ATTCTTTAATATCTCCTCACATTCTTAATTACCATAATAATTCTAGTTACAGCTAACAA	12173
Query	12177	CCTATTCCAACTTTTTATTGGGTGAGAAGGAGTCGGAATCATGTCATTTTTACTAATCGG	12236
Sbjct	12174	TCTATTCAGCTCTTCATTGGATGAGAAGGAGTTGGAATCATATCATTTCTATTAATCGG	12233
Query	12237	CTGATGATATGGCCGTACAGATGCAAATACCGCCGCTTACAAGCAATCCTATACAATCG	12296
Sbjct	12234	CTGATGATACGGTCGTACAGATGCAAACACCGCTGCCCTACAAGCAATCTTATACAATCG	12293
Query	12297	CATCGGAGATATTGGATTTATTATATCAATGGCATGATTCCTACTAAACGTAAACTCATG	12356
Sbjct	12294	CATTGGAGACATCGGATTTATTGTATCAATAGCGTGATTCCTACTGAATGTAAACTCATG	12353
Query	12357	GGATCTACAACAAATCTTTACTCAGCCCCGATCACACAAACCTCCCATTAATCGGACT	12416
Sbjct	12354	AGACCTACAACAAATCTTTATACTTAACCCTAACACACAAACCTCCCATTAATCGGACT	12413
Query	12417	CCTACTAGCTGCAACCGGAAAATCAGCCCAATTCGGCCTACACCCATGACTTCCATCAGC	12476
Sbjct	12414	CTTACTAGCTGCAACCGGAAAATCTGCCCAATTTGGCCTACACCCATGACTCCCATCAGC	12473
Query	12477	TATAGAAGGTCCTACACAGTATCAGCCTTACTTCACTCAAGTACAATAGTTGTGCGAGG	12536
Sbjct	12474	CATAGAAGGCCCTACACCGGTATCAGCCCTACTCCTCAAGCACAATAGTTGTGCGAGG	12533
Query	12537	TGTATTCCTATTAATCCGTTTCCACCCTTAAATAGAAATAACAAGCTAGCCCAAACACT	12596
Sbjct	12534	CGTATTCCTACTAATCCGTTTTACCCTTAAATAGAAACTAACAAGCTGGCCCAAACACT	12593
Query	12597	AACATTATGTTTAGGAGCTATTACCACACTATTTACAGCAATATGTGCACTCACCCAAA	12656
Sbjct	12594	GACACTGTGCTTAGGAGCTATTACCACACTATTTACAGCAATATGTGCACTTACCCAAA	12653
Query	12657	CGATATTaaaaaaTTATTGCCTTCTCTACCTCAAGCCAACCTAGGACTAATAATCGTAAC	12716
Sbjct	12654	CGACATCAAAAAAATTATTGCCTTCTCCACCTCAAGCCAACCTAGGACTAATGATTGTAAC	12713
Query	12717	TATTGGTATTAATCAACCATATCTAGCTTTCCTACACATCTGTACCCACGCATTCTTCAA	12776
Sbjct	12714	CATCGGTATTAATCAACCATATCTAGCTTTCCTACACGTCTGCCTCACGGCTTCTTCAA	12773
Query	12777	AGCCATATTATTTATATGCTCTGGATCCATCATCCACAACCTAAACAACGAACAAGACAT	12836
Sbjct	12774	AGCCATATTATTTATATGCTCTGGATCCATTATTCATAACTTAACAATGAGCAAGACAT	12833
Query	12837	CCGAAAAATAGGCGGCCTATTCAAAGCCATGCCTTTTACCACAACCTCACTCATCGTAGG	12896
Sbjct	12834	CCGAAAAATAGGCGGCCTATTCAAAGCCATGCCCTTACCACGACCTCACTCATCGTAGG	12893
Query	12897	TAGCTTGGCACTCACTGGCACACCATTCTCACAGGATTCTACTCTAAAGATTTAATTAT	12956
Sbjct	12894	CAGCCTAGCACTAACCGGCATACCATTCTCACAGGATTCTACTCAAAGACCTAATCAT	12953
Query	12957	CGAAACCGTAATATGTATATACCAACGCTGAGCCCTAACAATTACCCTAATTGCCAC	13016
Sbjct	12954	CGAAACTGCCAACACGTATACACCAACGCTGAGCCCTAACAATTACCCTAATTGCCAC	13013
Query	13017	ATCCTTAACGGCTGTATACAGCACCCGAATTATCTTCTTCGCACTCCTAGGACAGCCAG	13076
Sbjct	13014	ATCCCTAACAGCTGTATACAGCACTCGAATTATCTTCTTCGCACTCTAGGACAACCCAG	13073
Query	13077	CTTTCCCACTCTTATTATAATTAATGAAAACAACCCCTCATTAACTCTATCAAACG	13136
Sbjct	13074	CTTTCCCACTCTCATACAATCAACGAAAACAACCCCCACTAATGAACTCCATCAAACG	13133

Query	13137	ACTCCTAATTGGAAGCATCTTTGCCGGCTTCTTTATCTCTAACAATATCTATCCTACCAA	13196
Sbjct	13134	ACTTCTAATTGGAAGCATTCTGCCGGCTTCTTTATCTCTAACAACATCTATCCCACCAA	13193
Query	13197	TATCCAGAAATAACTATACCTTATTATCTAAAAACAACAGCCCTTGTCGTAACCATTTT	13256
Sbjct	13194	TATCCAGAAATAACCATACCTTATTATCTAAAAATAATAGCCCTCATCGTAACCATTCT	13253
Query	13257	AGGCTTCACCTTAGGACTTGAACCTAAGTTTAATAACATATAACTTAAACCTTAAACACCC	13316
Sbjct	13254	AGGCTTTACCTTAGGACTTGAACCTAAGCTAATAACATATAACTTAAACCTTAAACACCC	13313
Query	13317	CTCAAACCTATTTAACTTCTCAAATCTTCTCGGATATTTCCCAACCATCATACACCGACT	13376
Sbjct	13314	ATCAAACCTATTCAACTTTTCAAACCTCTCTCGGATATTTCCCAACTATCATGCACCGATT	13373
Query	13377	CTCTCCACTCATAAACCTATCTATAAGTCAAAAATCAGCATCCCTATTATTAGACTCAAT	13436
Sbjct	13374	CCCTTCACTCATAAACCTATCTATAGGCCAAAAATCAGCATCCCTATTATTAGACTCAAT	13433
Query	13437	CTGACTAGAAGCTATCTACCAAAATCCATCTCTCACCTTCAATTAAGCCTCAACACT	13496
Sbjct	13434	CTGATTAGAAGCTATATTACCAAAATCTATCTCCACCTTCAAATTAAGCCTCAACACT	13493
Query	13497	AATCTCAAACCAAAAAGGCCAATTAACCTATATTTCTCTCATTCTTATCACTATGAC	13556
Sbjct	13494	AACCTCAAACCAAAAAGGCCAATCAAGCTATATTTCTTTCAATCTCATCACCATGGC	13553
Query	13557	CCTAAGTCTACTCTACTTAACTCCACGGTAACCTCTAAAATAACTAAAACACCAATA	13616
Sbjct	13554	TCTAAGCCTACTTCTACTTAACTCCACGAGTAACCTCCAAAATAACTAAAACACCAATA	13613
Query	13617	AATAAAGATCAACCCGTAACAATAAATAACTCAAACCCATAACTATATAACGCAGCAACT	13676
Sbjct	13614	AACAAAGACCAACCTGTAACAATAACAACCTCAAACCCATAACTATACAATGCAGCAACT	13673
Query	13677	CCTATAAATTCCCCATTA AAAACCCCTGAATCCTCAGTATCACAAATTA CTCAATCCTCC	13736
Sbjct	13674	CCTATAATTTCTCCATTA AAAACCCCTGAATCACCAGTATCACAAATTA CTCAATCCTCC	13733
Query	13737	ACACCACTGAATCCAAGCATAA CTCTGTCTGTCATCTTTCAACTACTCA taaaaactaaa	13796
Sbjct	13734	ACAGCACTAAATCCAACATAATCTCTACCTCTCCACCTTTCAATACGTATA AAAACTAAA	13793
Query	13797	acaaaactccataaacaaccc taaaaataaatgcaccta aaaaacaacCATATTAGATACCCAC	13856
Sbjct	13794	ACGAACTCCATTAACAAGCCTAAAATAAACGCACCTAAAACAACCATATTAGATACCCAC	13853
Query	13857	ACTTCTGGTACTGCTCAGTAGCCATAGCCGTTGTATaacc aaaaaaccaccaacataccc	13916
Sbjct	13854	ACTTCGGTACTGCTCAGTAGCCATAGCAGTTGTATAACCAAAAACCACTAACATGCCT	13913
Query	13917	cccaataaatcaaaaatactatcaacccccaaaaatgacc cGCCAAAATTAAGATAATC	13976
Sbjct	13914	CCCAAATAAATTA AAAATATTATCAATCTAAAACGACCCGCCAAAATTAAGACAATT	13973
Query	13977	CCACAACCAACCCACCCTCACAATTAACACCAGGCCCCATAAATAGGAGAAGGCTTT	14036
Sbjct	13974	CCACAACCAACCCACCCTTACAATTAATACTAGACCCCATAAATAGGTGAAGGTTTT	14033
Query	14037	GAAGAAAACCTACAAAACCAATC AAAAAATAACACTCAAAATAAACACTACATACGTC	14096
Sbjct	14034	GAAGAAAACCTACAAAACCAACCAAAAAACAACACTCAAGATAAATACTACATATGTC	14093
Query	14097	ATCATTATTCCACATGGAATCTA ACCATGACTAATGACATGAAAAATCATCGTTGACT	14156
Sbjct	14094	ATCATTATTCTACATGGAATCTA ACCATGACTAATGACATGAAAAATCATCGTTGACT	14153
Query	14157	TCAACTACAAGAACCTTAATGACCAACATCCGAAAATCCACCCGCTAGCCAAAATTAT	14216
Sbjct	14154	TCAACTACAAGAACCTTAATGATCAACATCCGAAAATCCACCCACTAATCAAAATCAT	14213
Query	14217	CAATCACTCATTTATCGACCTACCAACCCATCAAACATCTCATCCTGATGAACTTTGG	14276
Sbjct	14214	TAATCACTCATTTATTGACCTACCAGCCCATCAAACATCTCATCCTGATGAACTTTGG	14273
Query	14277	CTCCCTTTTAGGAATATGCCTAATCTTACAAATTATCACGGGCCTATTCTTAGCCATACA	14336
Sbjct	14274	CTCTCTCCTAGGAATATGCCTAATCTTACAAATCATTACAGGCCTATTCTTAGCCATACA	14333
Query	14337	CTATACATCAGATACAACCACTGCCTTCTCATCCGTAGCCCATATTTGCCGAGATGTAAA	14396
Sbjct	14334	CTATACACCAGATACAACCTACCCTTTCTCATCCGTGACCCACATTTGCCGAGATGTAAA	14393
Query	14397	TTACGGATGAATTATCCGCTATCTCCACGCTAACGGAGCATCCATATTCTTCATCTGCCT	14456
Sbjct	14394	CTACGGATGAATCATCCGCTATCTCCACGCTAACGGAGCATCTATATTCTTCATCTGCCT	14453

Query	14457	ATTTATCCACGTAGGACGAGGTTTATACTATGGATCCTACACATTCTAGAAACATGAAA	14516
Sbjct	14454	ATTTATCCACGTAGGACGAGGCCTATATTATGGATCCTACACATTCTAGAAACATGGAA	14513
Query	14517	CATTGGAAATTATTCTACTATTACAGTGATGGCAACAGCATTATAGGCTACGTCCTGCC	14576
Sbjct	14514	CATCGGAATTGTCTACTATTACAGTGATAGCAACAGCATTATAGGCTATGTCTTACC	14573
Query	14577	GTGAGGACAAATGTCCTTCTGAGGAGCAACCGTAATCACAAACCTCCTCTCAGCCATCCC	14636
Sbjct	14574	ATGAGGCCAAATATCCTTCTGAGGAGCAACCGTAATTACAAACCTCCTCTCAGCCATCCC	14633
Query	14637	ATACATTGGCACCAACCTCGTTGAATGAATCTGAGGAGGCTTTTCCGTTGATAAAGCCAC	14696
Sbjct	14634	ATATATTGGCACCAACCTCGTTGAATGAATCTGAGGGGGCTTTTCCGTTGACAAAGCCAC	14693
Query	14697	CCTGACCCGATTTTTCGCATTCCACTTTATTCTCCATTATCATTACAGCCCTAGTAAT	14756
Sbjct	14694	CCTGACCCGATTTTTCGCGTTCACCTTTATCCTTCCATTATCATTACAGCCCTAGTAAT	14753
Query	14757	CGTACACCTTCTATTTCTCCATGAAACAGGATCTAACAACCCATCAGGAGTCTCATCTGA	14816
Sbjct	14754	CGTACACCTCCTATTTCTCCATGAAACAGGGTCTAACAACCCATCAGGAATCTCGTCTGA	14813
Query	14817	CATAGACAAATCCCATTCCACCCATATTACACAACCAAGACATTTTAGGCTTTCTAGT	14876
Sbjct	14814	CATAGACAAATCCCATTTCACCCTTACTATACAATCAAAGACATTCTAGGCCTCTTAGT	14873
Query	14877	ACTAATCTTAATACTCCTCATTCTAGTATTATTCTCACCCGACTTACTAGGGGACCCGGA	14936
Sbjct	14874	ACTAGTTCTAGTACTCCTTACTCTAACATTATTCTCACCCGACTTACTAGGAGATCCGGA	14933
Query	14937	TAACTACACACCAGCTAATCCTCTCAGCACTCCTCCACATATCAAACCAGAATGATATTT	14996
Sbjct	14934	TAACTATACACCAGCCAATCCCCTTAGTACTCCCCACATATTAGCCAGAATGATACTT	14993
Query	14997	TCTATTTGCCTACGCAATTCACGATCCATCCCCAACAACTAGGAGGCGTGCTAGCCCT	15056
Sbjct	14994	CCTATTTGCCTACGCAATTCACGATCTATTTCCAACAACTAGGAGGCGTACTAGCTCT	15053
Query	15057	TATCCTCTCCATTTTGATTCTAGCAATTATCCCAATACTCCATACATCAAACAACGAAG	15116
Sbjct	15054	AATCCTCTCTATCCTAATCCTGATAATTATCCCAATACTCCATACATCAAACAACGAAG	15113
Query	15117	CATAATGTTTCGACCACTAAGCCAATGTGTATTCTGACTATTAGTAGCAGACCTACTTAC	15176
Sbjct	15114	CATAATGTTTCGACCACTAAGCCAATGTATATTCTGACTGTTAGTAGCAGACCTACTCAC	15173
Query	15177	ACTCACATGAATCGGAGGACAACCAGTTGAACACCATTATTATTATCGGCCAACTAGC	15236
Sbjct	15174	ACTTACATGAATTGGAGGACAACCAGTAGAACACCATTATTATTATTGGCCAACCTAGC	15233
Query	15237	ATCAATCTTATACTTCTCCCTAATCTTAGTGTATACCCCTTGCAAGCGCAATCGAAAA	15296
Sbjct	15234	ATCAATCCTATATTTCTCCTAATCCTAGTATTATACCCCTCGTAGGCACAATCGAAAA	15293
Query	15297	CAACTTCCTAAAATGAAGAGTCTTTGTAGTATATCCATTACCCTGGTCTTGTAACCAGA	15356
Sbjct	15294	CAACTTCTAAAATGAAGAGTCTTTGTAGTATATCTATTACCCTGGTCTTGTAACCAGA	15353
Query	15357	AAAGGAGAATTAGCCTCTCCCTAAGACTTCAAGGAAGAAGCTATAGCCCACCATCAACA	15416
Sbjct	15354	AAAGGAGAATAA-CCTCTCCCAAGACTTCAAGGAAGAAGCCATAGCCCACCTTCAACA	15412
Query	15417	CCCAAAGCTGAAGTTCTACTTAAACTATTCTTGAACACCATTGAATGATAAACTGCTA	15476
Sbjct	15413	CCCAAAGCTGAAGTTCTACTTAAACTATTCTTGAACGCCATTAAACGATAAACTACCG	15472
Query	15477	TCGTGTGCATATCAGTATTAATAATTTTCTTTTTccccccccGTAGGTAC-CCCTA	15535
Sbjct	15473	TTGTGTACCATATCAGTATTAATAATTTTCTTTTTCCCCCCCC--GTA---ACGCCCTA	15527
Query	15536	TGTATATCGTGCAATTAATTTGTTTGCCCCATGCATATAAGCATGTACATTTGAATTATTA	15595
Sbjct	15528	TGTACATCGTGCAATTAATTTGTTTGCCCCATGCATATAAGCATGTACATTATAATTATTG	15587
Query	15596	ATCTTGATAAAAACATCGAATGATT--AATTCACATAATACTGGCAACCAACATGAA	15653
Sbjct	15588	ATCTTGATAAAAGGCATAATATGGTCGCAAT--AGCACAACTGGTAACTGAACATGAA	15645
Query	15654	TA-TCGTCACTCCAGGAATAGAATGGTTGATCTTACATAGTACATATTATTATTGGTCGT	15712
Sbjct	15646	TGCTTGTCTGCTCCAG-AC-AGAATGGTTGATTCACATAGTACATACTATTATTGATCGT	15703
Query	15713	ACATACCCCATTAAGTCAAATCATTCTCGACAACACGCATATCACCTCCCATGTAAATT	15772
Sbjct	15704	ACATACCCCATTAGGTCAAATCATTCTCGACAACATGCGTATCACCACCTACGTAGATT	15763

Query 15773 CTTAATTACCAACTCCCGAGAAATCACCAATCCTTGCGCGATCTGCATTTTCATTCTCGCT 15832  
 Sbjct 15764 CTTAACTACCAACTCCCGAGAAATCACCAACCCTTGCGAGACTTACATCCTATTCTCGCT 15823

Query 15833 CCGGGCCCATTA ACT-GTGGGGTAGTAAACTGAGCTGTATCCGGCATCTGGTTCTTAC 15891  
 Sbjct 15824 CCGGGCCCATTT-CTCGTGGGGTTCCTAAACTGAGCTGTATCCGGCATCTGGTTCTTAC 15882

Query 15892 TTCAGGGCCATCTCACCTAAAATCGCCTATTCTTTCTCTTAAATAAGACATCTCGATGG 15951  
 Sbjct 15883 TTCAGGGCCATCTCACCTAAAATCGCCTATTCTTTCTCTTAAATAAGACATCTCGATGG 15942

Query 15952 ACTAATGACTAATCAGCCCATGCTCACACATAACTGTGATGTCATGCAATTTGGTATTTTT 16011  
 Sbjct 15943 ACTAATGACTAATCAGCCCATGCTCACACATAACTGTGGTGTGTCATGCAATTTGGTATTTTT 16002

Query 16012 TATAATTTGGGGATGCTATGACTCAGCTATGGCCGTCTGAGGCCTTAACACATTCAAGCA 16071  
 Sbjct 16003 TATAATTTGGGGATGCTATGACTCAGCTATGGCCGTCTGAGGCCTTAACACATTCAAGCA 16062

Query 16072 AATTGTAGCTGGACTTAAATTGAACATGATTTACCCGCATCAGATAACCATAAGGTGTTA 16131  
 Sbjct 16063 AATTGTAGCTGGACTTAAATTGAACATGATTTATCCGCATCAGGTAACCATAAGGTGTTA 16122

Query 16132 TTCAGTCAATGGTcacaggacatacgtatacacacgcttacac--atac--gtatacaca 16187  
 Sbjct 16123 TTCAGTCAATGGTTACAGGACATACACACATACACACATACACGCATACACGCATACACGC 16182

Query 16188 cgcttacacatacgtatacacacgcttacaca--tac--gtatacacacgcttacac--a 16241  
 Sbjct 16183 CATACACACATACACACATACACGCATACACGCATACACGCATACACACG-T-ACACGCA 16240

Query 16242 tac--gtatacacacgcttacac--at--acgtatacacacgc 16278  
 Sbjct 16241 TACACGCATACACACG-T-ACACGCATACACGCATACACACGC 16281

Range 2: 16148 to 16566 [Graphics](#)

[▼ Next Match](#) [▲ Previous Match](#) [▲ First Match](#)

Score	Expect	Identities	Gaps	Strand
396 bits(214)	1e-111	359/427(84%)	18/427(4%)	Plus/Plus

Query 16281 acacatacgtatatacacgcttacac--at--acgtatatacacgcttacacatacgtat 16336  
 Sbjct 16148 ACACATACACACATACACGCATACACGCATACACGCATACACACATACACACATACACGC 16207

Query 16337 atacacgcttacaca--catacgtatatacacgcttacaca--catacgtatatacacgcttac 16392  
 Sbjct 16208 ATACACGCATACACGCATACACGCATACACGCATACACGCATACACGCATACACGC--AT 16265

Query 16393 acatacgtatacacacgcttacacacCCATTAAGTACATGATTATCTTAGCAAAcccccc 16452  
 Sbjct 16266 AC--ACGCATACACACGC--ACATACTTATCAAGTACACAATTACCTTAGCAAAACCCCC 16321

Query 16453 cttccccccATTAAACCTCGCGCCATGTATTCTCTAAAGCCTTGCCAAACCCCAAAAAC 16512  
 Sbjct 16322 CTTCCCCCATTAAACCTCGCGCCATGTATTCTCTAAAGCCTTGCCAAACCCCAAAAAC 16381

Query 16513 AAAGCAAATACACGAAGTCTACAAAGTTAACT-TTTTC-AATTAGCAAACCGCCCTA 16570  
 Sbjct 16382 AAAGCAAATACACGAAGTCTACAAAGTTAACTCTTTTTTAATTCAACAAACCGCCCTA 16441

Query 16571 AACTAATACAAACATGCTACTTCAACCAATAAAATTTATGTAGACAGACATCCCCCTAGA 16630  
 Sbjct 16442 AACTGATACAAACATGCTACTTCAACCAGTAAAAATTTATGTAGACAGACATCCCTCTAGA 16501

Query 16631 TCTGAAAAAAtttttttttAAAACGTTCTCAACTCACTTATAAAACAATAATAAACCCAAAA 16690  
 Sbjct 16502 TCTGAAAAAAtttttttttAACGTTCTCAACCCTTATAAAACAACAATAAATCTAAAA 16560

Query 16691 ATTA ACT 16697  
 Sbjct 16561 -TCA ACT 16566

Alineamiento pareado con BLAST (Madden, 2013) de la Subunidad 2 de la NADH Deshidrogenasa de la muestra obtenida en 2011 y 2023 de individuos silvestres de *T. bairdii*

	Score 1884 bits(1020)	Expect 0.0	Identities 1020/1020(100%)	Gaps 0/1020(0%)	Strand Plus/Plus	
Query	23					82
Sbjct	1					60
Query	83					142
Sbjct	61					120
Query	143					202
Sbjct	121					180
Query	203					262
Sbjct	181					240
Query	263					322
Sbjct	241					300
Query	323					382
Sbjct	301					360
Query	383					442
Sbjct	361					420
Query	443					502
Sbjct	421					480
Query	503					562
Sbjct	481					540
Query	563					622
Sbjct	541					600
Query	623					682
Sbjct	601					660
Query	683					742
Sbjct	661					720
Query	743					802
Sbjct	721					780
Query	803					862
Sbjct	781					840
Query	863					922
Sbjct	841					900
Query	923					982
Sbjct	901					960
Query	983					1042
Sbjct	961					1020

## Alineamiento pareado con BLAST (Madden, 2013) de la Subunidad 2 de la NADH

Score	Expect	Identities	Gaps	Strand
1576 bits(853)	0.0	1024/1109(92%)	2/1109(0%)	Plus/Plus
Query 1		CCATTATAACAACCGTTATCCTAGGAACCATAAATTGTAATAACAAGCTCTCACTGACTAA		60
Sbjct 3951		CCATCATAACAACCTGTTATTCTAGGAACCATAAATTGTAATAACAAGCTCCCACTGACTAA		4010
Query 61		TAATCTGAATCGGCTTCGAAATAAATCTTAGCCATCATCCAATCCTAATAAAAAAT		120
Sbjct 4011		TAATCTGAATTGGTTTCGAAATAAACCTTCTAGCTATCATCCAATCCTAATAAAAAAT		4070
Query 121		TTAACCCACGGACAATAGAAGCATCCACAAATACTTTTAAACACAAGCCACCGCATCCA		180
Sbjct 4071		TCAACCCACGGACAATAGAAGCATCCACAAATACTTTTAAACACAAGCCACCGCATCTA		4130
Query 181		TACTACTCATAATGGCAATTATTGTTAACCTCATATACTCAGGCCAATGAACTATTACGA		240
Sbjct 4131		TACTACTCATAATAGCAATCATTATTAACTCATATACTCAGGCCAATGAGCTATTACAA		4190
Query 241		AAATTTTAAATCCAGCAGCATCCATCATATAACAACAGCCCTAATCATAAAACTGGAT		300
Sbjct 4191		AAATTTTAAATCCAATAGCGTCCGTTATTATAACAACAGCTCTAATTTAAAACTGGAT		4250
Query 301		TATCTCCCTTCCACTTCTGAGTACCCGAAGTAACACAAGGCATCTCGCTTACCTCAGGCT		360
Sbjct 4251		TATCCCTTCCACTTCTGAGTACCCGAAGTAACACAAGGTATCTCACTTACTTACAGGCT		4310
Query 361		TAATTCTACTCACATGACAAAAATTAGCCCAATATCAATCTTATACCAAAATTCACCAT		420
Sbjct 4311		TAATTCTACTCACGTGACAAAAATTAGCCCAATATCAATCTTATACCAAAATTCACCAT		4370
Query 421		CCATCAATCCAAACATACTACTAAGTCTGCCATACTATCCGTTATAGTAGGAGGCTGAG		480
Sbjct 4371		CCATCAACCCAAACATACTACTAACCAGCGCTATTTATCCATTATAGTAGGAGGCTGAG		4430
Query 481		GAGGCCTAAATCAAACCTCAACTACGAAAAATCATAGCATACTCATCAATTGCTCACATAG		540
Sbjct 4431		GAGGCCTAAACCAAACCTCAACTACGAAAAATCATAGCATACTCATCAATTGCCACATAG		4490
Query 541		GATGAATAACAGCTATCCTAATTTACAACCCAACAATAACAATACTAAACATACTAATTT		600
Sbjct 4491		GATGAATAACAGCCATCCTAATTTACAACCCAACAATAACAATACTAAACATACTAATCT		4550
Query 601		ACATCATAACAACACTTACCATATTTAATATTATAAACCCTCTCAACCAACAATC		660
Sbjct 4551		ATATCATAACAACACTTACCATATTCATGATATTCATAACTCACTCCTCAACCAACAATC		4610
Query 661		TGTCACCTCAAAACAGTGAATAAAGACCCCTAATCACTACACTCATCTTAAACAACC		720
Sbjct 4611		TATCACTCTCAAATACATGAAATAAATCCCTAATTACCACACTTATCCTAACAAACC		4670
Query 721		TACTGTCCCTAGGTGGCCTCCCACCATTATCGGGATTATACCCAAATGACTAATTATCC		780
Sbjct 4671		TACTATCTAGGTGGCCTCCCACCCTATCAGGATTCATACCCAAATGACTAATCATCC		4730
Query 781		AAGAACTTACAAAAACAACAGCATTATTCTACCAACATTAATAGCTATTACAGCATTAC		840
Sbjct 4731		AAGAACTTACAAAAATAATAGCATTATTCTACCAACATTAATAGCCATTACAGCACTAC		4790
Query 841		TCAACCTGATTTTCTACATACGACTGGCCTACTCCACATCACTAACAAATATCCCATcaa		900
Sbjct 4791		TCAACCTATATTTCTACATACGACTGGCCTACTCCACATCACTAACAAATATCCCATCAA		4850
Query 901		caaacatacaaaaaataaaatgacaattcgaaaactcaaacataaatCCTCCTACCAA		960
Sbjct 4851		CAATAACATAAAAAATAAATGACAGTTGAAAACACAAAACAAATAATCCTCCTACCAA		4910
Query 961		CACTAATCACCATATCTACGCTCCTCCTACCCTAACACCAATAATTTCCATTCTAGAAT		1020
Sbjct 4911		TACTAATCACTATATCTACACTTCTCTTACCCTAACACCAATAATTTCCATCTAGAAT		4970
Query 1021		AGGAATTTAGGTTACGCAGACCAAGAGCCTTCAAAGCTTAAGCAGAAG-AAAAATCAC		1079
Sbjct 4971		AGGAATTTAGGTTACATAGACCAAGAGCCTTCAAAGCTTAAGCAGAAGAAAAATCAC		5030
Query 1080		TTAATTCCTGC-TACTAAGGGCTGCAAGA 1107		
Sbjct 5031		TTAATTCCTGCCTACTAAGGGCTGCAAGA 5059		

Deshidrogenasa con número de referencia NC\_063943.1y la muestra de 2023 de *T. bairdii*.

Alineamiento pareado con Blast de la Subunidad 2 de la NADH Deshidrogenasa con número de referencia NC\_063943.1y la muestra obtenida en 2023 de *T. bairdii*.

Score	Expect	Identities	Gaps	Strand
1471 bits(796)	0.0	960/1042(92%)	0/1042(0%)	Plus/Plus
Query 1	ATAAACCTCTAATCTTCAACAACCATATAACAACCGTTATCCTAGGAACCATAATTGTA	60		
Sbjct 3929	ATAAATCCTCTAATCTCCATAACCATCATAACAACGTTATTCTAGGAACCATAATTGTA	3988		
Query 61	ATAACAAGCTCTCACTGACTAATAATCTGAATCGGCTTCGAAATAAATCTCTTAGCCATC	120		
Sbjct 3989	ATAACAAGCTCCCCTGACTAATAATCTGAATGGTTTCGAAATAAACCTTCTAGCTATC	4048		
Query 121	ATCCCAATCCTAATaaaaaaaaTTAACCCACGGACAATAGAAGCATCCACCAAATACTTT	180		
Sbjct 4049	ATCCCAATCCTAATAAAAAAATCAACCCACGGACAATAGAAGCATCCACCAAATACTTT	4108		
Query 181	TTAACACAAGCCACCGCATCCATACTACTCATAATGGCAATTATTGTTAACCTCATATAC	240		
Sbjct 4109	CTAACACAAGCCACCGCATCTATACTACTCATAATAGCAATCATTATTAACCTCATATAC	4168		
Query 241	TCAGGCCAATGAACTATTACGAAAAATTTTAAATCCAGCAGCATCCATCATATAACAACA	300		
Sbjct 4169	TCAGGCCAATGAGCTATTACAAAAATTTTAAATCCAATAGCGTCCGTTATTATAACAACA	4228		
Query 301	GCCCTAATCATAAAACCTGGATTATCTCCCTTCCACTTCTGAGTACCCGAAGTAACACAA	360		
Sbjct 4229	GCTCTAATTTATAAAAACCTGGATTATCCCTTCTACTTCTGAGTACCCGAAGTAACACAA	4288		
Query 361	GGCATCTCGTTACCTCAGGCTTAATTCTACTCACATGACAAAAATTAGCCCAATATCA	420		
Sbjct 4289	GGTATCTCACTTACTCAGGCTTAATTCTACTCACGTGACAAAAATTAGCCCAATATCA	4348		
Query 421	ATCTTATACCAAATTTACCATCCATCAATCCAACATACTACTAAGTCTGCCATACTA	480		
Sbjct 4349	ATCTTATACCAAATTTACCATCCATCAACCCAACATACTACTAACCGCCGCTATTTTA	4408		
Query 481	TCCGTTATAGTAGGAGGCTGAGGAGGCCTAAATCAAACCTCAACTACGAAAAATCATAGCA	540		
Sbjct 4409	TCCATTATAGTAGGAGGCTGAGGAGGCCTAAACCAAACCTCAACTACGAAAAATCATAGCA	4468		
Query 541	TACTCATCAATTGCTCACATAGGATGAATAACAGCTATCCTAATTTACAACCCAACAATA	600		
Sbjct 4469	TATTCATCAATTGCCCACATAGGATGAATAACAGCCATCCTAATTTACAACCCAACAATA	4528		
Query 601	ACAATACTAAACATACTAATTTACATCATAACAACACTTACCATATTTATAATATTCATA	660		
Sbjct 4529	ACAATACTAAACATACTAATCTATATCATAACAACACTTACCATATTCATGATATTCATA	4588		
Query 661	ACCCACTCCTCAACCACAACCTGTCACTCTCAAACACGTGAAATAAAGACCCCTAATC	720		
Sbjct 4589	ACTCACTCCTCAACCACAACCTCTATCACTCTCAAATACATGAAATAAAAAATCCCTAATT	4648		
Query 721	ACTACACTCATCTTAAACAACCTACTGTCCCTAGGTGGCCTCCCACCATTATCGGGATTC	780		
Sbjct 4649	ACCACACTTATCCTAACAACCTACTATCTCTAGGTGGCCTCCCACCATTATCAGGATTC	4708		
Query 781	ATACCCAAATGACTAATTATCCAAGAACTTACAAAAACAACAGCATTATTCTACCAACA	840		
Sbjct 4709	ATACCCAAATGACTAATCATCCAAGAACTTACAAAAATAATAGCATTATTCTACCAACA	4768		
Query 841	TTAATAGCTATTACAGCATTACTCAACCTGTATTTCTACATACGACTGGCCTACTCCACA	900		
Sbjct 4769	TTAATAGCCATTACAGCACTACTCAACCTATATTTCTACATACGACTGGCCTACTCCACA	4828		
Query 901	TCACTAACAAATATCCCATcaacaaacaatacaaaaataaaatgacaattcgaaaactca	960		
Sbjct 4829	TCACTAACAAATATCCCATCAACAAATAACATAAAAAATAAATGACAGTTTGAAAAACACA	4888		
Query 961	aaacaaataaatCCTCTACCAACTAATCACCATATCTACGCTCCTCTACCCTTAACA	1020		
Sbjct 4889	AAACAAATAATCCTCTACCAACTAATCACTATATCTACACTTCTCTTACCCTTAACA	4948		
Query 1021	CCAATAATTTCCATTCTAGAAT 1042			
Sbjct 4949	CCAATAATTTCCATCTAGAAT 4970			

## Líneas de comando

```
FastQC: nohup fastqc ../TAP_1.fastq ../TAP_2.fastq -o . &
```

```
Trimmomatic: trimmomatic-0.39.jar PE ../TAP_1.fastq ../TAP_2.fastq  
TAP_1.Tp.fastq TAP_1.Tu.fastq TAP_2.Tp.fastq TAP_2.Tu.fastq  
ILLUMINACLIP:TrueSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3  
SLIDINGWINDOW:4:15 MINLEN:36 HEADCROP:15 &
```

ILLUMINACLIP: Corta el adaptador y otras secuencias específicas de Illumina de la lectura.

LEADING: Corta las bases del comienzo de una lectura si está por debajo de un umbral de calidad.

TRAILING: Corta las bases del final de una lectura, si está por debajo de un umbral de calidad

SLIDINGWINDOW: Realiza un enfoque de recorte de ventana deslizante. Comienza escaneando en el extremo 5 'y recorta la lectura una vez que la calidad promedio que cae dentro de la ventana se encuentra por debajo de un umbral.

MINLEN: Descarta la lectura si está por debajo de una longitud específica.

HEADCROP: Corta el número especificado de bases desde el inicio de la lectura.

```
BLAST: nohup  
/share/Part1/lserver/velvet_1.2.10/contrib/shuffleSequences_fasta/shuffleSequences_fastq.pl TAP_1.fastq TAP_2.fastq mixed.fastq &
```

```
nohup /share/Part1/lserver/IDBA/bin/fq2fa --paired --filter mixed.fastq mixed.fasta &
```

```
nohup formatdb -p F -i mixed.fasta &
```

```
nohup blastall -p blastn -m 8 -d mixed.fasta -i Tterrestris_NC_053962.fasta -o hits.txt &
```

```
cut -f2 hits.txt | sort | uniq >ids.txt
```

```
nohup tclsh fastqselect.tcl -infile TAP_1.fastq -name ids.txt -outfile mappingreads1.fastq &
```

```
nohup tclsh fastqselect.tcl -infile TAP_2.fastq -name ids.txt -outfile mappingreads2.fastq &
```

nohup

```
/share/Part1/lservin/velvet_1.2.10/contrib/shuffleSequences_fasta/shuffleSequences_fastq.pl mappingreads1.fastq mappingreads2.fastq mappingmixed.fastq &
```

```
nohup /share/Part1/lservin/IDBA/bin/fq2fa --paired --filter mappingmixed.fastq mappingmixed.fasta &
```

SPAdes

```
Set 1: nohup spades.py -k 49,59,69,79,89,99 --only-assembler -s mappingmixed.fasta -o /backup/Usuarios/.lservin/lservin/Chass/Bowtie_Tp &
```

Set 2:

```
nohup /share/Part1/lservin/SPAdes/bin/spades.py -k 39,49,59,69,79,89,99,109,119 --only-assembler -s mappingmixed.fasta -o /share/Part1/lservin/Chass/Lecturas/ensamble/
```

Set 3:

```
nohup spades.py --careful mappingmixed.fasta -o /backup/Usuarios/.lservin/lservin/Chass/Bowtie_Tp/Intento3/ &
```

Bowtie: bowtie2-build TerrestriRefSeq.fasta referencia

```
nohup bowtie2 --no-unal -x referencia -1 TAP_1_Tp.fastq -2 TAP_2_Tp.fastq -S bowtie.sam --al-conc /backup/Usuarios/.lservin/lservin/Chass/Llenar_GAPs &
```

```
nohup cut -f1 bowtie_tp.sam | sort | uniq >ids_tp.txt &
```

```
nohup tclsh fastqselect.tcl -infile TAP_1_Tp.fastq -name ids_tp.txt -outfile mappingreads1.fastq &
```

```
nohup tclsh fastqselect.tcl -infile TAP_2_Tp.fastq -name ids_tp.txt -outfile mappingreads2.fastq &
```

nohup

```
/backup/Usuarios/.lservin/lservin/velvet_1.2.10/contrib/shuffleSequences_fasta/shuffleSequences_fastq.pl mappingreads1.fastq mappingreads2.fastq mappingmixed.fastq &
```

```
nohup /backup/Usuarios/.lservin/lservin/IDBA/bin/fq2fa --paired --filter mappingmixed.fastq mappingmixed.fasta &
```

SPAdes ft bowtie:

```
INTENTO 1 nohup spades.py -k 49,59,69,79,89,99 --only-assembler -s mappingmixed.fasta -o /backup/Usuarios/.lservin/lservin/Chass/Bowtie_Tp &
```

```
INTENTO 2 nohup spades.py -k 39,49,59,69,79,89,99,109,119 --only-assembler -s mappingmixed.fasta -o /backup/Usuarios/.lservin/lservin/Chass/Bowtie_Nodos
```

```
INTENTO 3: nohup spades.py --careful mappingmixed.fasta -o  
/backup/Usuarios/.lservin/lservin/Chass/Bowtie_Tp/Intento3/ &
```

--only-assembler runs only assembling (without read error correction)

-s <filename> file with unpaired reads

--careful tries to reduce number of mismatches and short indels

-o <output\_dir> directory to store all the resulting files (required)

```
Quast: nohup quast.py Set_1.fast Set_2.fasta Set_3.fasta &
```

```
ORPA: perl ORPA.pl method=Gblocks Alineamientos_con_bairdii.aln  
seqdump_bairdii.txt salida.fasta --iqtree -rmdup
```