

**UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**  
**FACULTAD DE CIENCIAS MARINAS**  
**INSTITUTO DE INVESTIGACIONES OCEANOLÓGICAS**



**ENSAMBLE Y ANOTACIÓN DEL GENOMA MITOCONDRIAL**  
**DEL ÁGUILA REAL (*Aquila chrysaetos*) DEL NOROESTE DE**  
**BAJA CALIFORNIA, MÉXICO**

**T E S I S**

**QUE PARA CUBRIR PARCIALMENTE LOS REQUISITOS NECESARIOS**  
**PARA OBTENER EL GRADO DE**

**MAESTRO EN CIENCIAS EN ECOLOGÍA MOLECULAR y**  
**BIOTECNOLOGÍA**

**PRESENTA**

**JESÚS ROBERTO OYERVIDES FIGUEROA**

**DIRECTOR:**

**DR. GORGONIO RUIZ CAMPOS**

**CO DIRECTOR:**

**DR. FRANCISCO JAVIER GARCÍA DE LEÓN**

**SINODALES:**

**DR. DANTE ALBERTO MAGDALENO MONCAYO**

**DRA. ALICIA ABADÍA CARDOSO**

**ENSENADA, BAJA CALIFORNIA, MÉXICO. NOVIEMBRE 2019**

## RESUMEN

Los programas federales e información para evaluar el estado de conservación actual del águila real (*Aquila chrysaetos*) en México están sustentados principalmente en estudios de su distribución y abundancia, siendo los análisis genético-poblacionales prácticamente inexistentes. Para contribuir con lo anterior, en el presente trabajo se ensambló y anotó el genoma mitocondrial (mitogenoma) de seis individuos de águila real provenientes del noroeste de Baja California, basado en las estrategias por referencia y de novo. Se construyeron librerías de tipo “shotgun” con ADN total, implementándose dos plataformas de secuenciación masiva de nueva generación como son Illumina Miseq v.3 y Hiseq 4000. La longitud del mitogenoma del Águila real es de 17,472 pb (pares de bases) con una composición base de A 29.8%, C 32.5%, G 14.0% y T 23.6%. El mitogenoma contiene 13 genes codificantes para los complejos proteicos COX I-III, CYTB, ATP 6 y 8 y NADH 1-6 y, el orden de los genes es concordante con el modelo general del genoma mitocondrial observado en aves de la familia Accipitridae. Los resultados del ensamble y anotación del mitogenoma mostraron que los individuos estudiados muestran diferencias entre sí, pero todos difieren con respecto a la anotación previamente reportada en el GenBank (acceso KF905228.1) en tres características: (1) el tamaño de la molécula es 140 pb más grande que lo reportado; (2) la anotación de una región denominada pseudo región control; y (3) la anotación de la región codificante para la proteína NADH deshidrogenasa sub unidad 3 (NADH 3) en dos fracciones, la

cual es provocada por una inserción de un nucleótido que cambia el marco de lectura, siendo esto confirmado por PCR.

A partir de la reconstrucción filogenética de las secuencias del mitogenoma de los individuos estudiados, se determinó que, en Baja California existen dos linajes mitocondriales, uno de ellos se relaciona con individuos de California, Estados Unidos y el otro linaje se relaciona con individuos provenientes de distintas y muy dispersas regiones de la distribución del Águila real en Norteamérica entre ellas, Alaska, California, Idaho y Canadá.

Este trabajo aporta información novedosa para esta especie emblemática nacional con la perspectiva de ser un referente para futuros estudios genómicos poblacionales, y así promover la incorporación de información genética en los estudios encaminados a la conservación de la especie.

FACULTAD DE CIENCIAS MARINAS  
INSTITUTO DE INVESTIGACIONES OCEANOLÓGICAS  
POSGRADO EN ECOLOGÍA MOLECULAR y BIOTECNOLOGÍA

ENSAMBLE Y ANOTACIÓN DEL GENOMA MITOCONDRIAL  
DEL ÁGUILA REAL (*Aquila chrysaetos*) DEL NOROESTE DE  
BAJA CALIFORNIA, MÉXICO.

TESIS

QUE PARA CUBRIR PARCIALMENTE LOS REQUISITOS NECESARIOS  
PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS

PRESENTA

JESÚS ROBERTO OYERVIDES FIGUEROA

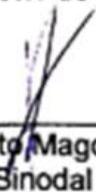
Aprobada por:



Dr. Gorgonio Ruiz Campos  
Director de tesis



Dr. Francisco Javier García de León  
Co director de tesis



Dr. Dante Alberto Magdaleno Moncayo  
Sinodal



Dra. Alicia Abadía Cardoso  
Sinodal

## **AGRADECIMIENTOS**

Agradezco al pueblo de México y al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca que me fue otorgada.

A la Universidad Autónoma de Baja California y la Facultad de Ciencias Marinas, por brindarme su apoyo institucional durante estos últimos años.

Al posgrado en Ecología Molecular y Biotecnología y mis profesores. Muchas gracias por los conocimientos y la formación como Maestro en Ecología Molecular y Biotecnología.

A los Dres. Gorgonio Ruíz Campos y Francisco Javier García de León, por permitirme trabajar con el Águila real, la especie más emblemática y representativa para los mexicanos. Además, por el apoyo que me brindaron y porque siempre confiaron en mi trabajo.

A la Dra. Alicia Abadía Cardoso, por sus comentarios y sabias recomendaciones.

Al Dr. Dante Alberto Magdaleno Moncayo, le agradezco especialmente, por compartirme su conocimiento y, porque para mí fue y sigue siendo, mi guía en el mundo de la bioinformática y la genómica.

Al Dr. Luís Manuel Enríquez Paredes, por siempre dar su más grande esfuerzo, aun cuando las cosas no marcharon bien. Le agradezco también, porque siempre tuvo para mí un buen consejo, por sus recomendaciones y ayuda en la pequeña etapa de laboratorio de mi proyecto de tesis y, por todo el apoyo que me ha brindado. Como persona, profesor y científico, tiene mi mayor admiración, es un ejemplo de lo que se debe ser y hacer.

A mis amigos Ari, José Luis, Luis Rodrigo, Graciela, Ramón, Maricza, Jade y, por supuesto a Melina, porque junto con ustedes he vivido y compartido momentos muy especiales. Gracias también, por escucharme, aconsejarme y en algún momento hacerme ver mis errores.

A mi familia, por creer en este proyecto personal y por siempre estar al pendiente de mí. Muchas gracias.

## DEDICATORIA

Esta tesis va dedicada a mi familia, ellos son los que me inspiran, me motivan y me ayudan a salir adelante. Mi hermano Juan Manuel, que con su disciplina en el trabajo y las metas claras que se ha trazado, ha sido para mí un ejemplo a seguir. Mi hermana Zitzi María, su perseverancia, entusiasmo y los logros que ha alcanzado me enorgullecen y hacen que quiera ser como ella. En especial y sobre todo, le dedico esta tesis a mis padres Juan Manuel y Esther, ustedes son mi mayor inspiración, el producto de su trabajo y amor hacia mí me han convertido en la persona que soy ahora. Orgulloso me siento de ustedes Familia.

Sin importar la distancia que me separa de mis hermanos y mis padres, yo se que el cariño y amor que nos tenemos, no se ha perdido con el paso de los años. Los amo.

*Desde la noche que me cubre,  
negra como el abismo de polo a polo,  
agradezco a cualquier dios que exista  
por mi alma inconquistable...*

*...No importa cuán estrecha sea la puerta,  
cuán cargada de castigos la sentencia.  
Soy el amo de mi destino:  
soy el capitán de mi alma.*

William E. Henley, *Invictus* (1875)

<b>INDICE</b>	
<b>RESUMEN</b> .....	<b>I</b>
<b>AGRADECIMIENTOS</b> .....	<b>IV</b>
<b>DEDICATORIA</b> .....	<b>VI</b>
<b>LISTA DE TABLAS</b> .....	<b>IX</b>
<b>LISTA DE FIGURAS</b> .....	<b>X</b>
<b>I. INTRODUCCIÓN</b> .....	<b>1</b>
<b>II. JUSTIFICACIÓN</b> .....	<b>14</b>
<b>III. HIPÓTESIS</b> .....	<b>15</b>
<b>IV. OBJETIVOS</b> .....	<b>15</b>
IV. 1 OBJETIVO GENERAL .....	15
IV. 2 OBJETIVOS PARTICULARES.....	15
<b>V. METODOLOGÍA</b> .....	<b>16</b>
V.1 Origen de las muestras .....	16
V.2 Extracción del ADN .....	17
V.3 Secuenciación del genoma de <i>Aquila chrysaetos</i> .....	18
V.3.1 Preprocesamiento .....	18
V.3.2 Ensamble por referencia del genoma mitocondrial de <i>Aquila chrysaetos</i> .....	22
V.3.3 Ensamble <i>de novo</i> del genoma mitocondrial de <i>Aquila chrysaetos</i> .....	23
V.3.4 Anotación del genoma mitocondrial de <i>Aquila chrysaetos</i> .....	24
V.3.5 Construcción del mapa del genoma mitocondrial de <i>Aquila chrysaetos</i> .....	25
V.4 Análisis del grado de sintenia .....	25
V.5 Comprobación experimental de la anotación .....	25
V.6 Llamado de SNP's e INDEL's (inserciones/deleciones) en el genoma mitocondrial de <i>Aquila chrysaetos</i> .....	27
V.7 Reconstrucción de relaciones filogenéticas de CO I, Cyt b, ND 2, $\psi$ RC, RC y el mitogenoma .....	28
V.8 Análisis de diversidad genética mitocondrial .....	30
<b>VI. RESULTADOS</b> .....	<b>31</b>
VI.1. Extracción del ADN .....	31
VI.2 Secuenciación .....	32
VI.3 Preprocesamiento de las secuencias .....	36
VI.4 Ensamble por referencia.....	37
VI.5 Ensamble <i>de novo</i> .....	40
VI.6 Anotación del genoma mitocondrial de <i>Aquila chrysaetos</i> .....	41

VI.7 Construcción del mapa del genoma mitocondrial de <i>Aquila chrysaetos</i> .....	44
VI.8 Comprobación experimental de la anotación de la región ND 3.....	45
VI.9 Análisis del grado de sintenia .....	47
VI.10 Llamado de SNP's e INDEL's en el mitogenoma de <i>Aquila chrysaetos</i> .....	48
VI.11 Reconstrucción de relaciones filogenéticas mediante el ADN mitocondrial.....	52
VI.12 Análisis de diversidad genética.....	53
<b>VII. DISCUSIÓN.....</b>	<b>58</b>
VII.1 Extracción del ADN .....	58
VII.2 Secuenciación .....	59
VII.3 Preprocesamiento de las secuencias .....	60
VII.4 Ensamble por referencia del mitogenoma .....	60
VII.5 Ensamble <i>de novo</i> del mitogenoma .....	62
VII.6 Anotación y reconstrucción del mapa del mitogenoma de <i>Aquila chrysaetos</i> .....	63
VII.7 Comprobación experimental de la anotación de la región ND 3.....	64
VII.8 Análisis del grado de sintenia .....	66
VII.9 Llamado de SNP's e INDEL's en el mitogenoma de <i>Aquila chrysaetos</i> .....	66
VII.10 Reconstrucción de las relaciones filogenéticas mediante el ADN mitocondrial .	67
VII.11 Análisis de diversidad genética mitocondrial .....	70
<b>VIII. CONCLUSIONES.....</b>	<b>75</b>
<b>Literatura citada.....</b>	<b>77</b>
<b>Anexo 1.....</b>	<b>95</b>

**LISTA DE TABLAS**

Tabla		Página
I	Genomas mitocondriales de especies de la familia Accipitridae.	12
II	Datos de la procedencia de los ejemplares de <i>Aquila chrysaetos</i> .	17
III	Resultados de la evaluación de la calidad del ADN de la extracción de las muestras de los individuos de <i>Aquila chrysaetos</i> y cocientes de la calidad de ADN.	31
IV	Estadísticos básicos de los archivos de secuenciación.	32
V	Número de secuencias crudas y tratadas mediante el control de calidad.	37
VI	Reporte de lecturas alineadas con el ensamble por referencia.	37
VII	Estadísticos básicos del ensamble de novo del genoma mitocondrial de <i>Aquila chrysaetos</i> .	41
VIII	Anotaciones hechas mediante MITOS webserver.	42
IX	Orden de los genes del genoma mitocondrial de referencia KF905228 y del individuo Ach 6 de <i>Aquila chrysaetos</i> .	47
X	Llamado de SNP's e INDEL's de los individuos Ach 1-6 contra el genoma de referencia KF905228 de <i>Aquila chrysaetos</i> .	49
XI	Diversidad genética mitocondrial de los distintos genes y regiones génicas utilizadas.	54

**LISTA DE FIGURAS**

Figura		Página
1	Mapa de distribución nacional de <i>Aquila chrysaetos</i> .	3
2	Localización geográfica de los individuos estudiados de <i>Aquila chrysaetos</i> .	16
3	Diagrama de flujo de los pasos a seguir y el software principal que se utilizará partiendo de las secuencias crudas hasta el llamado de variantes genéticas y la anotación y construcción del genoma mitocondrial.	22
4	Gráfico que muestra el valor de calidad del archivo de secuencias r1 (forward) del individuo Ach 1 de <i>Aquila chrysaetos</i> .	34
5	Gráfico que muestra el valor de calidad del archivo de secuencias r2 (reverse) del individuo Ach 1 de <i>Aquila chrysaetos</i> .	34
6	Gráfico que muestra el valor de la calidad del archivo de secuencias r1 (forward) del individuo Ach 6 de <i>Aquila chrysaetos</i> .	35
7	Gráfico que muestra el valor de la calidad del archivo de secuencias r2 (reverse) del individuo Ach 6 de <i>Aquila chrysaetos</i> .	35
8	Ensamble por referencia del genoma mitocondrial de <i>Aquila chrysaetos</i> , para el individuo Ach 1 utilizando como referencia KF905228 del GenBank.	39
9	Ensamble por referencia del genoma mitocondrial de <i>Aquila chrysaetos</i> , del individuo Ach 6 utilizando como referencia KF905228 del GenBank.	39
10	Mapa del mitogenoma anotado de novo del Águila real ( <i>Aquila chrysaetos</i> ) de Baja California, México.	45
11	Fotografía del gel de electroforesis que muestra el resultado de la amplificación de los fragmentos para ambos pares de cebadores.	46
12	Fracción del fragmento de la secuencia del ADN mitocondrial que codifica para la proteína ND 3 (NADH deshidrogenasa sub unidad 3) de <i>Aquila chrysaetos</i> .	46

- |    |  |    |
|----|--|----|
| 13 | Diagrama comparativo del orden de los elementos génicos de los genomas mitocondriales de referencia KF905228 y el ensamblado del individuo Ach 6 de <i>Aquila chrysaetos</i> .   | 48 |
| 14 | Árbol filogenético del mitogenoma, construido mediante el método bayesiano de inferencia filogenética a partir de la secuencia del mitogenoma de referencia KF905228 disponible en el GenBank, de cada uno de los individuos de <i>Aquila chrysaetos</i> que fueron secuenciados (Ach 1-6) y como grupo externo la especie <i>Aquila heliaca</i> KU646835. | 52 |

## I. INTRODUCCIÓN

La colección ornitológica de la Facultad de Ciencias de la Universidad Autónoma de Baja California es la más importante en el noroeste de México, con un acervo de más de 2,100 especímenes de esta región, contando con especímenes de Águila real de la parte noroeste de Baja California. A partir de esto surge la motivación de conocer la identidad genética de esta población mexicana y contribuir así al conocimiento de la especie desde un enfoque genético y genómico mitocondrial, que no ha sido implementado en el país.

El Águila real (*Aquila chrysaetos*) es una especie de ave rapaz moderna de las más grandes en América (Cornell University, 2015) y ampliamente distribuida en el mundo. A la vez que es una especie emblemática y de gran importancia cultural e histórica para los mexicanos, ya que forma parte de la identidad nacional al encontrarse en el escudo y bandera nacional.

A nivel internacional el estatus de conservación según IUCN (2016) es de preocupación menor y a nivel nacional, como una especie amenazada (SEMARNAT, 2010). La mayoría de los registros históricos y actuales confirmados de la distribución de *A. chrysaetos* en México ocurren dentro de los límites de sus áreas naturales protegidas (SEMARNAT, 2008); a pesar de ello se conoce poco sobre su biología, más allá de los clásicos monitoreos de los nidos o de las parejas reproductivas (Cubas y Fuentes, 2017).

Con el advenimiento de las tecnologías de secuenciación de nueva generación, nuevos estudios con especies no modelo empezaron a ser más comunes y con esto, se abre la posibilidad de estudiar las poblaciones de Águila real (*A. chrysaetos*) en México.

El objetivo principal de este estudio es ensamblar y anotar el genoma mitocondrial de seis individuos de *A. chrysaetos* del noroeste de Baja California y conocer la relación que existe con respecto al único mitogenoma reportado para la especie que corresponde a un individuo de California, Estados Unidos. Los objetivos particulares son conocer las relaciones filogenéticas con respecto a otros registros de secuencias mitocondriales (Citocromo oxidasa (CO I), Citocromo b (Cyt b), NADH deshidrogenasa sub unidad 2 (ND 2), pseudo región control ( $\psi$ RC), región control (RC) y el mitogenoma) existentes en el GenBank <sup>A</sup>, así como evaluar la diversidad genética. De aquí en adelante cualquier término que enseguida presente la siguiente "A", podrá ser consultado en el glosario del Anexo 1.

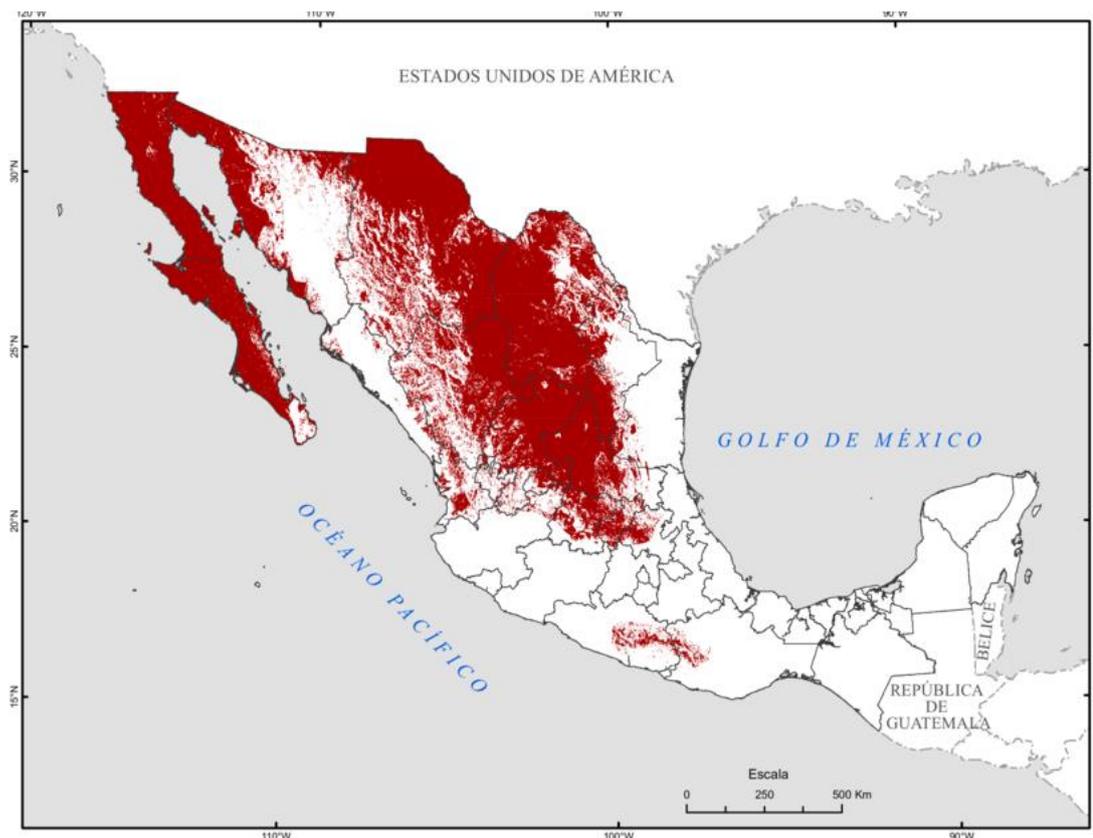
Este es el primer estudio genético para poblaciones mexicanas de *A. chrysaetos* que permitirá la mejor toma de decisiones para planes de manejo y conservación de la especie. Del mismo modo este estudio será una valiosa referencia para el análisis de otras poblaciones en México, Norteamérica y el mundo, dado que es una especie con un amplio ámbito de distribución geográfica.

#### *Distribución Geográfica*

El Águila real está presente en una gran variedad de ambientes desde el Ártico hasta los desiertos (Kochert et al., 2002). Su distribución es amplia (Watson, 2010), encontrándose en el hemisferio boreal (Holártico) a través de los continentes americano, eurasiático y africano (Kochert et al., 2002; Watson, 2010), y en América desde Alaska, Canadá, Estados Unidos, norte de México

incluyendo la península de Baja California, hasta la parte central de México (Howell y Webb, 1995; Kochert et al., 2002).

En México, el Águila real se distribuye principalmente en las regiones que abarcan los desiertos Chihuahuense y Sonorense y en sus regiones adyacentes (Figura 1).



*Figura 1:* Mapa de distribución nacional de *Aquila chrysaetos*, en rojo se resalta las áreas de distribución (tomado de Navarro-Sigüenza y Townsend-Peterson, 2007).

Es decir, en los estados de Chihuahua, Coahuila, Nuevo León, Tamaulipas, Sinaloa, Durango, Zacatecas, San Luis Potosí, Sonora, Baja California, Baja California Sur, Guanajuato, Querétaro, Michoacán, Hidalgo, Aguascalientes, Jalisco, Nayarit y México (Guerrero-Cárdenas et al., 2013; De León-Girón, 2017), además de registros visuales, pero sin la certeza de anidación en los estados de Puebla, Oaxaca y Veracruz. Farías et al. (2016)

mencionan que, en la década de 1930-1940, las poblaciones de Águila real en México se vieron afectadas por pesticidas y acciones de eliminación por representar una amenaza contra el ganado. Adicionalmente las poblaciones se han reducido producto de la pérdida de hábitat (Guerrero-Cárdenas et al., 2013).

En Baja California, la distribución histórica y reciente es muy similar (De León-Girón et al., 2016). Se menciona que en esta región convergen y conviven dos poblaciones ecológicas, una residente conformada por adultos reproductivos y otra migratoria compuesta por adultos y sub-adultos no reproductivos en busca de pareja y de un territorio en el cual establecerse. Esta última población ecológica parece venir de Oregón y del sur de California, Estados Unidos.

Otros estudios brindan evidencia de que la distribución de *A. chrysaetos* es compartida entre California y Baja California (Tracey et al., 2016; 2017), por ejemplo, en estudios de telemetría que rastrean a individuos capturados en el sur de California y en la frontera de Estados Unidos con México se observa que varios ejemplares incursionan al sur en la región noroeste de Baja California, inclusive algunos se llegan a encontrar a lo largo de toda la península, tan al sur como la región de Los Cabos en Baja California Sur.

### *Conservación*

La conservación de los depredadores tope en los ecosistemas, como es el caso del Águila real (*A. chrysaetos*), es sumamente importante, pues su

presencia repercute directamente en los procesos tróficos y por ende en la dinámica de las poblaciones de otras especies (De León-Girón, 2017).

En México, la generación de políticas públicas dirigidas a la conservación de las especies normalmente se realiza a partir de los resultados generados por estudios de tipo ecológico tradicional, colocando a las especies dentro de una categoría de conservación basado en la estimación de su abundancia total, tamaño de su hábitat o bien por las amenazas que pueden estar afectando a las especies (SEMARNAT, 2008), y en pocas ocasiones, mediante el sustento de estudios genéticos. En este orden de ideas, los estudios basados en el ADN mitocondrial (ADNmt), permiten determinar el estado actual de conservación de las especies, así como en la identificación de unidades evolutivas significativas y de manejo con sus aspectos demográficos (Moritz, 1994).

Si bien existen esfuerzos importantes para la conservación de la vida silvestre en el país, como el Programa de Acción para la Conservación de la Especie (PACE), es notoria la carencia de estudios genéticos que coadyuven a la protección de la fauna silvestre. Los estudios y datos genómicos realizados a partir de la secuenciación de nueva generación (NGS <sup>A</sup>, por sus siglas en inglés) representan una alternativa para estos objetivos.

Esta aproximación tiene grandes ventajas y ayuda a resolver ciertos problemas que son propios de la genética para la conservación. Entre ellos se puede mencionar: (1) la estimación no sesgada de parámetros de la genética de poblaciones con marcadores distribuidos a través de todo el genoma, que resulta ser muy complicado cuando se emplean unos cuantos marcadores

moleculares <sup>A</sup> o relativamente pocos; (2) la determinación de la neutralidad de los mismos y la observación de las interacciones entre el ambiente y los genes, y como esto se ve reflejado en el fenotipo y su aptitud (Angeloni et al., 2012); y (3) el conocer el estado actual de conservación de los ecosistemas donde se encuentran las especies bandera, como son los depredadores tope, y su presencia se manifiesta en una estabilidad del ecosistema permitiendo una diversidad de especies mayor (Sergio et al., 2006). La genómica para la conservación es una aproximación metodológica novedosa que ayuda con la identificación de unidades de conservación y detección de adaptaciones locales de las poblaciones (Garner et al., 2016). Un ejemplo claro de ello, es el estudio con salmón real (*Oncorhynchus tshawytscha*), en el cual se detectó, mediante el uso de 10,944 polimorfismos de un nucleótido (SNP's, por sus siglas en inglés) una estructura genética nunca antes observada, lo que aportó conocimiento fundamental para el manejo y monitoreo de las poblaciones (Larson et al., 2014).

El Águila real es una especie protegida por ambas naciones de Estados Unidos y México. En Estados Unidos se tiene el Acta de Protección y Conservación de la fauna silvestre del Águila calva y Águila real (Iraola, 2004) y en México la NOM-059-SEMARNAT-2010 (SEMARNAT, 2010).

#### *Escenario poblacional*

La situación poblacional del Águila real a nivel global es estable. En cambio a nivel regional los números poblacionales han incrementado en ciertas regiones, en otras han disminuido y en algunas se han mantenido estables. En algunas regiones de Europa la especie fue extirpada y en otras regiones,

como en Europa central, las poblaciones incrementaron (Suchentrunk et al., 1999). Suchentrunk et al. (1999) sugirieron que el Águila real permanezca dentro de las listas de especies amenazadas debido a su variabilidad genética reducida y por ser una especie rara en Europa, aunque, recientemente está presente en Europa central pero ausente en la mayor parte de Europa. Por otro lado, la especie sufrió un proceso demográfico que hizo desaparecer algunas poblaciones en las Islas Británicas. La especie se extinguió debido a la fragmentación de la población inicial de Europa continental que colonizó las Islas Británicas (Bourke et al., 2010). Sin embargo, en otras regiones de Europa, como en los Alpes italianos, la población de Águila real se ha incrementado, pasando de casi la desaparición, a reportarse 144 parejas en tan solo 36 años en el lapso de 1972-2008 (Fasce et al., 2011). Del mismo modo, en otras partes de Europa (España, Francia, Suecia y Suiza) la especie ha incrementado o bien mantenido sus poblaciones en la últimas décadas (Fasce et al., 2011).

En África, la población más tropical de Águila real está en Etiopía, donde se redujo por acción del cambio de uso de suelo y crecimiento de los asentamientos urbanos, pasando de ocupar siete sitios de anidamiento a solo tres reportados en el Parque Nacional Montañas Bale (Clouet y Barrau, 2015). En otras regiones al oeste de África en Mali y Níger, una pequeña y aislada población en el Sahara parece mantenerse, aunque sin registros de individuos juveniles (Thiollay, 2006).

El panorama en Asia en general es de poblaciones de Águila real disminuidas. En China la tendencia poblacional es de disminución rápida,

debido a diversos factores como la captura para la domesticación, alimento, envenenamiento por pesticidas, etc. (Ming et al., 2010). Mediante la evaluación de la reproducción por el conteo de nidos y polluelos, Ming et al. (2012) encontraron que la estabilidad de la población está comprometida y que, año con año, se reduce la densidad poblacional. En Japón, la población ha disminuido y su densidad es muy baja, reportándose alrededor de 175 parejas reproductivas (Takeuchi et al., 2006).

Por otro lado, la población de Águila real en el oeste de Estados Unidos, Alaska y Canadá se mantiene estable (Millsap et al., 2013). Sin embargo, en el este de Estados Unidos ha disminuido drásticamente (Bechard y McGrady, 2002). En México el escenario poblacional del Águila real es desalentador por la fragmentación del hábitat y las amenazas a las que está expuesta, como el envenenamiento, electrocución y choque con generadores eólicos. A pesar de esta situación, parece existir una recolonización de los estados sureños de Guerrero y Oaxaca (De León-Girón et al., 2016).

#### *Estudios genéticos y moleculares*

Uno de los primeros estudios genéticos realizados para Águila real es el de Masuda et al. (1998), quienes evaluaron la variación genética y la distribución de los haplotipos de la región control en Japón, Corea y China, y concluyeron que todos los individuos evaluados forman parte de una metapoblación que se fragmentó recientemente. Por otro lado, en Europa central, Suchentrunk et al. (1999) usaron 31 sistemas enzimáticos para analizar la diversidad genética dentro y entre poblaciones, y encontraron que el nivel de heterocigosidad esperada ( $H_e=0.034$ ) está dentro del intervalo que presentan otras especies

de aves que no se encuentran amenazadas ( $H_e=0.0-0.158$ ), aunque por debajo del promedio  $H_e=0.06$ . Sin embargo, la tasa de polimorfismo ( $P_{99\%}=10.8$ ) y el número promedio de alelos ( $A=1.11$ ) indican una variabilidad genética reducida en ese marcador. Aunque el valor de heterocigosidad en proporción de los sitios polimórficos (H/P; Hartl & Pucek, 1994) de la población estudiada es de 0.41, lo cual indica que la diversidad genética se ha conservado a pesar de un tamaño efectivo poblacional y de una densidad poblacional bajas. En las Islas Británicas, la especie sufrió un proceso demográfico que hizo desaparecer algunas poblaciones. En virtud de lo anterior, recientemente se realizó un análisis de la diversidad genética actual e histórica mediante 13 loci microsatelitales<sup>A</sup>, comparando 172 individuos contra 79 muestras de museo. Los resultados indican que la población sufrió un cuello de botella ancestral y se atribuyó a la fragmentación de la población inicial de Europa continental que colonizó las Islas Británicas (Bourke et al., 2010).

Por su parte y en un nivel más amplio de la distribución, Nebel et al. (2015) analizaron la filogeografía de la especie a través de diversas y amplias regiones de Europa, Asia continental, Japón y Norteamérica, mediante el uso de la RC del ADNmt y encontraron que existen dos haplo-grupos que corresponden a dos regiones geográficas (región Mediterránea y Holártica) y que tales linajes mitocondriales son producto de los dos refugios que permitieron la supervivencia de la especie durante la última era glacial.

En Norteamérica existen múltiples estudios para el Águila real que utilizan algunos marcadores moleculares. Sonsthagen et al. (2012) evaluaron

la diversidad genética en las islas del Canal en California, Estados Unidos, a partir de nueve microsatélites nucleares y de la RC del ADNmt, logrando determinar el sentido del flujo genético entre las poblaciones insulares y de la región continental obteniendo un mapa de distribución de los haplotipos, contribuyendo al conocimiento la distribución de la diversidad genética de la especie y en particular de la población norteamericana.

En años más recientes, Doyle et al. (2016) mediante el uso de 162 SNP's lograron identificar tres poblaciones de Águila real en Estados Unidos, una en California, otra en el este de Estados Unidos y una tercera en el oeste de Estados Unidos y Alaska, así también encontraron alta afinidad genética entre las poblaciones de Alaska y la de California. Los mismos autores, observaron que los valores de heterocigosidad de 159 SNP's autosómicos (de los 162 SNP's se descartaron dos que son mitocondriales y uno más que es un marcador de sexo) entre adultos y juveniles son distintos, siendo mayores en los adultos que en los juveniles. Los autores discuten que tal valor medio de heterocigosidad puede estar asociado con la viabilidad, pudiendo tener un efecto negativo en los juveniles.

Por su parte, Craig et al. (2016) empleando 36 microsatélites lograron diferenciar dos poblaciones de Águilas reales de Norteamérica a pesar de que dichas poblaciones presentan alta conectividad, pero la RC del ADNmt, no fue informativa, por lo que dichos resultados sugieren que la capacidad de detectar diferenciación entre poblaciones es dependiente del conjunto de marcadores moleculares. Los autores argumentan que el grado de flujo genético detectado pudo deberse a las reintroducciones en el este de

Norteamérica a partir de individuos provenientes del oeste, así como a la pérdida del hábitat que contribuye a una mayor dispersión y ampliación del hábito hogareño y, por tanto, mayor conectividad entre esas regiones. Además, estos autores sugieren que el número de marcadores genéticos nucleares debería ser mayor para posteriores análisis poblacionales y la distinción del origen de individuos entre residentes o migrantes durante el invierno.

Judkins y Van Den Bussche (2017), analizaron el patrón de variación genética y filogeográfica del linaje holártico (neártico) en Norteamérica y mediante la RC del ADNmt, muestran que existe una expansión poblacional y alta diversidad genética, así también un flujo genético moderado entre ambos linajes (holártico y mediterráneo).

A pesar de la existencia de estudios genéticos como los antes referidos para la región de Norteamérica, en ninguno de ellos se utilizaron muestras provenientes de México.

#### *Genomas mitocondriales de especies relacionadas*

Las nuevas tecnologías de secuenciación masiva son cada vez más utilizadas por grupos de investigación alrededor del mundo y aplicadas a distintas áreas de la genómica animal debido a su increíble poder de generar datos y de proyección de su uso en el futuro (Pareek et al., 2011). Los análisis utilizando el genoma mitocondrial completo, además de ser útiles para estudios filogenéticos y ayudar a resolver incertidumbres en las filogenias, permiten conocer los re-arreglos estructurales que presentan los mitogenomas, sirviendo estos a su vez como marcadores moleculares (Haring et al., 2001).

Una de las ventajas de la implementación del genoma mitocondrial completo en estudios de tipo ecológico-poblacional, es que brinda resultados más sólidos que si solamente se analizan unos pocos genes como comúnmente ocurre, al considerar una mayor cobertura del genoma y, por tanto, una mayor diversidad de nucleótidos (Angeloni et al., 2012).

En este contexto, existen varios estudios que reportan genomas mitocondriales completos para diversas especies de la familia Accipitridae, donde se incluye *A. chrysaetos*, y que se muestran en la Tabla I.

*Tabla I.* Genomas mitocondriales de especies de la familia Accipitridae, se muestra la longitud en pares de bases, el número de genes dentro del genoma mitocondrial, si contienen la región control y la pseudo región control, el método de obtención del genoma mitocondrial y la referencia.

<b>Especie</b>	<b>Longitud pb</b>	<b>Número de genes</b>	<b>RC y ψRC</b>	<b>Método</b>	<b>Referencia</b>
<i>Buteo buteo</i>	18,674	37	Si	PCR	Haring et al., 2001
<i>Accipiter virgatus</i>	17,952	37	Si	PCR	Song et al., 2015
<i>Aquila fasciata</i>	18,513	37	Si	PCR	Jiang et al., 2015
<i>Buteo lagopus</i>	18,559	37	Si	PCR	Jiang et al., 2015
<i>Accipiter gularis</i>	17,917	37	Si	PCR	Liu et al., 2017
<i>Gyps himalayensis</i>	17,381	37	Si	PCR	Jiang et al., 2019

Cabe mencionar que todos los genomas mitocondriales mencionados anteriormente fueron obtenidos mediante la amplificación por PCR<sup>A</sup> de poco más de una decena de fragmentos entre 11-13, que posteriormente fueron unidos para circularizar la molécula del genoma mitocondrial.

Doyle et al. (2014) obtuvieron por primera vez el genoma nuclear de un Águila real proveniente de la región central de California, Estados Unidos, mediante secuenciación de nueva generación (NGS por sus siglas en inglés)

en la plataforma ILLUMINA HiSeq 2000, sin embargo, dicho genoma nuclear no fue anotado. En el presente estudio, además de ensamblar y anotar el genoma mitocondrial, se comprobó el grado de sintenia, es decir, el orden conservado de bloques genómicos (Ghiurcuta y Moret, 2014). Se calculó y comparó la diversidad genética mitocondrial con las distintas secuencias disponibles en el GenBank (CO I, Cyt b, ND 2,  $\psi$ RC, RC y el mitogenoma) y, también, se determinaron las relaciones filogenéticas entre los individuos de Baja California con los registros publicados en el GenBank (CO I, Cyt b, ND 2,  $\psi$ RC, RC y el mitogenoma).

## II. JUSTIFICACIÓN

El Águila real es el ave representada en la bandera de los Estados Unidos Mexicanos y el escudo nacional, y pertenece a la parte prehispánica de la historia que da origen a esta nación. Esto convierte al Águila real en una especie con alto valor simbólico y cultural. Además, tiene gran importancia ecológica, ya que es un ave rapaz de gran tamaño y un depredador tope en las tramas tróficas del ecosistema, por lo que representa una especie indicadora de la calidad y estabilidad de los ecosistemas debido a la gran susceptibilidad al desequilibrio ecológico. De esta manera, resulta factible delimitar potenciales áreas naturales protegidas, en función del ámbito de distribución de la especie y de sus presas potenciales.

El Águila real presenta una distribución amplia en el hemisferio norte y las poblaciones mexicanas representan el límite más sureño de la distribución en el continente americano. Por la cercanía de las poblaciones de California, Estados Unidos y Baja California, México, la conectividad geográfica de los ecosistemas en ambos estados, la capacidad de migración de los individuos y el comportamiento de los individuos durante su etapa juvenil hace pensar que formen una misma población, sin embargo, no existen estudios que confirmen lo anterior. El presente estudio representa la generación de datos genéticos que ayudará a entender la relación entre los individuos de *A. chrysaetos* del noroeste de Baja California, México y los de California, Estados Unidos. Además, con la información generada se proporcionarán datos que ayuden a la elaboración de programas de conservación de la especie a nivel internacional.

### III. HIPÓTESIS

La alta capacidad de dispersión de las Águilas reales a través de California, Estados Unidos y de Baja California, México y la falta de barreras pre y pos-cigóticas, presuponen una conectividad genética entre los individuos de ambas regiones geográficas, por lo que se espera que conformen un mismo linaje mitocondrial.

### IV. OBJETIVOS

#### IV. 1 OBJETIVO GENERAL

- Ensamblar y anotar el genoma mitocondrial de seis individuos de Águila real (*Aquila chrysaetos*) provenientes del noroeste de Baja California, México y establecer sus relaciones filogenéticas con otras secuencias de genes mitocondriales y el mitogenoma reportados en el GenBank.

#### IV. 2 OBJETIVOS PARTICULARES

1. Ensamblar por referencia el genoma mitocondrial de *Aquila chrysaetos*.
2. Ensamblar *de novo* el genoma mitocondrial de *A. chrysaetos*.
3. Anotar los genes y elementos de los genomas mitocondriales obtenidos en *A. chrysaetos*.
4. Reconstruir las relaciones filogenéticas intra específicas de *A. chrysaetos* entre los mitogenomas obtenidos y los registros de secuencias mitocondriales de los genes CO I, Cyt b, ND 2,  $\psi$ RC y RC disponibles en el GenBank.
5. Evaluar la diversidad genética mitocondrial de *A. chrysaetos* a partir de los genes CO I, Cyt b, ND 2,  $\psi$ RC, RC y el mitogenoma.

## V. METODOLOGÍA

### V.1 Origen de las muestras

Las muestras de tejido de seis Águilas reales fueron obtenidas de la Colección de Ornitológica de la Facultad de Ciencias de la Universidad Autónoma de Baja California (Figura 2; Tabla II). Las seis muestras pertenecen a individuos catalogados dentro de la colección, de los cuales se posee información de las coordenadas exactas, localidades de recolecta, tipo de tejido de las muestras, identificadores de secuenciación y números de catálogo de la colección (Tabla II).



Figura 2: Localización geográfica de los individuos estudiados de *Aquila chrysaetos*.

*Tabla II.* Datos de la procedencia de los ejemplares de *Aquila chrysaetos*, números de identificación y tejidos utilizados para la extracción y posterior secuenciación del ADN, n-d no disponible.

Identificador de secuenciación	# catálogo UABC	Localidad de colecta	Fecha de colecta	Tejido para secuenciación	Coordenadas
Ach 1	UABC-820	Rosarito, B.C. (zona urbana)	15/oct/1995	Cojinete	32°22'14.2" N 117°02'52.8" W
Ach 2	UABC-092	Rosarito, B.C. (zona centro)	01/mar/1996	Cojinete	32°20'32" N 117°03'22" W
Ach 3	UABC-1989	Ensenada, B.C. (Ojos Negros camino a Ejido Uruapan)	18/sep/2013	Cojinete	31°51'37.7" N 116°16'16.3" W
Ach 4	UABC-1783	Ensenada, B.C. Bahía de San Quintín frente a Molino Viejo)	29/abr/2007	Cojinete	30°28'54.8" N 115°58'53.4" W
Ach 5	UABC-2101	Ensenada, B.C. (Ojos Negros camino al El Ranchito)	20/ene/2014	Cojinete	31°51'40.4" N 116°16'17.0" W
Ach 6	Sin número catalogo	Ensenada, B.C. (Rancho La Bellota)	n-d	Sangre/ Plumas	32°09'27.46" N 116°27'42.68" W

## V.2 Extracción del ADN

El tejido de los individuos Ach 1 a Ach 5 se obtuvo de los cojinetes de las patas (músculo), y el ADN se extrajo mediante el método de purificación de ácidos nucleicos por precipitación salina diferencial (Aljanabi y Martínez, 1997), mientras que para la muestra Ach 6 se utilizaron plumas y el ADN se extrajo

mediante el kit Qiagen Blood & Tissue. En ambos casos se evaluó la calidad y cantidad del ADN extraído mediante espectrofotometría con Nanodrop y, para las muestras Ach 1-5, adicionalmente mediante electroforesis<sup>A</sup> en gel de agarosa.

### V.3 Secuenciación del genoma de *Aquila chrysaetos*

Una vez extraído el ADN se prepararon las librerías para la secuenciación de tipo “shotgun”<sup>A</sup>, lo cual significa que se secuenció tanto el ADN nuclear como el mitocondrial. Los insertos obtenidos fueron de entre ~250 hasta ~700 pares de bases (pb). La secuenciación fue por extremos pareados o paired end<sup>A</sup> (en ambos sentidos, 5' a 3' y 3' a 5'), produciendo dos archivos por cada individuo secuenciado, es decir un archivo r1 de secuencias forward (en sentido 5' a 3') y otro archivo r2 de secuencias reverse (en sentido 3' a 5'). La secuenciación se llevó a cabo en dos plataformas distintas.

Las muestras Ach1-5 se secuenciaron en la plataforma ILLUMINA MiSeq V.3 (Georgia Genomics and Bioinformatics Core GGBC), mientras que para el ejemplar Ach 6 se secuenció en un ILLUMINA HiSeq 4000 (Oklahoma Medical Research Foundation Clinical Genomics Center). El rendimiento de producción de datos de secuenciación de ambas plataformas resulta ser adecuado para aplicaciones de ensamblaje<sup>A</sup> de genomas *de novo* (Glenn, 2011). La longitud de las secuencias de las muestras Ach 1-5 fue de entre 6 a 301 pb, mientras que la para la muestra Ach 6 fue de 35 a 151 pb.

#### V.3.1 Preprocesamiento

Todos los análisis se realizaron en un ordenador portátil con sistema operativo Ubuntu 18.04 (basado en Linux), un procesador de 8 núcleos, 4.1 GHz de

frecuencia máxima y 32 gb de memoria RAM, y 1 Tb de memoria de almacenamiento.

Los archivos de lecturas crudas (obtenidas directamente del secuenciador) están en un formato llamado fastq (.fastq), el cual es ampliamente utilizado en análisis bioinformáticos y permite su fácil tratamiento con un gran número de software libres. El formato fastq de secuencias (lecturas de secuenciación) se compone de cuatro líneas con información de cada una de las lecturas del ADN. La primera línea contiene el encabezado de la secuencia y comienza con el carácter "@", este carácter indica que, a partir de allí, hay un nuevo registro de secuencia, seguido de eso, una descripción e identificador de la secuencia. La segunda línea contiene la secuencia de nucleótidos que el secuenciador leyó en dicha secuencia de ADN, es decir las letras "A", "C", "G" ó "T". En caso de que no se haya leído correctamente una de las bases de la secuencia de ADN y el secuenciador no haya asignado una letra correspondiente a una de las bases nucleotídicas, entonces habrá una "N". La tercera línea de dicho formato comienza con un "+" seguido de un identificador de secuencia que no siempre se encuentra. Dicho carácter indica que allí termina la línea anterior y comienza la siguiente línea, que contendrá el mismo orden de caracteres que la anterior. Por último, la cuarta línea del formato contiene la información de calidad de secuencia de cada uno de sus nucleótidos leídos, dicha línea contiene múltiples caracteres codificados para cada uno de los valores del nivel de calidad en la escala Phred. Este valor indica la probabilidad de error de lectura, en otros términos, que dicha base no fue un error de secuenciación.

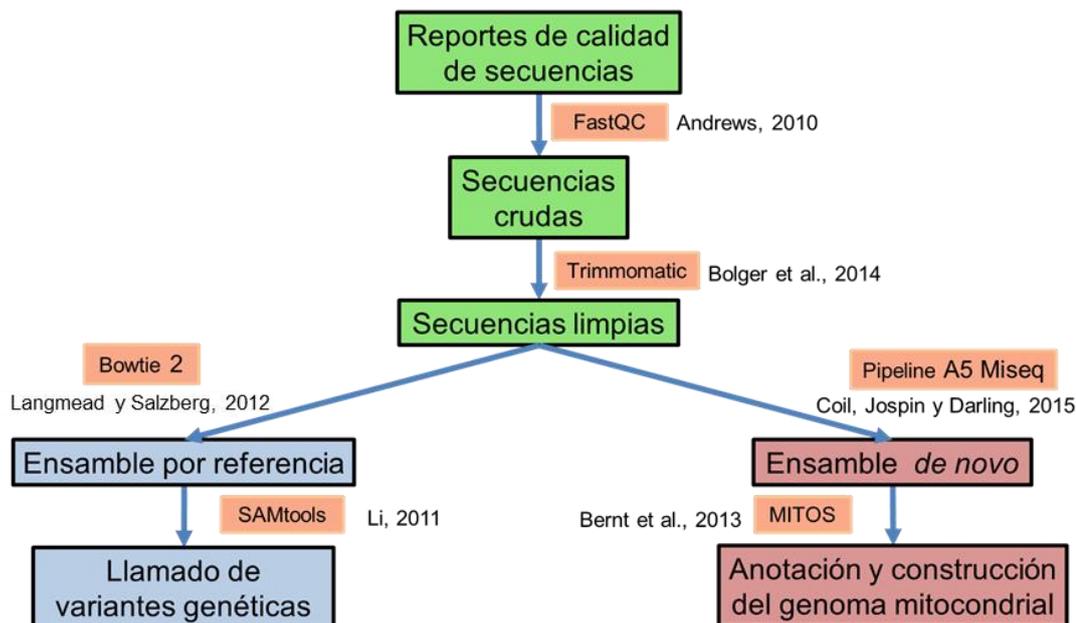
El primer paso del preprocesamiento de las secuencias consistió en el análisis de la calidad QS (Quality Score, por sus siglas en inglés) y longitud de las secuencias, la presencia de secuencias sobre-representadas, los adaptadores de secuenciación, etc. Los reportes de calidad de las secuencias crudas<sup>A</sup> se observaron con el programa FastQC (Andrews, 2010) y sirvieron como punto de partida para continuar con el tratamiento diseñado individualmente para cada uno de los archivos de secuencias provenientes de cada individuo de Águila real. Los reportes de calidad muestran un resumen de todas las categorías evaluadas y clasifica el estatus de calidad de cada categoría por un código de color, en el cual, el color verde corresponde a un estatus de calidad “adecuado”, anaranjado para un estatus de calidad de “atención” y en rojo para un estatus de calidad “malo” (Andrews, 2010).

El tratamiento posterior al análisis de los reportes de calidad consistió en filtrar las secuencias por su QS y longitud mínima, retirar los adaptadores de secuenciación y las secuencias sobre-representadas. Para alcanzar un nivel adecuado de calidad de las secuencias se aplicó un tratamiento estándar mediante el software Trimmomatic (Bolger, Lohse, y Usadel, 2014) que consistió de cuatro pasos. El primer paso retira los adaptadores de secuenciación, puede ser configurado para retirar las secuencias que se encuentran sobre-representadas y se realiza con la función ILLUMINACLIP. El segundo paso descarta aquellas secuencias cuyo valor medio de calidad QS se encuentre por debajo de 30 QS con la función AVGQUAL. El tercer paso evalúa la calidad media de las secuencias mediante una ventana deslizante que promedia el valor QS de dicha ventana y retira los fragmentos

que se encuentran por debajo de 25 de QS con la función SLIDINGWINDOW. El cuarto y último paso estándar MAXINFO hace una limpieza aún más estricta, balanceando entre conservar secuencias de una longitud muy grande, pero con poca exactitud o bien secuencias cortas con mucha exactitud de lectura. Este paso se configuró para conservar secuencias de, al menos 100 pares de bases y con una tendencia a conservar secuencias más largas con un valor de 0.3 (dicho valor en la función va de 0 a 1, donde valores cercanos a 0 favorecen secuencias largas y valores cercanos a 1 a secuencias con mucha exactitud).

El tratamiento aplicado al ejemplar Ach 6 de Águila real es una versión ampliada y modificada al tratamiento estándar debido a la baja calidad de secuencias que presentó aún con el tratamiento aplicado a las otras muestras y por la naturaleza de las lecturas producidas por provenir de una plataforma de secuenciación distinta. La modificación consistió en la aplicación de la función SLIDINGWINDOW, en la cual, el valor de calidad se modifica en la función MAXINFO a 27 de QS, y la tendencia de conservación de secuencias de gran longitud y muy exactas está equilibrada con un valor de 0.5. Además se añadieron las funciones LEADING y TRAILING; la primera, remueve bases que tengan un valor de calidad por debajo del especificado (27 QS) del inicio de la lectura; la segunda, remueve bases del extremo final de las lecturas cuyo valor de calidad se encuentre por debajo de 27 QS.

Posterior al preprocesamiento de las lecturas crudas, los archivos de secuencias se usan, en dos estrategias para el ensamble de los genomas mitocondriales, la estrategia por referencia y la *de novo* (Figura 3).



**Figura 3.** Diagrama de flujo de los pasos a seguir y el software principal que se utilizará partiendo de las secuencias crudas hasta el llamado de variantes genéticas y la anotación y construcción del genoma mitocondrial. En verde se ilustra el preprocesamiento de las secuencias, en azul la estrategia de ensamble por genoma de referencia y en rojo la estrategia de ensamble *de novo*.

## Objetivo 1

### V.3.2 Ensamble por referencia del genoma mitocondrial de *Aquila chrysaetos*

Los archivos de lecturas limpias<sup>A</sup>, después del preprocesamiento, se tomaron directamente para la realización de los ensamblajes por referencia, también llamados mapeos<sup>A</sup>. El programa utilizado fue Bowtie2 versión 2.3.4.2 (Langmead y Salzberg, 2012), el cual crea índices de los genomas de referencia (mediante el algoritmo<sup>A</sup> FM-Index<sup>A</sup>) y posteriormente mapea las lecturas paired-end (mediante el algoritmo de la transformada de Burrows-Wheeler<sup>A</sup>) de cada uno de los individuos. El genoma mitocondrial de referencia implementado fue de la misma especie *A. chrysaetos* con un número de acceso en GenBank de KF905228. Este programa permite realizar la tarea de ensamblar genomas por referencia utilizando equipo de cómputo

con relativo bajo poder, ya que resulta ser eficiente en el alineamiento<sup>A</sup> de secuencias cortas y ha sido implementado en la circularización de mitogenomas de vertebrados (Machado et al., 2018).

## **Objetivo 2**

### V.3.3 Ensamble *de novo* del genoma mitocondrial de *Aquila chrysaetos*

El ensamble *de novo* de los genomas mitocondriales de las muestras Ach 2 a Ach 6 se llevó a cabo con el pipeline<sup>A</sup> A5-miseq (Coil et al., 2015). El ensamble es un proceso automatizado que consta de 5 pasos; (1) limpieza de las lecturas: las lecturas son filtradas por su valor de calidad y, además, se remueven los adaptadores de secuenciación; (2) ensamble de contigs<sup>A</sup>: se toman las lecturas limpias y se crean contigs de una longitud mayor a la de las secuencias originales; (3) creación de scaffolds<sup>A</sup> (porciones de secuencias genómicas construidas a partir de contigs) crudos: se arman scaffolds a partir de los contigs creados anteriormente con parámetros de funcionamiento permisivos, es decir, con longitudes de insertos variables; (4) corrección de ensamblajes incorrectos: aquellos ensamblajes incorrectos son detectados a través de los pares de secuencias que no se encuentran mapeados respecto a una distancia esperada y entonces son descartados; (5) ensamble final de scaffolds: se arman nuevos scaffolds con parámetros astringentes, en este proceso se corrigen las regiones donde se encontraron los contigs incorrectos y por último se hace un resumen estadístico del ensamble.

La muestra Ach 1 se analizó con el software Velvet en su versión 1.1 (Zerbino y Birney, 2008). Tal software incluye un conjunto de algoritmos que construyen el ensamble de genomas a partir de los gráficos *de Bruijn*<sup>A</sup>, estos

gráficos ayudan a resolver errores y repeticiones. El utilizar un software distinto para realizar el ensamble *de novo* del individuo Ach 1 es consecuencia de que al implementar el software (pipeline A5-miseq), las capacidades del equipo de cómputo fueron superadas y el proceso se detuvo imprevistamente.

### **Objetivo 3**

#### V.3.4 Anotación del genoma mitocondrial de *Aquila chrysaetos*

La anotación<sup>A</sup> de los genomas mitocondriales fue a partir de los scaffolds obtenidos de los ensambles *de novo*. La determinación de los scaffolds que contienen información del genoma mitocondrial se realizó aplicando una búsqueda con la herramienta BLAST<sup>A</sup> (Basic Local Alignment Search Tool) de la página web del NCBI (National Center for Biotechnology Information) a los *scaffolds* de más de 4 Kpb (kilo pares de bases), posteriormente se alinearon en caso de ser necesario para alcanzar una mayor cobertura<sup>A</sup> del genoma mitocondrial, tomando en cuenta la longitud del genoma de referencia utilizado anteriormente (acceso GenBank KF905228). Después, se implementó RNAweasel (<http://megasun.bch.umontreal.ca/RNAweasel/>) para la identificación de ARNt's (ARN de transferencia), ARNr's (ARN ribosómico) e intrones y a su vez, MFannot (<http://megasun.bch.umontreal.ca/RNAweasel/>) para la identificación de proteínas y los marcos abiertos de lectura (secuencias cuya longitud es divisible entre tres y que están delimitadas por codones<sup>A</sup> de paro, (Sieber et al., 2018) de los genomas mitocondriales.

Adicional a la anotación anterior, se utilizó el servidor web MITOS (Bernt et al., 2013). Tal servidor, alojado en la Universidad Leipzig en Alemania, implementa un conjunto de procesos en flujo automatizados y

ayuda al usuario en la anotación de genomas mitocondriales de metazoos *de novo* y de forma estandarizada en cuanto a nombres de genes, estructuras secundarias de ARNt y ARNr y uso de código genético.

Al final del proceso de anotación, se hizo una curación manual de las anotaciones para revisar los seis marcos de lectura existentes, con el uso del software UGENE (Okonechnikov et al., 2012).

#### V.3.5 Construcción del mapa del genoma mitocondrial de *Aquila chrysaetos*

El software Genome Vx (Conant y Wolfe, 2008) se utilizó para crear el mapa del genoma mitocondrial del Águila real, a partir de la secuencia obtenida del ensamble *de novo* y de las coordenadas de inicio y termino de cada elemento génico anotado dentro del genoma mitocondrial del individuo Ach 6.

#### V.4 Análisis del grado de sintenia

Se realizó un análisis de la sintenia<sup>A</sup> (patrón de distribución de los genes en un cromosoma (Rödelsperger y Dieterich, 2010)) entre el genoma mitocondrial de Águila real reportado por Doyle et al., (2014) y el nuevo genoma mitocondrial ensamblado en este estudio. Dicho análisis se llevó a cabo con el software MAUVE (Darling, 2004), el cual identifica y alinea los elementos genómicos conservados y determina el arreglo global de la estructura de dichos elementos genómicos.

#### V.5 Comprobación experimental de la anotación

Debido a que en la anotación del ensamble *de novo* la proteína NADH 3 apareció dividida en dos fracciones y dicha característica no está reportada en el mitogenoma descrito por Doyle et al., (2014), se requirió una

comprobación experimental mediante una amplificación por PCR y su posterior secuenciación SANGER.

El diseño de cebadores se hizo a partir del ensamble *de novo* y usando el programa Primer-BLAST del NCBI. Se diseñaron dos pares: ND3Ach-1 (5'GCCTGATACTGGCACTTCGT3' y 3'CGGTGGGTCAGATTGATAGGG5') que genera un fragmento de 716 pb y ND3Ach-2 (5'CTTCTTCGTCGCTACAGGCT3' y 3'TTAAGTGGGTTCGGTGGGAAGG5') que genera un fragmento de 774 pb.

Para la amplificación por PCR se utilizaron ocho muestras de Águila real de la Colección Ornitológica de la Facultad de Ciencias de la Universidad Autónoma de Baja California. Tales muestras corresponden a los seis individuos utilizados para la secuenciación (aunque se tuvo que extraer nuevamente ADN porque todo fue utilizado para la secuenciación) y dos muestras extra que no forman parte en este estudio. Se tomaron muestras de tejido muscular de seis ejemplares, plumas de una muestra y sangre de otra. La extracción del ADN se llevó a cabo mediante el protocolo de extracción HOTSHOT (Truett et al., 2000) y se purificó con el kit DNA IQ System Promega siguiendo las especificaciones del fabricante.

Las condiciones de la reacción de PCR fueron las siguientes: 6 minutos 94°C, 35 ciclos de desnaturalización, alineamiento y extensión a 94°C durante 30 segundos, 51°C durante 30 segundos y a 72°C durante un minuto y medio respectivamente, una etapa de elongación final a 70°C por 7 minutos y una etapa final de conservación a 15°C.

Los fragmentos amplificados fueron enviados para secuenciación a la compañía SeqXcel.inc en San Diego, California y analizados con el equipo ADN ABI PRISM® 3130xl Genetic Analyzer.

#### V.6 Llamado de SNP's e INDEL's (inserciones/deleciones) en el genoma mitocondrial de *Aquila chrysaetos*

El llamado de SNP's<sup>A</sup> e INDEL's<sup>A</sup> (o llamado de variantes) es el proceso por el cual se identifican las variantes génicas de una secuencia por comparación contra un genoma de referencia (EMBL-EBI, 2019). El llamado de SNP's e INDEL's se hizo con el programa SAMtools (Li, 2011), que identifica cada SNP e INDEL y los filtra por su valor de calidad. El genoma de *A. chrysaetos* reportado por Doyle et al. (2014) se tomó como referencia para hacer el llamado de SNP's e INDEL's de los seis genomas ensamblados en este trabajo.

Una vez hecho el llamado de SNP's e INDEL's se filtraron por calidad del llamado de las variantes identificadas (Qs) descartando aquellas con un valor inferior a 20 Qs y por la profundidad de dicha variante, conservando aquellas con al menos 5X de profundidad (Li, 2011). A falta de una forma estandarizada de filtrado de SNP's e INDEL's, cada una fue evaluada manualmente en función del comportamiento del ensamble por referencia, conservando aquellos SNP's e INDEL's donde al menos la mitad de las lecturas alineadas presentara el cambio y corroborando su existencia en al menos otro individuo (Fridjonsson et al., 2011) cuando los valores de calidad del llamado Qs y/o la profundidad de la variante estuvieran cerca de los valores mínimos establecidos.

#### Objetivo 4

##### V.7 Reconstrucción de relaciones filogenéticas de CO I, Cyt b, ND 2, $\psi$ RC, RC y el mitogenoma

Las reconstrucciones filogenéticas se realizaron con secuencias recuperadas del GenBank y las propias de este estudio. Las secuencias del gen CO I utilizadas fueron 16 en total, seis procedentes de Baja California (propias de este estudio), cinco de Norteamérica (Estados Unidos y Canadá), cuatro de Europa y una de Asia. Las secuencias del gen Cyt b fueron 17 en total, seis de Baja California, tres de Norteamérica y ocho de Europa. Las secuencias del gen ND 2 fueron ocho en total, 6 de Baja California y dos de Norteamérica. Las secuencias de la  $\psi$ RC fueron 14 en total, seis de Baja California, una de Norteamérica, una de Europa y seis de Asia. Las secuencias de la RC fueron 76 en total, seis de Baja California, 13 de Norteamérica, 37 de Europa y 20 de Asia. Con el mitogenoma completo se utilizaron ocho secuencias en total, una secuencia es el genoma de referencia KF905228 (Doyle et al., 2014), seis son las obtenidas en este estudio y una secuencia más corresponde a *Aquila heliaca* (como grupo externo) con número de acceso GenBank KU6446835.

Se realizaron alineamientos múltiples para las secuencias del genoma mitocondrial y de las secuencias de cada uno de los genes y regiones génicas antes mencionados, todo ello mediante el algoritmo MAFFT 7 en su versión en línea con los parámetros default (Kato et al., 2017). Las regiones no conservadas entre las secuencias de los alineamientos múltiples fueron retiradas con el uso del programa en línea Gblocks con los parámetros default (Castresana, 2000).

Se determinó el modelo de sustitución nucleotídica y la tasa de mutación, mediante el programa JModelTest (Posada, 2008), tomando en cuenta el criterio de información bayesiano (BIC por sus siglas en inglés). Se emplearon tres métodos distintos para la construcción de las relaciones filogenéticas de cada uno de los genes mencionados anteriormente y del genoma mitocondrial completo.

El primer método es el de máxima verosimilitud<sup>A</sup> (MV; Felsenstein, 1981) realizado con el programa PhyML en su versión 3.0. (Guindon et al., 2010). Los parámetros del modelo de sustitución fueron modificados a partir de los resultados del programa JModelTest para cada gen y región y el genoma mitocondrial. En el apartado de búsqueda de topología del árbol, se configuró a la opción “best of NNI and SPR”, la cual estima la topología del árbol mediante dos algoritmos (NNI y SPR), el primero es un algoritmo de búsqueda de topologías de árbol que maximiza la probabilidad y su principal ventaja es que es más veloz, mientras que el segundo algoritmo poda y reescribe los sub-árboles hasta encontrar la mejor topología pero su desventaja es que es lento, También se modificó la opción del soporte de los nodos y, se estableció un valor de bootstrap<sup>A</sup> de 1000 iteraciones. El segundo método, el de inferencia filogenética bayesiana<sup>A</sup> fue realizado a través del programa MrBayes (Ronquist et al., 2012), implementando cuatro cadenas de Markov Monte Carlo (MCMC) con un total de seis millones de generaciones, muestreando cada 500 generaciones y un valor de burn in de 600,000 generaciones para alcanzar el estado estacionario (Lerner y Mindell, 2005). El tercer método, el de Neighbor-Joining<sup>A</sup> se realizó a través del software MEGA

X (Kumar et al., 2018), siendo el modelo de sustitución utilizado el de Kimura de dos parámetros, y se estableció un valor de bootstrap de 1000 réplicas.

Los resultados de los tres métodos se visualizaron con el programa FigTree 1.4 (Rambaut, 2009), se estimaron los valores de bootstrap de nodos para los métodos de máxima verosimilitud y Neighbor-Joining, así como el valor de probabilidad de los nodos para el método de inferencia filogenética bayesiano.

## **Objetivo 5**

### V.8 Análisis de diversidad genética mitocondrial

La diversidad genética se puede calcular en función de ciertos valores que expresan la riqueza de haplotipos<sup>A</sup>, es decir el conjunto de variantes de tipo SNP dentro de una secuencia y que a su vez varían de un individuo a otro (Nature Education, 2014) o bien la riqueza nucleotídica, que no es más que el número promedio de diferencias de nucleótidos entre pares de secuencias (Nei y Li, 1979).

Se utilizaron los alineamientos de secuencias antes descritos (en la sección V.7 de la reconstrucción de las relaciones filogenéticas) para realizar el análisis de diversidad genética por genes, regiones y mitogenoma. Los alineamientos de bloques conservados fueron la base para estimar los sitios segregantes, el número de haplotipos, la diversidad de haplotipos y nucleotídica y el número de diferencias entre secuencias. Tales valores de diversidad genética se estimaron con el programa DnaSP 6 (Rozas et al., 2017)

## VI. RESULTADOS

### VI.1. Extracción del ADN

Las muestras Ach1-5 presentaron una alta concentración de ADN, la más baja correspondió al individuo Ach 6 con 19.3 [ng/μl] y la más alta de 372.6 [ng/μl] al individuo Ach 5 (Tabla III). Las relaciones 260/280 y 260/230 se calcularon para todos los individuos excepto para Ach 6. El valor de la relación 260/280 fue adecuado (~1.8). Cuatro de los cinco individuos tienen una relación 260/230 por debajo de lo deseable (~2.0), como en el individuo Ach 5 con 1.32, lo que hace pensar que dichas muestras de ADN se encuentran ligeramente contaminadas por sales debido al método de extracción de ácidos nucleicos empleado. También se evaluó la integridad de ADN mediante gel de electroforesis (Figura 1 del Anexo 1).

*Tabla III.* Resultados de la evaluación de la calidad del ADN de la extracción de las muestras de los individuos de *Aquila chrysaetos* y cocientes de la calidad de ADN [ng/μl]: concentración en nanogramos por microlitro; 260/280: relación de absorbancia que indica la calidad de la muestra de ADN con respecto a contaminantes como proteínas; 260/230: relación de absorbancia que indica la calidad de las muestras de ADN con respecto a contaminantes como sales, fenoles y/o carbohidratos; n-d: no disponible.

<b>Individuo</b>	<b>[ng/μl]</b>	<b>260/280</b>	<b>260/230</b>
Ach 1	299.7	1.89	1.77
Ach 2	255.5	1.89	1.6
Ach 3	134.8	1.93	2.02
Ach 4	264.5	1.9	1.65
Ach 5	372.6	1.76	1.32
Ach 6	19.3	n-d	n-d

## VI.2 Secuenciación

La secuenciación de *A. chrysaetos* generó dos archivos paired-end en formato fastqc para cada uno de los seis individuos. En la Tabla IV se muestran estadísticos básicos producto de la secuenciación.

El rendimiento de la secuenciación fue muy dispar entre los individuos. En general, la plataforma de secuenciación ILLUMINA HiSeq 4000 (Ach 6) tuvo la calidad de secuenciación más baja de todos, como se observa en el reporte de calidad de secuencias generado (Tabla VI del Anexo 1). El número más alto de secuencias generadas es de 3,613,918 y correspondió al individuo Ach 1, mientras que el número más bajo de secuencias producidas (599,935) al individuo Ach 4. Se observaron diferencias claras entre las dos plataformas de secuenciación utilizadas a nivel de la longitud de secuencia producidas, donde las secuencias con ILLUMINA Miseq V.3 son de 301 pb e ILLUMINA HiSeq 4000 produjo secuencias que van de 35 a 151 pb.

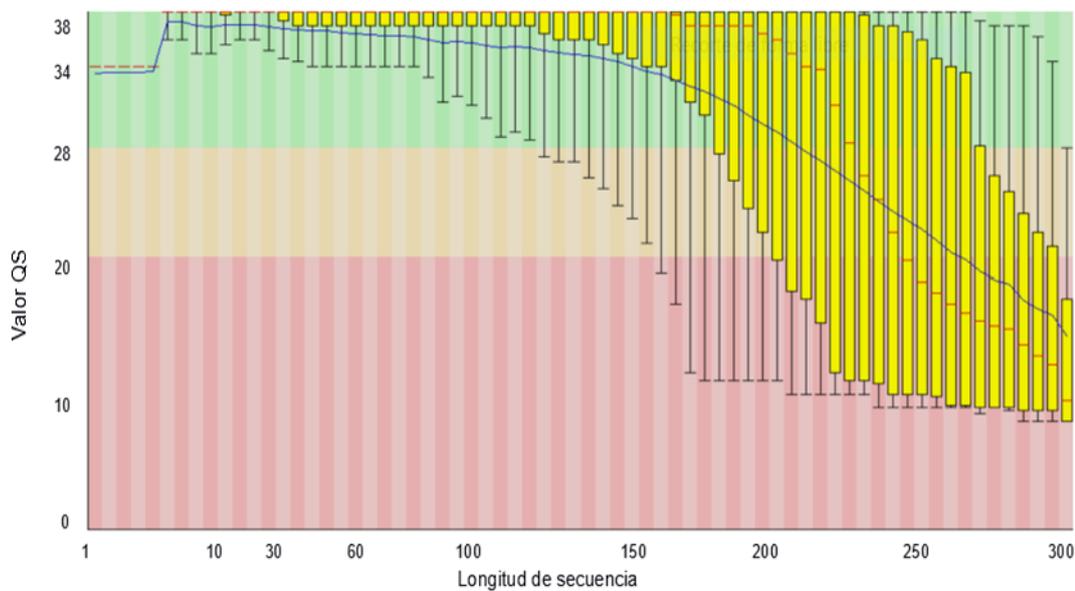
*Tabla IV.* Estadísticos básicos de los archivos de secuenciación. r1: archivo de secuencias forward en sentido 5'a 3'; r2: archivo de secuencias reverse en sentido 3'a 5'; %GC: porcentaje del contenido de guaninas y citocinas en las secuencias.

<b>Individuo</b>	<b>Lecturas totales</b>	<b>Longitud de lecturas</b>	<b>% GC</b>	<b>% de secuencias con adaptadores</b>
Ach 1	3,613,918	301	r1: 47 % r2: 48 %	r1: 80 % r2: 70 %
Ach 2	839,211	301	r1: 42 % r2: 43 %	r1: 40 % r2: 35 %
Ach 3	1,627,306	301	r1: 47 % r2: 48 %	r1: 68 % r2: 58 %
Ach 4	599,935	301	r1: 43 % r2: 44 %	r1: 58 % r2: 45 %
Ach 5	789,685	301	r1: 44 % r2: 45 %	r1: 55 % r2: 45 %

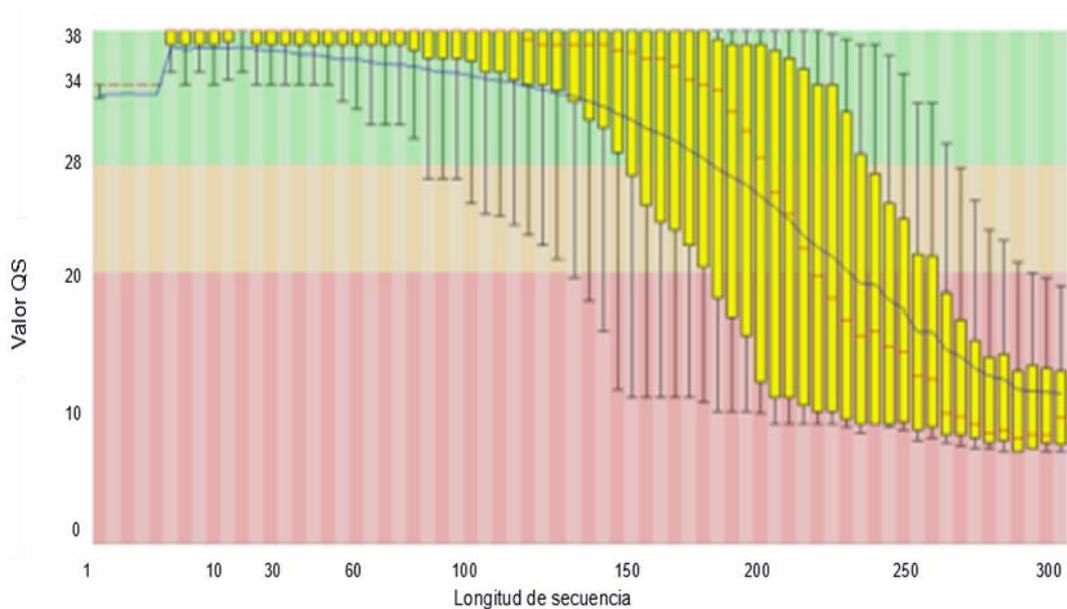
Ach 6	3,490,423	35-151	r1: 50 % r2: 50 %	r1: 0 % r2: 0 %
-------	-----------	--------	----------------------	--------------------

En las Figuras 4-7 se muestran los gráficos de calidad de las secuencias r1 y r2 de los individuos cuyos números de secuencias producidas fueron mayores (Ach 1 y Ach 6) y que representan las dos distintas plataformas de secuenciación implementadas.

El parámetro principal de la calidad de las lecturas producidas en la secuenciación es el valor de calidad QS, porque el comportamiento de los ensamblajes y análisis posteriores dependen mucho de él. La calidad de las secuencias de cada uno de los archivos de secuencias sin importar de que individuos se trate o bien de que archivo sea (r1 o r2), es muy variable a través de la distribución de la longitud de las secuencias. Los dos primeros tercios de la distribución de la longitud de las secuencias tienen valores de calidad que van de aceptable a muy buena, es decir, por arriba de 20 QS, pero el último tercio de la distribución de la longitud (las secuencias de mayor longitud y por ende las más deseables) tiene baja calidad y dicho valor está por debajo de 20 QS, incluso a un valor tan bajo como 2 QS encontrado en el individuo Ach 6 (Figuras 6 y 7). Los gráficos de calidad de ambos archivos (r1 y r2) de secuencias de los individuos Ach 2-5 se pueden observar a detalle en las Figuras 2-9 del Anexo 1.



**Figura 4.** Gráfico que muestra el valor de calidad del archivo de secuencias r1 (forward) del individuo Ach 1 de *Aquila chrysaetos*.



**Figura 5.** Gráfico que muestra el valor de calidad del archivo de secuencias r2 (reverse) del individuo Ach 1 de *Aquila chrysaetos*.

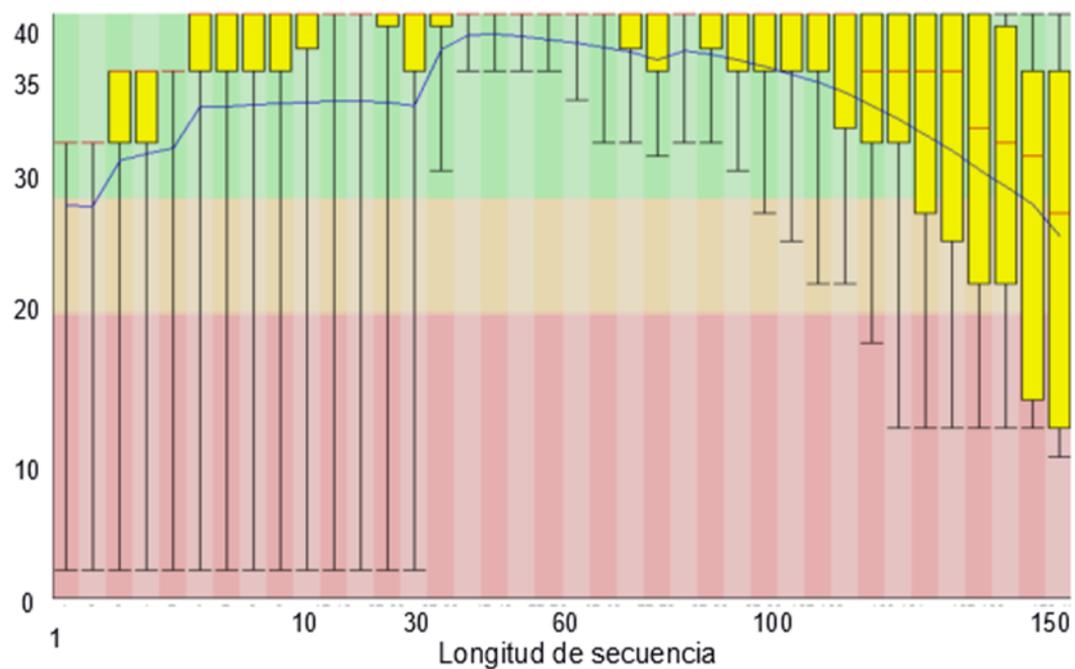


Figura 6. Gráfico que muestra el valor de la calidad del archivo de secuencias r1 (forward) del individuo Ach 6 de *Aquila chrysaetos*.

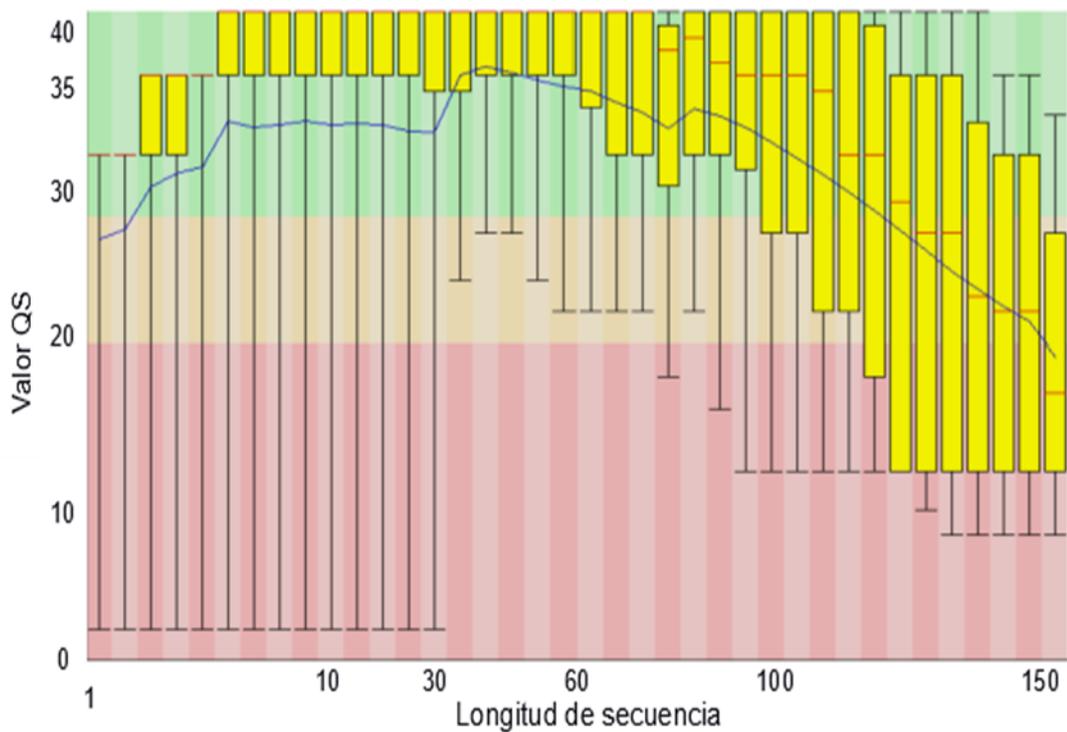


Figura 7. Gráfico que muestra el valor de la calidad del archivo de secuencias r2 (reverse) del individuo Ach 6 de *Aquila chrysaetos*.

### VI.3 Preprocesamiento de las secuencias

Una vez hecha la limpieza de las secuencias se mejoraron todos los apartados del reporte de calidad. Uno de los principales apartados es el de calidad de las secuencias por base. En dicho apartado, el estatus final para todos los individuos en ambos archivos (r1 y r2) de secuencias es adecuado (en color verde). Otros apartados importantes que influyen mucho en la calidad de los análisis posteriores son el apartado de secuencias sobre-representadas y el de contenido de adaptadores. El estatus de calidad después de la limpieza fue adecuado (Tablas I-VI que se encuentran en el Anexo 1).

En el apartado de distribución de la longitud de las secuencias algunos archivos se vieron afectados al retirar los adaptadores y las secuencias sobre-representadas, por lo que pasaron de tener un estatus de calidad adecuado a tener un estatus de atención. A pesar de que lo anterior parece ser un retroceso en la calidad de las secuencias, es preferible tener secuencias más pequeñas, pero sin el ruido que representan los adaptadores dentro de las lecturas y también las secuencias sobre-representadas. La Tabla V muestra el resultado final del proceso de limpieza y control de calidad de los archivos de secuencias. En ella se encuentran el número de pares de lecturas iniciales, es decir el número de secuencias pareadas totales generadas por el proceso de secuenciación, el número de secuencias que pasaron el control de calidad y el número de secuencias descartadas.

*Tabla V.* Número de secuencias crudas y tratadas mediante el control de calidad.

<b>Individuo</b>	<b>Secuencias pareadas iniciales</b>	<b>Pares de lecturas finales</b>	<b>Secuencias descartadas</b>
Ach 1	3,324,579	3,251,746	72,833 (2.19%)
Ach 2	738,931	718,629	20,302 (2.75%)
Ach 3	1,364,493	1,335,053	29,440 (2.16%)
Ach 4	539,825	526,875	12,950 (2.40%)
Ach 5	695,574	681,055	14,519 (2.09%)
Ach 6	2,714,901	2,656,332	58,569 (2.16%)

#### VI.4 Ensamble por referencia

La Tabla VI muestra el resultado del ensamble por referencia en términos de la profundidad media del ensamble y del número de lecturas alineadas, dicho en otras palabras, aquellas lecturas que forman parte del genoma mitocondrial y que coinciden con el genoma mitocondrial de referencia utilizado (acceso GenBank KF905228).

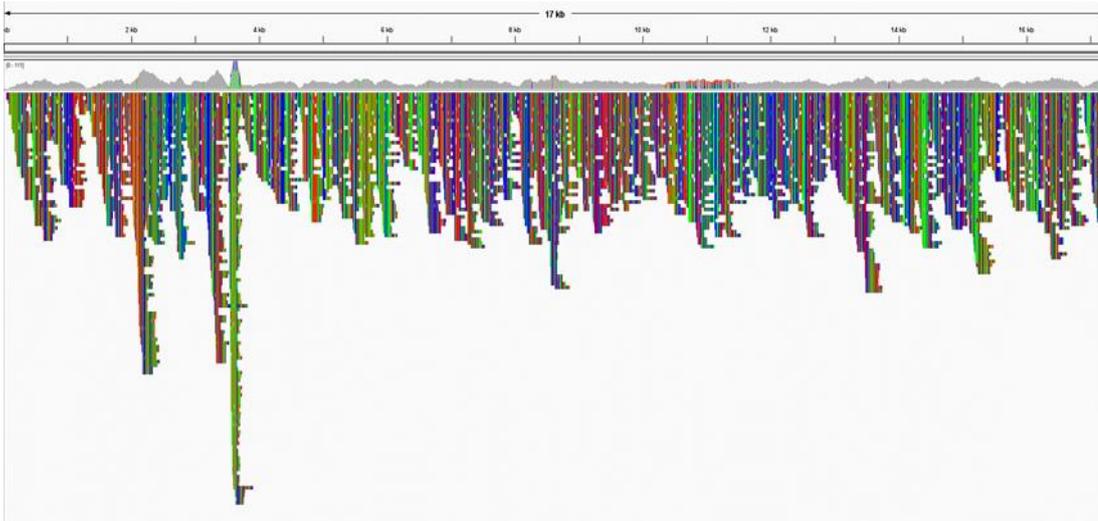
*Tabla VI.* Reporte de lecturas alineadas con el ensamble por referencia.

<b>Individuo</b>	<b>Lecturas totales</b>	<b>Parejas utilizadas</b>	<b>Tasa de alineamiento</b>	<b>Lecturas alineadas</b>	<b>Profundidad media</b>
Ach 1	2,977,821	5,952,996	0.04 %	2,661	30.5
Ach 2	620,569	1,240,200	0.08 %	938	11.6
Ach 3	1,163,205	2,325,548	0.04 %	873	11.9
Ach 4	462,693	924,052	0.15 %	1,350	18.0
Ach 5	585,488	1,169,902	0.09 %	1,083	14.6
Ach 6	2,290,738	4,557,654	0.52 %	23,931	168.3

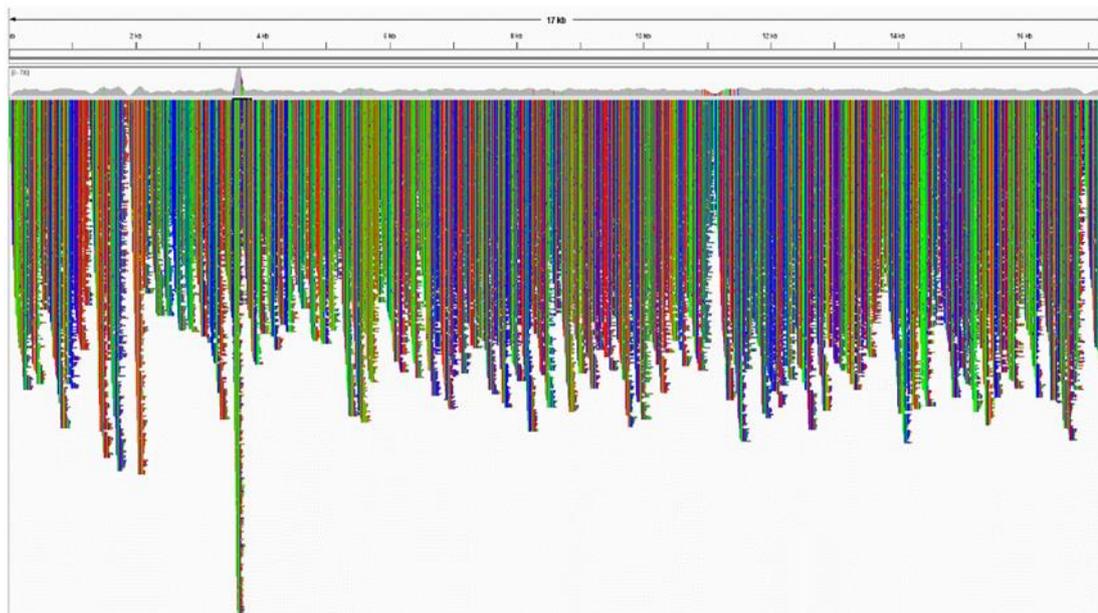
La Figura 8 muestra el ensamble por referencia para el individuo Ach 1. En la parte superior de la figura se muestra la longitud del genoma

mitocondrial de referencia de 17,332 pb. Por debajo de la longitud del genoma de referencia se observa una región sombreada en gris a lo largo de todo el genoma, dicha región presenta ciertas partes en colores que representan cambios puntuales con respecto al genoma de referencia. Estos sitios son altamente informativos por ser potenciales sitios de SNP's o INDEL's. De los individuos secuenciados por ILLUMINA MiSeq V.3, el ensamble por referencia con mayor número de lecturas corresponde al del individuo Ach 1, mientras que el ensamble por referencia con la mayor tasa de alineamiento de secuencias y con el mayor número de lecturas alineadas es aquél del individuo Ach 6 (secuenciado por ILLUMINA HiSeq 4000) con una tasa de 0.52% de secuencias alineadas y un total de 23,931 lecturas que conforman el ensamble.

En la Figura 9 se ofrece el ensamble por referencia del individuo Ach 6. En este caso, se obtuvo una cobertura del 100%, es decir, no hay región del genoma de referencia que no se encuentre representada por lecturas del individuo Ach 6. Además, la profundidad<sup>A</sup> media del ensamble fue de 168.3X, inclusive se observa una región con una profundidad de 727X.



*Figura 8.* Ensamble por referencia del genoma mitocondrial de *Aquila chrysaetos*, para el individuo Ach 1 utilizando como referencia KF905228 del GenBank.



*Figura 9.* Ensamble por referencia del genoma mitocondrial de *Aquila chrysaetos*, del individuo Ach 6 utilizando como referencia KF905228 del GenBank.

Al comparar los ensambles por referencia de los individuos Ach 1 y Ach 6 se observan grandes diferencias. Por ejemplo, la profundidad media del ensamble de Ach 1 es de 30.5X, mientras que la profundidad media del ensamble por referencia de Ach 6 es de 168.3X. El total de secuencias

alineadas es mayor en el individuo Ach 6 a pesar de que el individuo Ach 1 tuvo un mayor número de lecturas totales y de parejas de lecturas utilizadas para realizar el ensamble por referencia. De la misma manera que se comporta el ensamble por referencia de Ach 1, se comportan los individuos Ach 2-5, incluso muestran una menor profundidad, menor número de lecturas totales y de parejas de secuencias utilizadas (Figuras 10-13 del Anexo 1). En general, los ensambles por referencia de los individuos Ach 1-5 son poco confiables porque el número de secuencias que conforman los ensambles por referencia es bajo.

#### VI.5 Ensamble *de novo*

En la Tabla VII se muestra el resultado del ensamble *de novo* de los seis individuos secuenciados Ach 1-6. El individuo Ach 6 tuvo el mejor comportamiento en ambos ensambles. La profundidad media del ensamble *de novo* de Ach 6 es de 8.0X y el número de scaffolds armados es el tercero más bajo con 2,532. Además, el scaffold más largo generado (17,472 pb) resultó ser el que contenía todo el genoma mitocondrial de Ach 6, por lo que no fue necesario alinear varios scaffolds para recuperar la longitud total. Así, el valor N50 (valor de contigüidad del genoma, que indica la longitud del scaffold más pequeño de los scaffolds más largos que representan el 50% de la totalidad del genoma) fue de 572 siendo éste el segundo más bajo y solo después de la muestra Ach 1 con un N50=198.

El rendimiento de los datos al hacer el ensamble *de novo* para los individuos Ach 1-5 fue inferior y con ninguno de estos individuos se logró

recuperar el genoma mitocondrial completo por la falta de secuencias que sustenten los contigs y posteriormente a los scaffolds.

*Tabla VII.* Estadísticos básicos del ensamble *de novo* del genoma mitocondrial de *Aquila chrysaetos*. n-d: no disponible.

Individuo	Profundidad media	Scaffolds	Longitud del scaffold más grande	N50	Lecturas totales	% GC
Ach 1	2.0	2,600,450	10,270	198	5,955,642	n-d
Ach 2	2.0	4,645	9,927	711	1,241,138	41.4
Ach 3	5.0	2,948	12,235	756	2,326,410	51.5
Ach 4	2.0	1,805	16,761	713	925,386	42.8
Ach 5	3.0	1,938	19,231	729	1,170,976	46.6
Ach 6	8.0	2,532	17,472	572	4,581,476	47.1

#### VI.6 Anotación del genoma mitocondrial de *Aquila chrysaetos*

La anotación y curación del genoma mitocondrial se hizo a partir del scaffold más largo del ensamble *de novo* del individuo Ach 6, debido a que representa la totalidad del genoma mitocondrial. Esto se determinó por las búsquedas mediante la herramienta BLAST del GenBank. Si bien los ensambles *de novo* de los otros cinco individuos generaron scaffolds de tamaño considerable, ninguno contiene la totalidad del genoma mitocondrial en un solo scaffold y al unir estos scaffolds, tampoco se obtiene la longitud total de la molécula.

Las tres herramientas de anotación identificaron ARNt's. Además, los codones de dichos ARNt's coinciden, su longitud es similar y, los localizan en la misma cadena del ADN mitocondrial. Ninguna de las herramientas de anotación identificó intrones dentro de la secuencia.

El programa MITOS webserver (Tabla VIII) logró identificar la totalidad de los elementos génicos, como los 22 ARNt's, 2 ARNr's y los 13 genes codificantes de proteínas que conforman el genoma mitocondrial. Ya que la anotación hecha con esta herramienta resultó ser la más completa y congruente con los tamaños esperados de las sub unidades del ARN ribosómico, la curación se hizo a partir de esta anotación. También se determinó, manualmente, la ubicación de la RC del ADNmt.

*Tabla VIII.* Anotaciones hechas mediante MITOS webserver. (+) cadena pesada; (-) cadena ligera.

<b>Gen</b>	<b>Inicio</b>	<b>Termino</b>	<b>Cadena</b>	<b>Longitud</b>	<b>Codón ARNt</b>
ARNt F	1	69	+	69	ttc
ARNr S 12S	70	1,041	+	972	
ARNt V	1,042	1,113	+	72	gta
ARNr L 16S	1,114	2,709	+	1,596	
ARNt L	2,701	2,775	+	75	tta
ND 1	2,784	3,761	+	978	
ARNt I	3,760	3,831	+	72	atc
ARNt Q	3,844	3,915	-	72	caa
ARNt M	3,915	3,983	+	69	atg
ND 2	3,984	5,024	+	1,041	
ARNt W	5,023	5,094	+	72	tga
ARNt A	5,096	5,164	-	69	gca
ARNt N	5,168	5,239	-	72	aac
ARNt C	5,243	5,308	-	66	tgc
ARNt Y	5,309	5,380	-	72	tac
CO I	5,380	6,930	+	1,551	
ARNt S	6,922	6,996	-	75	tca
ARNt D	7,001	7,069	+	69	gac
CO II	7,072	7,755	+	684	

ARNt K	7,756	7,827	+	72	aaa
ATP 8	7,829	7,996	+	168	
ATP 6	7,987	8,670	+	684	
CO III	8,670	9,452	+	783	
ARNt G	9,454	9,522	+	69	gga
ND 3a	9,523	9,729	+	207	
ND 3b	9,731	9,874	+	144	
ARNt R	9,882	9,950	+	69	cga
ND 4L	9,952	10,248	+	297	
ND 4	10,242	11,648	+	1,407	
ARNt H	11,620	11,689	+	70	cac
ARNt S	11,690	11,755	+	66	agc
ARNt L	11,756	11,826	+	71	cta
ND 5	11,827	13,644	+	1,818	
Cyt b	13,656	14,798	+	1,143	
ARNt T	14,803	14,872	+	70	aca
ARNt P	16,031	16,101	-	71	cca
ND 6	16,114	16,632	-	519	
ARNt E	16,636	16,707	-	72	gaa
Región control	14,873	16,030	+	1,158	
Ψ región control	16,708	17,472	+	765	

---

A pesar de que RNAweasel y Mfanot (Tablas VIII y IX del Anexo 1) identificaron principalmente ARNt's y algunos otros elementos génicos del genoma mitocondrial, la anotación con estas dos herramientas no fue informativa.

### VI.7 Construcción del mapa del genoma mitocondrial de *Aquila chrysaetos*

El mapa del genoma mitocondrial de *Aquila chrysaetos* descrito en este estudio (Figura 10) mostró diferencias con respecto al mitogenoma reportado por Doyle et al. (2014). Por ejemplo, la fracción del gen que codifica para la proteína NADH 3 que Doyle et al. (2014) reportan como una sola, aquí se anotó dividida en dos segmentos (ND 3a y ND 3b). Además, aquí se reporta la anotación de una  $\psi$ RC que no fue reportada por Doyle et al. (2014). Otra diferencia consiste en el tamaño del genoma mitocondrial, siendo más grande (17,472 pb) el aquí descrito que el reportado por Doyle et al. (2014) (17,332 pb), con una diferencia de 140 pb.

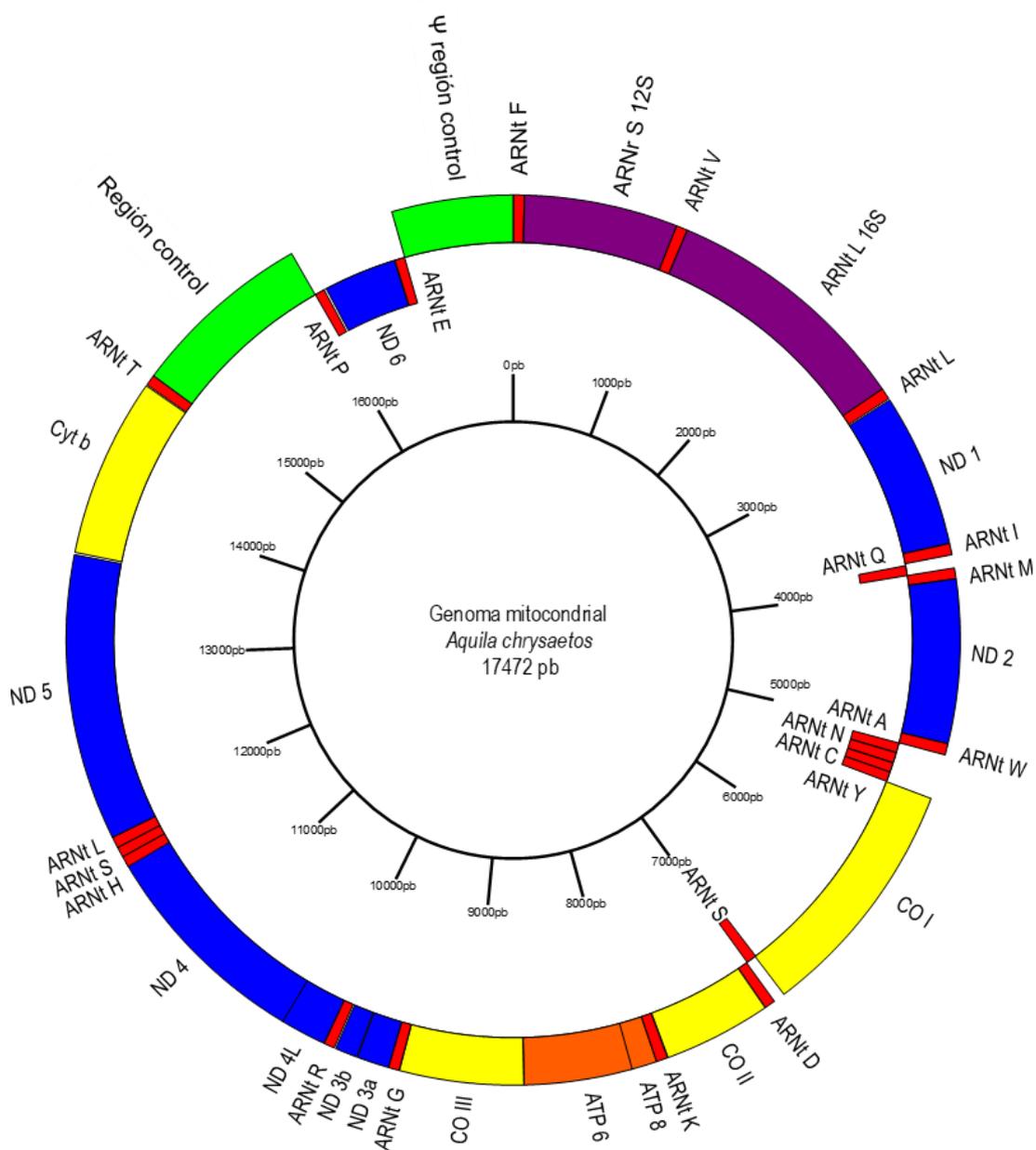
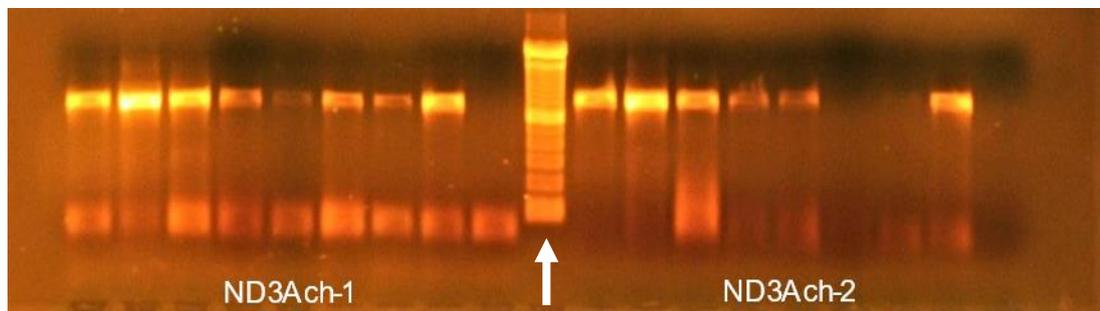


Figura 10. Mapa del mitogenoma anotado *de novo* del Águila real (*Aquila chrysaetos*) de Baja California, México. En la parte interior central del diagrama, en forma de anillo se muestra una escala de la longitud del genoma mitocondrial en sentido de las manecillas del reloj.

### VI.8 Comprobación experimental de la anotación de la región ND 3

La amplificación de los fragmentos de la región ND 3 fue exitosa en todas las muestras con el primer par de cebadores<sup>A</sup> (ND3Ach-1) y en seis de las ocho muestras para el segundo par de cebadores (ND3Ach-2). Además, el tamaño

de los fragmentos amplificados fue el tamaño esperado (entre 700 y 800 pb) de acuerdo con el marcador de peso molecular (Figura 11).



**Figura 11.** Fotografía del gel de electroforesis que muestra el resultado de la amplificación de los fragmentos para ambos pares de cebadores. ND3Ach-1 corresponde al primer par de cebadores diseñado, ND3Ach-2 corresponde al segundo par de cebadores y la flecha blanca indica el marcador de peso molecular.

Las secuencias de los fragmentos que codifican para la proteína ND 3 amplificados de las ocho muestras fue exitosa. Particularmente, se encontró una pequeña región con siete bases nucleotídicas que se compone de un codón de paro AGA de la proteína ND 3, enseguida una adenina (A) y por último un codón de inicio ATC que codifica para isoleucina (Figura 12). Esta región divide al fragmento ND 3 en las dos regiones aquí propuestas (Nd 3a y ND 3b).

```

AGTAGCCATCCTATTCCTGCTATTCGACCT AGA A ATC GCACTCTTACTCCCTCTGCCATGA

```

**Figura 12.** Fracción del fragmento de la secuencia del ADN mitocondrial que codifica para la proteína ND 3 (NADH deshidrogenasa sub unidad 3) de *Aquila chrysaetos*. Se muestra en rojo un codón de paro, en azul una adenina y en verde un codón de inicio que codifica para isoleucina de acuerdo con el código genético mitocondrial para vertebrados.

## VI.9 Análisis del grado de sintenia

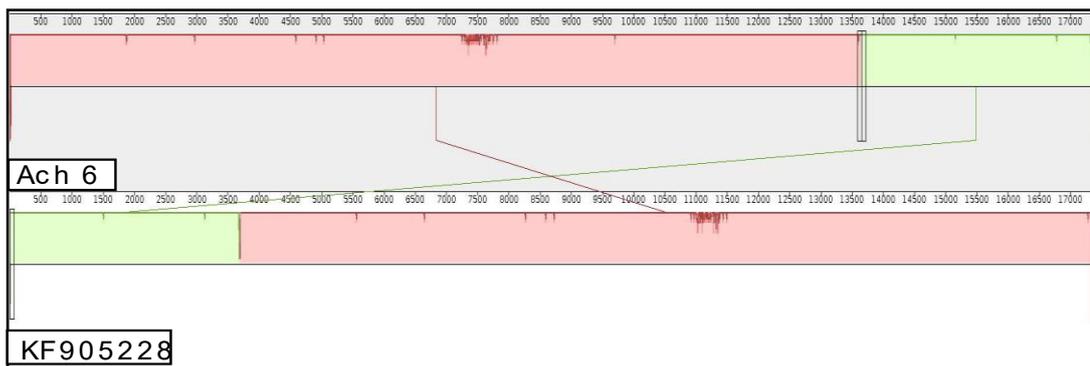
En el presente análisis, se detectó una división del gen que codifica para la sub unidad 3 de la proteína NADH deshidrogenasa (Tabla IX) pero es importante hacer notar que el análisis de sintenia no considera y excluye duplicaciones.

*Tabla IX.* Orden de los genes del genoma mitocondrial de referencia KF905228 y del individuo Ach 6 de *Aquila chrysaetos*. El símbolo + indica que dicha característica génica se encuentra en la cadena pesada del genoma mitocondrial y el símbolo – indica que dicha característica se encuentra en la cadena ligera.

<b>Genoma</b>	<b>Orden de genes</b>
KF905228	+Cyt b +ARNt T -ARNt P -ND 6 -ARNt E +ARNt F +ARNr S 12S +ARNt V +ARNr L 16S +ARNt L +ND 1 +ARNt I -ARNt Q +ARNt M +ND 2 +ARNt W -ARNt A -ARNt N -ARNt C -ARNt Y +CO I -ARNt S +ARNt D +CO II +ARNt K +ATP 8 +ATP 6 +CO III +ARNt G +ND 3 +ARNt R +ND 4L +ND 4 +ARNt H +ARNt S +ARNt L +ND 5
Ach 6	+ARNt F +ARNr S 12S +ARNt V +ARNr L 16S +ARNr L +ND 1 +ARNt I -ARNt Q +ARNt M +ND 2 +ARNt W -ARNt A -ARNt N -ARNt C -ARNt Y +CO I -ARNt S +ARNt D +CO II +ARNt K +ATP 8 +ATP 6 +CO III +ARNt G +ND 3a +ND 3b +ARNt R +ND 4L +ND 4 +ARNt H +ARNt S +ARNt L +ND 5 +Cyt b +ARNt T -ARNt P -ND 6 -ARNt E

El orden lineal del genoma mitocondrial de Ach 6 coincide con el genoma de referencia (Tabla IX). Aunque el orden no comienza en el mismo sitio en ambos mitogenomas, no significa que el orden sea distinto ya que esta molécula de ADN es circular. Por ejemplo, una fracción del genoma

mitocondrial del individuo Ach 6 que va desde el gen que codifica para la proteína Cyt b hasta el ARN de transferencia del ácido glutámico (ARNt E) que se encuentra al final; en cambio, dicha fracción se encuentra al inicio en el genoma de referencia (Figura 13), lo que indica que la arquitectura de los genomas mitocondriales del individuo Ach 6 y KF905228 es equivalente.



**Figura 13.** Diagrama comparativo del orden de los elementos génicos de los genomas mitocondriales de referencia KF905228 y el ensamblado del individuo Ach 6 de *Aquila chrysaetos*.

#### VI.10 Llamado de SNP's e INDEL's en el mitogenoma de *Aquila chrysaetos*

Potenciales SNP's e INDEL's fueron identificados dentro del mitogenoma de los individuos Ach 1-6 de *A. chrysaetos* con respecto al genoma de referencia KF905228 (Tabla X). Los SNP's e INDEL's identificados en el individuo Ach 6 (resaltados en negritas) se consideran confiables, dado que la profundidad del ensamble fue buena, con una media de 168.3X y la calidad del llamado de estos fue alta (Tablas X-XV del Anexo 1). Además, la gran mayoría de SNP's e INDEL's de este genoma se encuentran dentro de genes que codifican para alguna proteína, genes de ARN de transferencia y ribosómicos. Solamente un SNP se encuentra en la RC y un INDEL en la  $\psi$ RC, lo que es de esperarse

debido a que el genoma mitocondrial en las aves no contiene espaciadores intergénicos (Figura 10).

*Tabla X.* Llamado de SNP's e INDEL's de los individuos Ach 1-6 contra el genoma de referencia KF905228 de *Aquila chrysaetos*.

Posición	Referencia	Ach 1	Ach 2	Ach 3	Ach 4	Ach 5	Ach 6	Gen
<b>SNP's</b>								
1,505	G		A		A		<b>A</b>	RC
1,515	T				C			
3,130	G	A		A			<b>A</b>	
3,566	A				G			ND
3,695	T	A	A		A			6
3,696	T	A	A		A			
5,557	G	A	A	A	A		<b>A</b>	ARN rL
6,646	G	A	A				<b>A</b>	ND
7,141	A	G						1
8,269	C	T	A				<b>T</b>	ND
8,593	A	G		G	A		<b>G</b>	2
8,723	G	A	A	A			<b>A</b>	ARN tW
10,925	C	T		T	T	T	<b>T</b>	
10,963	T	C						
10,970	C	T		T	T	T	<b>T</b>	CO
10,975	T	C						II
11,003	A					G	<b>G</b>	
11,011	C						<b>T</b>	
11,205	A						<b>G</b>	

11,213	C						<b>T</b>	
11,237	C						<b>T</b>	
11,282	G						<b>A</b>	
11,284	G	T						
11,285	A						<b>C</b>	
11,351	T						<b>C</b>	
11,354	A						<b>G</b>	
11,378	C				T		<b>T</b>	
11,432	C	T	T	T	T		<b>T</b>	
11,489	A	C	C	C	C	C	<b>C</b>	
11,493	T			T				ARN
11,496	G			A				tK
11,505	T			G				
17,278	G		A				<b>A</b>	ND 5
<b>INDEL's</b>								
							<b>CTAA</b>	
3	CTAA						<b>CTTC</b>	Cyt
							<b>CAAA</b>	b
							<b>CTAA</b>	
		TGTG						
3,550	T	AACA						ψRC
		A						
		CCCA					<b>CCAA</b>	ARN
3,701	CAAA	CCAA					<b>CAAT</b>	tF
		TA					<b>AT</b>	
13,378	T	TC	TC	TC	TC	TC	<b>TC</b>	ND 3

---

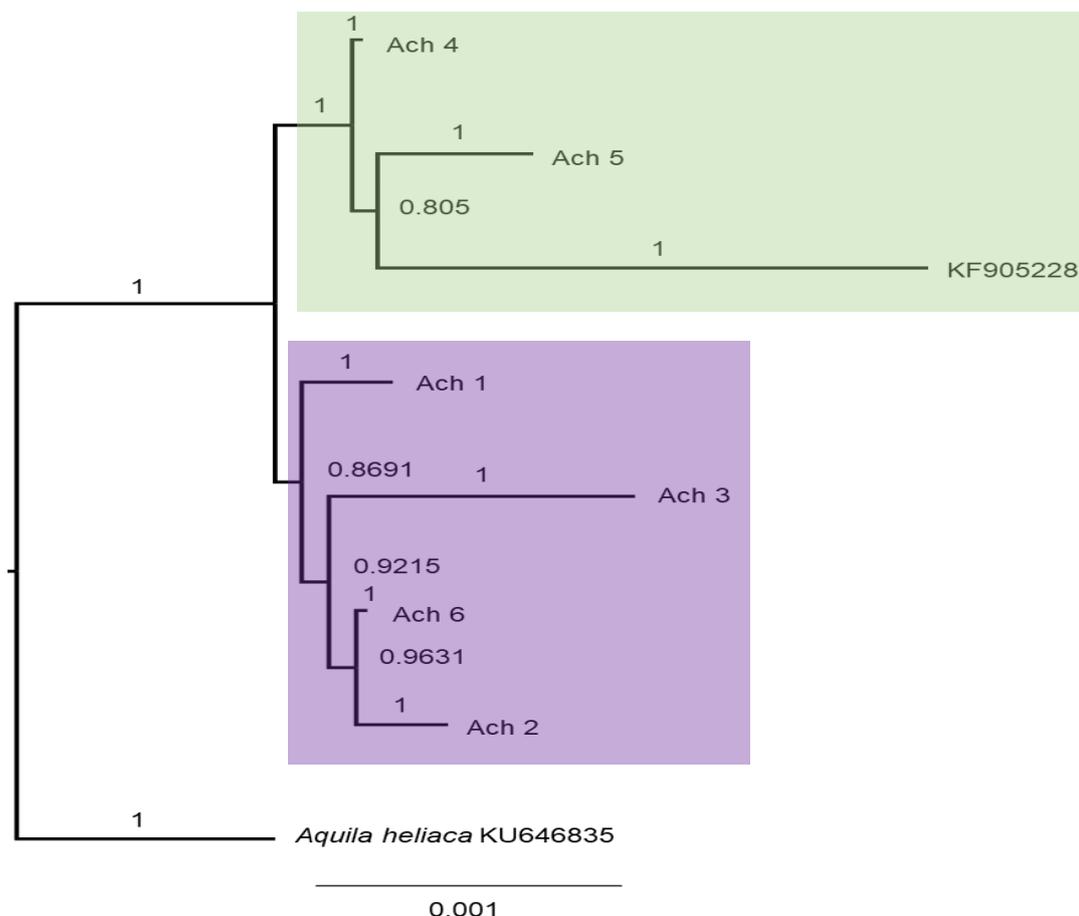
		<b>AACT</b>	
		<b>TCCA</b>	
17,325	AACTTCCA	<b>AACT</b>	Cyt
		<b>ACTT</b>	b
		<b>CCA</b>	

---

El gen CO II presentó la mayor cantidad de SNP's con respecto al genoma de referencia en los seis individuos, mientras que los genes Cyt b y ARNt F presentan los mayores cambios de longitud en los INDEL's. Además, en el gen ND 3 existe un cambio que es consistente en las secuencias de los seis individuos con respecto al genoma de referencia. Los SNP's e INDEL's llamados de los seis individuos deben ser probados experimentalmente en estudios poblacionales para conocer su poder informativo.

### VI.11 Reconstrucción de relaciones filogenéticas mediante el ADN mitocondrial

La reconstrucción filogenética del mitogenoma del Águila real por el método bayesiano de inferencia filogenética, muestra dos linajes de secuencias distintos. El primero de estos linajes (linaje 1) está formado por los individuos Ach 5, KF905228 y Ach 4, sustentado con un valor de probabilidad de la rama basal de 1.0. El linaje 2 está constituido por las secuencias Ach 2, Ach 6, Ach 3 y Ach 1, soportado con un valor de probabilidad de 0.8691 en su rama basal (Figura 14.)



**Figura 14.** Árbol filogenético del mitogenoma, construido mediante el método bayesiano de inferencia filogenética a partir de la secuencia del mitogenoma de referencia KF905228 disponible en el GenBank, de cada uno de los individuos de *Aquila chrysaetos* que fueron secuenciados (Ach 1-6) y como

grupo externo la especie *Aquila heliaca* KU646835. Se muestran dos linajes de secuencias, el linaje 1 en un recuadro verde y el linaje 2 en un recuadro morado, y el valor de probabilidad de cada una de las ramas  $\geq 0.5$ .

La reconstrucción filogenética por máxima verosimilitud y método bayesiano, y los fenogramas construidos por Neighbor-Joining no fueron informativos a nivel genealógico intraespecífico para los genes CO I, Cyt b, ND 2, la  $\psi$ RC y la RC. Para el mitogenoma, la reconstrucción filogenética por máxima verosimilitud y el fenograma por Neighbor-Joining tampoco fueron informativos (Figuras 14-30 del Anexo 1).

#### VI.12 Análisis de diversidad genética

Las Águilas reales de Baja California fueron analizadas de acuerdo con los resultados de la sección anterior, es decir, a través de los dos linajes de secuencias que fueron identificados luego de la reconstrucción de las relaciones filogenéticas y también por región geográfica. Ambos linajes presentan una diversidad haplotípica ( $H_d$ ) y nucleotídica ( $\pi$ ) nula para CO I, Cyt b y RC, con valores  $H_d=0$  y  $\pi=0$  y los seis individuos muestran un mismo haplotipo para ambos genes (Tabla XI).

El linaje 1 no presentó ningún sitio segregante entre las secuencias del gen codificante para la proteína ND 2, hubo un solo haplotipo y sus valores de diversidad haplotípica, nucleotídica y el número promedio de diferencias de nucleótidos entre secuencias fueron iguales a cero. El linaje 2 presentó un sitio segregante entre las secuencias y dos haplotipos. Además, su diversidad haplotípica fue moderada ( $H_d=0.5$ ), y la nucleotídica baja ( $\pi=0.00049$ ). El número promedio de diferencias de nucleótidos entre secuencias fue de  $K=0.5$ .

Por su parte, el análisis de diversidad genética para la  $\psi$ RC del linaje 1 no presentó sitios segregantes, hubo un haplotipo y sus valores de diversidad haplotípica, nucleotídica y el número promedio de diferencias de nucleótidos entre secuencias fueron iguales a cero. El linaje 2 presentó un sitio segregante, dos haplotipos, diversidad haplotípica moderada ( $H_d = 0.5$ ), nucleotídica baja ( $\pi = 0.00107$ ) y número promedio de diferencias de nucleótidos entre secuencias de  $K=0.5$ .

El análisis de diversidad genética del mitogenoma de las Águilas reales de Baja California del linaje 1 presentó 10 sitios segregantes, dos haplotipos distintos, diversidad haplotípica muy alta ( $H_d = 1.0$ ), nucleotídica muy baja ( $\pi = 0.00058$ ) y número promedio de diferencias de nucleótidos entre secuencias de 10. El mitogenoma del linaje 2 presentó 30 sitios segregantes, cuatro haplotipos, diversidad haplotípica también muy alta ( $H_d = 1.0$ ), nucleotídica muy baja ( $\pi = 0.00095$ ) y entre las secuencias hubo un promedio de  $K=15.2$  diferencias de nucleótidos.

*Tabla XI.* Diversidad genética mitocondrial de los distintos genes y regiones génicas utilizadas. Calculada a partir de secuencias propias y las obtenidas del GenBank de las distintas regiones de la distribución de *Aquila chrysaetos*, y para la región de Baja California ambos linajes por separado. Se muestra el número total de secuencias utilizadas (N), la longitud de las secuencias en pares de bases contempladas (pb), el número de sitios segregantes (S), el número de haplotipos (Nh), la diversidad haplotípica ( $H_d$ ), la diversidad nucleotídica ( $\pi$ ) y el número promedio de diferencias de nucleótidos entre secuencias (K).

Secuencia mitocondrial	Región geográfica	N	pb	S	Nh	$H_d$	$\pi$	K
CO I	Linaje 1	2	394	0	1	0.0	0.0	0.0
	Linaje 2	4	394	0	1	0.0	0.0	0.0
	Baja California	6	394	0	1	0.0	0.0	0.0

	Norteamérica	5	394	3	2	0.4	0.00305	1.2
	Europa	4	394	0	1	0.0	0.0	0.0
	Global	16	394	3	2	0.12	0.00095	0.375
Cyt b	Linaje 1	2	965	0	1	0.0	0.0	0.0
	Linaje 2	4	965	0	1	0.0	0.0	0.0
	Baja California	6	965	0	1	0.0	0.0	0.0
	Norteamérica	3	965	9	2	0.66	0.00622	6.0
	Europa	8	965	4	3	0.60	0.00752	1.857
	Global	17	965	4	3	0.40	0.00512	1.265
ND 2	Linaje 1	2	1038	0	1	0.0	0.0	0.0
	Linaje 2	4	1038	1	2	0.5	0.00049	0.5
	Baja California	6	1038	2	3	0.73	0.00111	1.113
	Norteamérica	2	1038	0	1	0	0	0
	Global	8	1038	2	3	0.67	0.00108	1.107
$\Psi$ región control	Linaje 1	2	467	0	1	0.0	0.0	0.0
	Linaje 2	4	467	1	2	0.5	0.00107	0.5
	Baja California	6	467	2	3	0.73	0.00186	0.867
	Asia	6	467	7	6	1.0	0.00544	2.533
	Global	14	467	9	8	0.85	0.00387	1.802
Región control	Linaje 1	2	291	0	1	0.0	0.0	0.0
	Linaje 2	4	291	0	1	0.0	0.0	0.0
	Baja California	6	291	1	2	0.53	0.00209	0.533
	Norteamérica	13	291	3	7	0.93	0.00481	1.4
	Europa	37	291	17	17	0.95	0.02968	5.937

	Asia	20	291	9	10	0.94	0.01125	3.274
	Global	76	291	17	24	0.92	0.02542	4.169
Mitogenoma	Linaje 1	2	17,407	10	2	1.0	0.00058	10.0
	Linaje 2	4	17,407	30	4	1.0	0.00095	15.167
	Baja California	6	17,407	43	6	1.0	0.00092	15.8
	Global	7	17,407	105	7	1.0	0.00193	32.857

Para CO I, la diversidad haplotípica y nucleotídica estimadas para las secuencias provenientes de Europa fue nula ( $H_d = 0$  y  $\pi = 0$ ), con ningún sitio segregante entre las secuencias analizadas, ninguna diferencia de nucleótidos entre secuencias y el único haplotipo que presentaron fue compartido con Baja California (Tabla XVI del Anexo 1). Para Norteamérica la diversidad haplotípica fue de  $H_d = 0.4$  y la nucleotídica de  $\pi = 0.00305$ , hubo tres sitios segregantes, el número promedio de diferencias de nucleótidos entre secuencias fue de  $K = 1.2$  y de los dos haplotipos encontrados, uno fue compartido con Europa y Baja California (Tabla XVI del Anexo 1).

La diversidad haplotípica y nucleotídica de Cyt b para Norteamérica y Europa fueron semejantes, con valores de  $H_d = 0.667$  y  $0.607$  y  $\pi = 0.00622$  y  $0.00752$ , respectivamente. El número de sitios segregantes para Norteamérica fue de nueve, representados en dos haplotipos, uno de ellos compartido con Baja California (Tabla XVII del Anexo 1). El promedio de diferencias de nucleótidos entre secuencias fue de  $K = 6$ . El número de sitios segregantes para Europa fue de cuatro, el número promedio de diferencias de nucleótidos entre secuencias arrojó una  $K = 1.85$  y tres haplotipos.

La diversidad genética estimada de ND 2 para Norteamérica no presentó sitios segregantes, el haplotipo presente fue compartido con Baja California (Tabla XVIII del Anexo 1) y la diversidad haplotípica, nucleotídica y el número promedio de diferencias de nucleótidos entre secuencias fueron nulos.

El análisis de diversidad genética de la  $\psi$ RC de Asia mostró que hay siete sitios segregantes, seis haplotipos distintos y uno de éstos fue compartido con individuos de Baja California (Tabla XIX del Anexo 1), la diversidad haplotípica fue de  $H_d = 1.0$ , la nucleotídica de  $\pi = 0.00544$  y un número promedio de diferencias de nucleótidos entre secuencias con  $K = 2.53$ .

En el caso de la RC, las regiones de Norteamérica, Europa y Asia presentaron valores de diversidad haplotípica y nucleotídica altas ( $H_d = 0.933$ ,  $0.952$  y  $0.942$  y  $\pi = 0.00481$ ,  $0.02968$  y  $0.01125$ , respectivamente), mientras que para Norteamérica hubo tres sitios segregantes, el número promedio de diferencias de nucleótidos entre secuencias fue de  $K = 1.4$  y siete haplotipos. Para Asia hubo nueve sitios segregantes,  $3.27$  diferencias de nucleótidos entre secuencias promedio y seis haplotipos. Para Europa hubo 17 sitios segregantes,  $5.97$  diferencias de nucleótidos entre secuencias, 15 haplotipos y uno de ellos es compartido con Asia, Norteamérica y Baja California. Este haplotipo fue el más frecuente y ampliamente distribuido (Tabla XX del Anexo 1).

## VII. DISCUSIÓN

### VII.1 Extracción del ADN

La cantidad y calidad del ADN extraído es un punto clave para un proyecto de secuenciación, pues de esto depende mucho el rendimiento del secuenciador y, a su vez, la calidad de las secuencias generadas. Si bien, la cantidad de mitocondrias dentro de las células es variable y depende del tipo de tejido, el número de copias de ADNmt dentro de las células es de alrededor de 1,000 copias (Snustad y Simmons, 2015). Por lo que, en teoría, la totalidad de las muestras tienen asegurada una cantidad adecuada de ADNmt. Además, esta molécula tiene un grado de integridad más alto que el ADN nuclear en muestras de museo como es el caso del actual estudio. Por tanto, el origen y tipo de muestras no debió influir de manera negativa en la cantidad de ADN extraído.

La calidad del ADN extraído con respecto a las relaciones de absorbancia 260/280 y 260/230 fue adecuada en los cinco individuos aquí evaluados (Ach 1-5) para el criterio de contaminación por proteínas pero de mala a regular para contaminación por sales, fenoles y carbohidratos (Desjardins y Conklin, 2010).

El método de extracción (precipitación salina diferencial) para los individuos Ach 1-5 pudo haber influido de manera negativa en la calidad del ADN extraído, pues el rendimiento del proceso de secuenciación fue muy dispar y en general muy bajo.

## VII.2 Secuenciación

El rendimiento del proceso de secuenciación fue dispar en la plataforma de secuenciación ILLUMINA MiSeq V.3 (Ach 1-5). Tal como se observó en los resultados de la extracción del ADN (Figura 1 del Anexo 1), la calidad del ADN no es excelente, ya que existe cierto grado de contaminación por sales, carbohidratos y/o fenoles y su integridad genómica está comprometida por ser muestras de individuos conservados de museo. La diferencia en el tamaño de secuencias producidas al comparar las dos plataformas de secuenciación (ILLUMINA HiSeq 4000 y MiSeq V.3) es debido al funcionamiento propio de cada una de ellas.

El rendimiento de la secuenciación en general es relativamente bajo. Tal comportamiento puede ser debido a la baja calidad del ADN que influye de manera importante en la preparación de las librerías de secuencias. A su vez, esto produce rendimientos de secuenciación bajos y posteriormente a secuencias de baja calidad (Van Dijk et al., 2014). Además, la baja calidad del ADN también afecta la optimización de las muestras que se secuenciaron al mismo tiempo que las de Águila real dentro del secuenciador, ya que el fabricante estima rendimientos de 25 millones de lecturas 2 X 300 pb por corrida para MiSeq V.3 y de 5,000 millones de lecturas 2 x 150 pb por corrida para HiSeq 4000 (ILLUMINA, 2019). A pesar de ello fue posible realizar los ensamblajes por referencia del genoma mitocondrial de los seis individuos, aunque algunos de ellos (Ach 1-5) con regiones sin consenso.

### VII.3 Preprocesamiento de las secuencias

El preprocesamiento previo de las secuencias realizado influyó positivamente en el rendimiento de los ensamblajes por referencia, *de novo* y el llamado de SNP's, ya que la calidad de los archivos de secuencias de los seis individuos mejoró considerablemente después del preprocesamiento en todos los aspectos de los reportes de calidad.

El preprocesamiento es un paso fundamental para un proyecto genómico de ensamblaje por referencia o *de novo*, ya que mejora sustancialmente el proceso de ensamblaje y acorta los tiempos de ejecución y los recursos computacionales utilizados para el ensamblaje (Del Fabbro et al., 2013). Bolger et al. (2014) recomiendan un paso de preprocesamiento de las lecturas antes de cualquier análisis, ya sea ensamblaje por referencia o *de novo* y menciona que, de no ser retirados, los adaptadores provenientes de la preparación de las librerías y adaptadores de identificación para la secuenciación pueden ser incorporados en el ensamblaje final. Esto es un riesgo, pues reducen la longitud de los contigs y, por tanto, los scaffolds formados. Además, afecta el valor N50 de los ensamblajes *de novo* haciendo que se pierda continuidad en el ensamblaje.

### VII.4 Ensamblaje por referencia del mitogenoma

Entre los ensamblajes por referencia generados, el más confiable es el del individuo Ach 6 debido a su tasa más alta de alineamiento, su mayor profundidad del ensamblaje y el 100% de cobertura del genoma de referencia, por lo que las variantes nucleotídicas observadas (SNP's) con respecto al genoma de referencia KF905228 son reales y no artefactos del ensamblaje o

falsos positivos. Los ensamblajes de los otros individuos Ach 1-5 no fueron tan confiables, ya que, su profundidad es pobre, contienen regiones sin cobertura y sus tasas de alineamiento de lecturas son bajas, además de que la secuencia consenso generada en estos ensamblajes presenta huecos con respecto al genoma de referencia, esto debido a que en dichas regiones del mitogenoma de referencia no hay un alineamiento de secuencias que permita crear tal consenso.

En este sentido, la tasa de alineamiento es un parámetro clave que da una idea general de qué tan confiable y bueno es un ensamblaje por referencia (Conesa et al., 2016), pues dicha tasa de alineamiento es el porcentaje de lecturas del total de lecturas generadas por el secuenciador que alinearon en alguna parte del genoma de referencia utilizado. Además, las lecturas utilizadas son cortas (150 pb y 300 pb) y por su naturaleza pueden alinear correctamente en múltiples sitios del genoma de referencia (zonas repetitivas).

Otro aspecto clave del ensamblaje por referencia es la elección del algoritmo que usa el programa de análisis, es decir, elegir el algoritmo de alineamiento que mejor funcione para el tipo de lecturas que se tienen. En este caso, Bowtie2 implementa una estrategia de alineamiento adecuada basado en los algoritmos FM-Index y transformada de Burrows-Wheeler, pues se ha demostrado que programas de alineamiento como el implementado en el actual estudio funcionan correctamente para aplicaciones como descubrimiento de INDEL's (Lindner y Friedel, 2012) y trabajan bien al alinear secciones largas del genoma de referencia (Thankaswamy-Kosalai et al.,

2017), lo que ayuda a tener mejores resultados en cuanto a cobertura se refiere.

#### VII.5 Ensamble *de novo* del mitogenoma

Idealmente, cuando se ensambla un genoma, el mejor resultado posible es cuando toda la molécula de ADN está contenida en un solo y único scaffold (Hunt et al., 2014), por ejemplo un scaffold por toda la molécula de ADN mitocondrial o por cada cromosoma en el caso del genoma nuclear. Por lo anterior, al evaluar la calidad de los ensamblados *de novo* del genoma mitocondrial del Águila real, el mejor ensamble corresponde al individuo Ach 6 pues la totalidad del genoma mitocondrial se encuentra dentro de un único scaffold. Además, esto es soportado por otras métricas que permiten estimar la calidad del ensamble, por ejemplo, la profundidad del ensamble de la cual debe ser de al menos 8.0X (ocho copias de cada parte del genoma) (Baker, 2012); siendo el de los otros individuos (Ach 1-5) entre 2.0-5.0X. Otra métrica es el número de contigs o scaffolds generados y la longitud del contig o scaffold más grande generado (Gurevich et al., 2013); en el caso del individuo Ach 6, el número de scaffolds formados es el tercero más bajo con 2,532. La longitud del scaffold más grande generado en el individuo Ach 6 es de 17,472 pb (este scaffold contiene todo genoma mitocondrial). Además, el valor N50 (Bradnam et al., 2013) da una idea de la contigüidad que tiene el ensamble, el valor N50=572 del individuo Ach 6 es el segundo más bajo, el cual pudo haber sido influenciado por la degradación que sufrió el ADN durante el proceso de extracción, sin embargo, no se evaluó la calidad del ADN de dicha muestra.

## VII.6 Anotación y reconstrucción del mapa del mitogenoma de *Aquila chrysaetos*

La anotación del mitogenoma de *Aquila chrysaetos* hecha con RNAweasel no fue exitosa, ya que no logró identificar los ARN ribosómicos a pesar de identificar algunos ARN de transferencia. Por otro lado, Mfannot, aunque identifica las dos sub unidades de ARN ribosómico, no logra identificar todos los genes codificantes para proteínas. Con estas dos herramientas de anotación, la identificación de los ARN de transferencia y ribosómicos se realiza a partir del algoritmo ERPIN (Gautheret y Lambert, 2001), el cual funciona a partir de la identificación de la estructura secundaria de las secuencias.

El programa MITOS webserver (Bernt et al., 2013) proporcionó una anotación exitosa, ya que primero identifica las secuencias candidatas para genes codificantes para proteínas y las compara con la base de datos RefSeq 39, alinea las secuencias de cada gen identificado y corrige los alineamientos mediante el porcentaje de empalme. De la misma manera, compara los ARN de transferencia con secuencias curadas manualmente, además de considerar la estructura secundaria de tales secuencias. Finalmente, a partir de la búsqueda de los codones de inicio y paro para cada gen codificante para proteína, en caso de ser necesario, reubica los ARN de transferencia y también los ARN ribosómicos.

Debido al éxito de anotación que tuvo MITOS webserver al identificar todos los elementos génicos (proteínas, ARN de transferencia y ribosómicos y regiones reguladoras) se decidió realizar la curación manual a partir de esta

anotación. Al realizar esta curación manual de la anotación hay claras diferencias con respecto al genoma mitocondrial previamente reportado (Doyle et al., 2014), por ejemplo, el tamaño del genoma anotado en este estudio es mayor por 140 pb. De igual forma, se anotó una región no codificante reguladora (la pseudo región control), la cual es una región relativamente conservada dentro del grupo de los falconiformes (Song et al., 2015) y que se encuentra en la misma posición entre el ARNt-E y el ARNt-F, tal como lo muestra (Liu et al., 2017) con el mitogenoma de *Accipiter gularis*, otra especie de la misma familia Accipitridae. Además, se observó una anotación diferencial en la región que codifica para la proteína ND 3, en la cual se optó por dividir la región en dos fracciones y nombrarlas ND 3a y ND 3b. Doyle et al. (2014) no reportan tal anotación. En el presente estudio, se acepta esta anotación y se sustenta por el reconocimiento de los codones de inicio y paro para ambas fracciones en el mismo marco de lectura y, adicionalmente, su búsqueda y comparación de la secuencia de nucleótidos a secuencia traducida en aminoácidos a partir de la herramienta BLASTx del GenBank. Además, se realizó una comprobación experimental por PCR.

#### VII.7 Comprobación experimental de la anotación de la región ND 3

Así como se observa en el gel de electroforesis, los fragmentos de ADN amplificados mediante PCR corresponden con la longitud de secuencia esperada. Ambos pares de cebadores que se diseñaron fueron adecuados y permitieron corroborar la secuencia de nucleótidos buscada que demuestra experimentalmente la anotación de la región ND 3.

Realizar la reconstrucción del mitogenoma es un campo importante para estudios filogenéticos, al igual que el descubrimiento de nuevos marcadores moleculares como SNP's e INDEL's. Haring et al. (2001) comentan que los arreglos estructurales del mitogenoma pueden servir también como marcadores moleculares. Por tanto, es de gran relevancia el comprobar toda característica peculiar de una anotación de mitogenoma, como en este caso, la anotación de la proteína ND 3. El motivo por el cual la anotación de la proteína ND 3 se hizo en dos fragmentos, es porque, en el marco de lectura donde se encuentra anotada la proteína, existe una región conformada por siete nucleótidos, entre el codón de inicio ATC que da comienzo a la proteína y el codón de paro TAA que marca el fin de la proteína. Los primeros tres de esta región de siete nucleótidos, codifican para un nuevo codón de paro AGA, seguidos de un nucleótido Adenina (A) y tres nucleótidos que codifican para un nuevo codón de inicio ATC. Lo anterior hace que coexistan dos porciones codificantes que son anotadas como ND 3, separadas por un solo nucleótido que cambia el marco de lectura haciendo que surja la segunda porción con sus codones de inicio y paro. Es importante resaltar que esta división en dos fracciones no ha sido mencionada en los mitogenomas anotados para el grupo de los falconiformes. Por su parte Mindell et al. (1998) reportan un nucleótido que no es traducido dentro de la secuencia de la proteína ND 3 y Eberhard y Wright (2016) mencionan que tal característica se observa en todo el orden de los Psitaciformes (cotorras, loros y afines). Esto refuerza el resultado de la anotación en dos fracciones para la

región ND 3 del mitogenoma del Águila real (*Aquila chrysaetos*) y la comprobación experimental realizadas en el presente estudio.

#### VII.8 Análisis del grado de sintenia

El mitogenoma de *Aquila chrysaetos* ensamblado no presentó cambios en el orden de los genes con respecto al de referencia (KF905228). Tal como era de esperarse, tampoco presentó cambios en el orden de los genes con respecto a otras especies del mismo grupo de aves (Accipitridae) como *Buteo buteo* (Haring et al., 2001), *Accipiter virgatus* (Song et al., 2015), *Aquila fasciata* (Jiang et al., 2015) y *Accipiter gularis* (Liu et al., 2017). Además, es concordante con uno de los modelos generales de mitogenoma presentado por Eberhard y Wright, (2016). Tal modelo se caracteriza por una duplicación de la región control y de otros genes adyacentes, que posteriormente se degradaron, dando origen a la pseudo región control y a dicho modelo general del mitogenoma (Eberhard y Wright, 2016).

#### VII.9 Llamado de SNP's e INDEL's en el mitogenoma de *Aquila chrysaetos*

Es común en los estudios de NGS que la profundidad de secuenciación sea de baja cobertura con alrededor de 5X y, por tanto, no sea exitosa al realizar el llamado de SNP's (Nielsen et al., 2011). Para los SNP's e INDEL's encontrados en el mitogenoma del individuo Ach 6 con respecto del genoma de referencia KF905228, además de haber sido filtradas por su valor de calidad (probabilidad de no ser un error), cumplen con uno de los aspectos más delicados del proceso que es el de profundidad (Li, 2011). Por tanto, los SNP's e INDEL's encontrados en los seis individuos son potenciales marcadores que deberán ser probados en futuros estudios poblacionales,

tomando en cuenta ciertos aspectos técnicos y así evitar el llamado de falsos positivos (The 1000 Genomes Project Consortium, 2015).

#### VII.10 Reconstrucción de las relaciones filogenéticas mediante el ADN mitocondrial

La topología de los fenogramas para la  $\psi$ RC y la RC muestra consistentemente dos linajes, aunque con bajo valor de soporte de bootstrap. Por un lado, el linaje 1 formado por Ach 4, Ach 5 de Baja California y KF905228 de California y, por otro lado, el linaje 2 formado por Ach 1, Ach 2, Ach 3 y Ach 6 de Baja California. La topología del fenograma del gen ND 2 muestra dos linajes, el linaje 1 formado por Ach 1, Ach 4, Ach 5 de Baja California y KF905228 de California y, por otro lado, el linaje 2 formado por Ach 2, Ach 3, Ach 6 de Baja California.

El fenograma de la RC muestra dos clados, los cuales, corresponden a los linajes 1 y 2 identificados para los individuos de Baja California. Los individuos del linaje 1 de Baja California presentan el mismo haplotipo reportado por Sonsthagen et al. (2012) (JQ246417 acceso GenBank) y Craig et al. (2016) (KX687708 acceso GenBank), ambos provenientes de individuos de California. Además, en ambos estudios se menciona que dicho haplotipo es el más frecuente y ampliamente distribuido en Norteamérica a través de las regiones de Alaska, California, Canadá e Idaho.

El haplotipo correspondiente al linaje 2 que presentan los individuos de Baja California es el mismo que el reportado por Craig et al. (2016) (KX687706 acceso GenBank) y Sonsthagen et al. (2012) (JQ246420 acceso GenBank).

En ambos estudios se menciona que dicho haplotipo se encuentra en California en frecuencia baja.

El consenso de las filogenias (construidas por máxima verosimilitud y por inferencia bayesiana) y el fenograma (construidos por Neighbor-Joining) muestran dos clados consistentes entre los tres métodos implementados. Los individuos Ach 4, Ach 5 de Baja California y KF905228 de California forman un clado bien sustentado denominado linaje 1, El segundo clado, denominado linaje 2, está formado por los individuos Ach 1, Ach 2, Ach 3 y Ach 6 de Baja California. La separación de los dos linajes está bien soportada por el método bayesiano. En cambio, la parte basal de la filogenia construida con el método de máxima verosimilitud y del fenograma construido con Neighbor-joining no tienen fuerte soporte de bootstrap.

Considerando todo lo encontrado en los distintos genes, regiones génicas y el mitogenoma, es claro que los individuos de Baja California forman parte de dos grupos genéticos (linajes mitocondriales), no obstante la posición del individuo Ach 1 es incierta dentro de tales grupos para el gen ND 2, dicha asignación incierta está directamente influenciada por la secuencia consenso del individuo Ach 1, la cual presentó baja profundidad de ensamble en la región del gen ND 2 y por tanto, algunos nucleótidos no pudieron ser asignados como una base nitrogenada específica. Por otro lado, dependiendo de la región utilizada para la reconstrucción filogenética y debido a que las tasas de evolución y los niveles de variación genética son distintos (unos más altos que otros) entre genes y regiones del ADN mitocondrial (Bates et al., 2003, Dawnay et al., 2007 y Väli et al., 2019), se considera adecuado emplear

el ADN mitocondrial en estudios filogenéticos y genealógicos (Rubinoff y Holland, 2005).

La existencia de dos grupos genéticos mitocondriales para los individuos de Baja California parece obedecer al comportamiento de la especie y su patrón de migración en el oeste (McIntyre y Lewis, 2016) y noroeste de Estados Unidos de Norteamérica y a través de Canadá (Bedrosian et al., 2018). Estos linajes mitocondriales podrían una evidencia molecular de lo descrito por De León-Girón et al. (2016) para Baja California, pues comentan que coexisten dos poblaciones ecológicas en la zona. La primera población es migratoria, con su origen al oeste de Norteamérica (Oregón). Esta población corresponde al linaje 1 de este estudio y según Craig et al. (2016), representa al haplotipo mitocondrial más ampliamente distribuido en Norteamérica a través de Alaska, California, Idaho y Canadá. La segunda población es residente en Baja California y está representada por el linaje 2 de este trabajo. De acuerdo con Craig et al. (2016), esta población tiene uno de los haplotipos mitocondriales menos frecuente y restringido a California.

Lo anterior encuentra soporte con la gran capacidad de adaptación a ambientes (Judkins y Van Den Bussche, 2017), las grandes distancias de dispersión entre Estados Unidos y México de los individuos de la especie (Tracey et al., 2016; 2017), tanto reproductivos como no reproductivos (De León-Girón et al., 2016), y al amplio ámbito hogareño de las parejas reproductivas (D'Addario et al., 2019). La presencia de estos grupos de individuos (reproductivos y no reproductivos) en Baja California es parte del comportamiento reproductivo (Watson, 2010), al proceso de sucesión y

sustitución de las parejas reproductivas de la región (De León-Girón et al., 2016).

#### VII.11 Análisis de diversidad genética mitocondrial

Las variantes genéticas (SNP's e INDEL's) son el resultado de la variabilidad genética que existe entre al menos dos secuencias (secuencia de referencia contra secuencia a analizar) y permite conocer el sitio de cambio de una base nucleotídica y/o la inserción/delección de alguna o varias bases.

Los valores de diversidad genética mitocondrial calculados en el presente estudio para los genes CO I y Cyt b de manera global para Baja California y para ambos linajes son bajos. Este resultado era de esperarse, teniendo en cuenta que son genes que codifican para proteínas que tienen regiones altamente conservadas (Dawnay et al., 2007) y están sujetas a limitaciones bioquímicas que provocan altos niveles de homoplasia (Faria et al., 2007).

La diversidad haplotípica del gen mitocondrial ND 2 para el linaje 1 de los individuos de Baja California es cero, mientras que para el linaje 2 (que contiene dos haplotipos distintos) es moderada. De manera global para Baja California, la diversidad genética es moderada y concuerda con lo reportado anteriormente. Por ejemplo, Bates et al. (2003) mencionan que ND 2 es genéticamente más diverso que Cyt b y Faria et al. (2007) afirman que ND 2 es uno de los genes mitocondriales más variables dentro de las aves y por ende regularmente implementado para estimar diversidad genética.

En ocasiones la diversidad genética es comparable entre la  $\psi$ RC y la RC (Väli et al., 2019). La diversidad haplotípica de la  $\psi$ RC de los individuos

de Baja California (de manera global) es alta y mayor a la que presento la RC de Baja California y, tanto para la  $\psi$ RC y la RC, los valores de diversidad nucleotídica son similares y bajos. La diversidad genética de la  $\psi$ RC de los individuos de Asia es semejante a la del linaje 1 de Baja California. En cambio, la región de Asia es ligeramente más diversa que el linaje 2 y que la región de Baja California (de manera global), probablemente porque la región de Asia incluye individuos de Japón, Corea y China (Masuda et al., 1998) y la extensión de muestreo es más amplia en comparación con los individuos restringidos a Baja California.

En el presente estudio, los valores de diversidad haplotípica y nucleotídica de los linajes 1 y 2 y global para Baja California son moderadamente altos con respecto a lo reportado para el Águila moteada (*Clanga clanga*, familia Accipitridae) y se encuentran dentro del intervalo que presentó dicha especie ( $H_d = 0.62-0.91$   $\pi = 0.0017-0.0056$ ). Los valores de diversidad haplotípica y nucleotídica globales calculados en este estudio son mayores al compararse con *Clanga clanga* y por ende son similares o ligeramente mayores que otras especies de la familia Accipitridae (Väli et al., 2019).

La diversidad genética mitocondrial de la RC de los individuos de Baja California para ambos linajes por separado es nula, es decir, que no existen sitios segregantes. Regionalmente, la diversidad genética de Baja California es moderada ( $H_d = 0.533$  y  $\pi = 0.002$ ) y similar a la estimada por Sonsthagen et al. (2012) para la población de California ( $H_d = 0.548$  y  $\pi = 0.002$ ). Esta similitud puede deberse a la cercanía de los sitios de muestreo entre los dos

estudios. El sitio de muestreo de los individuos de California corresponde a la zona continental de California frente a las Islas del Canal (Sonsthagen et al., 2012). Esta se encuentra muy próxima a la región noroeste de Baja California de la cual provienen todos los individuos analizados aquí. Por otro lado, los datos que reportan Craig et al. (2016) y Judkins y Van Den Bussche, (2017) son un poco diferentes. En ambos estudios la diversidad haplotípica de los individuos de Baja California es menor a la de California, Oregón e Idaho, sin embargo, la diversidad nucleotídica es similar entre las regiones (Tabla XI). En contraste con lo anterior, la diversidad haplotípica y nucleotídica de Baja California resultó ser mayor que la de Alaska, la cual presentó los niveles de diversidad genética más bajos y el menor número de haplotipos (Craig et al., 2016).

Las diferencias de diversidad genética entre lo calculado por otros (Sonsthagen et al., 2012; Craig et al., 2016 y Judkins y Van Den Bussche, 2017) para California y lo estimado en el presente estudio, puede ser debido a la magnitud de la región de muestreo y al número limitado de individuos analizados en este estudio (seis individuos). Es decir, que mientras que Sonsthagen et al. (2012) realizaron un muestreo regional cercano a las islas del Canal en California, en contraparte Craig et al. (2016) cubrió una región más extensa que abarcó Alaska, California e Idaho, y el de Judkins y Van Den Bussche (2018) que incluyó individuos de otros estados de Estados Unidos.

En el presente estudio, se encontraron dos linajes representados por dos haplotipos. Estos haplotipos han sido previamente reportados por otros autores para la región de California, siendo el haplotipo 1 (que corresponde al

linaje 1 de este estudio) el más frecuente y ampliamente distribuido en otras regiones de Norteamérica (Craig et al., 2016; Sonsthagen et al., 2012).

La diversidad genética de la RC de *Aquila chrysaetos* presenta resultados interesantes cuando se compara contra otras especies. Por ejemplo, el valor de diversidad haplotípica es muy similar al de *Harpia harpyja*, la cual es una especie con alta diversidad genética (Lerner et al., 2009). En cambio, la diversidad nucleotídica es mucho mayor en *Aquila chrysaetos*. Esta diferencia tan grande probablemente sea producto de la historia demográfica de ambas especies. Nebel et al. (2015) mencionan que los altos niveles de diversidad haplotípica y bajos niveles de diversidad nucleotídica (como es el caso de *Harpia harpyja*) indican un cuello de botella seguido de una expansión poblacional, mientras que altos niveles de diversidad haplotípica y nucleotídica (como es el caso de *Aquila chrysaetos*) se relacionan con grandes poblaciones demográficamente estables, tanto históricas como recientes.

Los resultados de diversidad genética de la  $\psi$ RC observados en *A. chrysaetos* en las regiones de Asia y Baja California, así como en otras especies, y el hecho de tener niveles de diversidad genética semejantes a la RC, demuestran que la  $\psi$ RC es un marcador genético adecuado para estudios de genética poblacional y reconstrucción filogenética a nivel intraespecífico.

El número promedio de diferencias de nucleótidos entre secuencias calculado es de  $32.86 \approx 33$ , lo que explica por qué el número de haplotipos del mitogenoma sea igual al número de individuos analizados (seis provenientes del actual estudio más el genoma de referencia KF905228),

provocando que la diversidad haplotípica sea de 1.0. La diversidad nucleotídica es baja con 105 sitios segregantes en 17,472 pb de longitud total del genoma.

A partir de los resultados generados en este estudio se puede decir que los mitogenomas ensamblados y anotados del Águila real del noroeste de Baja California son distintos al mitogenoma de referencia proveniente de un ejemplar de California (Doyle et al., 2014). Por otra parte, las reconstrucciones filogenéticas y genealógicas intraespecíficas permiten distinguir dos linajes en Baja California. Sin embargo, cuando se evalúa la diversidad genética de la RC, salta a la vista que dichos clados encontrados en Baja California concuerdan con un haplotipo que parece ser regional de California (linaje 2) y otro que es frecuente en gran parte de Norteamérica (linaje 1) (Craig et al., 2016; Sonsthagen et al., 2012). Estos dos linajes podrían relacionarse con las poblaciones ecológicas descritas por De León-Girón et al. (2016), por lo que los individuos del noroeste de Baja California, México comparten al menos dos linajes mitocondriales con los individuos de California, Estados Unidos.

## VIII. CONCLUSIONES

1. Se ensambló y anotó el genoma mitocondrial de *Aquila chrysaetos* del noroeste de Baja California, México.
2. Los mitogenomas construidos a partir de un genoma de referencia de algunos individuos no presentan el 100% de cobertura del genoma de referencia, en algunas zonas la cobertura es baja y tienen huecos que no permitieron completar la longitud total de la molécula.
3. El mitogenoma ensamblado *de novo* de *Aquila chrysaetos* contiene los 37 genes presentes en el genoma mitocondrial de vertebrados, 22 ARN's de transferencia, 2 ARN's ribosómicos, 13 genes codificantes para proteínas y dos regiones reguladoras y tiene una longitud mayor a la presentada por el mitogenoma de referencia de la especie.
4. Existen tres diferencias en la anotación del mitogenoma de *Aquila chrysaetos* con respecto a la anotación del mitogenoma de referencia. La primera es un aumento en la longitud de la secuencia de 140 pares de bases, la segunda es el reconocimiento de la pseudo región control y la tercera es la división en dos fracciones de la región que codifica para la proteína ND 3.
5. Existe concordancia en la sintenia del mitogenoma de *Aquila chrysaetos* con el genoma de referencia y en general con el modelo de arquitectura del mitogenoma que presentan los individuos de la familia Accipitridae.

6. Se llamaron y filtraron 33 polimorfismos de un solo nucleótido (SNP's) en conjunto de los seis individuos de *Aquila chrysaetos* contra el mitogenoma de referencia de la especie.
7. Con el gen ND 2, la  $\psi$ RC, la RC y el mitogenoma, se encontró que existen dos grupos genéticamente distintos entre las seis muestras analizadas, uno de ellos con una estrecha relación con individuos de California y el otro relacionado con individuos de distintas regiones de Norteamérica entre ellas, Alaska, California, Idaho y Canadá.
8. La diversidad genética mitocondrial evaluada para Baja California es nula para los genes CO I y Cyt b. La diversidad genética de ND 2 y  $\psi$ RC son moderadamente altas y similares entre sí, aunque ligeramente más bajas que lo reportado para la especie y otras especies relacionadas. La RC presentó niveles de diversidad genética moderados, aunque menores con respecto a otras poblaciones de *Aquila chrysaetos*.
9. En suma, las regiones de California, Estados Unidos de América y noroeste de Baja California, México, comparten los dos linajes mitocondriales de *Aquila chrysaetos* identificados en este estudio, uno de ellos ampliamente distribuido en Norteamérica y otro regional para California y Baja California.

**Literatura citada**

- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Aljanabi, S., y Martínez, I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR- based techniques. *Nucleic Acids Research*. 25(22): 4692-4693. <https://doi.org/10.1093/nar/25.22.4692>
- Angeloni, F., Wagemaker, N., Vergeer, P., y Ouborg, J. 2012. Genomic toolboxes for conservation biologists: Genomic toolboxes for conservation biologists. *Evolutionary Applications*. 5(2): 130-143. <https://doi.org/10.1111/j.1752-4571.2011.00217.x>
- Baker, M. 2012. De novo genome assembly: What every biologist should know. *Nature Methods*. 9(4): 333-337. <https://doi.org/10.1038/nmeth.1935>
- Barker, F. K., Benesh, M. K., Vandergon, A. J., y Lanyon, S. M. 2012. Contrasting evolutionary dynamics and information content of the avian mitochondrial control region and ND2 gene. *PLOS ONE*. 7(10): e46403. <https://doi.org/10.1371/journal.pone.0046403>
- Bates, J. M., Tello, J. G., y Silva, J. M. C. 2003. Initial assessment of genetic diversity in ten bird species of South American Cerrado. *Studies on Neotropical Fauna and Environment*. 38(2): 87-94. <https://doi.org/10.1076/snfe.38.2.87.15924>
- Bechard, M. J., y McGrady, M. J. 2002. Preface: status and conservation of golden eagles. *Journal of Raptor Research*. 36(1): 2

- Bedrosian, B. E., Domenech, R., Shreading, A., Hayes, M. M., Booms, T. L., y Barger, C. R. 2018. Migration corridors of adult Golden Eagles originating in northwestern North America. *PLOS ONE*. 13(11): e0205204. <https://doi.org/10.1371/journal.pone.0205204>
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritsch, G., Stadler, P. F. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*. 69(2): 313-319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Bolger, A. M., Lohse, M., y Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15): 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bourke, B. P., Frantz, A. C., Lavers, C. P., Davison, A., Dawson, D. A. y Burke, T. A. 2010. Genetic signatures of population change in the British golden eagle (*Aquila chrysaetos*). *Conservation Genetics*. 11(5): 1837-1846. <https://doi.org/10.1007/s10592-010-0076-x>
- Bradnam, K. R., Fass, J. N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I. y Korf, I. F. 2013. Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*. 2(1): <https://doi.org/10.1186/2047-217X-2-10>
- Cadahía, L., Pinsker, W., Negro, J. J., Pavlicev, M., Urios, V. y Haring, E. 2009. Repeated sequence homogenization between the control and pseudo-control regions in the mitochondrial genomes of the subfamily Aquilinae. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. 312(3): 171-185.

- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*. 17(4): 540-552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Clouet, M., y Barrau, C. 2015. Decline of the Golden Eagle (*Aquila chrysaetos*) in Ethiopia. *Journal of Raptor Research*. 49(2): 222-226. <https://doi.org/10.3356/rapt-49-02-222-226.1>
- Coil, D., Jospin, G., y Darling, A. E. 2015. A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*. 31(4): 587-589. <https://doi.org/10.1093/bioinformatics/btu661>
- Conant, G. C., y Wolfe, K. H. 2008. GenomeVx: Simple web-based creation of editable circular chromosome maps. *Bioinformatics*. 24(6): 861-862. <https://doi.org/10.1093/bioinformatics/btm598>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A. y Mortazavi, A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology*. 17(1): 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Cornell University, 2015. The Cornell Lab of Ornithology all about birds. Consultado en [https://www.allaboutbirds.org/guide/Golden\\_Eagle/lifehistory](https://www.allaboutbirds.org/guide/Golden_Eagle/lifehistory)
- Craig, E. H., Adams, J. R., Waits, L. P., Fuller, M. R., y Whittington, D. M. 2016. Nuclear and mitochondrial DNA analyses of Golden Eagles (*Aquila chrysaetos canadensis*) from three areas in western North America; initial results and conservation implications. *PLOS ONE*. 11(10): e0164248. <https://doi.org/10.1371/journal.pone.0164248>

- Cubas, F., y Fuentes, C. L. 2017. El águila real, un símbolo en peligro. Fondo Mexicano Para La conservación De La Naturaleza, A.C.
- D'Addario, M., Monroy-Vilchis, O., Zarco-González, M. M., y Santos-Fita, D. 2019. Potential distribution of *Aquila chrysaetos* in Mexico: Implications for conservation. *Avian Biology Research*. 12(1): 33-41. <https://doi.org/10.1177/1758155918823424>
- Darling, A. C. E. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*. 14(7): 1394-1403. <https://doi.org/10.1101/gr.2289704>
- Dawnay, N., Ogden, R., McEwing, R., Carvalho, G. R., y Thorpe, R. S. 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science International*. 173(1): 1-6. <https://doi.org/10.1016/j.forsciint.2006.09.013>
- De León-Girón, G., Rodríguez-Estrella, R., y Ruiz-Campos, G. 2016. Current distribution status of Golden Eagle (*Aquila chrysaetos*) in Northwestern Baja California, Mexico. *Revista Mexicana de Biodiversidad*. 87(4): 1328-1335. <https://doi.org/10.1016/j.rmb.2016.10.003>
- De León-Girón, G. 2017. Abundancia, ecología reproductiva dieta, uso de hábitat y amenazas del Águila real (*Aquila chrysaetos canadiensis*) en Baja California, México. Propuestas para su conservación. Tesis doctoral. Doctorado en Ciencias en Uso, Manejo y Preservación de los Recursos Naturales. Centro de Investigaciones Biológicas del Noreste, la Paz, Baja California Sur, México: 7.

- Del Fabbro, C., Scalabrin, S., Morgante, M., y Giorgi, F. M. 2013. An extensive evaluation of read trimming effects on illumina NGS data analysis. PLOS ONE. 8(12): e85024. <https://doi.org/10.1371/journal.pone.0085024>
- Desjardins, P. y Conklin, D. 2010. NanoDrop microvolume quantitation of nucleic acids. Journal of Visualized Experiments. 45(1): e2565. <https://doi.org/10.3791/2565>
- Doyle, J. M., Katzner, T. E., Bloom, P. H., Ji, Y., Wijayawardena, B. K. y DeWoody, J. A. 2014. The genome sequence of a widespread apex predator, the Golden Eagle (*Aquila chrysaetos*). PLOS ONE. 9(4): e95599. <https://doi.org/10.1371/journal.pone.0095599>
- Doyle, J. M., Katzner, T. E., Roemer, G. W., Cain, J. W., Millsap, B. A., McIntyre, C. L. y DeWoody, J. A. 2016. Genetic structure and viability selection in the golden eagle (*Aquila chrysaetos*), a vagile raptor with a Holarctic distribution. Conservation Genetics. 17(6): 1307-1322. <https://doi.org/10.1007/s10592-016-0863-0>
- Eberhard, J. R., y Wright, T. F. 2016. Rearrangement and evolution of mitochondrial genomes in parrots. Molecular Phylogenetics and Evolution. 94: 34-46. <https://doi.org/10.1016/j.ympev.2015.08.011>
- EMBL-EBI. 2019. Definition: variant calling. Consultado en <https://www.ebi.ac.uk/training/online/glossary/variant-calling>
- Faria, P. J., Baus, E., Morgante, J. S., y Bruford, M. W. 2007. Challenges and prospects of population genetic studies in terns (Charadriiformes,

- Aves). *Genetics and Molecular Biology*. 30(3): 681-689.  
<https://doi.org/10.1590/S1415-47572007000400029>
- Farías, V., Hernández, O., Arizmendi, M. del C., Téllez, O., Botello, F., Olivares, S. J. y Correa, M. 2016. Registro notable de Águila real (*Aquila chrysaetos*) en la Reserva de la Biosfera Tehuacán-Cuicatlán, Puebla, México. *Revista Mexicana de Biodiversidad*. 87(3): 1153-1158.  
<https://doi.org/10.1016/j.rmb.2016.06.001>
- Fasce, P., Fasce, L., Villers, A., Bergese, F. y Bretagnolle, V. 2011. Long-term breeding demography and density dependence in an increasing population of Golden Eagles *Aquila chrysaetos*: Golden Eagle population dynamics. *International Journal of Avian Science*. 153(3). 581-591. <https://doi.org/10.1111/j.1474-919X.2011.01125.x>
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*. 17(6): 368-376.  
<https://doi.org/10.1007/BF01734359>
- Fridjonsson, O., Olafsson, K., Tompsett, S., Bjornsdottir, S., Consuegra, S., Knox, D. y Hjorleifsdottir, S. 2011. Detection and mapping of mtDNA SNPs in Atlantic salmon using high throughput DNA sequencing. *BMC Genomics*. 12(1). 179. <https://doi.org/10.1186/1471-2164-12-179>
- Garner, B. A., Hand, B. K., Amish, S. J., Bernatchez, L., Foster, J. T., Miller, K. M. y Luikart, G. 2016. Genomics in conservation: Case studies and bridging the gap between data and application. *Trends in Ecology & Evolution*. 31(2): 81-83. <https://doi.org/10.1016/j.tree.2015.10.009>

- Gautheret, D. y Lambert, A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *Journal of Molecular Biology*. 313(5): 1003-1011. <https://doi.org/10.1006/jmbi.2001.5102>
- Ghiurcuta, C. G. y Moret, B. M. E. 2014. Evaluating synteny for improved comparative studies. *Bioinformatics*. 30(12): i9-i18. <https://doi.org/10.1093/bioinformatics/btu259>
- Glenn, T. C. 2011. Field guide to next-generation DNA sequencers: field guide to next-gen sequencers. *Molecular Ecology Resources*. 11(5): 759-769. <https://doi.org/10.1111/j.1755-0998.2011.03024.x>
- Guerrero-Cárdenas, I., Galina-Tessaro, P., Álvarez-Cárdenas, S. y Mesa-Zavala, E. 2013. Avistamientos recientes de Águila real (*Aquila chrysaetos*) en la sierra El Mechudo, Baja California Sur, México. *Revista Mexicana de Biodiversidad*. 84(1): 397-401. <https://doi.org/10.7550/rmb.26780>
- 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. <https://doi.org/10.1093/sysbio/syq010>
- Gurevich, A., Saveliev, V., Vyahhi, N. y Tesler, G. 2013. QUASt: Quality assessment tool for genome assemblies. *Bioinformatics*. 29(8): 1072-1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Haring, E., Kruckenhauser, L., Gamauf, A., Riesing, M. J. y Pinsker, W. 2001. The complete sequence of the mitochondrial genome of *Buteo buteo*

- (Aves, Accipitridae) indicates an early split in the phylogeny of raptors. *Molecular Biology and Evolution*. 18(10): 1892-1904. <https://doi.org/10.1093/oxfordjournals.molbev.a003730>
- Haring, E., Kvaløy, K., Gjershaug, J. O., Røv, N. y Gamauf, A. 2007. Convergent evolution and paraphyly of the hawk-eagles of the genus *Spizaetus* (Aves, Accipitridae) phylogenetic analyses based on mitochondrial markers. *Journal of Zoological Systematics and Evolutionary Research*. 45(4): 353-365. <https://doi.org/10.1111/j.1439-0469.2007.00410.x>
- Hartl, G. B. y Pucek, Z. 1994. Genetic depletion in the european Bison (*Bison bonasus*) and the significance of electrophoretic heterozygosity for conservation. *Conservation Biology*. 8(1): 167-174.
- Hebert, P. D., Stoeckle, M. Y., Zemplak, T. S. y Francis, C. M. 2004. Identification of birds through DNA barcodes. *PLOS Biology*. 2(10): e312.
- Helbig, A. J., Kocum, A., Seibold, I. y Braun, M. J. 2005. A multi-gene phylogeny of aquiline eagles (Aves: Accipitriformes) reveals extensive paraphyly at the genus level. *Molecular phylogenetics and evolution*. 35(1): 147-164.
- Howell, S.N.G. y Webb, S. 1995. A guide of the birds of Mexico and northern Central America. Oxford University Press. Oxford: 851.
- Hunt, M., Newbold, C., Berriman, M. y Otto, T. D. 2014. A comprehensive evaluation of assembly scaffolding tools. *Genome Biology*. 15(3): R42. <https://doi.org/10.1186/gb-2014-15-3-r42>

- Illumina. 2019. Sequencing platforms. Consultado en <https://www.illumina.com/systems/sequencing-platforms.htm>
- Iraola, R. 2004. The bald and golden eagle protection act. *Albany Law Review*. 68: 973.
- IUCN (Unión Internacional para la Conservación de la Naturaleza). 2016. The IUCN red list of threatened species 2016. Consultado en: <https://www.iucnredlist.org/species/22696060/93541662>. Fecha de consulta: 06 de septiembre 2019.
- Jiang, Lan, Chen, J., Wang, P., Ren, Q., Yuan, J., Qian, C. y Kan, X. 2015. The mitochondrial genomes of *Aquila fasciata* and *Buteo lagopus* (Aves, Accipitriformes): Sequence, structure and phylogenetic analyses. *PLOS ONE*. 10(8): e0136297. <https://doi.org/10.1371/journal.pone.0136297>
- Jiang, Lichun, Peng, L., Tang, M., You, Z., Zhang, M., West, A. y Merilä, J. 2019. Complete mitochondrial genome sequence of the Himalayan Griffon, *Gyps himalayensis* (Accipitriformes: Accipitridae): Sequence, structure, and phylogenetic analyses. *Ecology and Evolution*. 9(15): 8813-8828. <https://doi.org/10.1002/ece3.5433>
- Johnsen, A., Rindal, E., Ericson, P. G., Zuccon, D., Kerr, K. C., Stoeckle, M. Y. y Lifjeld, J. T. 2010. DNA barcoding of Scandinavian birds reveals divergent lineages in trans-Atlantic species. *Journal of Ornithology*. 151(3): 565-578.
- Judkins, M. E. y Van Den Bussche, R. A. 2017. Holarctic phylogeography of golden eagles (*Aquila chrysaetos*) and evaluation of alternative North

- American management approaches. *Biological Journal of the Linnean Society*. 123(2): 471-482. <https://doi.org/10.1093/biolinnean/blx138>
- Katoh, K., Rozewicki, J. y Yamada, K. D. 2017. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*. <https://doi.org/10.1093/bib/bbx108>
- Kochert, M., Steenhof, K., McIntyre, C., Craig, E., Poole, A., y Gill, F. 2002. *The birds of North America*. No. 684. The academy of natural sciences Philadelphia, Philadelphia, PA, and the American Ornithologists Union, Washington. DC Google Scholar.
- Kumar, S., Stecher, G., Li, M., Knyaz, C. y Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*. 35(6): 1547-1549. <https://doi.org/10.1093/molbev/msy096>
- Langmead, B. y Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9(4): 357-359. <https://doi.org/10.1038/nmeth.1923>
- Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D. y Seeb, J. E. 2014. Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*. 7(3): 355-369. <https://doi.org/10.1111/eva.12128>
- Lerner, H. R. L., Johnson, J. A., Lindsay, A. R., Kiff, L. F. y Mindell, D. P. 2009. It's not too late for the Harpy Eagle (*Harpia harpyja*): High levels of genetic diversity and differentiation can fuel conservation programs.

PLOS ONE. 4(10): e7336.

<https://doi.org/10.1371/journal.pone.0007336>

Lerner, H. R. L. y Mindell, D. P. 2005. Phylogeny of eagles, old world vultures, and other Accipitridae based on nuclear and mitochondrial DNA. *Molecular Phylogenetics and Evolution*. 37(2): 327-346. <https://doi.org/10.1016/j.ympev.2005.04.010>

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 27(21): 2987-2993. <https://doi.org/10.1093/bioinformatics/btr509>

Lindner, R. y Friedel, C. C. 2012. A comprehensive evaluation of alignment algorithms in the context of RNA-Seq. *PLOS ONE*. 7(12): e52403. <https://doi.org/10.1371/journal.pone.0052403>

Liu, G., Li, C., Du, Y. y Liu, X. 2017. The complete mitochondrial genome of Japanese sparrowhawk (*Accipiter gularis*) and the phylogenetic relationships among some predatory birds. *Biochemical Systematics and Ecology*. 70: 116-125. <https://doi.org/10.1016/j.bse.2016.11.007>

Machado, D., Janies, D., Brouwer, C. y Grant, T. 2018. A new strategy to infer circularity applied to four new complete frog mitogenomes. *Ecology and Evolution*. 8(8): 4011-4018. <https://doi.org/10.1002/ece3.3918>

Masuda, R., Noro, M., Kurose, N., Nishida-Umehara, C., Takechi, H., Yamazaki, T. y Yoshida, M. C. 1998. Genetic characteristics of endangered Japanese golden eagles (*Aquila chrysaetos japonica*) based on mitochondrial DNA D-loop sequences and karyotypes. *Zoo*

Biology. 17(2): 111-121. [https://doi.org/10.1002/\(SICI\)1098-2361\(1998\)17:2<111::AID-ZOO6>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1098-2361(1998)17:2<111::AID-ZOO6>3.0.CO;2-C)

McIntyre, C. L. y Lewis, S. B. 2016. Observations of migrating Golden Eagles (*Aquila chrysaetos*) in eastern interior Alaska offer insights on population size and migration monitoring. *Journal of Raptor Research*. 50(3): 254-264. <https://doi.org/10.3356/JRR-15-13.1>

Millsap, B. A., Zimmerman, G. S., Sauer, J. R., Nielson, R. M., Otto, M., Bjerre, E. y Murphy, R. 2013. Golden eagle population trends in the western United States: 1968-2010: Golden Eagle population trends. *The Journal of Wildlife Management*. 77(7): 1436-1448. <https://doi.org/10.1002/jwmg.588>

Mindell, D. P., Sorenson, M. D. y Dimcheff, D. E. 1998. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Molecular Biology and Evolution*. 15(11): 1568-1571. <https://doi.org/10.1093/oxfordjournals.molbev.a025884>

Ming, M. y Ding, P., Li, W., Chen, Y. y Hu B. 2010. Breeding ecology and survival status of the Golden Eagle in China. *Raptors Conservation*. 19: 75-87

Ming, M., Zhang, T., Ding, P., Bayahen, K., Xing, R., Zhao, X. y Huang, Y. 2012. Golden Eagle in the north-western China. *Raptors Conservation*. 25: 70-78

Moritz, C. 1994. Applications of mitochondrial DNA analysis in conservation: A critical review. *Molecular Ecology*. 3(4): 401-411. <https://doi.org/10.1111/j.1365-294X.1994.tb00080.x>

- Navarro-Sigüenza, A. G. y A. Townsend Peterson. 2007. Mapas de las aves de México basados en WWF. Informe final SNIB-CONABIO proyecto No. CE015. México D. F.
- Nebel, C., Gamauf, A., Haring, E., Segelbacher, G., Villers, A. y Zachos, F. E. 2015. Mitochondrial DNA analysis reveals Holarctic homogeneity and a distinct Mediterranean lineage in the Golden eagle (*Aquila chrysaetos*). *Biological Journal of the Linnean Society*. 116(2): 328-340. <https://doi.org/10.1111/bij.12583>
- Nei, M. y Li, W. H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*. 76(10): 5269-5273. <https://doi.org/10.1073/pnas.76.10.5269>
- Nielsen, R., Paul, J. S., Albrechtsen, A. y Song, Y. S. 2011. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*. 12(6): 443-451. <https://doi.org/10.1038/nrg2986>
- Okonechnikov, K., Golosova, O. y Fursov, M. 2012. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*. 28(8): 1166-1167. <https://doi.org/10.1093/bioinformatics/bts091>
- Pareek, C. S., Smoczynski, R. y Tretyn, A. 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*. 52(4): 413-435. <https://doi.org/10.1007/s13353-011-0057-x>
- Parson, W., Pegoraro, K., Niederstätter, H., Föger, M. y Steinlechner, M. 2000. Species identification by means of the cytochrome b gene. *International journal of legal medicine*. 114(1-2): 23-28.

- Posada, D. 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*. 25(7): 1253-1256.  
<https://doi.org/10.1093/molbev/msn083>
- Rambaut A. 2009. FigTree: Tree figure drawing tool. disponible en línea en:  
<http://tree.bio.ed.ac.uk/software/figtre?e/>
- Rödelsperger, C. y Dieterich, C. 2010. CYNTENATOR: Progressive gene order alignment of 17 vertebrate genomes. *PLOS ONE*. 5(1): e8861.  
<https://doi.org/10.1371/journal.pone.0008861>
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S. y Huelsenbeck, J. P. 2012. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*. 61(3): 539-542.  
<https://doi.org/10.1093/sysbio/sys029>
- Rozas, J., Ferrer-Mata, A., Sánchez-Del Barrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E. y Sánchez-Gracia, A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*. 34(12): 3299-3302.  
<https://doi.org/10.1093/molbev/msx248>
- Rubinoff, D. y Holland, B. S. 2005. Between two extremes: Mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*. 54(6): 952-961.  
<https://doi.org/10.1080/10635150500234674>

- Rudnick, J. A., Katzner, T. E., Bragin, E. A. y De Woody, J. A. 2007. Species identification of birds through genetic analysis of naturally shed feathers. *Molecular Ecology Notes*. 7(5): 757-762.
- Saitoh, T., Sugita, N., Someya, S., Iwami, Y., Kobayashi, S., Kamigaichi, H. y Nishiumi, I. 2015. DNA barcoding reveals 24 distinct lineages as cryptic bird species candidates in and around the Japanese Archipelago. *Molecular Ecology Resources*. 15(1): 177-186.
- Scitable by Nature Education. 2014. Definition: haplotipo. Consultado en <https://www.nature.com/scitable/definition/haplotype-haplotypes-142>
- Seibold, I., Helbig, A. J., Meyburg, B. U., Negro, J. J. y Wink, M. 1996. Genetic differentiation and molecular phylogeny of European Aquila eagles according to cytochrome b nucleotide sequences. *Eagle studies. Birds of Prey Bulletin*: 1Á/15.
- SEMARNAT. 2008. Programa de acción para la conservación de la especie: Águila Real (*Aquila chrysaetos*). México, D.F.: Comisión Nacional de Áreas Naturales Protegidas. [https://www.biodiversidad.gob.mx/especies/especies\\_priori/fichas/pdf/pace\\_aguila.pdf](https://www.biodiversidad.gob.mx/especies/especies_priori/fichas/pdf/pace_aguila.pdf)
- SEMARNAT. (2010). Norma Oficial Mexicana NOM-059-SEMARNAT-2010, Protección ambiental-especies nativas de México de flora y fauna silvestres-categorías de riesgo y especificaciones para su inclusión, exclusión o cambio-lista de especies en riesgo. *Diario oficial de la federación*. 110.

- Sergio, F., Newton, I., Marchesi, L. y Pedrini, P. 2006. Ecologically justified charisma: Preservation of top predators delivers biodiversity conservation: Top predators and biodiversity. *Journal of Applied Ecology*. 43(6): 1049-1055. <https://doi.org/10.1111/j.1365-2664.2006.01218.x>
- Sieber, P., Platzer, M. y Schuster, S. 2018. The definition of open reading frame revisited. *Trends in Genetics*. 34(3): 167-170. <https://doi.org/10.1016/j.tig.2017.12.009>
- Snustad, D. P., y Simmons, M. J. 2015. Principles of genetics. binder ready version. John Wiley y Sons.
- Song, X., Huang, J., Yan, C., Xu, G., Zhang, X. y Yue, B. 2015. The complete mitochondrial genome of *Accipiter virgatus* and evolutionary history of the pseudo-control regions in Falconiformes. *Biochemical Systematics and Ecology*. 58: 75-84. <https://doi.org/10.1016/j.bse.2014.10.013>
- Sonsthagen, S. A., Coonan, T. J., Latta, B. C., Sage, G. K. y Talbot, S. L. 2012. Genetic diversity of a newly established population of golden eagles on the Channel Islands, California. *Biological Conservation*. 146(1): 116-122. <https://doi.org/10.1016/j.biocon.2011.11.031>
- Suchentrunk, F., Haller, H. y Ratti, P. 1999. Gene pool variability of a golden eagle (*Aquila chrysaetos*) population from the Swiss Alps. *Biological Conservation*. 90(2): 151-155. [https://doi.org/10.1016/S0006-3207\(99\)00025-7](https://doi.org/10.1016/S0006-3207(99)00025-7)
- Takeuchi, T., Shiraki, S., Nashimoto, M., Matsuki, R., Abe, S. y Yatake, H. 2006. Regional and temporal variations in prey selected by Golden

Eagles *Aquila chrysaetos* during the nestling period in Japan: Prey selection in Golden Eagles. *Ibis*. 148(1): 79-87. <https://doi.org/10.1111/j.1474-919X.2006.00495.x>

Thankaswamy-Kosalai, S., Sen, P. y Nookaew, I. 2017. Evaluation and assessment of read-mapping by multiple next-generation sequencing aligners based on genome-wide characteristics. *Genomics*. 109(3-4): 186-191. <https://doi.org/10.1016/j.ygeno.2017.03.001>

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature*. 526(7571). 68-74. <https://doi.org/10.1038/nature15393>

Thiollay, J.M. 2006. Severe decline of large birds in the Northern Sahel of West Africa: A long-term assessment. *Bird Conservation International*. 16(4): 353-365. <https://doi.org/10.1017/S0959270906000487>

Tracey, J. A., Madden, M. C., Sebes, J. B., Bloom, P. H., Katzner, T. E., y Fisher, R. N. 2016. Biotelemetry data for golden eagles (*Aquila chrysaetos*) captured in coastal southern California, November 2014–February 2016. US Geological Survey. 994

Tracey, J. A., Madden, M. C., Sebes, J. B., Bloom, P. H., Katzner, T. E. y Fisher, R. N. 2017. Biotelemetry data for Golden Eagles (*Aquila chrysaetos*) captured in coastal Southern California, February 2016–February 2017. US Geological Survey. 46

Truett, G. E., Heeger, P., Mynatt, R. L., Truett, A. A., Walker, J. A. y Warman, M. L. 2000. Preparation of PCR-quality mouse genomic DNA with Hot

Sodium Hydroxide and Tris (HotSHOT). *BioTechniques*. 29(1): 52-54.

<https://doi.org/10.2144/00291bm09>

Väli, Ü., Dombrovski, V., Dzmitranok, M., Maciorowski, G. y Meyburg, B.U.

2019. High genetic diversity and low differentiation retained in the European fragmented and declining Greater Spotted Eagle (*Clanga clanga*) population. *Scientific Reports*. 9(1): 3064.

<https://doi.org/10.1038/s41598-019-39187-1>

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. y Thermes, C. 2014. Ten years of next-generation sequencing technology. *Trends in genetics*. 30(9): 418-426.

Watson, J. 2010. *The golden eagle*, 2<sup>a</sup> ed. New Haven: Yale University Press.

<https://books.google.com.mx/books?id=aj5MNvCkun0C>

Wink, M. y Sauer-Gürth, H. 2004. Phylogenetic relationships in diurnal raptors based on nucleotide sequences of mitochondrial and nuclear marker genes. *Raptors worldwide*. Berlin: World working group on birds of prey and owls: 483-498.

Zerbino, D. R. y Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. 18(5): 821-829.

<https://doi.org/10.1101/gr.074492.107>

## Anexo 1

### Tablas

#### Preprocesamiento

*Tabla XII.* Estatus de los parámetros de los reportes de calidad generados con FastQC de Ach 1 de las secuencias crudas (tomadas directas del secuenciador) y tratadas (limpias). Estatus adecuado (color verde), estatus atención (anaranjado), estatus malo (rojo).

<b>Categoría</b>	<b>Crudas r1</b>	<b>Tratadas r1</b>	<b>Crudas r2</b>	<b>Tratadas r2</b>
Estadísticos básicos	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Calidad de las secuencias por base	MALO	ADECUADO	MALO	ADECUADO
Calidad de secuenciación por cuadro	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Valor de calidad por secuencia	ADECUADO	ADECUADO	ATENCIÓN	ADECUADO
Contenido de bases por secuencia	ATENCIÓN	ATENCIÓN	MALO	ATENCIÓN
Contenido GC por secuencia	ATENCIÓN	ADECUADO	ATENCIÓN	ADECUADO
Contenido de bases N	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Distribución de la longitud de las secuencias	ADECUADO	ATENCIÓN	ADECUADO	ATENCIÓN
Niveles de duplicación de secuencias	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Secuencias sobre-representadas	ATENCIÓN	ADECUADO	ATENCIÓN	ADECUADO
Contenido de adaptadores	MALO	ADECUADO	MALO	ADECUADO

*Tabla XIII.* Estatus de los parámetros de los reportes de calidad generados con FastQC de Ach 2 de las secuencias crudas (tomadas directas del secuenciador) y tratadas (limpias). Estatus adecuado (color verde), estatus atención (anaranjado), estatus malo (rojo).

<b>Categoría</b>	<b>Crudas r1</b>	<b>Tratadas r1</b>	<b>Crudas r2</b>	<b>Tratadas r2</b>
Estadísticos básicos	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Calidad de las secuencias por base	MALO	ADECUADO	MALO	ADECUADO
Calidad de secuenciación por cuadro	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Valor de calidad por secuencia	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Contenido de bases por secuencia	ATENCIÓN	ADECUADO	MALO	ATENCIÓN

Contenido GC por secuencia	MALO	MALO	ATENCIÓN	MALO
Contenido de bases N	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Distribución de la longitud de las secuencias	ADECUADO	ATENCIÓN	ADECUADO	ATENCIÓN
Niveles de duplicación de secuencias	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Secuencias sobre-representadas	ATENCIÓN	ADECUADO	ATENCIÓN	ADECUADO
Contenido de adaptadores	MALO	ADECUADO	MALO	ADECUADO

*Tabla XIV.* Estatus de los parámetros de los reportes de calidad generados con FastQC de Ach 3 de las secuencias crudas (tomadas directas del secuenciador) y tratadas (limpias). Estatus adecuado (color verde), estatus atención (anaranjado), estatus malo (rojo).

<b>Categoría</b>	<b>Crudas r1</b>	<b>Tratadas r1</b>	<b>Crudas r2</b>	<b>Tratadas r2</b>
Estadísticos básicos	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Calidad de las secuencias por base	MALO	ADECUADO	MALO	ADECUADO
Calidad de secuenciación por cuadro	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Valor de calidad por secuencia	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Contenido de bases por secuencia	MALO	ATENCIÓN	MALO	ATENCIÓN
Contenido GC por secuencia	ATENCIÓN	ADECUADO	ATENCIÓN	ADECUADO
Contenido de bases N	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Distribución de la longitud de las secuencias	ADECUADO	ATENCIÓN	ADECUADO	ATENCIÓN
Niveles de duplicación de secuencias	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Secuencias sobre-representadas	ATENCIÓN	ADECUADO	ATENCIÓN	ADECUADO
Contenido de adaptadores	MALO	ADECUADO	MALO	ADECUADO

*Tabla XV.* Estatus de los parámetros de los reportes de calidad generados con FastQC de Ach 4 de las secuencias crudas (tomadas directas del secuenciador) y tratadas (limpias). Estatus adecuado (color verde), estatus atención (anaranjado), estatus malo (rojo).

<b>Categoría</b>	<b>Crudas r1</b>	<b>Tratadas r1</b>	<b>Crudas r2</b>	<b>Tratadas r2</b>
Estadísticos básicos	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Calidad de las secuencias por base	MALO	ADECUADO	MALO	ADECUADO

Calidad de secuenciación por cuadro	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Valor de calidad por secuencia	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Contenido de bases por secuencia	ATENCIÓN	ATENCIÓN	MALO	ATENCIÓN
Contenido GC por secuencia	ATENCIÓN	MALO	ATENCIÓN	MALO
Contenido de bases N	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Distribución de la longitud de las secuencias	ADECUADO	ATENCIÓN	ADECUADO	ATENCIÓN
Niveles de duplicación de secuencias	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Secuencias sobre-representadas	ATENCIÓN	ADECUADO	ATENCIÓN	ADECUADO
Contenido de adaptadores	MALO	ADECUADO	MALO	ADECUADO

*Tabla XVI.* Estatus de los parámetros de los reportes de calidad generados con FastQC de Ach 5 de las secuencias crudas (tomadas directas del secuenciador) y tratadas (limpias). Estatus adecuado (color verde), estatus atención (anaranjado), estatus malo (rojo).

<b>Categoría</b>	<b>Crudas r1</b>	<b>Tratadas r1</b>	<b>Crudas r2</b>	<b>Tratadas r2</b>
Estadísticos básicos	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Calidad de las secuencias por base	MALO	ADECUADO	MALO	ADECUADO
Calidad de secuenciación por cuadro	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Valor de calidad por secuencia	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Contenido de bases por secuencia	ATENCIÓN	ATENCIÓN	MALO	ATENCIÓN
Contenido GC por secuencia	ATENCIÓN	MALO	ATENCIÓN	MALO
Contenido de bases N	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Distribución de la longitud de las secuencias	ADECUADO	ATENCIÓN	ADECUADO	ATENCIÓN
Niveles de duplicación de secuencias	ADECUADO	ADECUADO	ADECUADO	ADECUADO
Secuencias sobre-representadas	ATENCIÓN	ADECUADO	ATENCIÓN	ADECUADO
Contenido de adaptadores	MALO	ADECUADO	MALO	ADECUADO





ARNt L	11,902	11,972	+	71	cta
ND 5	12,003	13,790	+	1,788	
Cyt b	13,802	14,944	+	1,143	
ARNt P	16,177	16,246	-	70	cca
ND 6	16,260	16,781	-	522	

### Llamado de SNP's e INDEL's en el mitogenoma de *Aquila chrysaetos*

*Tabla XXI.* Llamado de SNP's e INDEL's del individuo Ach 1 contra el genoma de referencia KF905228. Se muestra la posición donde se encuentra el cambio con respecto al genoma de referencia, el nucleótido de referencia, el nucleótido alternativo, la calidad (Qs) del llamado del SNP o INDEL y la profundidad que tiene dicho cambio.

Posición	Referencia	Alternativa	Calidad Qs	Profundidad
<b>SNP's</b>				
3130	G	A	222	28
3695	T	A	118	50
3696	T	A	123	49
5557	G	A	222	35
6646	G	A	214	13
7141	A	G	149	10
8269	C	T	219	14
8593	A	G	222	51
8723	G	A	222	30
10925	C	T	225	27
10963	T	C	113	9
10970	C	T	106	21
10975	T	C	183	15

11284	G	T	199	10
11432	C	T	118	17
11489	A	C	176	16
<b>INDEL's</b>				
3550	T	TGTGAACAA	30.5	5
3701	CAAA	CCCACCAATA	104	2
13378	T	TC	214	41

*Tabla XXII.* Llamado de SNP's e INDEL's del individuo Ach 2 contra el genoma de referencia KF905228. Se muestra la posición donde se encuentra el cambio con respecto al genoma de referencia, el nucleótido de referencia, el nucleótido alternativo, la calidad (Qs) del llamado del SNP o INDEL y la profundidad que tiene dicho cambio.

<b>Posición</b>	<b>Referencia</b>	<b>Alternativa</b>	<b>Calidad Qs</b>	<b>Profundidad</b>
<b>SNP's</b>				
1505	G	A	167	7
3695	T	A	43	12
3696	T	A	43	12
5557	G	A	222	11
6646	G	A	210	9
8269	C	A	190	8
8723	G	A	147	6
11432	C	T	71.5	6
11489	A	C	73.5	6
17278	G	A	113	7
<b>INDEL's</b>				

13378	T	TC	106	4
-------	---	----	-----	---

*Tabla XXIII.* Llamado de SNP's e INDEL's del individuo Ach 3 contra el genoma de referencia KF905228. Se muestra la posición donde se encuentra el cambio con respecto al genoma de referencia, el nucleótido de referencia, el nucleótido alternativo, la calidad (Qs) del llamado del SNP o INDEL y la profundidad que tiene dicho cambio.

Posición	Referencia	Alternativa	Calidad Qs	Profundidad
<b>SNP's</b>				
3130	G	A	157	8
5557	G	A	131	5
8593	A	G	169	7
8723	G	A	131	5
10925	C	T	198	18
10970	C	T	36	13
11432	C	T	21	7
11489	A	C	30	7
11493	T	C	105	9
11496	G	A	89	9
11505	T	G	66	7
<b>INDEL's</b>				
13378	T	TC	168	7

*Tabla XXIV.* Llamado de SNP's e INDEL's del individuo Ach 4 contra el genoma de referencia KF905228. Se muestra la posición donde se encuentra el cambio con respecto al genoma de referencia, el nucleótido de referencia,

el nucleótido alternativo, la calidad (Qs) del llamado del SNP o INDEL y la profundidad que tiene dicho cambio.

<b>Posición</b>	<b>Referencia</b>	<b>Alternativa</b>	<b>Calidad Qs</b>	<b>Profundidad</b>
<b>SNP's</b>				
1505	G	A	155	8
1515	T	C	158	10
3566	A	G	55	11
3695	T	A	105	40
3696	T	A	75	40
5557	G	A	217	13
8593	A	G	86	7
10925	C	T	127	6
10970	C	T	76	9
11378	C	T	65.5	15
11432	C	T	115	13
11489	A	C	138	12
<b>INDEL's</b>				
13378	T	TC	214	11

*Tabla XXV.* Llamado de SNP's e INDEL's del individuo Ach 5 contra el genoma de referencia KF905228. Se muestra la posición donde se encuentra el cambio con respecto al genoma de referencia, el nucleótido de referencia, el nucleótido alternativo, la calidad (Qs) del llamado del SNP o INDEL y la profundidad que tiene dicho cambio.

<b>Posición</b>	<b>Referencia</b>	<b>Alternativa</b>	<b>Calidad Qs</b>	<b>Profundidad</b>
<b>SNP's</b>				

10925	C	T	48	14
10970	C	T	43	13
11003	A	G	41	15
11489	A	C	179	11
<b>INDEL's</b>				
13378	T	TC	214	11

*Tabla XXVI.* Llamado de SNP's e INDEL's del individuo Ach 6 contra el genoma de referencia KF905228. Se muestra la posición donde se encuentra el cambio con respecto al genoma de referencia, el nucleótido de referencia, el nucleótido alternativo, la calidad (Qs) del llamado del SNP o INDEL y la profundidad que tiene dicho cambio.

<b>Posición</b>	<b>Referencia</b>	<b>Alternativa</b>	<b>Calidad Qs</b>	<b>Profundidad</b>
<b>SNP's</b>				
1505	G	A	222	227
3130	G	A	222	152
5557	G	A	222	209
6646	G	A	222	175
8269	C	T	222	185
8593	A	G	222	145
8723	G	A	222	138
10925	C	T	222	176
10970	C	T	222	169
11003	A	G	114	143
11011	C	T	19.1	134

11205	A	G	18.1	110
11213	C	T	22	132
11237	C	T	95	148
11282	G	A	77.5	168
11285	A	C	77.5	170
11351	T	C	85	175
11354	A	G	82	173
11378	C	T	222	187
11432	C	T	222	170
11489	A	C	222	211
17278	G	A	222	84

---

**INDEL's**

---

3	CTAA	CTAACTTCCAAAC TAA	104	3
3701	CAAA	CCAACAATAT	177	1
13378	T	TC	214	175
17325	AACTTCCA	AACTTCCAAACTA CTTCCA	110	4

---

### **Análisis de diversidad genética**

Los haplotipos encontrados producto del análisis de diversidad genética de las secuencias de los genes, regiones génicas utilizadas y del mitogenoma se muestran en las siguientes tablas.

*Tabla XXVII.* Haplotipos del gen CO I (citocromo oxidasa sub unidad 1) mitocondrial obtenidas del GenBank y las reportadas en el presente estudio.

<b>Haplotipo</b>	<b>Número de secuencias</b>	<b>Secuencias</b>	<b>Región geográfica</b>	<b>Referencia</b>
		KR017975	Norteamérica	No publicada
		AY666434	Norteamérica	Hebert et al., 2004
		AY666481	Norteamérica	Hebert et al., 2004
		AB842562	Asia	Saitoh et al., 2015
		Ach 1	B.C. México	Este estudio
		Ach 2	B.C. México	Este estudio
		Ach 3	B.C. México	Este estudio
		Ach 4	B.C. México	Este estudio
COI 1	15	Ach 5	B.C. México	Este estudio
		Ach 6	B.C. México	Este estudio
		KF905228	Norteamérica	Doyle et al., 2014
		GU571265	Europa	Johnsen et al., 2010
		GU571264	Europa	Johnsen et al., 2010
		GU571737	Europa	Johnsen et al., 2010
		GU571738	Europa	Johnsen et al., 2010
COI 2	1	DQ834328	Norteamérica	Rudnik et al., 2007

*Tabla XXVIII.* Haplotipos del gen Cyt b (citocromo b) mitocondrial obtenidas del GenBank y las reportadas en el presente estudio.

<b>Haplotipo</b>	<b>Número de secuencias</b>	<b>Secuencias</b>	<b>Región geográfica</b>	<b>Referencia</b>
		AJ604485	Europa	Helbig et al., 2005
		Z73462	Europa	Siebold et al., 1996
		EU345473	Europa	Wink et al., 2004

Cyt b 1	13	KF905228	Norteamérica	Doyle et al., 2014
		Ach 1	B.C. México	Este estudio
		Ach 2	B.C. México	Este estudio
		Ach 3	B.C. México	Este estudio
		Ach 4	B.C. México	Este estudio
		Ach 5	B.C. México	Este estudio
		Ach 6	B.C. México	Este estudio
		AY987294	Norteamérica	Lerner y Mindell, 2005
		EU345513	Europa	Wink et al., 2004
		EU345511	Europa	Wink et al., 2004
Cyt b 2	3	AJ604486	Norteamérica	Helbig et al., 2005
		X86740	Europa	Haring et al., 2007
		EU345512	Europa	Wink et al., 2004
Cyt b 3	1	AF172375	Europa	Parson et al., 1999

*Tabla XXIX.* Haplotipos del gen ND 2 (NADH deshidrogenasa sub unidad 2) mitocondrial obtenidas del GenBank y las reportadas en el presente estudio.

Haplotipo	Número de secuencias	Secuencias	Región geográfica	Referencia
ND2 1	4	KF905228	Norteamérica	Doyle et al., 2014
		AY987116	Norteamérica	Lerner y Mindell, 2005
		Ach 4	B.C. México	Este estudio
		Ach 5	B.C. México	Este estudio
ND2 2	1	Ach 1	B.C. México	Este estudio

		Ach 2	B.C. México	Este estudio
ND2 3	3	Ach 3	B.C. México	Este estudio
		Ach 6	B.C. México	Este estudio

*Tabla XXX.* Haplotipos de la pseudo región control del ADN mitocondrial obtenidas del GenBank y las reportadas en el presente estudio.

Haplotipo	Número de secuencias	Secuencias	Región geográfica	Referencia
		Ach 1	B.C. México	Este estudio
		Ach 3	B.C. México	Este estudio
pRC 1	5	Ach 6	B.C. México	Este estudio
		AF435099	Europa	Vali, 2002
		AB004679	Asia	Masuda et al., 1998
pRC 2	1	LC146689	Asia	No publicada
		Ach 4	B.C. México	Este estudio
pRC 3	3	Ach 5	B.C. México	Este estudio
		KF905228	Norteamérica	Doyle et al., 2014
pRC 4	1	Ach 2	B.C. México	Este estudio
pRC 5	1	AB004683	Asia	Masuda et al., 1998
pRC 6	1	AB004682	Asia	Masuda et al., 1998

pRC 7	1	AB004680	Asia	Masuda et al., 1998
pRC 8	1	AB004681	Asia	Masuda et al., 1998

*Tabla XXXI.* Haplotipos de la región control del ADN mitocondrial obtenidas del GenBank y las reportadas en el presente estudio.

Haplotipo	Número de secuencias	Secuencias	Región geográfica	Referencia	
RC 1	6	Ach 1	B.C. México	Este estudio	
		Ach 2	B.C. México	Este estudio	
		Ach 3	B.C. México	Este estudio	
		Ach 6	B.C. México	Este estudio	
		JQ246420	Norteamérica	Sonsthagen et al., 2012	
		KX687706	Norteamérica	Craig et al., 2016	
	RC 2	16	Ach4	B.C. México	Este estudio
			Ach5	B.C. México	Este estudio
			JQ246417	Norteamérica	Sonsthagen et al., 2012
			KR259264	Asia	Nebel et al., 2015
			KR336779	Asia	Nebel et al., 2015
			KF905228	Norteamérica	Doyle et al., 2014
			KX687708	Norteamérica	Craig et al., 2016
			KX687707	Norteamérica	Craig et al., 2016
			KR259260	Europa	Nebel et al., 2015
			KR336769	Europa	Nebel et al., 2015
KR336767	Europa	Nebel et al., 2015			
KR259263	Asia	Nebel et al., 2015			
KR336788	Asia	Nebel et al., 2015			
KR259275	Europa	Nebel et al., 2015			

---

		KR336768	Europa	Nebel et al., 2015
		KR336777	Europa	Nebel et al., 2015
		KX687711	Norteamérica	Craig et al., 2016
		KR259261	Asia	Nebel et al., 2015
		KR336770	Asia	Nebel et al., 2015
RC 3	5	KR259262	Asia	Nebel et al., 2015
		KR336789	Asia	Nebel et al., 2015
		KX687705	Norteamérica	Craig et al., 2016
		JQ246421	Norteamérica	Sonsthagen et al., 2012
		KR336772	Asia	Nebel et al., 2015
		KR259265	Asia	Nebel et al., 2015
		LC146690	Asia	No publicada
RC 4	8	KR259266	Asia	Nebel et al., 2015
		KR336787	Asia	Nebel et al., 2015
		LC146691	Asia	No publicada
		KX687709	Norteamérica	Craig et al., 2016
RC 5	2	JQ246419	Norteamérica	Sonsthagen et al., 2012
		KR259270	Europa	Nebel et al., 2015
RC 6	2	KR336775	Europa	Nebel et al., 2015
		KR259271	Europa	Nebel et al., 2015
		KR336764	Europa	Nebel et al., 2015
		FJ627044	Europa	Cadahia et al., 2009
RC 7	5	KR259272	Europa	Nebel et al., 2015
		EF459579	Europa	Haring et al., 2007

---

---

RC 8	2	KR259276	Europa	Nebel et al., 2015
		KR336776	Europa	Nebel et al., 2015
RC 9	2	KR259274	Europa	Nebel et al., 2015
		KR336766	Europa	Nebel et al., 2015
RC 10	1	KX687710	Norteamérica	Craig et al., 2016
RC 11	2	KR259269	Europa	Nebel et al., 2015
		KR336765	Europa	Nebel et al., 2015
RC 12	1	JQ246418	Norteamérica	Sonsthagen et al., 2012
RC 13	2	KR259273	Asia	Nebel et al., 2015
		KR336773	Asia	Nebel et al., 2015
RC 14	2	KR259267	Asia	Nebel et al., 2015
		KR336790	Asia	Nebel et al., 2015
RC 15	2	KR259268	Asia	Nebel et al., 2015
		KR336786	Asia	Nebel et al., 2015
RC 16	4	KR259251	Europa	Nebel et al., 2015
		KR336783	Europa	Nebel et al., 2015
		hKR259259	Europa	Nebel et al., 2015
		KR336771	Europa	Nebel et al., 2015
RC 17	1	KR259252	Europa	Nebel et al., 2015

---

RC 18	2	KR259253	Europa	Nebel et al., 2015
		KR336784	Europa	Nebel et al., 2015
RC 19	2	KR259254	Europa	Nebel et al., 2015
		KR336780	Europa	Nebel et al., 2015
RC 20	2	KR259255	Europa	Nebel et al., 2015
		KR336778	Europa	Nebel et al., 2015
RC 21	2	KR259256	Europa	Nebel et al., 2015
		KR336774	Europa	Nebel et al., 2015
RC 22	1	KR336782	Europa	Nebel et al., 2015
RC 23	2	KR259257	Europa	Nebel et al., 2015
		KR336781	Europa	Nebel et al., 2015
RC 24	2	KR259258	Europa	Nebel et al., 2015
		KR336785	Europa	Nebel et al., 2015

*Tabla XXXII.* Haplotipos del mitogenoma obtenido del GenBank y los reportados en el presente estudio.

<b>Haplotipo</b>	<b>Número de secuencias</b>	<b>Secuencias</b>	<b>Región geográfica</b>	<b>Referencia</b>
MG 1	1	KF905228	Norteamérica	Doyle et al., 2014
MG 2	1	Ach 1	B.C. México	Este estudio
MG 3	1	Ach 2	B.C. México	Este estudio
MG 4	1	Ach 3	B.C. México	Este estudio

---

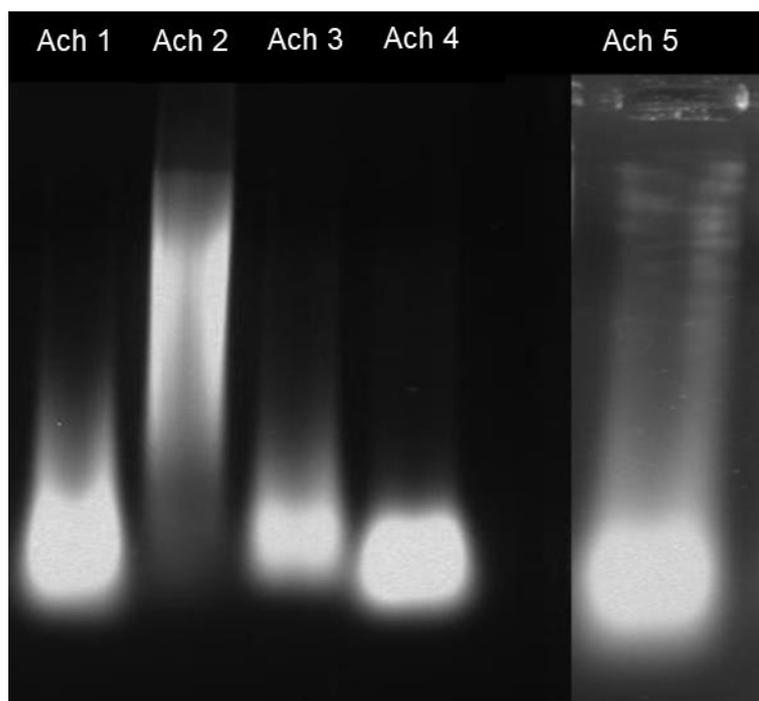
MG 5	1	Ach 4	B.C. México	Este estudio
MG 6	1	Ach 5	B.C. México	Este estudio
MG 7	1	Ach 6	B.C. México	Este estudio

---

## Figuras

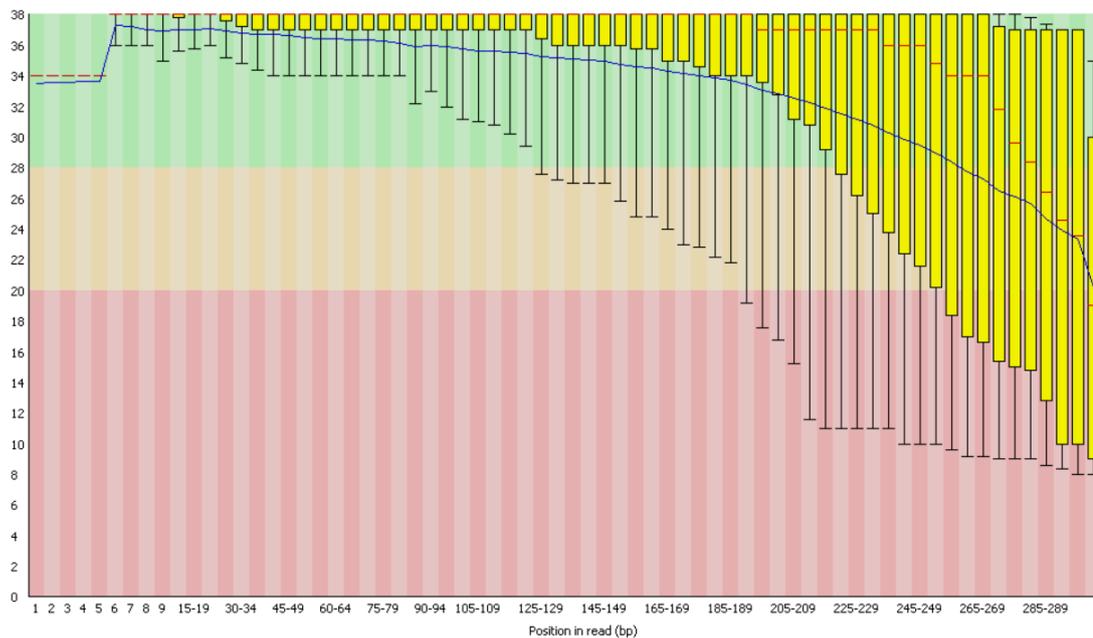
### Extracción del ADN

En la figura 1 se muestra el grado de integridad del ADN extraído, es decir la calidad relativa del ADN en términos del tamaño de las moléculas de ADN, no es extraño encontrar ADN degradado en muestras de museo como es el caso de estas muestras de Águila real (*Aquila chrysaetos*) pertenecientes a la colección ornitológica de la Facultad de Ciencias de la Universidad Autónoma de Baja California. El individuo Ach 2 es el que presentó el grado más bajo de degradación del ADN, observándose una gran mancha que representa ADN degradado, pero la mayoría de este ADN de mediano peso molecular. Mientras que los demás individuos presentan ADN muy degradado con moléculas de bajo peso molecular.

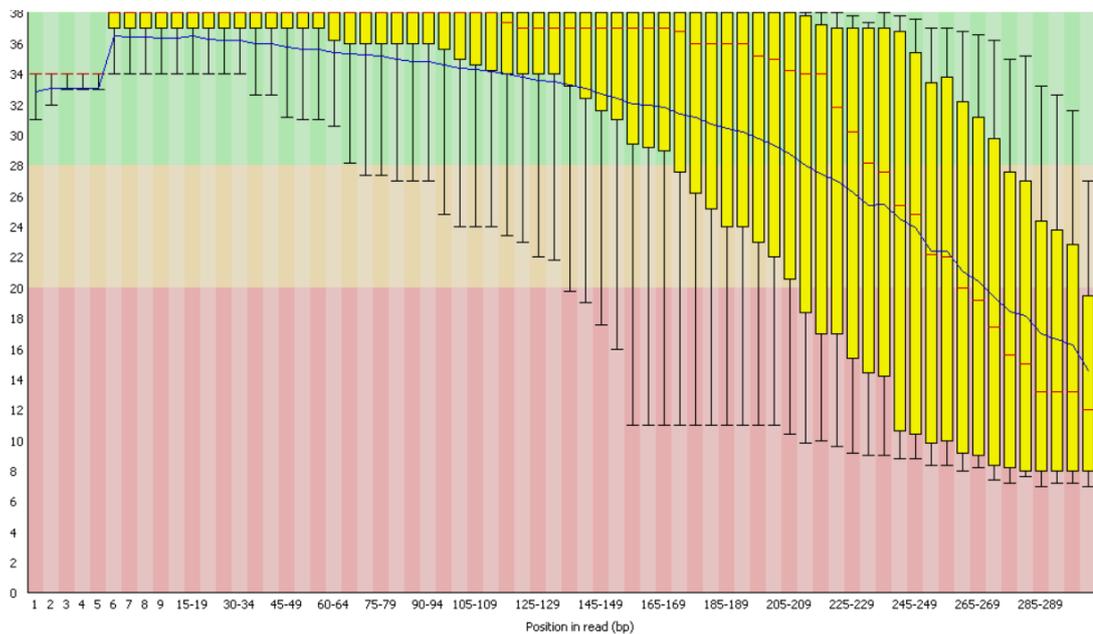


*Figura 15.* Imagen de la electroforesis del ADN genómico de los individuos Ach 1-5 que fue secuenciado.

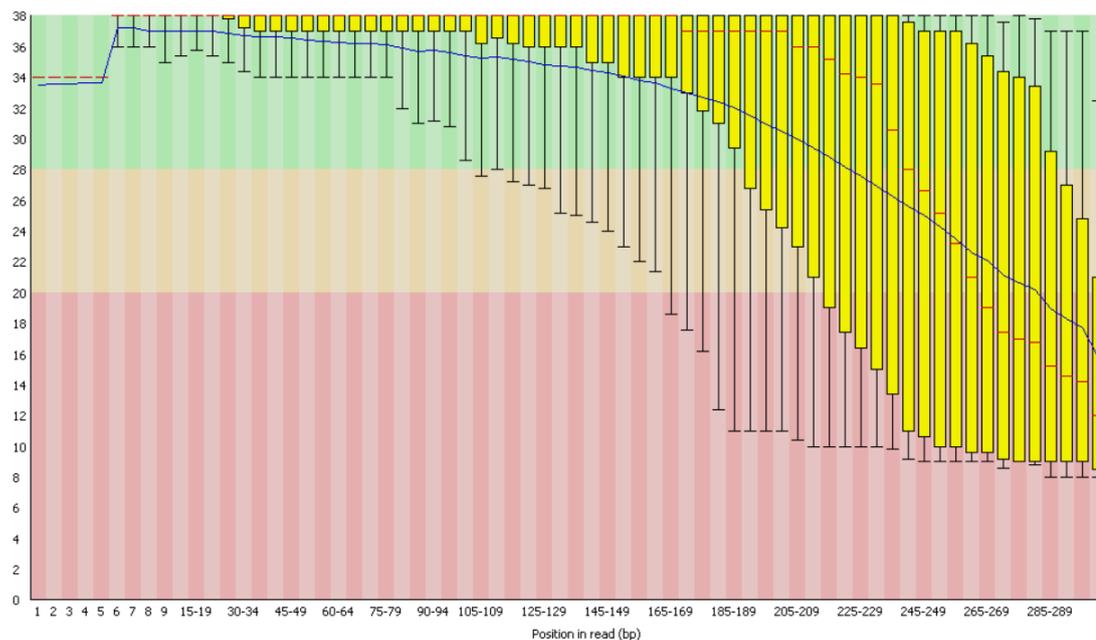
## Secuenciación



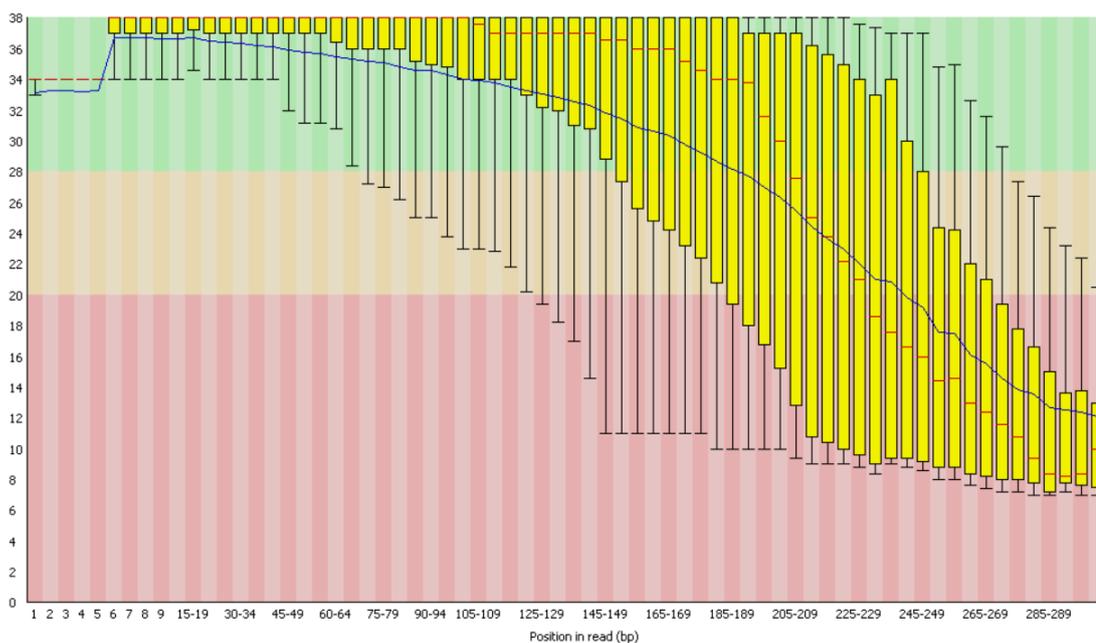
*Figura 16.* Gráfico del valor de calidad del archivo de secuencias r1 (forward) del individuo Ach 2.



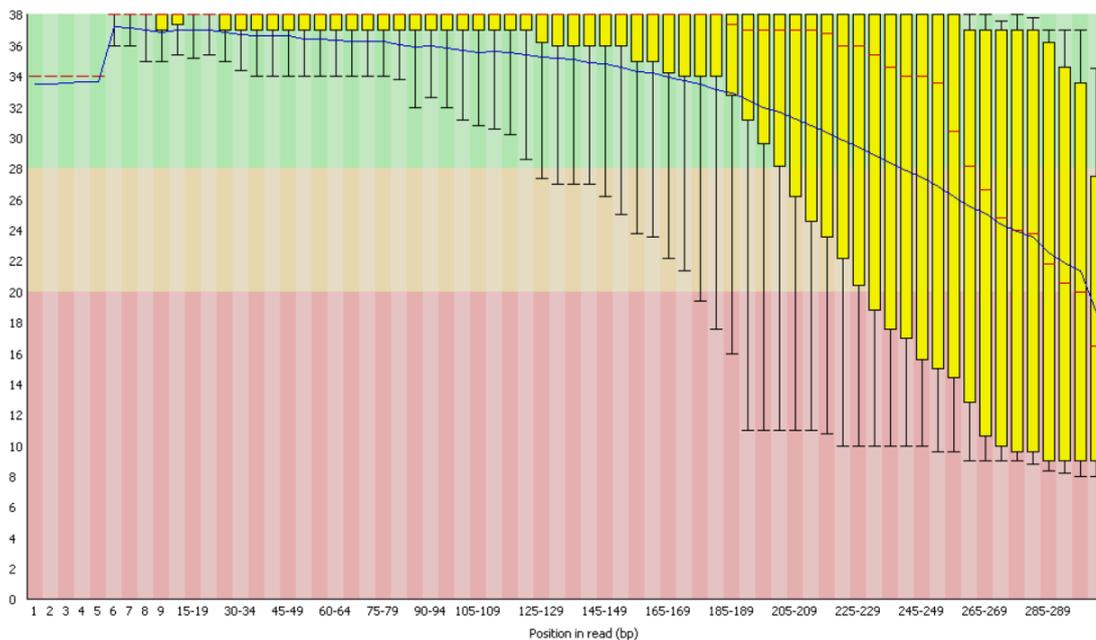
*Figura 17.* Gráfico del valor de calidad del archivo de secuencias r2 (reverse) del individuo Ach 2.



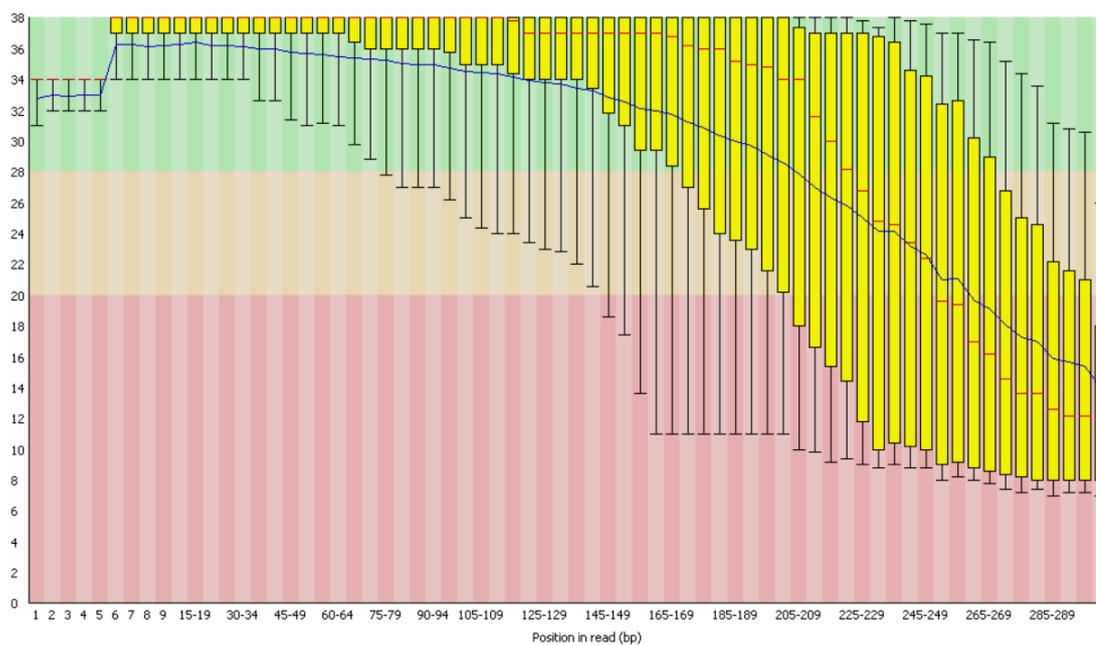
**Figura 18.** Gráfico del valor de calidad del archivo de secuencias r1 (forward) del individuo Ach 3.



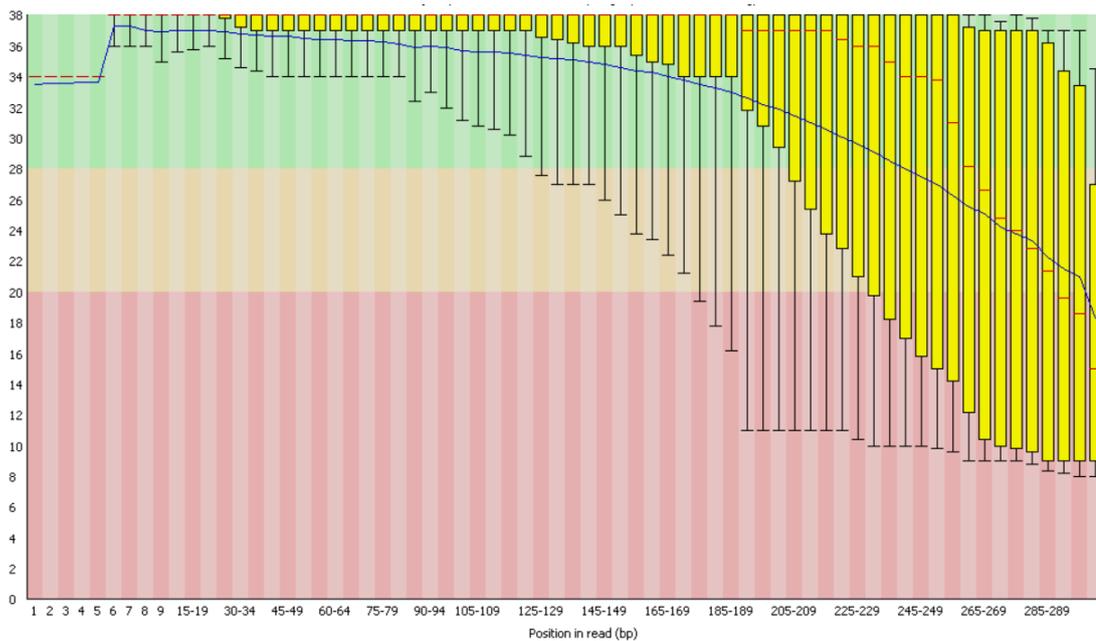
**Figura 19.** Gráfico del valor de calidad del archivo de secuencias r2 (reverse) del individuo Ach 3.



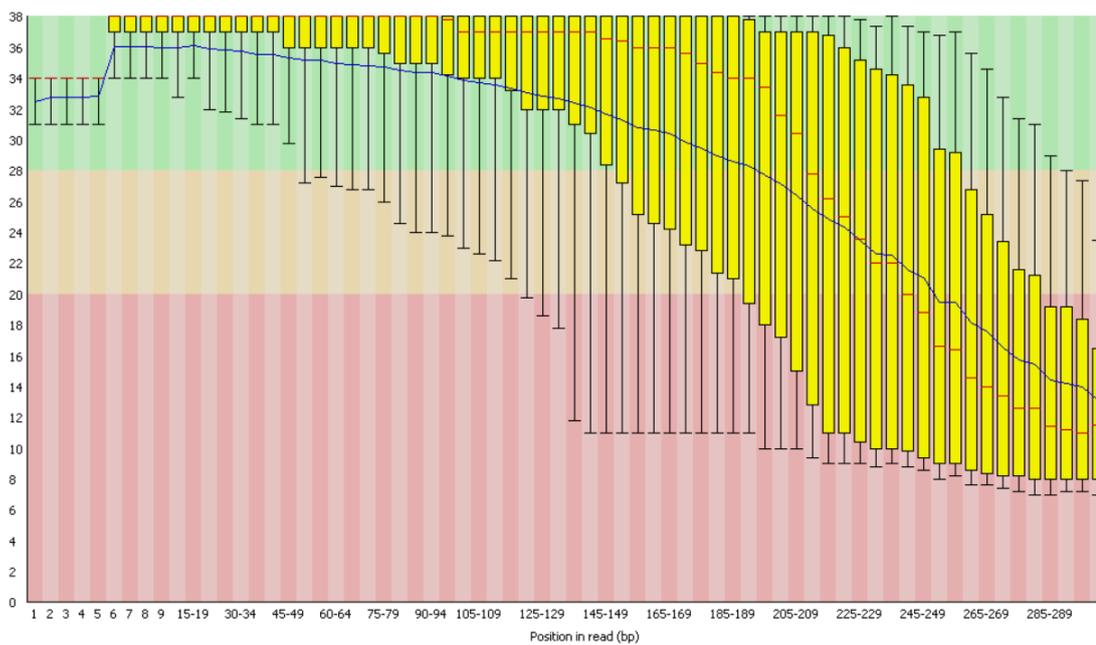
**Figura 20.** Gráfico del valor de calidad del archivo de secuencias r1 (forward) del individuo Ach 4.



**Figura 21.** Gráfico del valor de calidad del archivo de secuencias r2 (reverse) del individuo Ach 4.



**Figura 22.** Gráfico del valor de calidad del archivo de secuencias r1 (forward) del individuo Ach 5.



**Figura 23.** Gráfico del valor de calidad del archivo de secuencias r2 (reverse) del individuo Ach 5.

### Ensamble por referencia del mitogenoma de *Aquila chrysaetos*

A continuación, se muestran los ensambles por referencia de los individuos Ach 2-5 de *Aquila chrysaetos*.

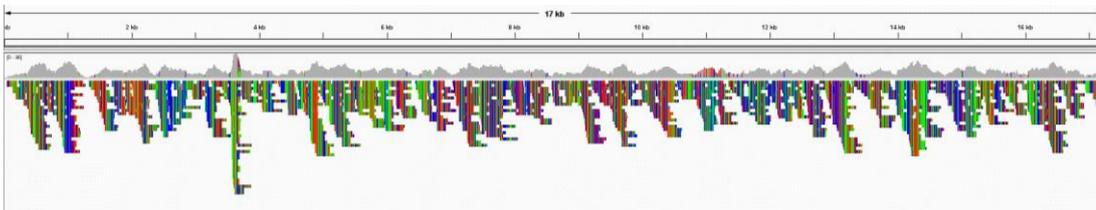


Figura 24. Ensamble por referencia del genoma mitocondrial de *Aquila chrysaetos*, del individuo Ach 2 utilizando como referencia KF905228.

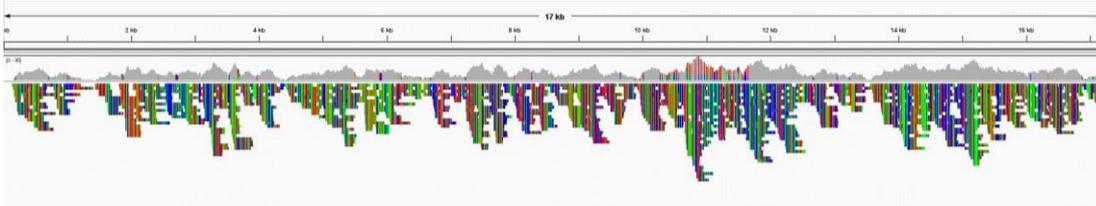


Figura 25. Ensamble por referencia del genoma mitocondrial de *Aquila chrysaetos*, del individuo Ach 3 utilizando como referencia KF905228.

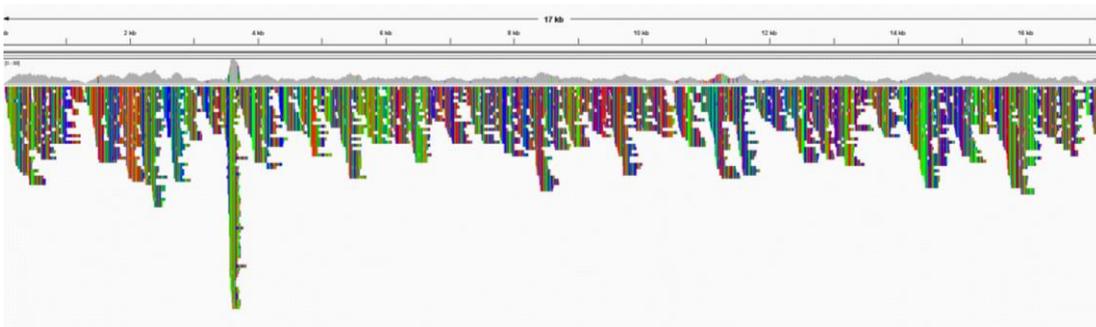
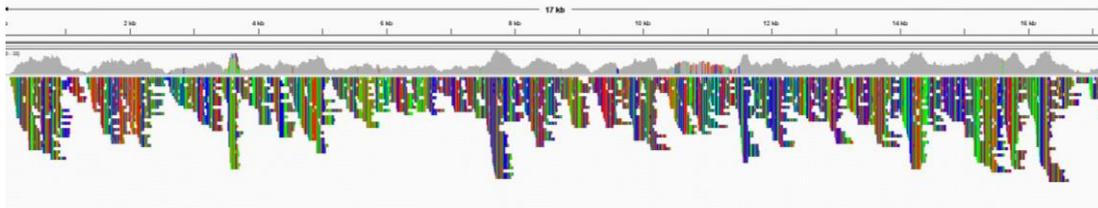


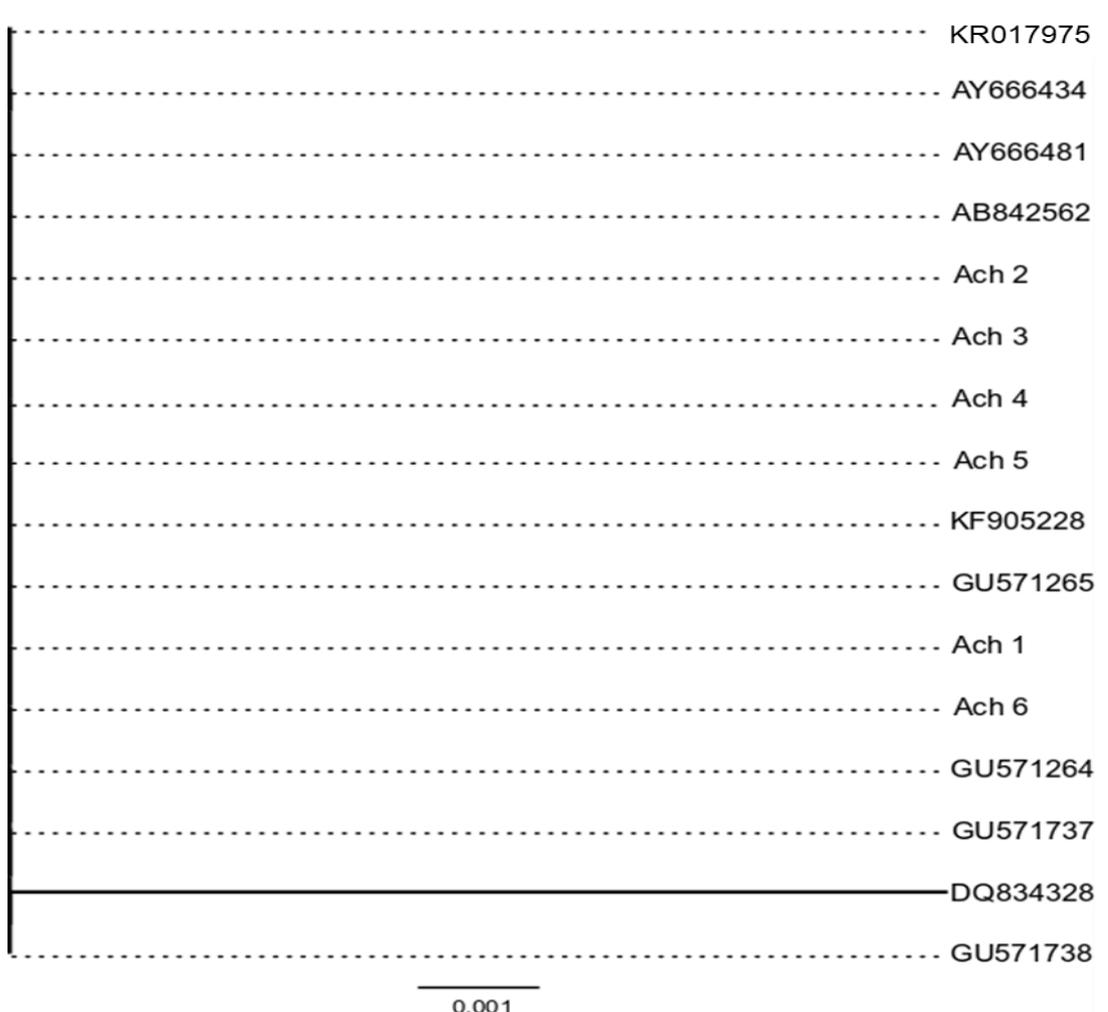
Figura 26. Ensamble por referencia del genoma mitocondrial de *Aquila chrysaetos*, del individuo Ach 4 utilizando como referencia KF905228.



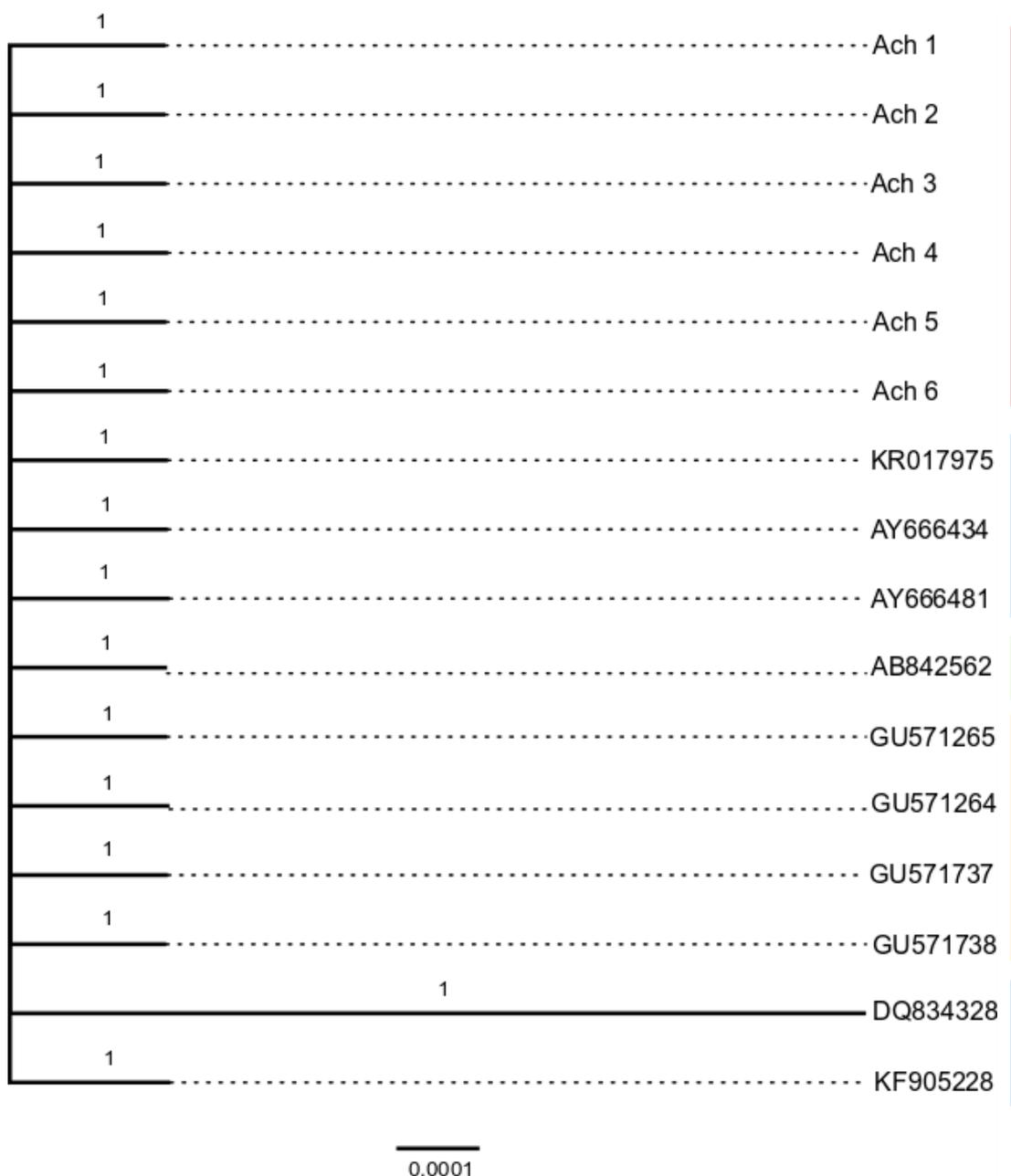
*Figura 27.* Ensamble por referencia del genoma mitocondrial de *Aquila chrysaetos*, del individuo Ach 5 utilizando como referencia KF905228.

## Reconstrucción filogenética

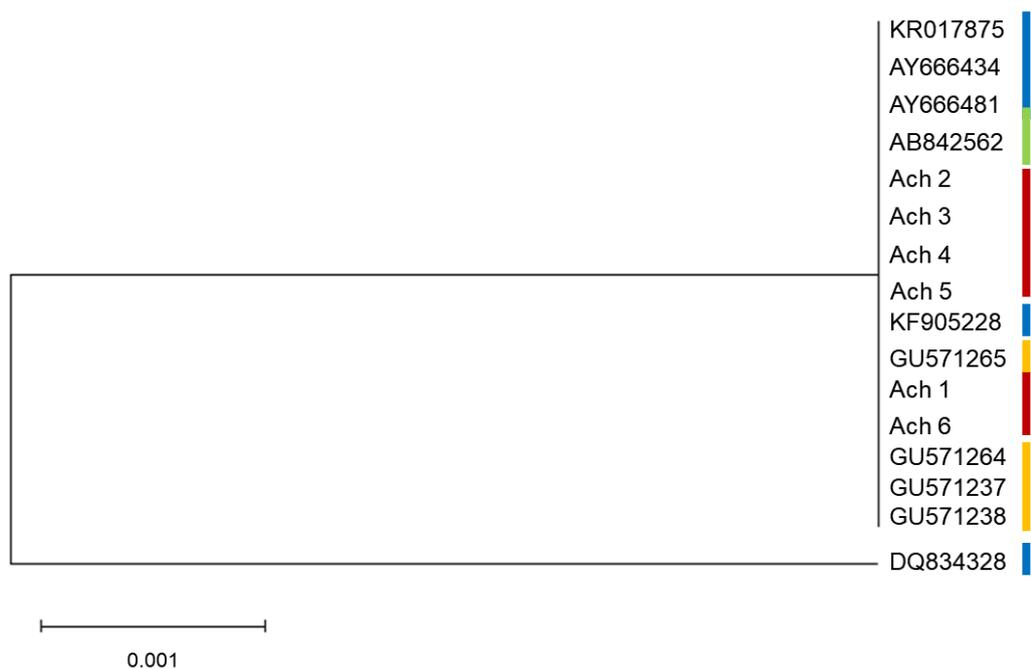
A continuación, se presentan las filogenias construidas (tanto por el método de máxima verosimilitud, inferencia filogenética bayesiana) y los fenogramas hechos mediante Neighbor joining a partir de las secuencias de los individuos secuenciados (Ach 1-6) y las disponibles en el GenBank para los genes CO I, Cyt B, ND 2 y la  $\psi$ RC y la RC, así también la filogenia por máxima verosimilitud y el fenograma hecho mediante Neighbor joining para el mitogenoma.



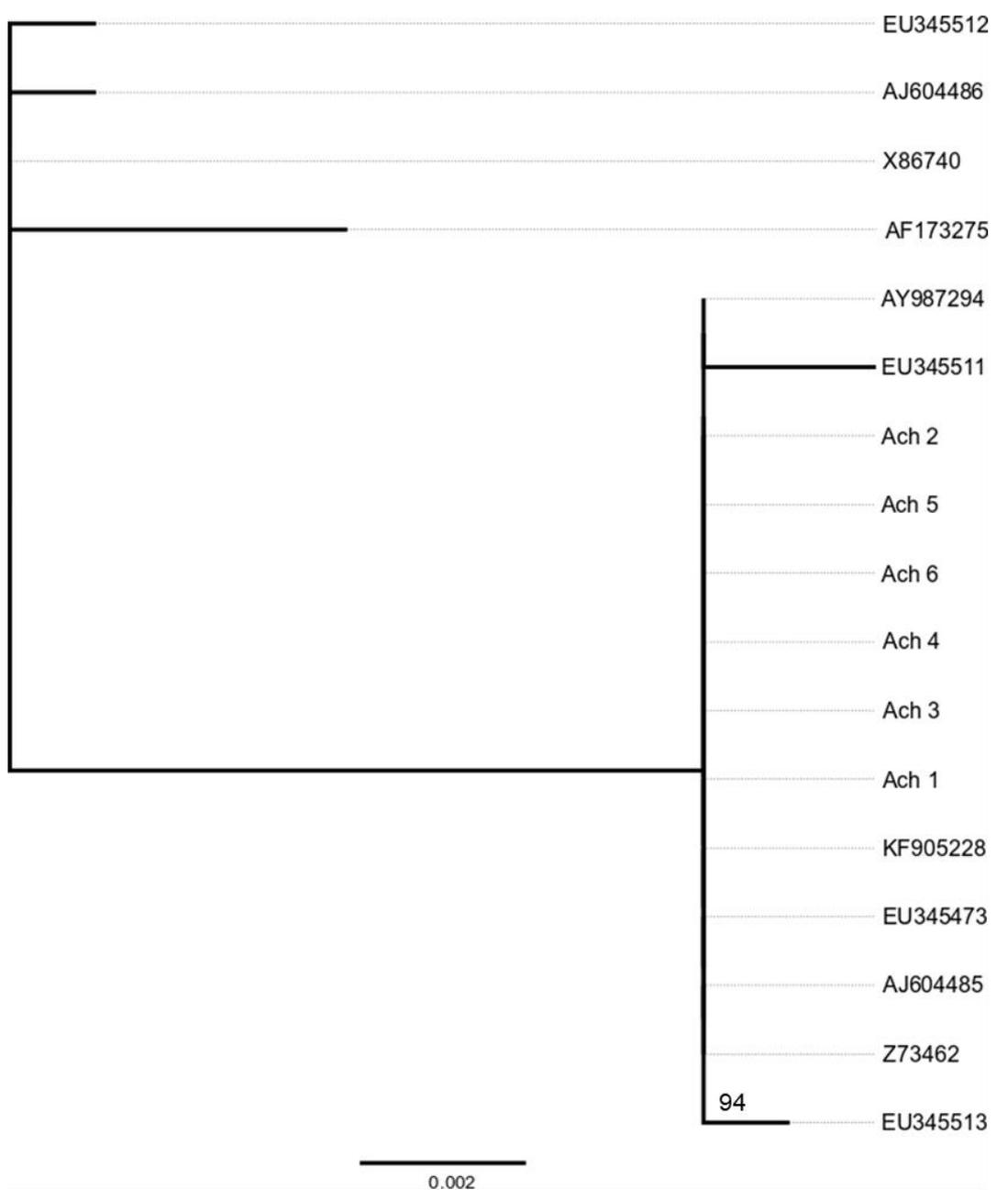
*Figura 28.* Árbol filogenético del gen CO I (citocromo oxidasa sub unidad 1), construido mediante el método de máxima verosimilitud a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.



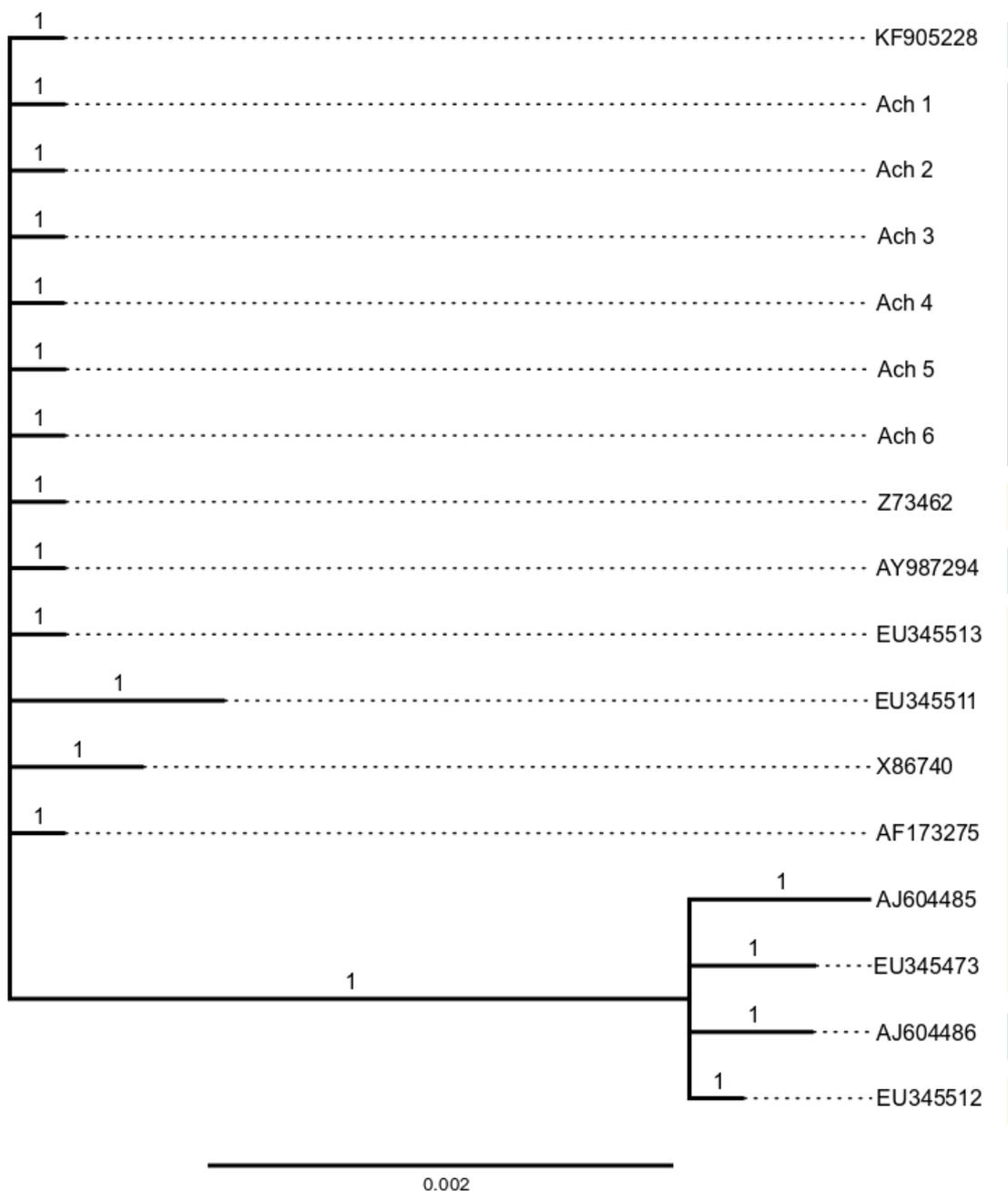
*Figura 29.* Árbol filogenético del gen CO I (citocromo oxidasa sub unidad 1), construido mediante el método bayesiano de inferencia filogenética a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor de probabilidad de cada una de las ramas. Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.



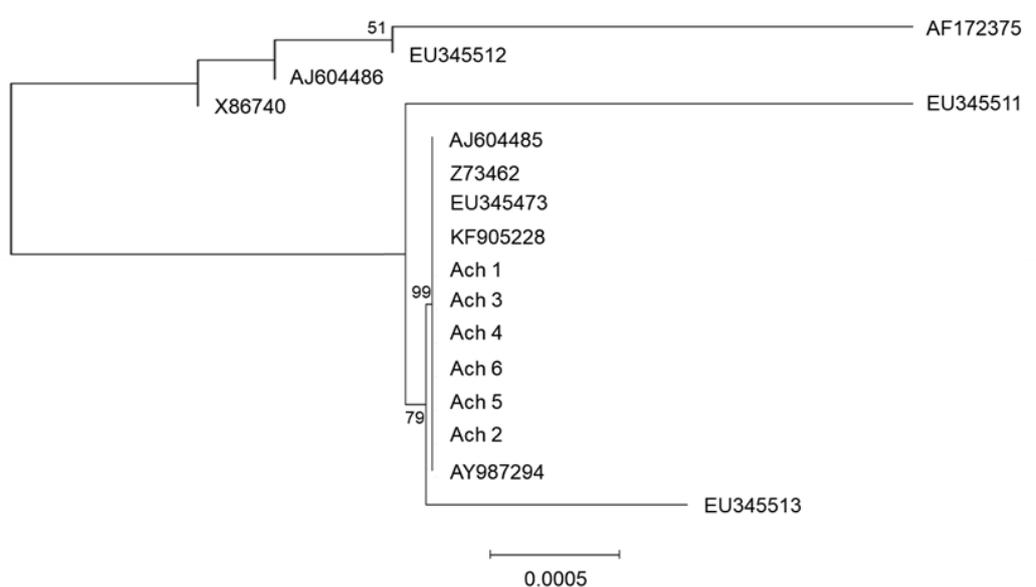
*Figura 30.* Fenograma del gen CO I (citocromo oxidasa sub unidad 1), construido mediante el método de Neighbor joining a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos secuenciados (Ach 1-6). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.



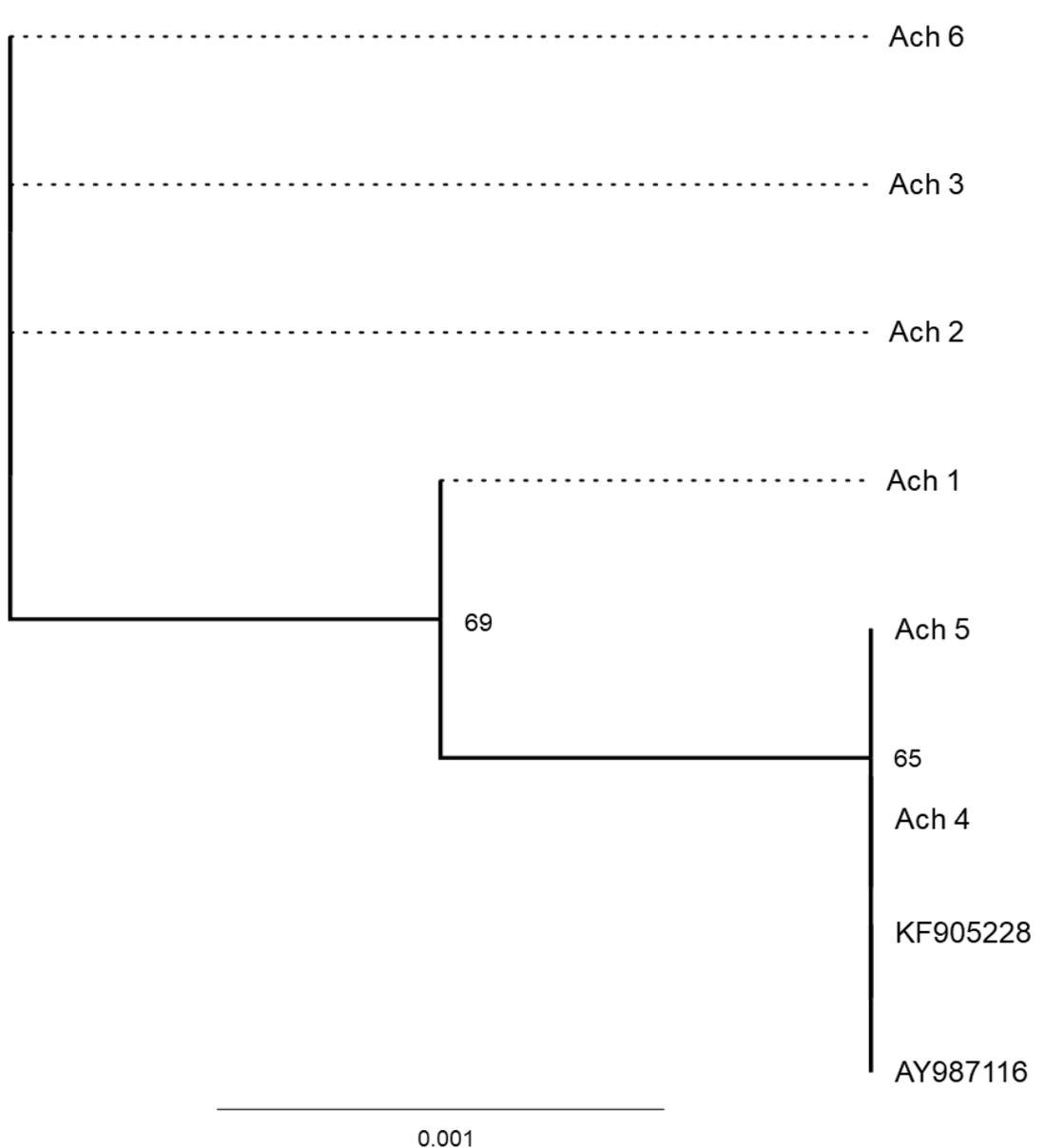
*Figura 31.* Árbol filogenético del gen Cyt b (citocromo b), construido mediante el método de máxima verosimilitud a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica y en amarillo Europa.



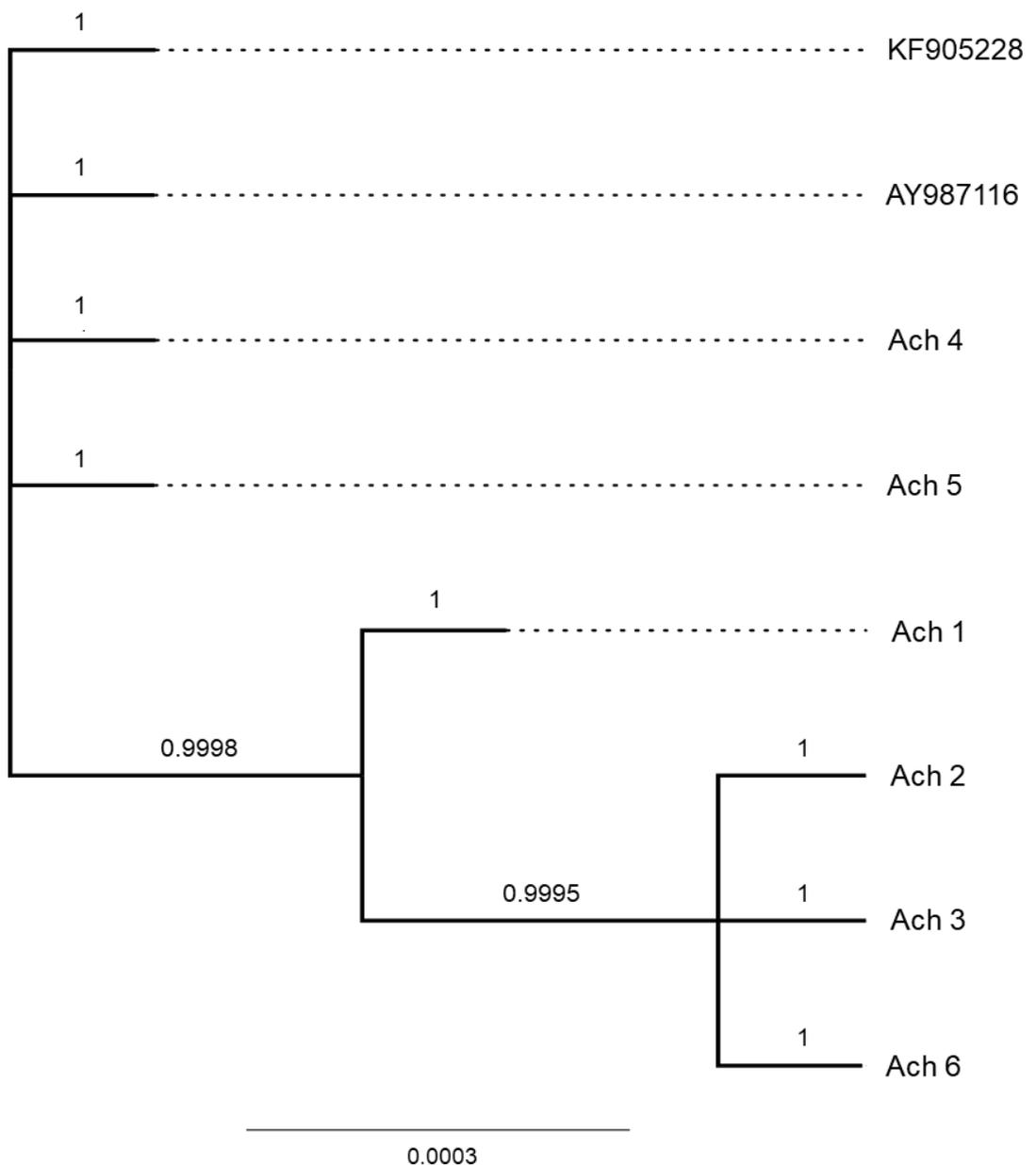
*Figura 32.* Árbol filogenético del gen Cyt b (citocromo b), construido mediante el método bayesiano de inferencia filogenética a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor de probabilidad de cada una de las ramas. Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica y en amarillo Europa.



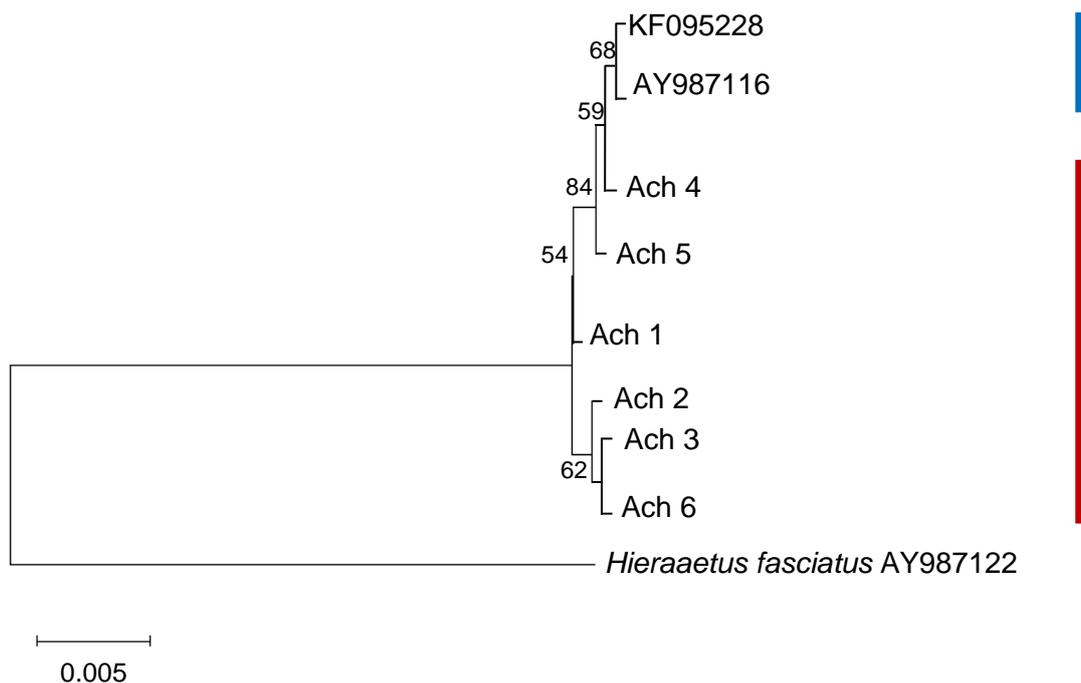
*Figura 33.* Fenograma del gen Cyt b (citocromo b), construido mediante el método de Neighbor joining a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos secuenciados (Ach 1-6). Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica y en amarillo Europa.



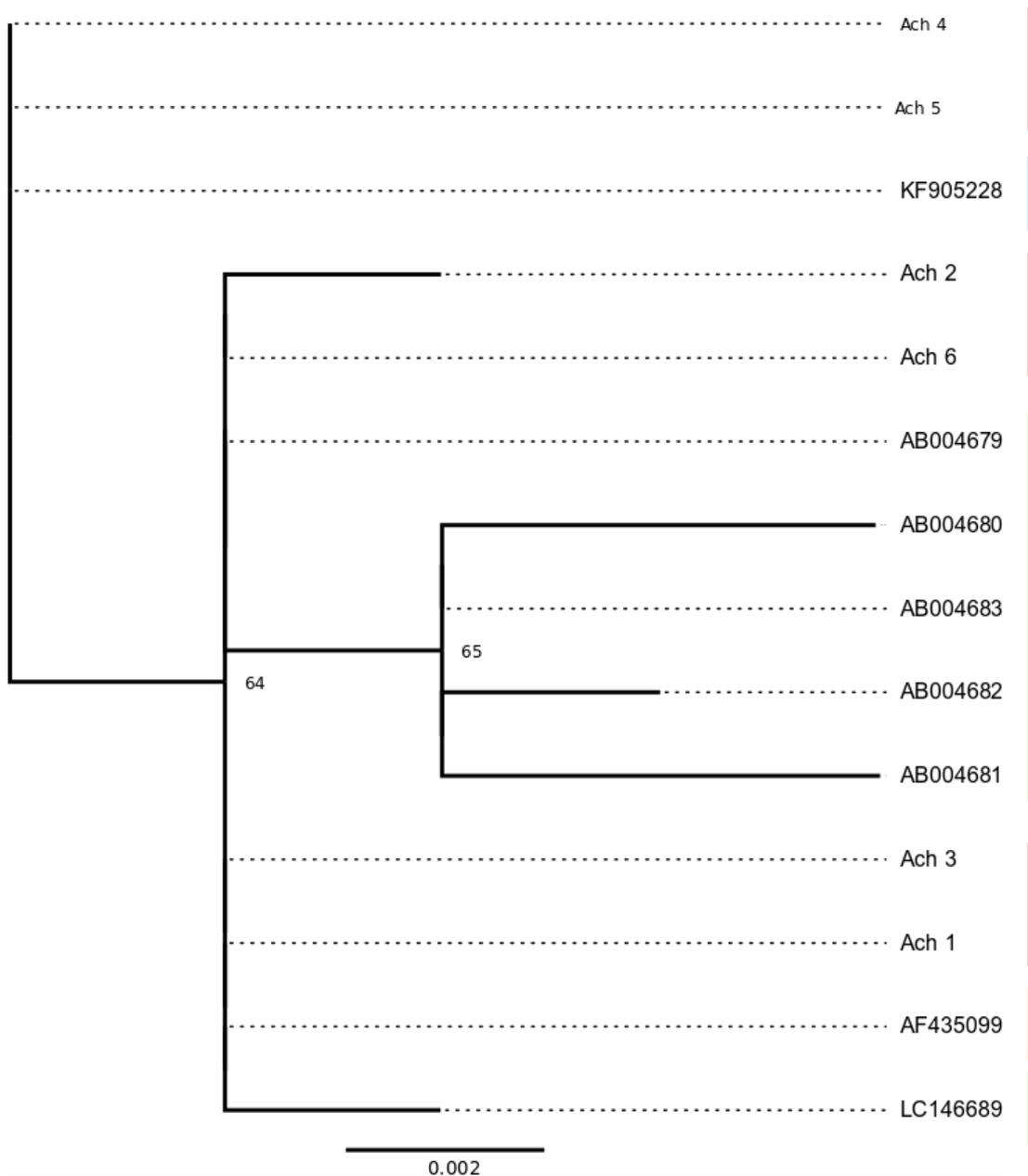
*Figura 34.* Árbol filogenético del gen ND 2 (NADH deshidrogenasa sub unidad 2), construido mediante el método de máxima verosimilitud a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California y en azul Norteamérica.



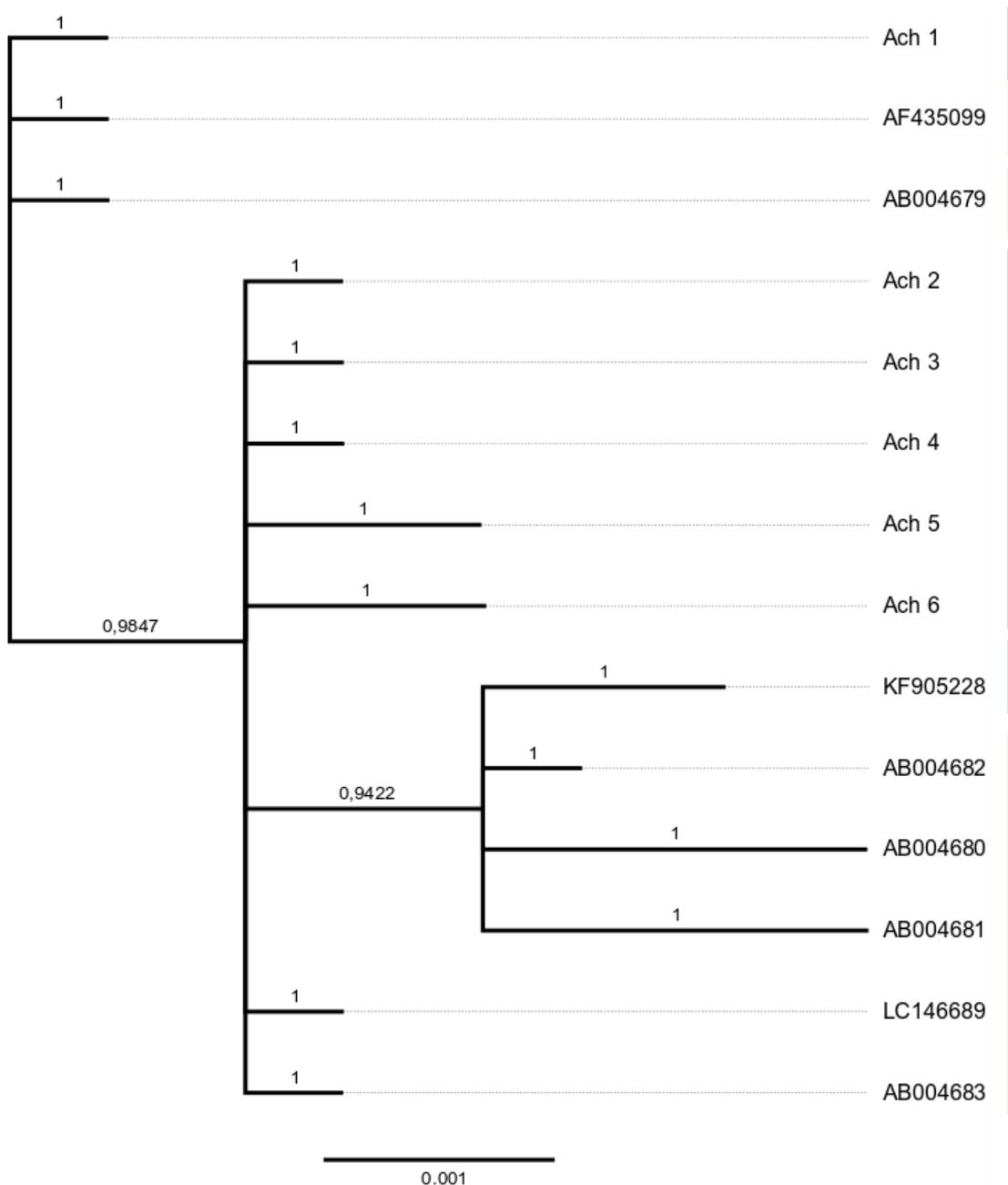
**Figura 35.** Árbol filogenético del gen ND 2 (NADH deshidrogenasa sub unidad 2), construido mediante el método bayesiano de inferencia filogenética a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor de probabilidad de cada una de las ramas. Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California y en azul Norteamérica.



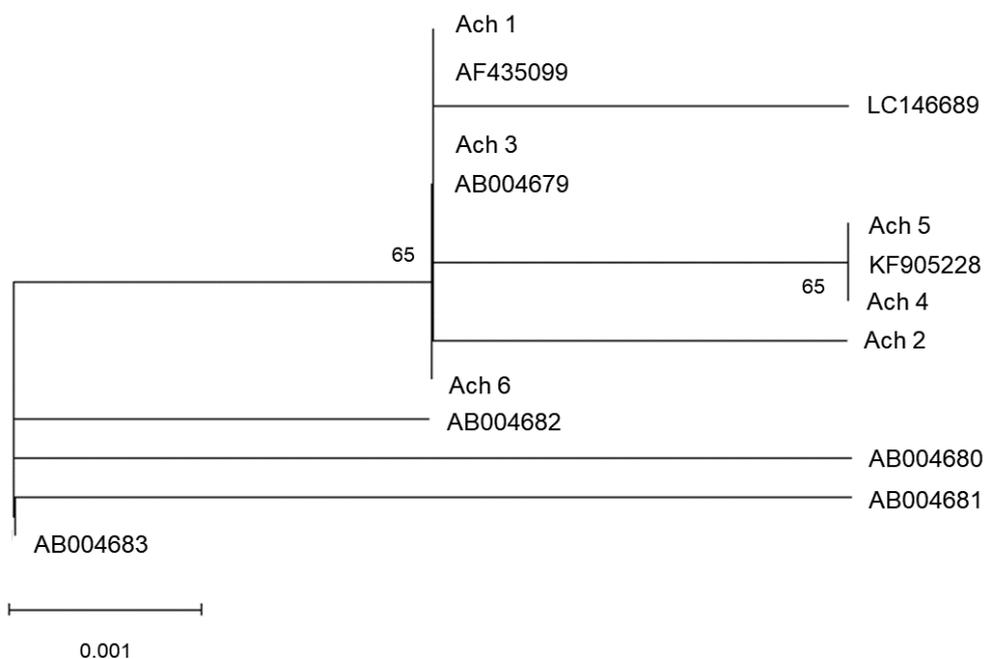
**Figura 36.** Fenograma del gen mitocondrial ND 2(NADH deshidrogenasa sub unidad 2), construido mediante el método de Neighbor joining a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos secuenciados (Ach 1-6) y como grupo externo la especie *Hieraetus fasciatus*. Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California y en azul Norteamérica.



*Figura 37.* Árbol filogenético de la pseudo región control, construido mediante el método de máxima verosimilitud a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.



*Figura 38.* Árbol filogenético de la pseudo región control, construido mediante el método bayesiano de inferencia filogenética a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor de probabilidad de cada una de las ramas. Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.



*Figura 39.* Fenograma de la pseudo región control, construido mediante el método de Neighbor joining a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos secuenciados (Ach 1-6). Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.

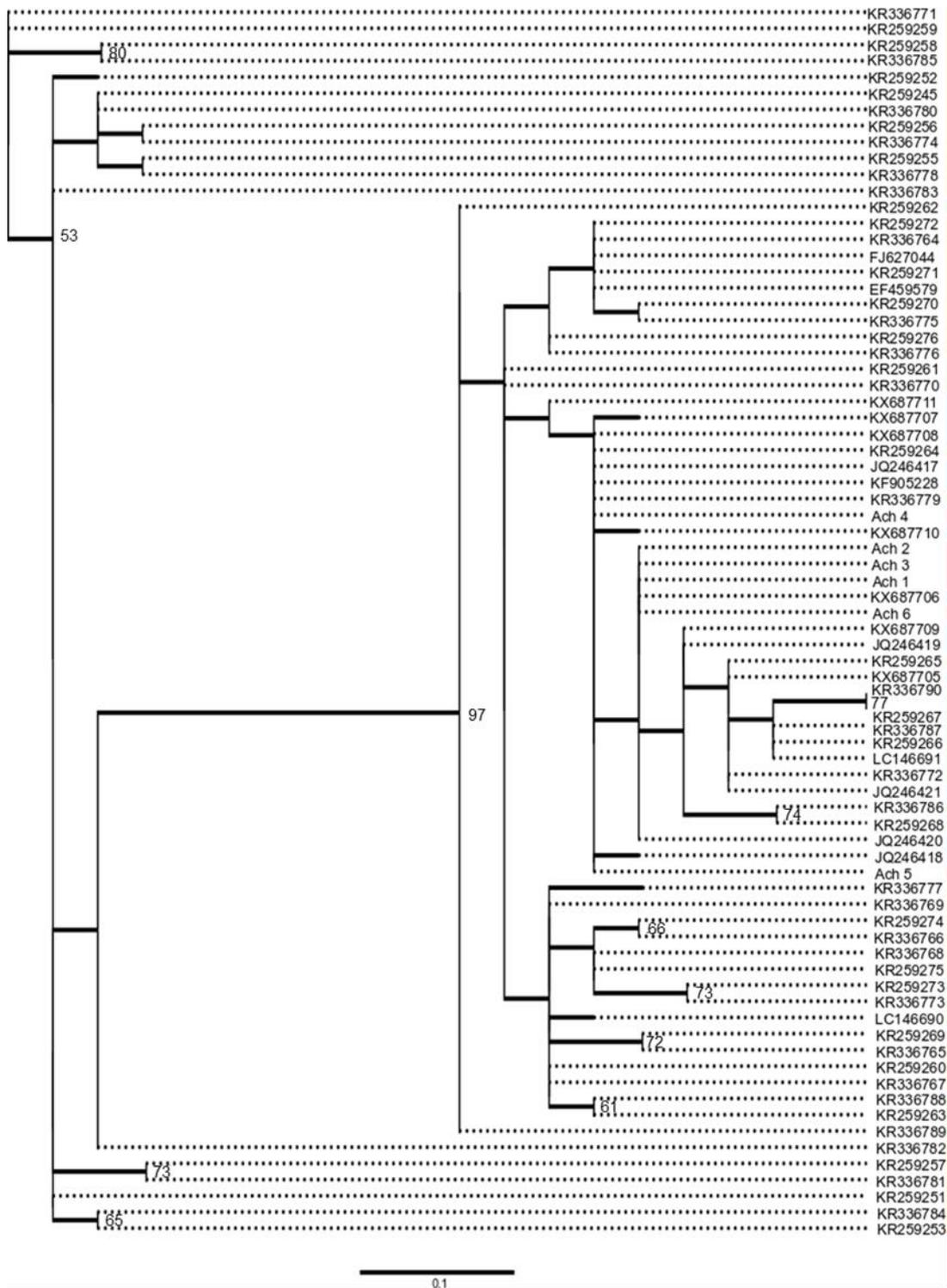


Figura 40. Árbol filogenético de la región control, construido mediante el método de máxima verosimilitud a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.

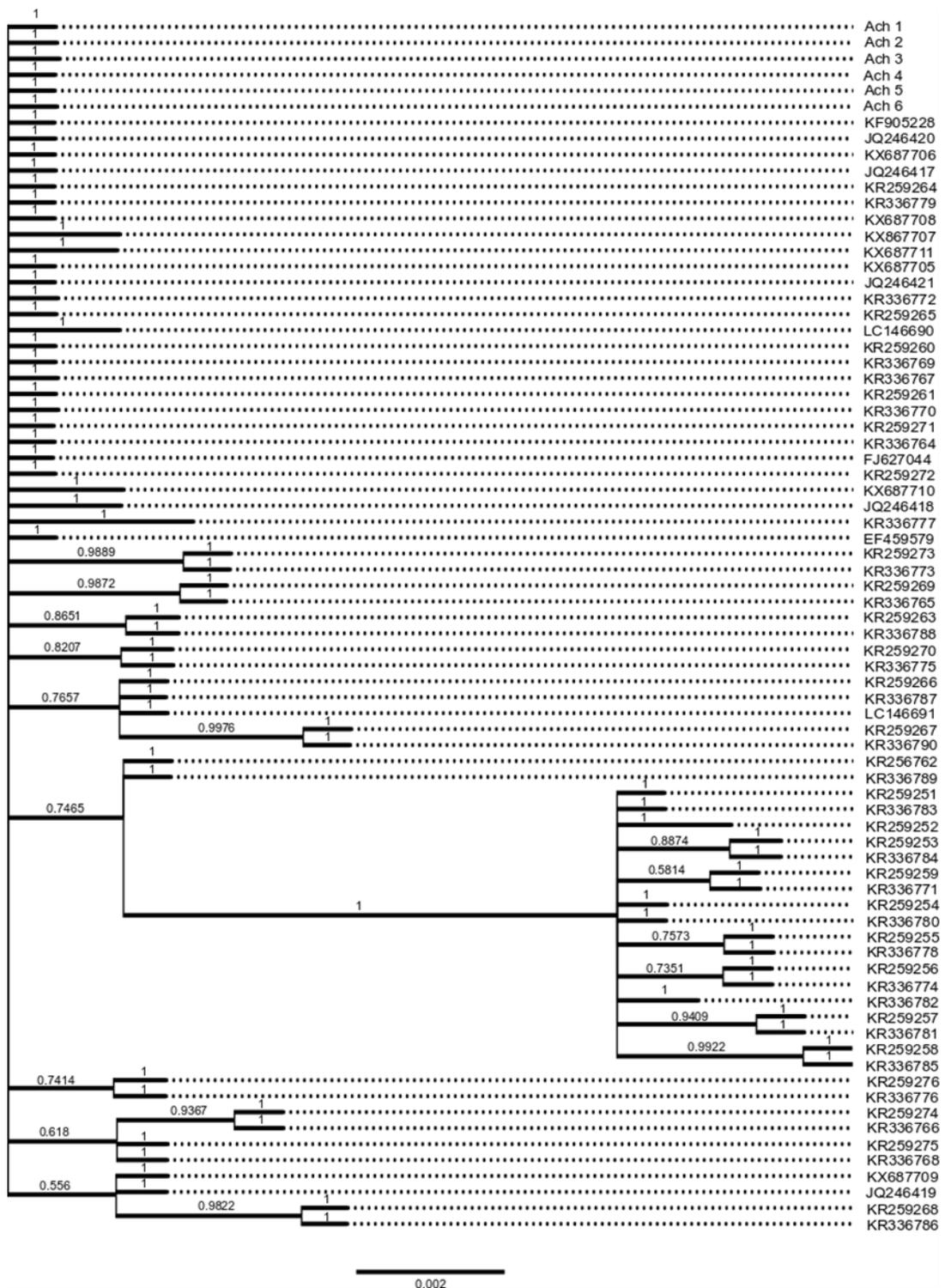
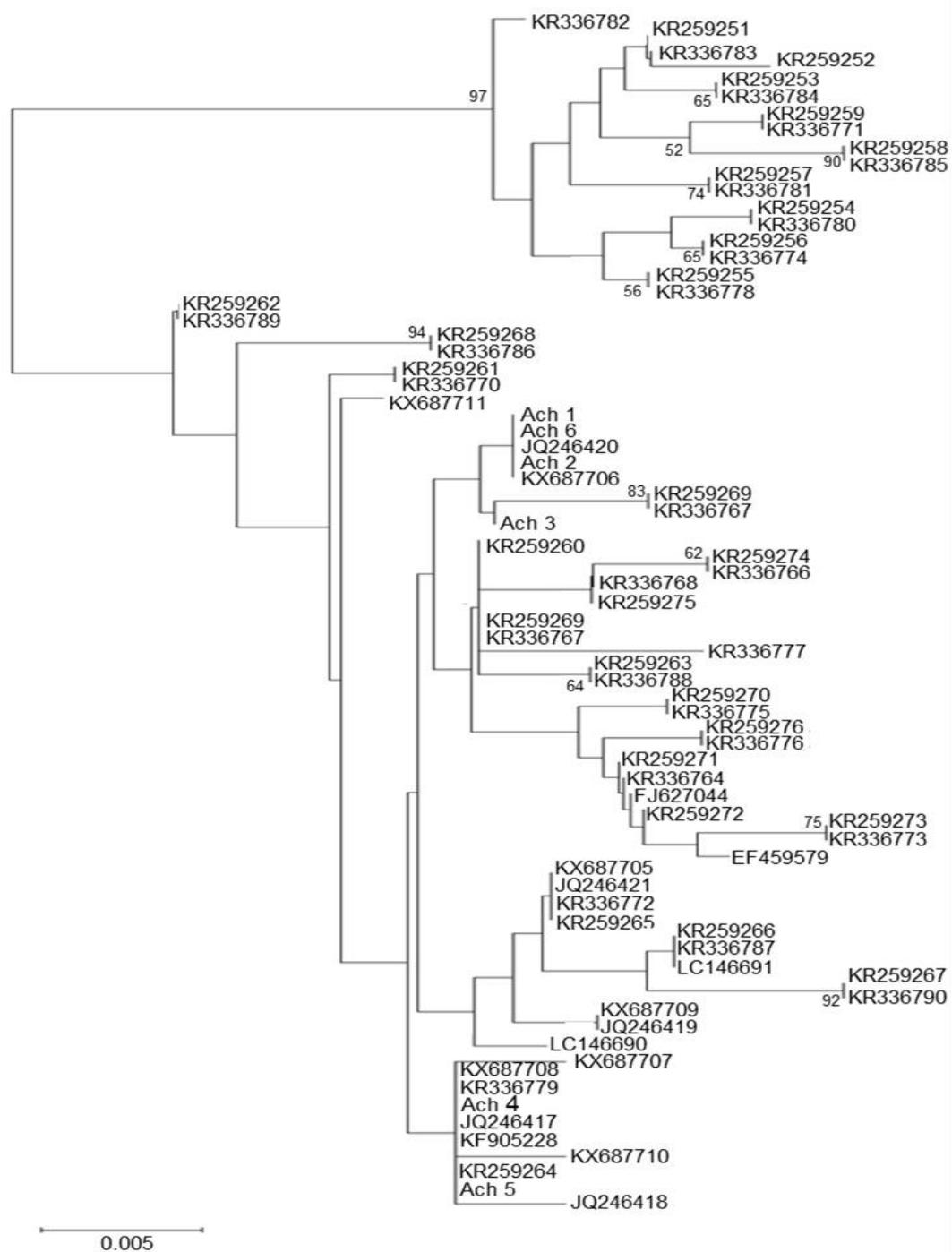
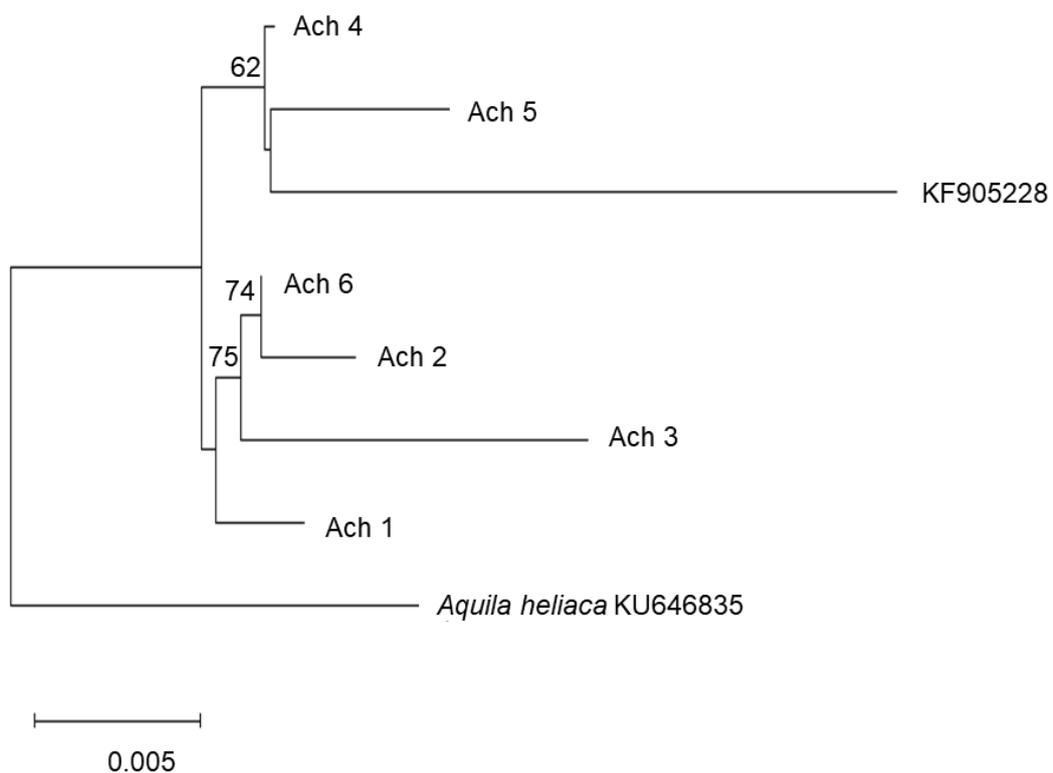


Figura 41. Árbol filogenético de la región control, construido mediante el método bayesiano de inferencia filogenética a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6). Se muestra el valor de probabilidad de cada una de las ramas. Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.



*Figura 42.* Fenograma de la región control, construido mediante el método de Neighbor joining a partir de las secuencias disponibles en el GenBank y de cada uno de los individuos secuenciados (Ach 1-6). Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%). Las barras de color representan la región de la distribución a la que pertenece cada secuencia; en rojo Baja California, en azul Norteamérica, en verde Asia y en amarillo Europa.





*Figura 44.* . Árbol filogenético del mitogenoma, construido mediante el método de máxima verosimilitud a partir de la secuencia KF905228 disponible en el GenBank y de cada uno de los individuos que fueron secuenciados (Ach 1-6) y como grupo externo la especie *Aquila heliaca* KU646835. Se muestra el valor del porcentaje de bootstrap de los nodos (> 50%).

**Glosario****ALGORITMO:**

Serie de pasos que definen un procedimiento o fórmula para resolver un problema, que puede codificarse en un lenguaje de programación y ejecutarse. Los algoritmos bioinformáticos típicamente son utilizados para procesar, almacenar, analizar, visualizar y hacer predicciones de datos biológicos <sup>1</sup>.

**ALINEAMIENTO:**

El resultado de la comparación de dos o más secuencias de genes o proteínas para determinar el grado de similitud de bases o aminoácidos. Los alineamientos de secuencias son utilizados para determinar la similitud, homología, función u otro grado de relación entre dos o más genes o producto de estos <sup>1</sup>.

**ANOTACIÓN:**

Es la combinación de comentarios, notas, referencias y citas, ya sea de formato libre o utilizando un vocabularios controlado, que juntos describen toda la información experimental e inferida acerca de un gen o proteína. Las anotaciones también pueden aplicarse a la descripción de otros sistemas biológicos. La anotación automatizada por lotes de secuencia biológica masiva es uno de los usos clave de las herramientas bioinformáticas <sup>1</sup>.

**BLAST:**

Herramienta de búsqueda y alineamiento básico local. Un alineamiento de comparación de secuencias optimizado para la velocidad utilizada para buscar bases de datos de secuencias para alineaciones locales óptimas a una secuencia de consulta <sup>2</sup>.

**BOOTSTRAP:**

Valor producto del bootstrapping, porcentaje de ocasiones que muestran el mismo resultado. En filogenia, el valor (en porcentaje) obtenido de ver cuantos árboles de las réplicas tienen el mismo nodo en cuestión <sup>13</sup>.

**BOOTSTRAPPING:**

Es un método estadístico basado en muestreos repetidamente aleatorios con reemplazamiento de una muestra original para brindar una colección de nuevas muestras pseudo replicadas, de las cuales la varianza puede ser estimada <sup>6</sup>.

**CEBADORES:**

Cadena corta de polinucleótidos preexistente a la que la ADN polimerasa puede agregar nuevos desoxirribonucleótidos <sup>3</sup>.

**COBERTURA:**

Porción del genoma a analizar o bien la porción que se encuentra representado por secuencias en un ensamble de ADN.

**CODÓN:**

Una secuencia de tres nucleótidos adyacentes que designan un aminoácido específico, sitio de inicio o paro para la transcripción <sup>1</sup>.

**CONTIG:**

Una secuencias de ADN contigua ensamblada utilizando secuencias de ADN superpuestas <sup>2</sup>.

**ELECTROFORESIS:**

Es una técnica de laboratorio utilizada para separar ADN, ARN o proteínas, basada en su tamaño y carga eléctrica. Una corriente eléctrica es usada para mover las moléculas para ser separadas a través de un gel. Unos poros en el gel funcionan como un tamiz, permitiendo a las moléculas más pequeñas moverse más rápido que las grandes. Las condiciones usadas durante la electroforesis pueden ser ajustadas para separar moléculas de un rango de tamaño deseado <sup>4</sup>.

**ENSAMBLE:**

Compilación de secuencias sobrelapadas de uno o más genes relacionados que han sido agrupados juntos basado en el grado de identidad o similaridad de las secuencias. El ensamble de secuencias puede ser utilizado para unir fragmentos de secuenciación "shotgun" basados en el sobrelapamiento de enzimas de restricción, o puede ser usado para identificar e indexar genes nuevos <sup>5</sup>.

**GENBANK:**

Base de datos de secuencias genéticas operado por una división del Instituto Nacional de la Salud (NIH) <sup>1</sup>.

La base de datos de secuencias de ácidos nucleicos alojada del NCBI (National Center for Biotechnology Information) <sup>2</sup>.

**GRÁFICOS DE BRUIJN:**

Es una gráfica en donde los nodos son secuencias de símbolos de algún alfabeto y las siglas de los nodos describen como tales secuencias puedes traslapar para formar otras secuencias <sup>7</sup>.

**HAPLOTIPO:**

Una de las formas alternativas del genotipo de un gen complejo. Este término es aplicado a un gen complejo más que el término alelo, que se refiere a una de las formas alternativas de un solo gen <sup>2</sup>.

**INDEL's:**

Mutación que involucra la adición/delección de material genético. Una inserción es una mutación que puede ser pequeña, involucra un par de bases extras, o un fragmento más grande, o una pieza más grande de un cromosoma <sup>4</sup>. Una delección es lo contrario a una inserción.

**ÍNDICE FM:**

Es una estructura de datos comprimidos para la búsqueda de patrones de texto completo <sup>11</sup>.

Es un índice de subcadena de texto completo comprimido basado en la transformada de Burrows-Wheeler, con algunas similitudes con la matriz de sufijos.

Estructura de datos oportunista, que permite la compresión del texto de entrada al tiempo que permite consultas rápidas de subcadenas <sup>12</sup>.

**INFERENCIA FILOGENÉTICA BAYESIANA:**

La reconstrucción de la filogenia mediante un método matemático para calcular la probabilidad de la dicotomía de ramas que combinan varias topologías independientes.

**MAPEO:**

También llamado ensamble por referencia, que en una estrategia de ensamble en el cual, un genoma "de referencia" sirve como templete para el solapamiento de las secuencias.

**MARCADOR MOLECULAR GENÉTICO:**

Cualquier gen que puede ser fácilmente reconocido por su efecto fenotípico, y que puede ser utilizado como un marcador para una célula, cromosoma o individuo portador de dicho gen. Además, cualquier polimorfismo detectable usado para identificar un gen específico <sup>1</sup>.

**MÁXIMA VEROSIMILITUD:**

Es un criterio para la estimación de un parámetro de datos observados bajo un modelo explícito. En el análisis filogenético, el árbol óptimo bajo el criterio de máxima verosimilitud es el árbol que es más probable que haya ocurrido dados los datos observados y el supuesto modelo de evolución <sup>6</sup>.

**MICROSATÉLITES:**

Marcador molecular altamente polimórfico comprendido por mononucleótidos, dinucleótidos, trinucleótidos o tetra nucleótidos que son repetidos en arreglos en tándem y distribuidos a través del genoma <sup>3</sup>.

**NEIRGHBOR JOINING:**

Es un algoritmo de búsqueda heurística para encontrar el árbol con la mínima evolución <sup>6</sup>.

**NGS:**

Secuenciación de Nueva Generación

**PAR DE BASES:**

Un par de bases nitrogenadas, que se mantienen juntas por puentes de hidrógeno, que forman en núcleo del ADN y ARN, es decir, las interacciones A:T, G:C y A:U <sup>1</sup>.

**PAIRED END:**

Tipo de Secuenciación de ADN que involucra la secuenciación en ambos sentidos de la cadena de ADN, es decir en el sentido 5´ a 3´ y en sentido inverso 3´ a 5´.

**PCR:**

Reacción en Cadena de Polimerasa es una técnica de laboratorio usada para amplificar secuencias de ADN. El método involucra usar secuencias cortas de ADN llamadas cebadores para seleccionar la porción del genoma a ser amplificada. La temperatura de la muestra es repetidamente elevada y descendida para ayudar a la enzima de replicación de ADN a copiar la secuencia de ADN objetivo. La técnica puede producir millones de copias de la secuencia deseada en tan solo unas horas <sup>4</sup>.

**PIPELINE:**

Es un conjunto de elementos procesadores de datos conectados en serie, en donde la salida de un elemento es la entrada del siguiente. Los elementos del pipeline son generalmente ejecutados en paralelo, en esos casos, debe haber un almacenamiento tipo buffer insertado entre elementos <sup>10</sup>.

**PROFUNDIDAD:**

La profundidad es el número de veces que se encuentra representado un fragmento de una cadena de ADN en un ensamble por las secuencias a ensamblar.

**SCAFFOLD:**

Secuencia de tamaño mayor al de un contig, ya que es formado a partir de la unidos de estos. Estos no son consecutivos y suelen tener huecos entre contigs y son formados gracias a contigs adicionales <sup>8</sup>.

**SECUENCIAS CRUDAS:**

Lecturas de secuencias tomadas directamente del secuenciador si ningún tratamiento de limpieza previo.

**SECUENCIAS LIMPIAS:**

Lecturas de secuencias pasadas por un proceso de limpieza de adaptadores, secuencias sobre representadas, secuencias de baja calidad, etc.

**SECUENCIACIÓN SHOTGUN:**

Es una técnica de laboratorio para determinar la secuencia de ADN del genoma de un organismo. El método involucra fragmentar el genoma en una colección de pequeños fragmentos de ADN que son secuenciados

individualmente. Un programa de computadora busca superposiciones en las secuencias de ADN y las usa para colocar los fragmentos individuales en su orden correcto para reconstruir el genoma.

Para secuenciar un fragmento grande de ADN la manera más eficiente de hacerlo es rompiendo aleatoriamente el ADN en muchos fragmentos más pequeños, más o menos como una escopeta, con cada una de esas piezas luego secuenciadas individualmente <sup>4</sup>.

#### SINTENIA:

El estado de estar en el mismo cromosoma. También se dice que un gen es sinténico para un cromosoma particular si se sabe que está ubicado en ese cromosoma, pero por lo demás no está mapeado <sup>2</sup>.

La presencia física en el mismo cromosoma de dos o más loci genéticos, si están o no lo suficientemente cerca como para demostrar la vinculación <sup>3</sup>.

#### SNP's:

Polimorfismos de un solo nucleótido son un tipo de polimorfismo que implica la variación de un solo par de bases <sup>4</sup>.

#### TRANSFORMADA BURROWS.WHEELER:

Es un algoritmo de transformación que reestructura los daos de tal manera que el mensaje transformado es más compresible. Técnicamente, es una permutación lexicológica reversible de los caracteres de una cadena <sup>9</sup>.

#### Fuentes de definiciones

1. [http://www.ufrgs.br/lbma/animacoes/bioinf/\\_glossary300.htm](http://www.ufrgs.br/lbma/animacoes/bioinf/_glossary300.htm)
2. <http://www.informatics.jax.org/glossary/blast>
3. <http://atlasgeneticsoncology.org/Educ/GlossaryID30028ES.html#B>
4. <https://www.genome.gov/genetics-glossary/Electrophoresis>
5. <https://www.biosyn.com/bioinformatics.aspx>
6. <http://www.bcu.ubc.ca/~otto/EvolDisc/Glossary.html>
7. <http://delta.cs.cinvestav.mx/~mcintosh/oldweb/s1996/jesus/node11.html>
8. <http://conogasi.org/articulos/ensamble-de-genoma/>
9. <https://www.geeksforgeeks.org/burrows-wheeler-data-transform-algorithm/>
10. <https://gerryariel.wordpress.com/2015/02/26/definicion-y-caracteristicas-depipeline/>
11. <https://arxiv.org/abs/1506.04896>
12. <https://people.unipmn.it/manzini/papers/focs00draft.pdf>
13. [https://www.google.com.mx/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&ved=2ahUKEwjpu7qf8rzkAhWTjp4KHXF8AjQQFjAGegQIABAC&url=http%3A%2F%2Fwebs.um.es%2Falopez%2Fmiwiki%2Flib%2Fexe%2Ffetch.php%3Fid%3Dcursofilogenia%26cache%3Dcache%26media%3Di\\_cursillo\\_de\\_filogenia\\_para\\_usuarios.pdf&usg=AOvVaw1YdHHHg8YCh71\\_0WmFRgW0](https://www.google.com.mx/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&ved=2ahUKEwjpu7qf8rzkAhWTjp4KHXF8AjQQFjAGegQIABAC&url=http%3A%2F%2Fwebs.um.es%2Falopez%2Fmiwiki%2Flib%2Fexe%2Ffetch.php%3Fid%3Dcursofilogenia%26cache%3Dcache%26media%3Di_cursillo_de_filogenia_para_usuarios.pdf&usg=AOvVaw1YdHHHg8YCh71_0WmFRgW0)