

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

**FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y
CREATIVIDAD**

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

**Análisis de genómica comparativa entre dos grupos de *Atelopus ignescens*,
tomando como referencia el genoma de otra especie del mismo género:
Exploración de Variantes Genéticas Asociadas a Resistencia y Adaptabilidad**

Autor: MSc. Fernando David Velalcázar Guerrero

Director del trabajo de titulación: Dr. Andrés Merino Viteri

Quito, 2025

AGRADECIMIENTOS

Quiero expresar mis agradecimientos a la Pontificia Universidad Católica del Ecuador y a los docentes de la maestría por proporcionarme los recursos académicos, así como todas sus enseñanzas que hicieron posible la realización de este trabajo de titulación.

También quisiera agradecer a mi tutor el Doctor Andrés Merino Viteri quien siempre ha creído en mí y me ha brindado su apoyo y conocimiento desde la licenciatura, siendo un pilar fundamental para la elaboración de mis tesis. Además, agradezco al proyecto Alianza Jambato y a las personas que lo conforman, por permitirme utilizar las secuencias RADseq utilizadas en este estudio.

Finalmente, quiero expresar mi profundo agradecimiento a mi padre, que con su enseñanza y ejemplo me demostró a no rendirme jamás a pesar de las dificultades de la vida. Mi gratitud también va para Devy y Jenifer, quienes me han brindado siempre su apoyo, y me han dado fuerzas para seguir siempre adelante no solo en la titulación sino también en la vida.

DEDICATORIA

Dedico este trabajo a mi padre, que, aunque ahora no este físicamente conmigo, siempre lo llevo en mis pensamientos.

ÍNDICE GENERAL

RESUMEN:	7
ABSTRACT:	8
1. INTRODUCCIÓN:	9
2. METODOLOGÍA:	11
2.1. Obtención y preparación de datos:.....	11
2.1.1. Secuencias del Centro Nacional para la Información Biotecnológica (NCBI): ...	11
2.1.2. Descarga de datos:	12
2.2. Análisis con Galaxy (Entorno bioinformático interactivo de código abierto):	12
2.2.1. Preprocesamiento de secuencias:	12
2.2.2. Alineamiento con Map with minimap 2 - Galaxy:	12
2.2.3. Llamado de Variantes Genéticas con bcftools – Galaxy.....	13
2.2.4. Anotación del genoma de referencia con BREAKER 3 – Galaxy	14
2.2.5. Anotación funcional de variantes con bcftools csq - Galaxy.....	14
2.3. Análisis de genómica comparativa (R - studio):.....	15
2.3.1. Preparación de datos y análisis estadístico:.....	15
2.3.2. Carga de archivos VCF comprimidos:.....	16
2.3.3. Comparación de variantes genéticas:.....	16
2.3.4. Clasificación y conteo de variantes genéticas por tipo SNP (Polimorfismo de un solo nucleótido) vs INDEL (inserciones o deleciones):	16
2.3.5. Visualización y conteo de contigs más variantes mediante mapa de calor:	17
2.3.6. Estimación de heterocigosidad observada por cada muestra y por grupo:.....	17
2.3.7. Estructura y estabilidad genotípica entre grupos:.....	18
2.3.8. PCA (análisis de componentes principales)	18
2.4. Visualización Genómica de Variantes en JBrowse y Anotaciones funcionales en UNIPROT	19
3. RESULTADOS	21
3.1. Preprocesamiento de secuencias:.....	21
3.2. Alineamiento con minimap 2 - Galaxy:	21
3.3. Análisis de genómica comparativa en R:.....	22
3.3.1. Comparación de variantes genéticas entre muestras actuales e históricas de <i>A. ignescens</i> :	22
3.3.2. Clasificación y conteo de variantes genéticas por tipo SNP (Polimorfismo de un solo nucleótido) vs INDEL (inserciones o deleciones):	23
3.3.3. Visualización y conteo de contigs más variantes mediante mapa de calor:	24

3.3.4.	Heterocigosidad observada en las muestras de <i>A. ignescens</i> :	25
3.3.5.	Estructura y estabilidad genotípica entre grupos:	27
3.3.6.	PCA (análisis de componentes principales):	28
3.4.	Anotación funcional de <i>A. ignescens</i> y visualización de genes variantes:	29
4.	DISCUSIÓN:	32
4.1.	Validación del alineamiento de las secuencias RADseq de <i>A. ignescens</i> con el genoma de referencia de <i>A. laetissimus</i> :	32
4.2.	Exploración y caracterización de las variantes genéticas de los grupos actuales e históricos:	34
4.3.	Limitaciones de la anotación funcional por la cobertura parcial del genoma de referencia	38
4.4.	Implicaciones genéticas para la conservación de <i>Atelopus ignescens</i> estrategias de manejo futuro	39
5.	CONCLUSIONES Y RECOMENDACIONES:	40
6.	REFERENCIAS:	42

ÍNDICE DE FIGURAS

Figura 1. Pipeline de la obtención y preparación de datos en Galaxy para el análisis genómico de las secuencias.....	15
Figura 2. Pipeline del análisis de genómica comparativa realizado en R-studio a partir de los archivos obtenidos en Galaxy	19
Figura 3. Comparación de variantes genéticas entre muestras actuales e históricas de <i>Atelopus ignescens</i>	22
Figura 4. Proporción de variantes genéticas por grupo y por tipo SNP (Polimorfismo de un solo nucleótido) vs INDEL (inserciones o deleciones).....	23
Figura 5. Mapa de calor de la distribución de variantes por grupo en los 20 contigs con mayor acumulación de variantes.	25
Figura 6. Heterocigosidad observada (Ho) por cada individuo de <i>Atelopus ignescens</i>	26
Figura 7. Heterocigosidad observada (Ho) por grupo actuales e históricos.	27
Figura 8. Comparación genotípica entre muestras actuales e históricas de <i>A. ignescens</i> en diferentes loci.....	28
Figura 9. PCA genotípico de las muestras actuales e históricas de <i>A. ignescens</i>	29
Figura 10. Genes de <i>A. ignescens</i> con mayor cantidad de variantes que caen en regiones codificantes (CDS)	30
Figura 11. Visualización de la región del gen 1407 en Jbrowse, en donde se logran apreciar algunas de las variantes genéticas principalmente SNPs.	30
Figura 12. Visualización de la región del gen 74 en Jbrowse, en donde se logran apreciar algunas de las variantes genéticas principalmente	31
Figura 13. Visualización de la región del gen 131 en Jbrowse, en donde se logran apreciar algunas de las variantes genéticas principalmente SNPs.	31

ÍNDICE DE TABLAS

Tabla 1. Identificación de las muestras RADseq en NCBI del Bioproyecto: “Amphibian rediscoveries provide insights into species persistence”	12
Tabla 2. Clasificación y conteo de variantes genéticas por tipo SNP (Polimorfismo de un solo nucleótido) vs INDEL (inserciones o deleciones):.....	¡Error! Marcador no definido.
Tabla 3. Cantidad de variantes genéticas: variantes exclusivas actuales e históricas y variantes compartidas, de los 20 contigs con más variantes.....	¡Error! Marcador no definido.
Tabla 4: Heterocigosidad observada (H_o) por muestra de <i>A. ignescens</i>	¡Error! Marcador no definido.
Tabla 5. Heterocigosidad observada (H_o) por grupo actuales e históricas de <i>A. ignescens</i> .	¡Error! Marcador no definido.
Tabla 6. Comparación de genotipos entre muestras actuales e históricas	¡Error! Marcador no definido.
Tabla 7: Lista de componentes que explican el 80% de la variación genética de las muestras.	¡Error! Marcador no definido.

RESUMEN:

Atelopus ignescens, es un sapo arlequín endémico de los Andes ecuatorianos, que se encuentra en peligro crítico de extinción, debido a su alarmante declive poblacional asociado a factores como la pérdida de hábitat, el cambio climático y la quitridiomicosis. Por tal motivo, es imprescindible abordar la problemática no solo desde un punto de vista ecológico, sino también desde el punto de vista genético que nos permitiría inferir los mecanismos de la especie que le permite resistir a amenazas ambientales y biológicas. Por lo tanto, el presente estudio analizó la variabilidad genética entre grupos actuales e históricas de *Atelopus ignescens*, utilizando datos RADseq vs el genoma de referencia de *A. laetissimus*, mediante análisis informático y estadístico para realizar un análisis de variantes, heterocigosidad y una prospección de anotación funcional. Esto con el objetivo de tratar de generar una explicación biológica, del porqué de la presencia de la especie en su única población remanente reportada hasta el momento. El análisis de la estructura genética, combinado con la prospección de la anotación funcional aportaron información relevante sobre los procesos evolutivos y demográficos que han modelado estas poblaciones, para poder generar información relevante para su conservación.

Palabras clave: RADseq, alineamiento, *Atelopus ignescens*, genómica funcional, variantes genéticas.

ABSTRACT:

Atelopus ignescens is an endemic harlequin toad from the Ecuadorian Andes, currently classified as critically endangered due to its alarming population decline. This decline has been linked to factors such as habitat loss, climate change, and chytridiomycosis. Therefore, it is essential to address this issue not only from an ecological perspective but also from a genetic standpoint, which could help infer the mechanisms that allow the species to resist environmental and biological threats. In this context, the present study analyzed the genetic variability between current and historical groups of *Atelopus ignescens* using RADseq data aligned against the reference genome of *A. laetissimus*. Bioinformatic and statistical analyses were performed to assess genetic variants, heterozygosity, and conduct a functional annotation prospecting. The objective was to provide a biological explanation for the persistence of the species within its only remaining reported population to date. The analysis of genetic structure, combined with the functional annotation prospecting, provided relevant insights into the evolutionary and demographic processes that have shaped these populations, aiming to generate valuable information to support conservation strategies for the species.

Keywords: RADseq, alignment, *Atelopus ignescens*, functional genomics, genetic variants

1. INTRODUCCIÓN:

Atelopus ignescens, conocido como el “jambato negro”, es un sapo arlequín endémico de los Andes ecuatorianos, que habitaba los páramos y bosques montanos entre 2.800 y 4.200 metros de altitud, y era bastante común en las provincias de Imbabura, Pichincha, Cotopaxi, Tungurahua, Chimborazo y Bolívar, en la década de los 80 (Vega-Yanez et al., 2024). Sus avistamientos no eran solo en zonas prístinas, sino también en zonas disturbadas por el hombre, como pastizales, potreros e incluso dentro de los parques de las ciudades de Quito, Latacunga y Ambato. El último registro confirmado de su presencia fue en el año de 1988 en la localidad de Oyacachi, posteriormente a pesar de los esfuerzos de búsqueda, se la catalogó como extinta en la lista roja de la UICN (Ron et al., 2003).

Tras tres décadas en el año 2016, la especie fue redescubierta en la localidad de Angamarca - Cotopaxi. Lo cual, reavivó los esfuerzos de investigación y conservación de la especie, debido a que sigue siendo una de las tantas especies que están en peligro crítico de extinción dado su alarmante declive poblacional, que se ha asociado a factores como la pérdida de hábitat, el cambio climático y la quitridiomycosis (Ron et al., 2003; Harvey et al., 2013; Vredenburg et al., 2010; IUCN, 2021; Vega-Yanez et al., 2024). En cuanto a la investigación, el análisis de genómica utilizando herramientas bioinformáticas nos podría ayudar a inferir los mecanismos que le han permitido a la especie a resistir ciertas amenazas ambientales y biológicas, explicando así gran parte de la problemática que enfrenta y enfrentó en el pasado *A. ignescens* (Frankham, 2005).

Estos análisis genéticos y bioinformáticos utilizan la genómica comparativa, la cual es una herramienta esencial para caracterizar variantes genéticas que puedan estar implicadas en la adaptabilidad y resistencia de una especie (Ellegren, 2014). Una de las técnicas que se usa para poder analizar estas variantes, es la secuenciación de ADN asociada a sitios de restricción (RADseq), gracias a esta técnica es posible identificar regiones del genoma altamente informativas sin requerir la secuenciación completa, lo cual es valioso para especies con genomas complejos (Baird et al., 2008). Esta técnica ha sido aplicada exitosamente en anfibios para revelar estructuras poblacionales, diferenciación genética y patrones de selección local (Davey & Blaxter, 2011; Hohenlohe et al., 2010).

En el caso de *A. ignescens*, la falta de información sobre un genoma de referencia limita ciertos tipos de análisis. Sin embargo, la disponibilidad de genomas completos de especies pertenecientes al mismo género, puede ser útil, siempre y cuando se considere el error que conlleva la comparación de genomas de especies distintas (Ekblom & Galindo, 2011). Aun así, estas comparaciones podrían revelar potenciales genes conservados y variantes potencialmente funcionales, los cuales estarían involucrados en procesos biológicos esenciales como la respuesta inmunológica, la regulación del equilibrio hídrico o la adaptación a condiciones térmicas extremas, entre otros aspectos vinculados a la supervivencia en hábitats de alta montaña (Bolaños et al., 2018; Jaynes et al., 2022; Savage & Zamudio, 2011). En este contexto, estudios en otras especies de anuros han identificado genes asociados con estas funciones adaptativas, como por ejemplo genes del complejo mayor de histocompatibilidad que puede estar relacionado con la resistencia a la quitridiomycosis, genes que codifican proteínas de choque térmico como el HSP90AA1, y genes asociados al equilibrio hídrico como los receptores de vasotocina y mesotocina. Encontrar este tipo de variantes reforzaría la hipótesis de que en *Atelopus ignescens* podrían existir variantes funcionales que le confieran ventajas adaptativas en su entorno andino (Kohn et al., 2003; Savage y Zamudio, 2011; Lu et al., 2020).

Este tipo de estudios se los puede realizar en la plataforma bioinformática Galaxy en donde se encuentran herramientas como: Blast o Minimap 2 las cuales permite la alineación de lecturas cortas o fragmentadas contra genomas de referencia, permitiendo un mapeo rápido y confiable incluso entre especies no idénticas (Afgan et al., 2018; Li et al., 2009). A partir de los alineamientos obtenidos, se pueden identificar variantes genéticas como: SNPs e indels y genes que podrían ser conservados, lo que permite explorar la distribución de dichas variantes y genes en el contexto del genoma de referencia (Li et al., 2009).

La importancia de este estudio radica en la detección de estas variantes ya que son clave para la historia evolutiva y el potencial adaptativo de *A. ignescens*, permitiendo discernir entre variantes neutras y aquellas con posible relevancia funcional. Además, el análisis de genes conservados a partir de las regiones alineadas puede revelar componentes genéticos esenciales para la supervivencia, los cuales han sido mantenidos a lo largo de la evolución del linaje *Atelopus* (Koonin et al., 2006; Jones et al., 2012).

En conjunto, este enfoque integrador proporciona una base genómica para orientar estrategias de conservación, considerando tanto la diversidad genética actual como la capacidad evolutiva de la especie para enfrentar amenazas futuras (Scheele et al., 2019; Springborn et al., 2022). Por tal motivo dentro de este estudio se pretende: 1) Analizar las secuencias RADseq de *Atelopus ignescens* comparándolas con el genoma de referencia de *Atelopus laetissimus*, mediante la utilización de la herramienta MInimap 2. 2) de ser el caso, determinar la presencia de genes conservados, en *Atelopus ignescens*, 3) identificar posibles variantes genéticas en *Atelopus ignescens* que potencialmente estén asociadas con la resistencia y la adaptación al entorno y 4) plantear estrategias de conservación, considerando tanto la diversidad genética actual como la capacidad evolutiva de la especie para enfrentar amenazas futuras.

2. METODOLOGÍA:

Este estudio sigue un diseño de investigación cuantitativo, exploratorio y comparativo orientado en la caracterización genética entre dos poblaciones una actual y otra histórica de la especie *Atelopus ignescens*. Las muestras tomadas de los individuos actuales fueron realizadas en la única población existente, en la localidad de Angamarca, sin embargo, la fecha de recolección no es un dato público. Por otro lado, las muestras históricas son provenientes de tejidos congelados de museo, recolectados a finales de 1980, conservados en la universidad de Kansas (Jaynes et al., 2022). Se utilizaron herramientas bioinformáticas en Galaxy y análisis estadísticos en R studio para la detección y comparación de variantes genéticas a partir de secuencias RADseq. El enfoque metodológico realiza una prospección de anotación funcional de genes para extrapolar potenciales implicaciones biológicas de la supervivencia de la especie.

2.1. Obtención y preparación de datos:

2.1.1. Secuencias del Centro Nacional para la Información Biotecnológica (NCBI):

Se accedió a la base de datos del NCBI y se obtuvo 10 secuencias RADseq (SRR21124650 hasta SRR21124660) del Bioproyecto: “Amphibian rediscoveries provide insights into species persistence” de *Atelopus ignescens*, 2 secuencias históricas (SRR21124661 y SRR21124662) del año 1990 que son correspondientes a muestras

preservadas en museos del Instituto de Biodiversidad de la Universidad de Kansas (Tabla 1), y por último se descargó el genoma completo de 1 especie del mismo género *Atelopus* disponible que corresponde a la especie *Atelopus laetissimus* (Afgan et al., 2018; Baird et al., 2008; Davey & Blaxter, 2011; Jaynes et al., 2022).

Tabla 1. Identificación de las muestras RADseq en NCBI del Bioproyecto: “Amphibian rediscoveries provide insights into species persistence”

	Código SRR	Grupo	Enlace NCBI individual
Muestra 1	SRR21124650	Actuales	SRR21124650
Muestra 2	SRR21124651	Actuales	SRR21124651
Muestra 3	SRR21124652	Actuales	SRR21124652
Muestra 4	SRR21124653	Actuales	SRR21124653
Muestra 5	SRR21124654	Actuales	SRR21124654
Muestra 6	SRR21124655	Actuales	SRR21124655
Muestra 7	SRR21124657	Actuales	SRR21124657
Muestra 8	SRR21124658	Actuales	SRR21124658
Muestra 9	SRR21124659	Actuales	SRR21124659
Muestra 10	SRR21124660	Actuales	SRR21124660
Muestra histórica 1	SRR21124661	Históricas	SRR21124661
Muestra histórica 2	SRR21124662	Históricas	SRR21124662

2.1.2. Descarga de datos:

Se descargaron las secuencias en formato FASTA y FASTQ, estos formatos permitieron la manipulación bioinformática posterior (Fujita et al., 2021).

2.2. Análisis con Galaxy (Entorno bioinformático interactivo de código abierto) (Figura 1):

2.2.1. Preprocesamiento de secuencias:

Se verificó el control de calidad de las secuencias con la herramienta FastQC - Galaxy para asegurar la integridad y confiabilidad de los datos (Andrews, 2010; Hohenlohe et al., 2010).

2.2.2. Alineamiento con Map with minimap 2 - Galaxy:

2.2.2.1. Alineamiento:

Se alinearon las 10 secuencias RADseq de la población actual y las 2 secuencias RADseq históricas de *A. ignescens*, de manera independiente, con la secuencia del genoma de

referencia de *A. laetissimus* usando la herramienta Map with minimap 2 – Galaxy (Afgan, et al., 2018; Li, 2018). Esta herramienta es eficiente para secuencias RADseq, ya que permite manejar secuencias restringidas (Afgan, et al., 2018).

2.2.2.2.Ordenamiento de Alineamientos:

Al finalizar el proceso previo (sección 2.2.2.1) se generaron los archivos BAM (Binary Alignment Map) de cada secuencia RADseq los cuales fueron ordenados con la herramienta Samtools sort - Galaxy (Afgan, et al., 2018; Nabeel, 2023). Esta herramienta permitió organizar las lecturas alineadas de acuerdo con sus coordenadas genómicas, en el genoma de referencia, lo cual es un paso indispensable para los análisis de variantes (Li et al., 2009; Nabeel, 2023).

2.2.2.3.Verificación de la calidad del alineamiento:

Los archivos generados anteriormente sirvieron de entrada para la herramienta samtools stats- Galaxy la cual evaluó la calidad de los alineamientos obteniendo métricas como: calidad de lecturas mapeadas, error rate, tasa de mismatches y cobertura (Afgan, et al., 2018; Li et al., 2009).

2.2.2.4.Filtrado de alineamientos:

Los archivos de la sección 2.2.2.2 se colocaron como entrada en la herramienta samtools view – Galaxy, para conservar únicamente las lecturas alineadas, eliminando las secuencias no alineadas y las duplicadas (Afgan, et al., 2018). Finalmente, se volvió a verificar la calidad de los alineamientos filtrados con la herramienta de la sección 2.2.2.3 y se volvió a verificar el ordenamiento de todos los archivos filtrados con la herramienta de la sección 2.2.2.2.

2.2.3. Llamado de Variantes Genéticas con bcftools – Galaxy

2.2.3.1.Generación de pilas de alineamiento:

Se usaron los archivos BAM generados en la sección 2.2.2.4 (archivos sometidos a ordenamiento), los cuales fueron utilizados como entrada en la herramienta bcftools mpileup - Galaxy (Afgan, et al., 2018; Li, 2011). Esta herramienta calcula la información de cobertura por posición genómica y permite evaluar la calidad de los alineamientos y lecturas para cada base usando el genoma de referencia (Danecek, 2021).

2.2.3.2.Llamado de variantes:

Con los archivos BCF (Binary Call Format) generado en la sección anterior, se usó la herramienta bcftools call – Galaxy que aplica modelos de probabilidad para detectar posiciones polimórficas, diferenciando entre variantes homocigotas y heterocigotas (Danecek, 2021). Generando los archivos VCF (Variant Call Format) (Afgan, et al., 2018; Li, 2011).

2.2.3.3.Filtrado de variantes:

Después del llamado de las variantes, se utilizó la herramienta bcftools filter – Galaxy, la cual selecciona únicamente las variantes con buena calidad, valores de profundidad aceptables y una baja probabilidad de error (Afgan et al., 2018).

Aquí se tomó en cuenta en los parámetros: Targets – Include con el comando (%QUAL \geq 30 & FORMAT/DP \geq 10), esto quiere decir que: La calidad debe ser \geq 30 (este valor 30 genera un error de 1 en 1000 generando un 99.9% de confianza) y una profundidad por muestra que debe ser \geq 10, esto garantiza que las variantes se aproximen a la realidad (Danecek, 2021). Posteriormente a este filtrado los 10 archivos actuales y los 2 históricos se compilaron en un solo archivo utilizando la herramienta Bcftools merge- Galaxy (Afgan et al., 2018, Danecek, 2021).

2.2.4. Anotación del genoma de referencia con BREAKER 3 – Galaxy

Esta sección nos sirvió para preparar los datos para la anotación funcional. *Atelopus ignescens* no posee una base de datos de un genoma anotado disponible, por tal motivo, se creó una anotación funcional para el genoma completo de *A. laetissimus*, para esto se usó la herramienta BRAKER3, la cual predijo estructuras genéticas de novo, a partir de secuencias genómicas y modelos de predicción como Genemark y AUGUSTUS, y generó un archivo GFF3 el cual fue utilizado en el siguiente paso (Brúna, et al, 2021; Cingolani et al., 2012).

2.2.5. Anotación funcional de variantes con bcftools csq - Galaxy

En esta sección se intentó predecir las consecuencias funcionales de las variantes detectadas en las secuencias RADseq de *Atelopus ignescens*, para esto se utilizó la herramienta bcftools csq- Galaxy, que permitió asociar variantes genéticas a efectos

funcionales sobre genes anotados, como variantes sinónimas, no sinónimas, de cambio de marco de lectura o variantes intergénicas (Danecek, 2021; Afgan, et al., 2018; Cingolani et al., 2012; Blum et al., 2021). Los archivos usados fueron: el archivo de la sección 2.2.3.3, el archivo de la sección 2.2.4 y el genoma de referencia de *A. laetissimus*.



Figura 1. Pipeline de la obtención y preparación de datos en Galaxy para el análisis genómico de las secuencias.

2.3. Análisis de genómica comparativa (R - studio) (Figura 2):

2.3.1. Preparación de datos y análisis estadístico:

Se empleó el entorno R Studio (v4.4.2) (R Core Team, 2021), específicamente el ecosistema Bioconductor (Huber et al., 2015) con sus paquetes: vcfR (Knaus & Grünwald, 2017), Tidyverse (Wickham et al., 2019), ggplot2 (Wickham, 2016), dplyr (Wickham et al., 2023), tidyr (Wickham et al., 2024), adegenet (Jombart, 2008), los cuales se usaron para analizar muestras históricas y actuales de *A. ignescens* mediante un análisis comparativo, de variantes exclusivas y compartidas con el fin de ver patrones de diferenciación .

2.3.2. Carga de archivos VCF comprimidos:

Se utilizó el archivo VCF obtenido en Galaxy en la sección 2.2.3.3 y su índice, los cuales se descargaron en una versión comprimida (.vcf_bgzip) y en formato .tbi respectivamente. Estos fueron leídos en R studio por la función *read.vcfR* del paquete *vcfR* (Knaus & Grünwald, 2017). Esta herramienta permite leer y manipular los archivos Galaxy.

2.3.3. Comparación de variantes genéticas:

Utilizando el archivo de la sección 2.3.2 y el paquete *vcfR*, se extrajo una tabla de variantes del VCF mediante la función *extract.gt*. Esta función permite extraer objetos o campos obligatorios generados que contienen información de cromosomas o contigs (CHROM), la posición de la variante en el cromosoma (POS). Luego, se construyó una clave única (CHROM_POS) para representar de manera específica cada posición variante en el genoma de referencia (Knaus & Grünwald, 2017).

Se definieron dos subconjuntos de columnas correspondientes a las muestras actuales e históricas, esta estructura permitió comparar los datos posición por posición en los dos grupos. Para cada posición, se evaluó si al menos una muestra del grupo de datos tenía un genotipo válido distinto de "./." (indicador de genotipo ausente o no determinado). Esta evaluación se hizo usando la función *apply* para iterar sobre filas de la matriz de genotipos y se aplicó los siguientes criterios: Variante compartida (la posición tiene un genotipo válido en ambos grupos), variante exclusiva de actuales (la posición tiene genotipo válido en al menos una muestra actual y ninguno en las históricas) y variante exclusiva de históricas (la posición tiene genotipo válido en al menos una muestra histórica y ninguno en las actuales), las posiciones que no presentaban información fueron descartadas del análisis (Wickham et al., 2019; Wickham et al., 2023; Wickham et al., 2024). Cada posición fue clasificada en una de las tres categorías principales. Se generó un resumen con el número total de variantes en cada categoría, el cual fue visualizado posteriormente mediante un gráfico de barras con *ggplot2* (Wickham, 2016).

2.3.4. Clasificación y conteo de variantes genéticas por tipo SNP (Polimorfismo de un solo nucleótido) vs INDEL (inserciones o deleciones):

Se caracterizaron las variantes genéticas según SNPs e INDELS para explorar patrones de mutación y proporción de variabilidad estructural en cada grupo (Wickham et al.,

2019). Para esto se usó los paquetes `dplyr` y `vcfR`. Este procedimiento utilizó el mismo conjunto de variantes descrito en la sección 2.3.3. Para clasificar cada variante, se evaluó la longitud de los alelos de referencia (REF) y alternativos (ALT) extraídos del VCF: si ambos tenían longitud de un solo nucleótido, se clasificó como SNP; en caso contrario, se consideró un INDEL (Knaus & Grünwald, 2017; Wickham et al., 2023).

Posteriormente, se analizó la presencia o ausencia de estas variantes en cada grupo poblacional (actuales e históricas), utilizando la matriz de genotipos (GT). Se consideró que una variante estaba presente en un grupo si al menos una muestra de ese grupo tenía un genotipo válido (distinto de "."). Las variantes fueron clasificadas en tres categorías según su presencia: Exclusivas de actuales e históricas, y compartidas. Una vez definidas estas categorías, se realizó un conteo de SNPs e INDELS por grupo, y se calculó su frecuencia relativa (%) dentro de cada categoría (Wickham et al., 2023). Finalmente, los resultados fueron visualizados mediante un gráfico de barras con proporciones, utilizando el paquete `ggplot2` (Wickham et al., 2016).

2.3.5. Visualización y conteo de contigs más variantes mediante mapa de calor:

Para identificar las regiones genómicas más variables entre poblaciones, se construyó un mapa de calor centrado en los 20 contigs con mayor densidad de variantes. Se utilizó el paquete `vcfR` para importar y manipular los datos genómicos, y `dplyr` para agrupar y contar las variantes por contig (CHROM) en cada uno de los tres grupos: actuales, históricas y compartidas. Posteriormente, se utilizó la función `slice_max` ($n = 20$) para seleccionar los contigs con el mayor número total de variantes, estos fueron utilizados para filtrar y organizar los datos. Finalmente, se utilizó `ggplot2` para construir el mapa de calor, representando los contigs en el eje y, los grupos en el eje x y la intensidad de color en función del número de variantes por contig (Wickham, 2016).

2.3.6. Estimación de heterocigosidad observada por cada muestra y por grupo:

Se determinó la heterocigosidad por muestra con el fin de evaluar los niveles de diversidad genética entre las muestras históricas y actuales de *Atelopus ignescens* (Allendorf et al., 2013). Para esto se usaron los archivos VCF y los paquetes `vcfR` y

adegenet. El archivo VCF combinado se transformó en un objeto *genind* usando la función *vcfR2genind* (Jombart, 2008).

Se estimó la heterocigosidad observada (H_o) como la proporción de loci heterocigotos por individuo. Debido al gran número de loci (aproximadamente 961,000), se realizó un muestreo aleatorio de 10,000 loci por repetición durante 50 repeticiones para obtener estimaciones estables. En cada repetición: se seleccionaron aleatoriamente 10,000 loci, se eliminaron loci monomórficos, se calculó H_o con la función *Hs* del paquete *adegenet* (Jombart, 2008). y finalmente se promediaron los valores por individuo y grupo utilizando el paquete *stats* (R Core Team, 2021). Se utilizó la prueba de suma de rangos de Wilcoxon para comparar la heterocigosidad observada entre las dos poblaciones, debido a que los datos no cumplían con supuestos de normalidad y los tamaños muestrales eran desiguales (10 a 2). También se generó un gráfico de caja (boxplot) que muestra la distribución de H_o por grupo y por individuo con sus desviaciones estándar, utilizando el paquete *ggplot2* (Frankham et al., 2010; Wickham, 2016; R Core Team, 2021).

2.3.7. Estructura y estabilidad genotípica entre grupos:

Se extrajeron matrices de genotipo con la función *extract.gt* para obtener la matriz de genotipos (GT), en la que cada posición está codificada como "0/0" (Homocigoto referencia), "0/1" (Heterocigoto), "1/1" (Homocigoto alternativo) o "." (Genotipo ausente). Se eliminaron posiciones sin genotipos válidos (NA o ".") en al menos uno de los grupos para evitar sesgos por datos faltantes. Para cada posición con datos válidos, se compararon los genotipos históricos vs. actuales y se clasificaron en una de las siguientes categorías: "Igual": si todos los genotipos coinciden entre grupos, "Hetero en históricos, Homo en actuales", "Hetero en actuales, Homo en históricos", "Otro": si se encuentran combinaciones diferentes o conflictivas (Knaus & Grünwald, 2017). Finalmente, se calculó el número y porcentaje de variantes por categoría y se representaron en un gráfico de barras horizontales (Wickham, 2016).

2.3.8. PCA (análisis de componentes principales)

Para identificar patrones de estructura genética entre las poblaciones actuales e históricas, se realizó un análisis de componentes principales (PCA) a partir de los genotipos (GT) de las variantes filtradas. Esto permite reducir la dimensionalidad de datos multivariantes y detectar agrupamientos naturales sin requerir modelos poblacionales previos (Jombart et

al., 2010). Se utilizó el paquete *adegenet* para transformar el objeto *genind* utilizado anteriormente en *genlight*, el cual optimiza la eficiencia computacional en matrices grandes (Jombart, 2008).

El análisis PCA se llevó a cabo con la función *glPca* de *adegenet*, solicitando los dos primeros ejes principales ($nf=2$). Para visualizar los resultados, se utilizó *ggplot2*, destacando los grupos poblacionales y los porcentajes de varianza explicada por PC1 y PC2. Adicionalmente, se calculó la varianza acumulada de todos los componentes principales para determinar cuántos ejes eran necesarios para representar al menos el 80% de la variabilidad genética (Jombart, 2008; Jombart et al., 2010).

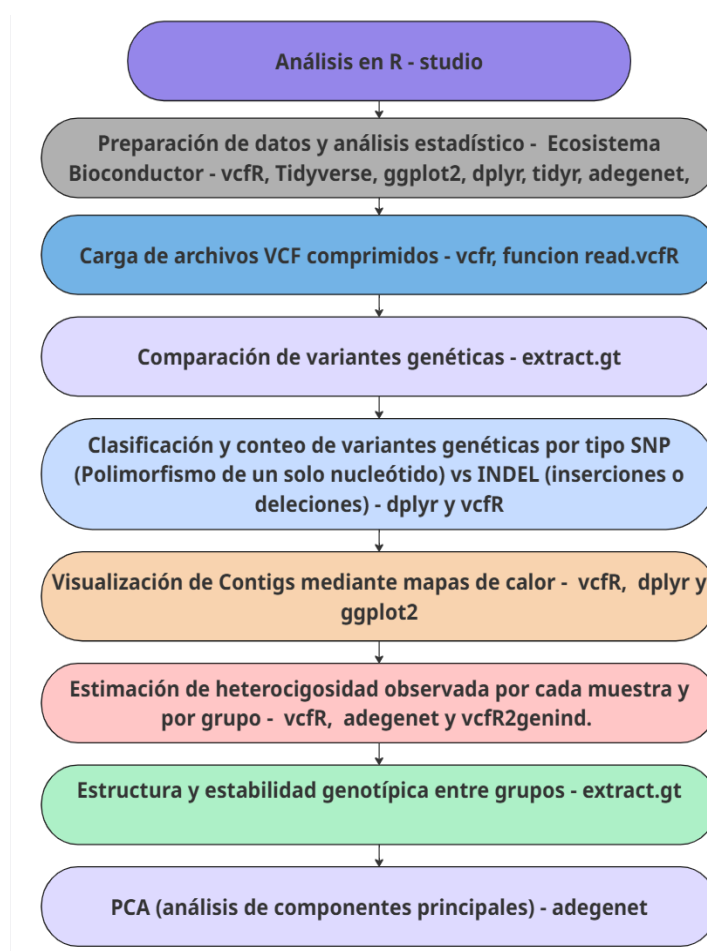


Figura 2. Pipeline del análisis de genómica comparativa realizado en R-studio a partir de los archivos obtenidos en Galaxy

2.4. Visualización Genómica de Variantes en JBrowse y Anotaciones funcionales en UNIPROT

Para poder observar las variantes detectadas y su ubicación en relación con las regiones anotadas del genoma de referencia, se empleó la herramienta JBrowse, la cual es un

navegador genómico interactivo utilizado para la exploración gráfica (Buels, et al., 2016; Skinner et al., 2009).

Se cargaron los siguientes archivos de entrada: El archivo VCF de la sección 2.2.3.3., el archivo GFF3 de la sección 2.2.4 y el archivo FASTA del genoma de referencia.

Para la anotación funcional posicional se utilizó en primera instancia el programa R studio, usado anteriormente, para determinar el número y la ubicación de las variantes dentro del genoma anotado de *A. laetissimus* (R Core Team, 2021). Para ello, se emplearon los paquetes *vcfR*, *GenomicRanges* (Lawrence et al., 2013), *rtracklayer* (Lawrence et al., 2009), *dplyr* (Wickham et al., 2023) y *ggplot2* (Wickham, 2016). Inicialmente, se evaluó la distribución de las variantes identificadas en el archivo VCF, de la sección 2.2.5, sobre distintas características genómicas anotadas, tales como gen, mRNA y CDS, utilizando el archivo GFF3 de la sección 2.2.4.

El archivo VCF fue convertido a un objeto *GRanges*, mientras que el archivo GFF3 fue leído con la función *import* del paquete *rtracklayer*. A continuación, se identificó la superposición entre las variantes genéticas y las características anotadas, usando la función *findOverlaps*. Se identificaron cuántas variantes afectaban regiones CDS y se agruparon según sus transcritos y genes parentales, extrayendo los metadatos de la columna Parent del GFF3 para establecer la jerarquía gen–transcrito–CDS (Lawrence et al., 2009; Lawrence et al., 2013). Posteriormente, se construyeron visualizaciones en forma de gráfico de barras horizontales con *ggplot2*, de los genes con mayor cantidad de regiones codificantes afectadas por variantes (Wickham, 2016).

Posteriormente se tomó los cuatro genes que presentaban mayor cantidad de variantes en sus CDS y mediante el paquete *dplyr* se obtuvo la ubicación del CDS dentro del gen (Lawrence et al., 2013). Finalmente, se extrajo la secuencia nucleotídica completa del CDS correspondiente y se ingresó manualmente en la herramienta BLAST (*blastx*) disponible en <https://www.uniprot.org/blast>. Esta búsqueda permitió identificar posibles homologías funcionales con proteínas caracterizadas en otras especies, dominios conservados, descripciones funcionales, y rutas metabólicas asociadas (UniProt Consortium, 2023).

3. RESULTADOS

3.1. Preprocesamiento de secuencias:

Las 12 secuencias encontradas para *A. ignescens* presentaron un tamaño de entre 140-150 pb, con una media de secuencias comprendido entre 8.1 ± 0.36 millones de lecturas. El resultado del control de calidad demostró que la mayoría de las muestras presentaron un elevado Phred score de calidad por base con un valor de $Q98.43 \pm 0.23\%$ sin presentar contenido de nucleótidos sin identificación. El contenido de GC (guanina-citosina) presentó valores con una media entre $47.87 \pm 0.11\%$ lo cual generó algunas advertencias leves en Galaxy para el resultado del FASTQC. También, se pudo verificar que 4 de las 12 secuencias presentaron duplicación de secuencias con una media de $45.09 \pm 2.09\%$. Adicionalmente, las secuencias presentaban, al inicio y final, presencia de adaptadores residuales que generaban algo de ruido, debido a que no son parte del ADN biológico. Para intentar mejorar las secuencias, se decidió cortar los extremos que generaban ruido, pero el trimado no generó ninguna mejoría en los valores de los parámetros antes mencionados.

3.2. Alineamiento con minimap 2 - Galaxy:

El resultado del alineamiento de las secuencias RADseq, actuales e históricas, de *A. ignescens* contra el genoma de referencia de *Atelopus laetissimus*, después de ser ordenados, generó 2 archivos BAM, uno por cada población, con los siguientes resultados: Número total de secuencias evaluadas 89,448,068 actuales y 10,857,838 históricas, lecturas mapeadas (número de secuencias alineadas correctamente al genoma de referencia) 41,087,784 actuales y 6,473,021 históricas con un porcentaje de mapeo de 45.9% y 59.6% respectivamente, longitud promedio de lecturas actuales e históricas 145 pb para ambas poblaciones, bases mapeadas (CIGAR) o bases con alineamientos confiables 5,602,641,656 actuales y 847,556,688 históricas, mismatches (número de bases que no coincidieron) 370,356,331 actuales y 56,518,771 históricas, tasa de error 6.61% actuales y 6.67% históricas en base a los mismatches/bases mapeadas y por último calidad promedio Phred score para las dos poblaciones fue 255.

Con el afán de mejorar aún más estos resultados se generó un filtrado de los archivos BAM que arrojó los siguientes resultados: Número total de secuencias evaluadas

13,830,048 en actuales y 2,014,730 en históricas, número de lecturas mapeadas 13,830,048 en actuales y 2,014,730 históricas en un porcentaje de mapeo del 100% para ambas poblaciones, longitud promedio de lecturas actuales e históricas de 145 pb para las dos poblaciones, bases mapeadas (CIGAR) 1,922,873,027 actuales y 271,006,849 históricas, mismatches 119,893,616 en actuales y 16,296,334 en históricas, tasa de error 6.24% actuales y 6.01% históricas en base a los mismatches/bases mapeadas y por último calidad promedio Phred para las dos poblaciones fue 255.

3.3. Análisis de genómica comparativa en R:

3.3.1. Comparación de variantes genéticas entre muestras actuales e históricas de *A. ignescens*:

Del total de las variantes analizadas en ambos grupos de datos, se determinó que 743,331 variantes fueron exclusivas del grupo de muestras actuales, lo que representa la mayor proporción del total analizado. Por otro lado, se identificaron 164,675 variantes compartidas entre las muestras actuales e históricas. Finalmente, 53,596 variantes fueron exclusivas de las muestras históricas, reflejando señales potenciales de pérdida o reemplazo genético a lo largo del tiempo (Figura 3).

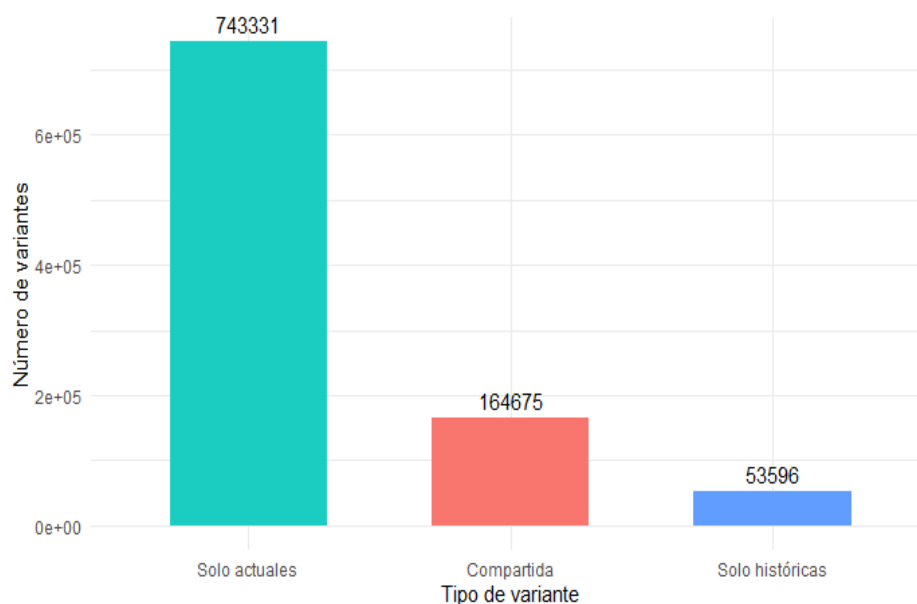


Figura 3. Comparación de variantes genéticas entre muestras actuales e históricas de *Atelopus ignescens*.

3.3.2. Clasificación y conteo de variantes genéticas por tipo SNP (Polimorfismo de un solo nucleótido) vs INDEL (inserciones o deleciones):

Del total de variantes encontradas, se determinó que en las muestras actuales 87,010 variantes eran indels (11.71%) y 656,321 variantes eran SNPs (88.29%), en las muestras históricas se encontraron 4,885 indels (7.99%) y 49,311 SNPs (92.01%). Finalmente se observó que de todas las variantes compartidas entre ambos grupos 45,463 (27.61%) son indels y 119,212 (72.39%) son SNPs (Tabla 2, Figura 4).

Tabla 2. Clasificación y conteo de variantes genéticas por tipo SNP (Polimorfismo de un solo nucleótido) vs INDEL (inserciones o deleciones):

Grupo	Tipo	N	Porcentaje
Actuales	INDEL	87010	11.71
Actuales	SNP	656321	88.29
Compartida	INDEL	45463	27.61
Compartida	SNP	119212	72.39
Históricas	INDEL	4285	7.99
Históricas	SNP	49311	92.01

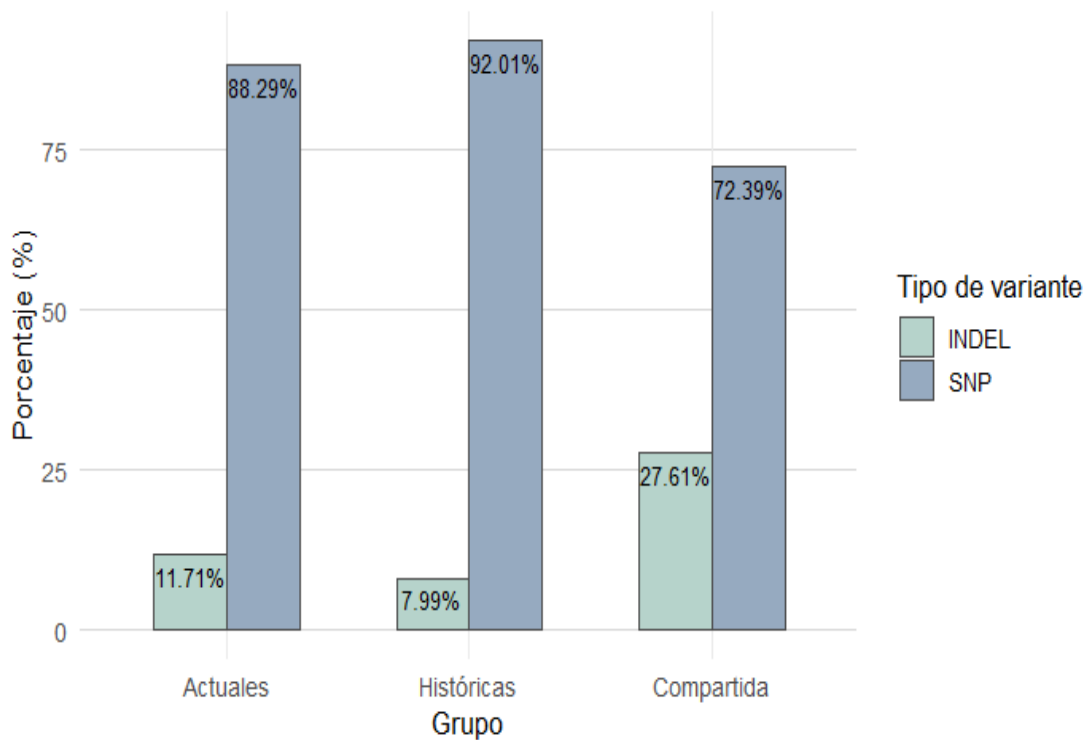


Figura 4. Proporción de variantes genéticas por grupo y por tipo SNP (Polimorfismo de un solo nucleótido) vs INDEL (inserciones o deleciones).

3.3.3. Visualización y conteo de contigs más variantes mediante mapa de calor:

El análisis de las 20 regiones genómicas (contigs), con mayor acumulación de variantes, evidenció una distribución claramente sesgada hacia las muestras actuales. El contig con mayor número de variantes fue JBGDNC010000009.1, con 27,670 variantes exclusivas del grupo actuales, 6,303 variantes compartidas, y 1,944 variantes exclusivas de históricas. Este patrón se repitió a lo largo de los 20 contigs, donde en todos los casos las variantes exclusivas de las muestras actuales superan tanto a las compartidas como a las exclusivas de históricas (Tabla 3). La visualización mediante el mapa de calor refuerza este patrón, mostrando una mayor intensidad de color en el grupo actuales, especialmente en los primeros contigs, lo que indica una mayor carga de variación genética reciente (Figura 5).

Tabla 3. Cantidad de variantes genéticas: variantes exclusivas actuales e históricas y variantes compartidas, de los 20 contigs con más variantes.

	Contig	Actuales	Compartidas	Históricas
1	JBGDNC010000009.1	27670	6303	1944
2	JBGDNC010000027.1	21924	4823	1583
3	JBGDNC010000067.1	19739	4422	1492
4	JBGDNC010000052.1	15123	3550	998
5	JBGDNC010000024.1	14674	3339	1096
6	JBGDNC010000013.1	14659	3264	1011
7	JBGDNC010000107.1	14261	3118	995
8	JBGDNC010000022.1	13955	2891	1080
9	JBGDNC010000001.1	13761	2942	1029
10	JBGDNC010000036.1	13056	2869	916
11	JBGDNC010000085.1	13082	2676	809
12	JBGDNC010000120.1	12395	2606	931
13	JBGDNC010000159.1	11818	2651	789
14	JBGDNC010000126.1	11638	2375	902
15	JBGDNC010000041.1	11425	2283	914
16	JBGDNC010000042.1	11323	2371	840
17	JBGDNC010000057.1	11081	2660	704
18	JBGDNC010000048.1	11090	2207	673
19	JBGDNC010000031.1	10880	2303	766
20	JBGDNC010000010.1	10514	2509	765

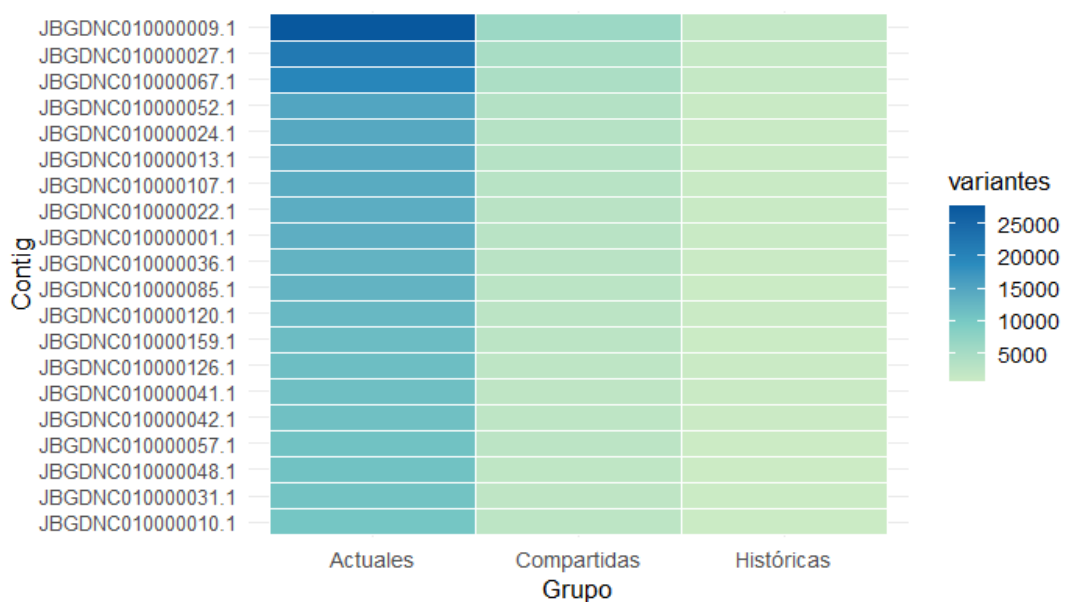


Figura 5. Mapa de calor de la distribución de variantes por grupo en los 20 contigs con mayor acumulación de variantes.

3.3.4. Heterocigosidad observada en las muestras de *A. ignescens*:

En la tabla 4 se puede observar la heterocigosidad observada y la desviación estándar por cada individuo, se denota que la mayoría de muestras pertenecientes a la población actual presentan un promedio de heterocigosidad mayor que los de la población histórica, siendo la muestra SRR21124657 (actuales) la que presenta la media de heterocigosidad más alta (0.296 ± 0.011), y la muestra SRR21124650 (actuales) la que presenta la media de heterocigosidad más baja (0.0123 ± 0.002). Las muestras históricas de Chaucán y Pachancho presentaron medias de heterocigosidad bastante similares entre sí (0.116 ± 0.008 y 0.115 ± 0.009 respectivamente) (Figura 6).

Tabla 4: Heterocigosidad observada (Ho) por muestra de *A. ignescens*

	Muestra	Grupo	Ho_media	Ho_sd
1	SRR21124650	Actuales	0.01231197	0.002653324
2	SRR21124651	Actuales	0.20820860	0.010948965
3	SRR21124652	Actuales	0.12056463	0.009648251
4	SRR21124653	Actuales	0.11174294	0.008700990
5	SRR21124654	Actuales	0.25399129	0.009727103
6	SRR21124655	Actuales	0.29102143	0.013571733
7	SRR21124657	Actuales	0.29603708	0.011325431
8	SRR21124658	Actuales	0.21106027	0.009283284

9	SRR21124659	Actuales	0.09701055	0.007547557
10	SRR21124660	Actuales	0.01574684	0.003106209
11	SRR21124661_chaucan	Históricas	0.11609906	0.007735649
12	SRR21124662_pachancho	Históricas	0.11474558	0.009331440

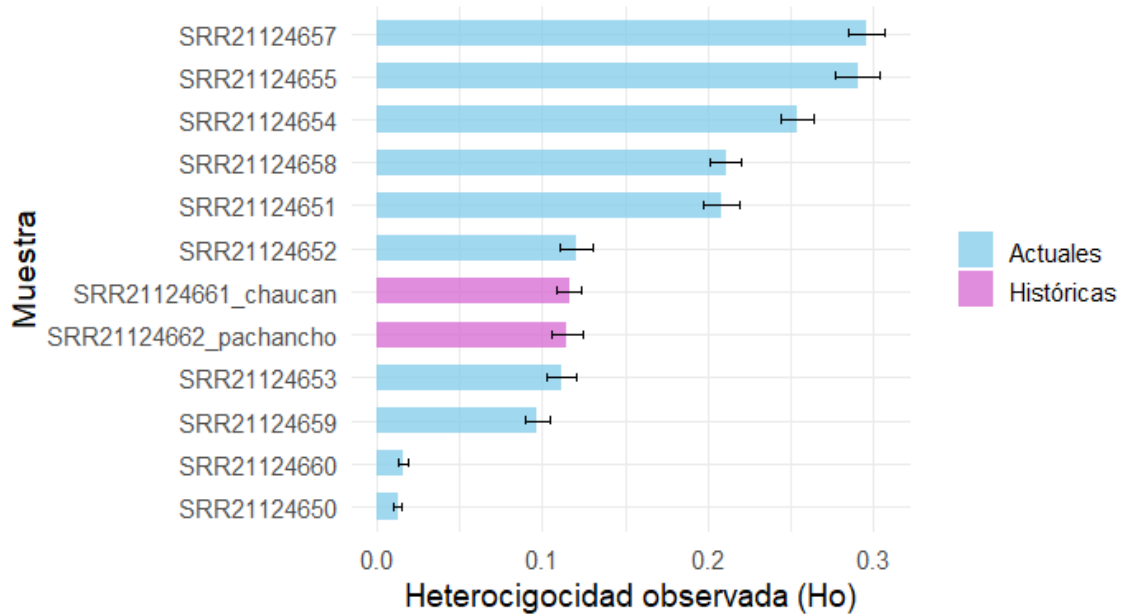


Figura 6. Heterocigosidad observada (Ho) por cada individuo de *Atelopus ignescens*. En color celeste se encuentran las muestras actuales y en color rosa se encuentran las muestras históricas. Las barras corresponden a la desviación estándar (sd) de cada individuo.

Al analizar las medias por grupo se obtuvo una media de 0.162 ± 0.105 para el grupo de muestras actuales y una media de 0.115 ± 0.0009 para las muestras históricas. El boxplot refleja que la población actual tiene mayor variabilidad en Ho, en comparación con la población histórica (Tabla 5; Figura 7). Sin embargo, el test de Wilcoxon arrojó un p-valor de 0.758 al comparar ambos grupos, lo que indica que las dos poblaciones no difieren significativamente en términos de diversidad genotípica.

Tabla 5. Heterocigosidad observada (Ho) por grupo actuales e históricas de *A. ignescens*

Grupo	N°	Media Ho	Mediana Ho	Sd Ho	Min Ho	Max Ho
Actuales	10	0.162	0.164	0.1053	0.0123	0.296
Históricas	2	0.115	0.115	0.0009	0.115	0.116

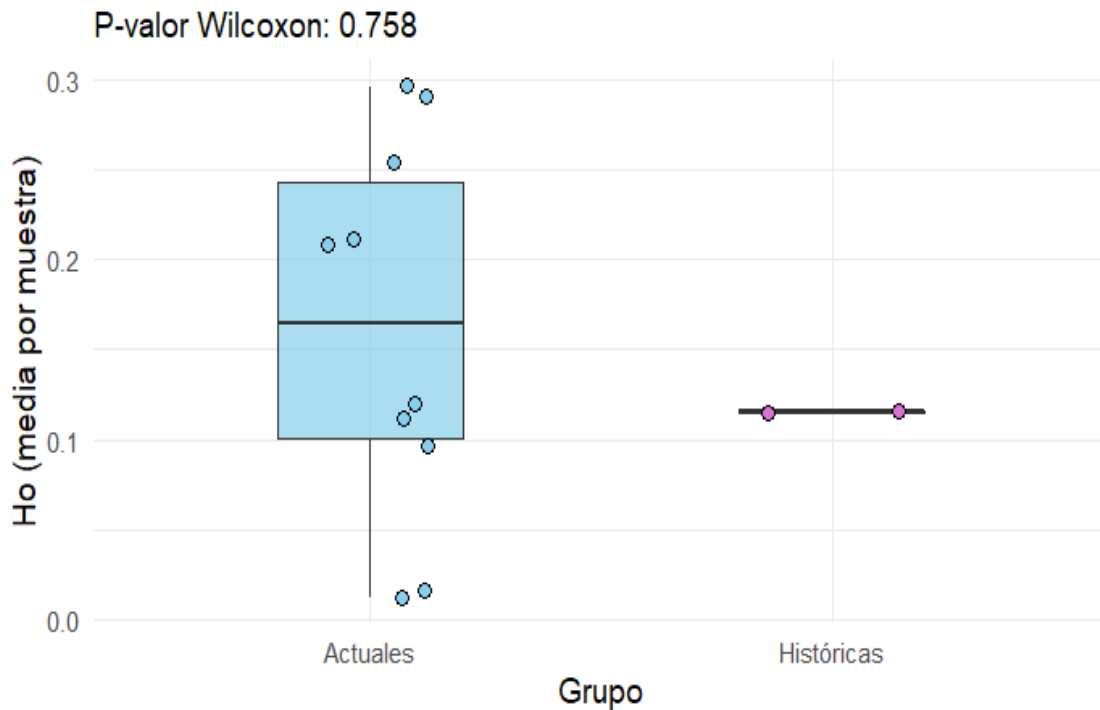


Figura 7. Heterocigosidad observada (H_o) por grupo actuales e históricos. En el grupo actual el primer cuartil (Q1) ubicado en la parte inferior de la caja (~ 0.0970) indica que el 25% de los datos se encuentra por debajo de este valor. La línea media corresponde a la mediana de los datos (~ 0.164) y representa al segundo cuartil (Q2). El tercer cuartil (Q3) ubicado en la parte superior de la caja (~ 0.2539) indica que el 75% de los datos se encuentran por debajo de este valor. Los bigotes representan los límites superiores e inferiores de los datos. En el grupo de muestras históricas no se pudo determinar cuartiles ni bigotes debido a la poca cantidad de datos.

3.3.5. Estructura y estabilidad genotípica entre grupos:

Se compararon un total de 164,675 posiciones genotipadas entre las muestras actuales e históricas. Del total de sitios analizados: 133,035 variantes (80.79%) presentaron genotipos idénticos entre ambas poblaciones. En 12,632 posiciones (7.67%), los individuos actuales mostraron un genotipo heterocigoto, mientras que los históricos fueron homocigotos. En 6,691 posiciones (4.06%), ocurrió el patrón inverso: heterocigosidad en históricos y homocigosidad en actuales. Finalmente, 12,317 variantes (7.48%) se clasificaron en la categoría "Otro", incluyendo combinaciones complejas o discrepantes (como homocigotos diferentes o presencia de datos faltantes) (Tabla 6; Figura 8).

Tabla 6. Comparación de genotipos entre muestras actuales e históricas

Clasificación	N	Porcentaje
Hetero en actual, Homo en histórico	12632	7.67
Hetero en histórico, Homo en actual	6691	4.06
Igual	133035	80.79
Otro	12317	7.48
TOTAL	164675	100

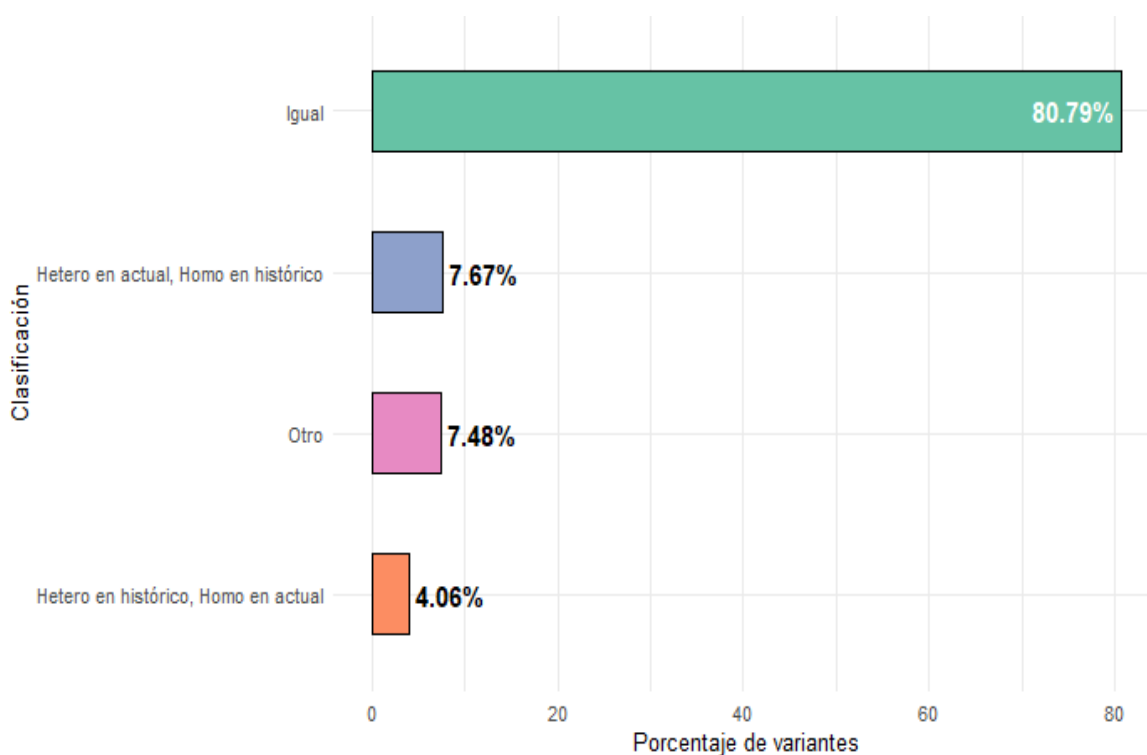


Figura 8. Comparación genotípica entre muestras actuales e históricas de *A. ignescens* en diferentes loci.

3.3.6. PCA (análisis de componentes principales):

El análisis de componentes principales permitió evidenciar que las muestras históricas de *A. ignescens* presentan una leve separación respecto a las actuales en el espacio de PC1 y PC2, aunque con cierto grado de solapamiento (Figura 9). La varianza explicada acumulada por los primeros 7 componentes sugiere una diversidad genética distribuida en múltiples dimensiones (Tabla 7).

Tabla 7: Lista de componentes que explican el 80% de la variación genética de las muestras.

	PC	Prop varianza	Varianza acumulada
1	PC1	0.2318	0.2318
2	PC2	0.1427	0.3745
3	PC3	0.1278	0.5023
4	PC4	0.1115	0.6138
5	PC5	0.0975	0.7113
6	PC6	0.0730	0.7843
7	PC7	0.0648	0.8491

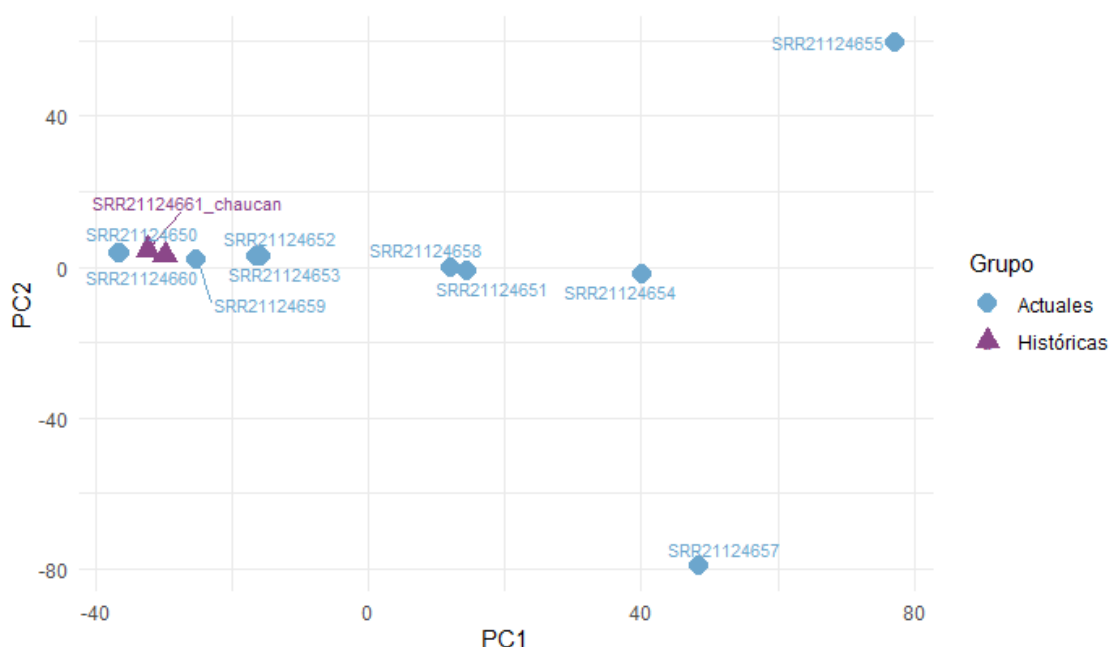


Figura 9. PCA genotípico de las muestras actuales e históricas de *A. ignescens*.

3.4. Anotación funcional de *A. ignescens* y visualización de genes variantes:

En base a la anotación funcional realizada a partir del genoma de *A. laetissimus*, se encontró que *A. ignescens* presenta 3,431 variantes que corresponden a mRNA, 2,569 variantes se encuentran en alguna posición de un locus de un gen y 1,451 variantes que corresponden a regiones codificantes de los genes (CDS). Dentro de los CDS mayormente afectados por variantes se encuentran el CDS 15 y 16 del gen 1407 con 39 variantes entre ambos (Figuras 10 y 11), en UNIPROT se constató que estas regiones podrían estar codificando isoformas de la proteína con dominio de transcriptasa inversa, característica de retroelementos o de la proteína con dominio de endonucleasa/exonucleasa/fosfatasa.

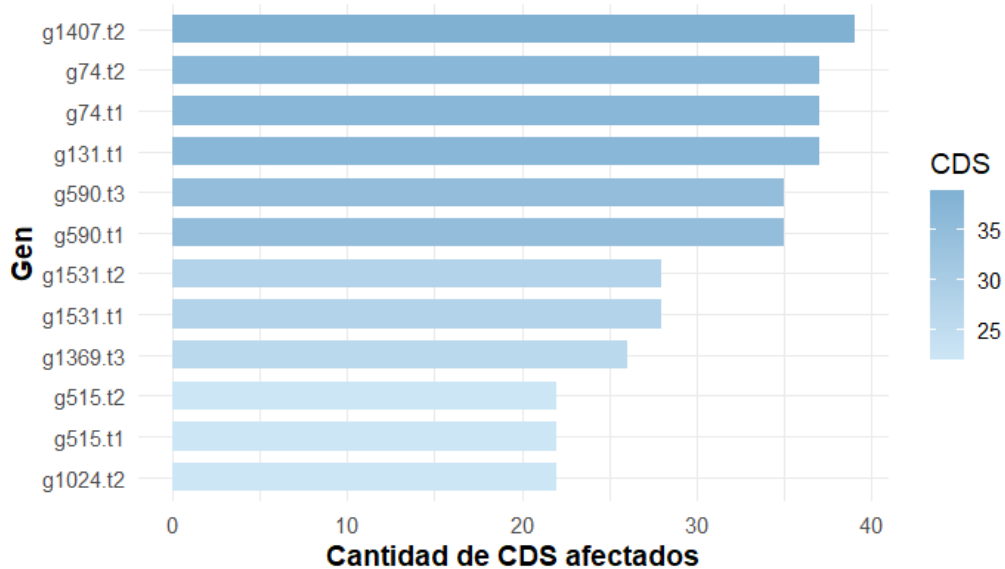


Figura 10. Genes de *A. ignescens* con mayor cantidad de variantes que caen en regiones codificantes (CDS)

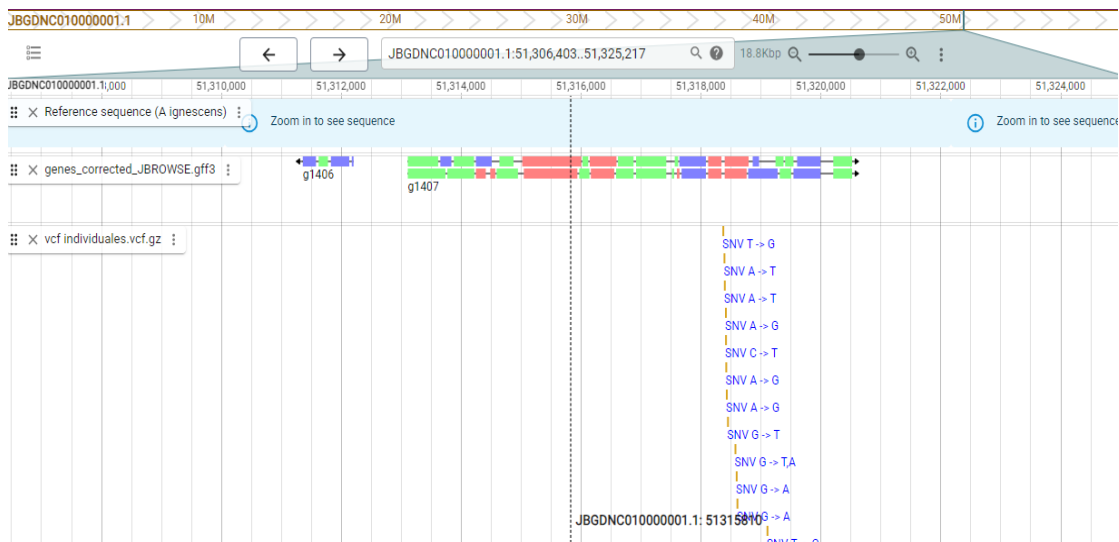


Figura 11. Visualización de la región del gen 1407 en Jbrowse, en donde se logran apreciar algunas de las variantes genéticas principalmente SNPs.

En segundo y tercer lugar se encontró al CDS 1 del gen 74 transcripto 1 y el CDS 3 del gen 74 transcripto 2 con 37 variantes cada uno (Figuras 10 y 12), según UNIPROT estas regiones codifican la proteína FREM2, la cual contiene múltiples dominios Calx-beta involucrados en la adhesión epitelial y la formación de matriz extracelular durante el desarrollo.

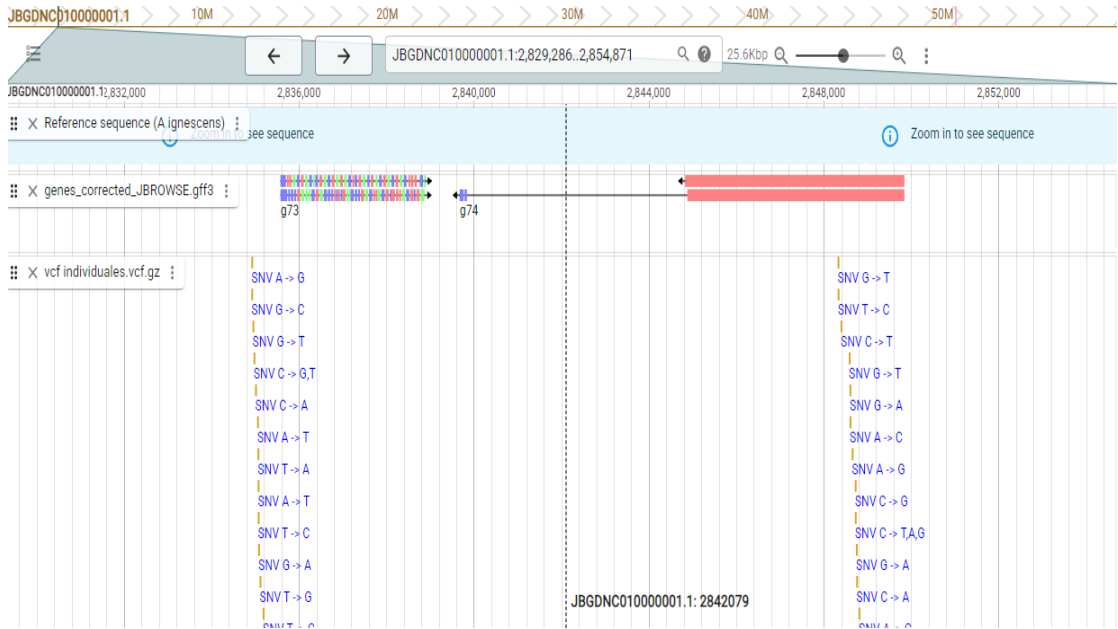


Figura 12. Visualización de la región del gen 74 en Jbrowse, en donde se logran apreciar algunas de las variantes genéticas principalmente

Por último, analizamos el cuarto lugar que contenía igualmente 37 variantes, el CDS 1 del gen 131 (Figuras 10 y 13), esta región podría codificar una proteína con dominio RRM derivada de un elemento transponible LINE-1 (L1). Este tipo de proteínas forma parte de retroelementos autónomos que pueden replicarse dentro del genoma.

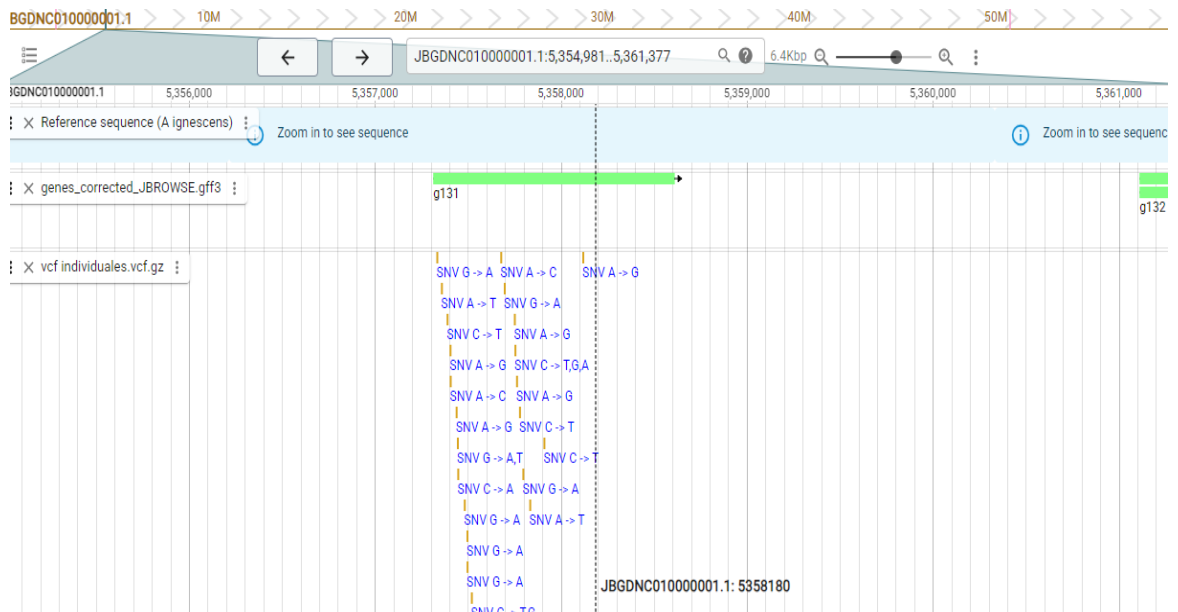


Figura 13. Visualización de la región del gen 131 en Jbrowse, en donde se logran apreciar algunas de las variantes genéticas principalmente SNPs.

4. DISCUSIÓN:

4.1. Validación del alineamiento de las secuencias RADseq de *A. ignescens* con el genoma de referencia de *A. laetissimus*:

Los análisis generados de las secuencias RADseq con la herramienta de verificación FASTQC demostraron que el tamaño de las secuencias concuerda con lo descrito por Baird y colaboradores (2008). Este estudio describe la técnica de secuenciación de ADN basada en sitios de restricción (RADseq) para la identificación rápida de SNPs en dos organismos modelo (*Gasterosteus aculeatus* y *Neurospora crassa*). Para esto Baird y colaboradores trabajaron con la plataforma de secuenciación Illumina, obteniendo secuencias con un tamaño aproximado de 300 a 700 pb, que al pasar por un proceso de trimado generan librerías con tamaños parecidos al que se obtuvieron en este estudio (140-150 pb).

Así mismo, el número de lecturas reportado en el presente estudio es esperado (8.1 ± 0.36 millones de lecturas), si lo comparamos con estudios previos de RADseq en donde se reportan valores de 1 millón de lecturas por muestra, demostrando que estos valores varían dependiendo de la cantidad de ADN y la eficacia de la técnica usada (Peterson et al., 2012), lo que se corrobora con el estudio de Ekblom y Galindo (2011) al realizar análisis de ecología molecular en poblaciones de especies no modelo. También, se pudo evidenciar que las secuencias usadas en este trabajo presentan una excelente calidad de lecturas por base ($Q98.43 \pm 0.23\%$) siendo este un buen comienzo para el estudio de mapeo genético, ensamblaje y alineamiento. Esta afirmación se encuentra respaldada por la literatura donde se menciona que el valor Phred score, que mide la probabilidad de que una base haya sido llamada incorrectamente, debe generalmente encontrarse entre valores $>Q20$ para considerarse secuencias aceptables y $>Q30$ que son secuencias recomendadas para análisis de variantes (Ewing & Green; 1998; Andrews, 2010).

Por otro lado, el resultado de la cantidad de GC ($47.87 \pm 0.11\%$) generó una advertencia leve, este resultado sugiere que nuestro contenido de GC difiere del modelo estadístico esperado para dichos nucleótidos, por ende, se asume una distribución aleatoria de bases, con un porcentaje del 25%. Sin embargo, al ser únicamente una advertencia leve no es un valor que represente problemas técnicos en la calidad de las secuencias utilizadas (Andrews, 2010).

La presencia de duplicación de secuencias ($45.09 \pm 2.09\%$) también es algo que se esperaba, puesto que estamos trabajando con un genoma reducido o incluso estas duplicaciones se podría atribuir al sobre-secuenciamiento de las RADseq (Kozarewa et al; 2009). Por último, se notó presencia de adaptadores residuales (secuencias que se generan de manera artificial y sirven para que el ADN se una al flow cell de las plataforma de secuenciación) que generalmente se mantienen en algún porcentaje de la secuencia, para no comprometer el genoma biológico, en el momento de la generación de las librerías RADseq, o se mantienen cuando los ciclos de la plataforma illumina se configuran automáticamente y las secuencias tienen tamaños menores a la configuración pre-establecida (Kircher et al., 2012). Por tales motivos, se determinó que las 12 secuencias en base a los parámetros de estudios anteriores se encontraban en condiciones aceptables para continuar con la fase de alineamiento.

Los resultados obtenidos en el proceso de alineamiento tanto en los datos en bruto como en los filtrados fueron consistentes y fiables. Esta afirmación se basa en la cantidad de lecturas obtenidas en este análisis pre-filtrado (41,087,784 actuales y 6,473,021 históricas) y post-filtrado (13,830,048 actuales y 2,014,730 históricas), lo que sugiere una adecuada cobertura para analizar variantes sobre todo en el grupo de muestras actuales. Tomando en cuenta que en estudios similares se han reportado incluso valores menores, con medias entre 1-3 millones por muestra. Demostrando que, a pesar de la falta de disponibilidad de un genoma de referencia propio de *A. ignescens*, el genoma completo de una especie perteneciente al mismo género como *A. laetissimus*, puede servir como referencia para un alineamiento confiable (Peterson et al., 2012; Rochette et al., 2019).

Los resultados obtenidos en el porcentaje de mapeo de este estudio (pre-filtrado 45.9% actuales y 59.6% históricas), tomando en cuenta que el alineamiento se realizó con una especie del mismo género, corroboran que el mapeo fue exitoso y aceptable para continuar con el resto de análisis. Dado que para que se considere un alineamiento exitoso el porcentaje de mapeo idóneo, por lo general, tienen que estar entre el 70% si se alinea con especies modelo y entre el 40 - 60% si se alinea con genomas de referencia de especies cercanas (Davey & Blaxter, 2011; Ekblom & Galindo, 2011),

Por otro lado, los valores obtenidos en mismatches y tasas de error (pre-filtrado 6.61% actuales y 6.67% históricas) también fueron esperados en este estudio puesto que no se

usó el genoma de referencia de la misma especie, como ya se mencionó anteriormente, por tal motivo, se realizó un filtrado del alineamiento para reducir el ruido de estas diferencias de genomas disminuyendo un poco estos valores (post-filtrado 6.24% actuales y 6.01% históricas) y que de esta manera la precisión del llamado de variantes aumente sobre todo en un grupo de datos tan reducido como es el grupo de las muestras históricas (Hohenlohe et al. 2010; Davey et al., 2013; Shafer et al., 2017).

4.2. Exploración y caracterización de las variantes genéticas de los grupos actuales e históricos:

Al realizar el análisis de las variantes genéticas llamadas (bcftools-call), se pudo evidenciar que el grupo de muestras actuales tienen un mayor número de variantes exclusivas (743,331) en comparación con las muestras históricas (53,596), esto sugiere que posiblemente existieron cambios genéticos a través del tiempo, sin embargo, esto no indica que necesariamente haya una diversificación real dentro de la especie, sino más bien podría ser únicamente una acumulación de mutaciones como respuesta a cambios ambientales, puesto que esta especie ha tenido un declive importante que la llevó al borde de la extinción e incluso se la consideró extinta en 1988 (Byrne et al., 2021; Ellegren, 2014; Ron et al., 2003).

Otra hipótesis que también podría explicar esta diferencia entre ambas poblaciones, en especial el número tan bajo de variantes exclusivas en las muestras históricas, podría ser la pérdida de variantes originales o ancestrales de la especie, que posiblemente ya se encontraban ausentes en los individuos históricos de finales de los 80, esta reducción pudo ser causada por eventos de cuello de botella, deriva génica, aumento de endogamia y reducción del potencial adaptativo que sufría la especie en ese entonces y que llevó a *A. ignescens* hasta casi su extinción (Frankham, 2005; Allendorf et al., 2013; Frankham et al., 2010; Ellengren, 2014).

A pesar de las diferencias encontradas entre ambas poblaciones, se identificó un elevado número de variantes compartidas (164,675), lo que sugiere que aún existen un gran número de variantes potencialmente conservadas a través del tiempo, que tal vez ya se encuentran fijadas en el genoma de la especie. Estas variantes podrían representar polimorfismos ancestrales estables, que han sido mantenidos a pesar de procesos como la deriva génica y que les ha ayudado a persistir pese a las presiones ambientales y

demográficas a las que se ha visto sometida la especie, como se ha podido evidenciar en estudios previos en otras especies de anuros (Premachandra et al.,2023; Trumbo et al.,2023).

Dentro de todas estas variantes genéticas, ya mencionadas, se identificó que en los tres grupos de variantes (exclusivas de individuos actuales e históricos y compartidas entre grupos), los SNPs (polimorfismo de un solo nucleótido) son las variantes con mayor porcentaje de presencia, dejando con un menor porcentaje a los indels (inserciones y deleciones) (Figura 4). Una razón que justifique esta mayor presencia de SNPs dentro de las secuencias sería la técnica utilizada durante la secuenciación, puesto que, en trabajos previos se menciona que técnicas RADseq son muy útiles para encontrar variantes pequeñas como los SNPs, debido a que esta técnica secuencia únicamente porciones pequeñas de todo el genoma, lo que reduce la probabilidad de capturar mutaciones de mayor tamaño (Uricaru et al., 2020).

Desde una perspectiva biológica, esta predominancia de SNPs podría estar relacionada no solo con las características propias de la técnica, sino también con la dinámica evolutiva de una especie. En este sentido, se cree que mutaciones puntuales, como los cambios de un solo nucleótido, pueden acumularse con mayor facilidad en el genoma de la especie, concordando con lo mencionado por Sherlock et al. (2024) en un estudio previo donde realiza un análisis genómico para la filogenia de cecílicos y verifica que los SNPs son mucho más comunes que los INDELS dentro de las poblaciones; especialmente tras eventos como el cuello de botella, donde se da una reducción drástica del tamaño poblacional, lo que favorece una pérdida de alelos raros a través del tiempo cambiando la composición genética de la población (Nei et al. 1975; Frankham, 2005).

Los resultados obtenidos en el conteo de contigs con más variantes aportan evidencia complementaria que respalda la idea de que las variantes más pequeñas (SNPs), identificadas anteriormente, son predominantes en las secuencias analizadas. Tanto el conteo como la visualización, por mapa de calor (Figura 5), denotan un claro patrón de acumulación de variantes en regiones reducidas del genoma (contigs) (Hohenlohe, 2010; Fritsche et al., 2020; Chen et al., 2024), sobre todo en las muestras actuales, lo cual no solo sería un resultado esperado por el uso de la técnica RADseq (Uricaru et al., 2020), porque de ser así se encontraría ese mismo patrón en los grupos de variantes históricas y

compartidas, sino también podría ser un posible fenómeno biológico, remarcando aún más la hipótesis de que los individuos actuales han sufrido ciertos procesos evolutivos como la acumulación de mutaciones puntuales tras eventos de cuello de botella, donde la población remanente amplificó variantes de nuevo origen debido a la pérdida de alelos ancestrales. Sin embargo, estas mutaciones o variantes de la población actual no se distribuyen aleatoriamente en el genoma, sino que se concentra en ciertas regiones, donde el mapa de calor es más encendido, esto podría tener ciertas implicaciones para comprender procesos adaptativos y de resiliencia genética de la población actual (Ørsted et al., 2019; Sparks et al., 2023).

Así mismo, los resultados de heterocigosidad observada, que mide la proporción de loci heterocigotos en una muestra y se basa en un rango donde, < 0.1 representa baja diversidad genética, entre 0.1 a 0.3 diversidad genética moderada y > 0.3 alta diversidad genética (Frankham et al., 2010), demuestran que hay una diferencia visible entre los individuos actuales e históricos (Figura 7), inclusive dentro del grupo de muestras actuales se logra apreciar que hay diferencias en la heterocigosidad entre cada individuo analizado (Figura 6). Este fenómeno podría deberse a eventos recientes de dispersión, efecto fundador, fragmentación, flujo génico desigual o selección natural, lo que resulta en la acumulación o pérdida de diversidad genética a nivel individual, como se ha podido observar en otras especies de rana como la rana arbórea *Hyla arborea* (Krug et al., 2022). Por otro lado, también podría deberse a la técnica RADseq que no pudo cubrir las mismas regiones en todos los individuos lo que podría contribuir a la variabilidad en las estimaciones de heterocigosidad (Cummins et al., 2019; Trumbo et al., 2023).

Sin embargo, a pesar de estas diferencias observables dentro y entre grupos, se puede observar que en ambas poblaciones, actuales e históricas, la mayor parte de las muestras analizadas denotan una variabilidad moderada (Tabla 4) y que según el test de Wilcoxon no hay una diferencia estadísticamente significativa que evidencie grandes diferencias entre los grupos en estudio, lo que podría significar que la diversidad genética, en base a la heterocigosidad, se ha mantenido relativamente estable en el tiempo. Algunos estudios mencionan que aunque se presenten cuellos de botella y deriva génica en una especie, los niveles de heterocigosidad se pueden mantener estables, siempre y cuando haya existido algún tipo de flujo génico o cuando en esa población no ha existido por mucho tiempo este cuello de botella, lo que sugiere que a pesar de los cambios demográficos la

diversidad génica global de la especie se mantiene, a pesar de que en una de las poblaciones haya mayor cantidad de variantes (Frankham et al., 2010; Ellegren, 2014; Coloma et al., 2004).

Los resultados obtenidos al realizar el análisis comparativo entre las muestras actuales e históricas, en donde se evaluaron 164,675 posiciones genotipadas, permiten profundizar la interpretación de las diferencias y similitudes de heterocigosidad observadas previamente. El hecho de que se haya podido identificar un 80.79% de variantes con genotipos idénticos entre ambas poblaciones, respalda la idea de que existe un fondo genético compartido y conservado a través del tiempo, es decir que hay una preservación del acervo génico de sus antecesores (Skoglund, 2014). No obstante, el que haya posiciones donde la heterocigosidad es diferencial entre las muestras actuales e históricas refuerza la hipótesis de dinámicas evolutivas que podrían haber modulado la diversidad genética actual. Específicamente hablando de ese 7.67% de posiciones heterocigotas en los individuos actuales, que eran homocigotas en los históricos, lo que sugiere una posible acumulación de variantes nuevas, que se ha observado en especies con procesos de expansión poblacional posterior a un cuello de botella. Por otro lado, el 4.06% de posiciones que presentan el patrón inverso (Heterocigotas en históricos- homocigotas en actuales), podrían corroborar el hecho de que hay pérdida de diversidad de variantes ancestrales por deriva génica como ya se mencionó anteriormente dentro de esta discusión (Leigh et al., 2019; Hoelzel et al., 2014).

Cuando realizamos la técnica estadística del análisis de componentes principales (PCA), se intentó reducir las dimensiones de la variabilidad genética encontrada en las secuencias analizadas en este estudio, en pocas variables nuevas llamadas componentes principales, sin perder demasiada información. Esta técnica ha sido ampliamente utilizada con datos de alta complejidad o que son muy extensos, justamente como los obtenidos en este estudio (Lever et al., 2017).

Cada componente de un PCA es una combinación lineal de las variantes originales (posiciones de variantes genómicas), PC1 es la dirección que explica la mayor cantidad de varianza en los datos, PC2 es la segunda dirección de variación ortogonal a PC1 y así sucesivamente con el resto de direcciones hasta llegar a PC7, que en el caso de este estudio es el componente en donde se explica la mayoría de la varianza acumulada

(0.8491) (Greenacre et al., 2022). Esto significa que existen diferencias genéticas, distribuidas en varias dimensiones, entre las poblaciones históricas y actuales, sin embargo, genéticamente ambas poblaciones no son completamente distintas ya que se ve un solapamiento entre muestras, puesto que comparten algunas características genéticas, conservadas en el tiempo, como SNPs o INDELS, lo que denota una diferenciación genética moderada entre ambas poblaciones (Sirsi et al., 2024). Resultados similares han sido reportados en estudios previos como el de Antunes y colaboradores (2021), que, al analizar la diferenciación genética entre subespecies de salamandras de fuego en la península ibérica, en donde emplearon PCA, encontraron una separación parcial en los componentes principales. Estos análisis confirman que la separación parcial de variantes genéticas posiblemente fue influenciada por eventos demográficos del pasado, o un posible flujo génico intermitente o cuellos de botella, reafirmando aún más lo planteado anteriormente en este estudio (Sirsi et al., 2024).

4.3. Limitaciones de la anotación funcional por la cobertura parcial del genoma de referencia

La anotación funcional de las variantes detectadas en *A. ignescens*, basada en la anotación del genoma de referencia de *A. laetissimus*, nos permitió encontrar en qué partes del genoma se encontraban las variantes genéticas, ya sea en mRNAs (3,431), loci de genes (2,569) y regiones codificantes (CDS) (1,451). Sin embargo, es importante recalcar que este análisis estuvo condicionado por la cobertura incompleta de la anotación del genoma de *A. laetissimus*, generada por Braker3, lo que limitó la identificación de todas las regiones codificantes, posiblemente dejando fuera variantes funcionalmente importantes. Esto afecta directamente la precisión de la asignación funcional de las variantes encontradas en este estudio (Holt & Yandell, 2011).

A pesar de estas limitaciones, fue posible identificar regiones del genoma con alta carga de variantes en genes funcionales. Entre las variantes funcionales con mayor predominancia se encuentran aquellas localizadas en los CDS 15 y 16 del gen 1407, que según UNIPROT, corresponden a regiones que codifican isoformas de proteínas con dominios de transcriptasa inversa y endonucleasa/exonucleasa/fosfatasa, que son características en retroelementos. Este tipo de elementos móviles es conocido por su capacidad de inducir reorganizaciones genómicas, promover la inserción de nuevas variantes y forman parte de procesos de regulación génica, siendo un posible aporte para

la variabilidad estructural en una especie (Chuong et al., 2017; Lanciano & Mirouze, 2018).

Así mismo, se identificaron variantes en los CDS 1 y 3 del gen 74, transcripto 1 y 2 respectivamente, ambos codifican para isoformas de la proteína FREM2, este tipo de proteínas contienen múltiples dominios Calx-beta, los cuales están involucrados en la organización de la matriz extracelular y la adhesión epitelial. Estas funciones son esenciales durante el desarrollo embrionario y la estabilidad de los tejidos. La variación de estos genes podría tener implicaciones en la integridad estructural de las células o en procesos de adaptación ambiental (Yu et al., 2018).

Finalmente, también se encontraron un gran número de variantes en el gen 131, que codifica isoformas de una proteína con dominio RRM (RNA recognition motif), que está asociada a elementos transponibles tipo LINE-1, que pueden modificar la expresión génica y la regulación post-transduccional mediante inserciones o elementos transponibles (Goodier, 2016).

Aunque estos resultados obtenidos son únicamente una prospección a la anotación funcional de *A. ignescens*, debido a las limitaciones, aportan una visión preliminar sobre las posibles consecuencias funcionales que pueden tener las variantes genéticas en la especie.

4.4. Implicaciones genéticas para la conservación de *Atelopus ignescens* estrategias de manejo futuro

Los resultados obtenidos en este estudio revelan posibles patrones de cambio entre las poblaciones históricas y actuales de *A. ignescens*, a pesar de las limitaciones que existieron durante el estudio, como la falta de un genoma de referencia propio de la especie y el número tan limitado de muestras históricas, se encontraron resultados que podrían tener implicaciones importantes para la conservación de la población remanente de esta especie. Resultados como la reducción en la heterocigosidad de algunas muestras actuales, junto con la pérdida de variantes que se encuentran presentes en las muestras históricas, sugiere que la especie ha pasado por eventos de cuello de botella, que estarían asociados con su aparente extinción a finales del siglo XX. Estos eventos seguramente han disminuido la cantidad de variantes originales de la especie y por ende podrían haber

disminuido la capacidad adaptativa de la misma (Frankham, 2005; Leigh et al., 2019). A pesar de ello, el alto número de variantes exclusivas del grupo de muestras actual y compartidas, indica que aún existe un fondo genético común y que posiblemente se encuentra en adaptación, lo que podría ser clave para plantear estrategias de restauración genética y reintroducción (Frankham, 2005).

Estos hallazgos, junto con la distribución no aleatoria de las variantes dentro del genoma y la diferencia de heterocigosidad entre los individuos actuales, indica que se puede utilizar estos datos para plantear e implementar medidas de conservación genética como: implementación de programas de cría en cautiverio que maximicen la heterocigosidad, mediante selección de individuos que presenten mayor variabilidad y de esta manera se minimice la consanguinidad y por otro lado, el monitoreo genético continuo de la población remanente mediante marcadores informativos como SNPs y regiones codificantes clave para la supervivencia de la especie (Hoban et al., 2021; Trumbo et al., 2023).

5. CONCLUSIONES Y RECOMENDACIONES:

Se realizó el alineamiento de las 12 secuencias de *A. ignescens* con el genoma de referencia de *A. laetissimus* con un porcentaje de mapeo aceptable del 45.9% para las muestras actuales y del 59.6% para las muestras históricas, considerándose un mapeo exitoso al tratarse de especies no modelo.

Así mismo, el uso del genoma cercano de una especie del mismo género como *A. laetissimus*, como genoma de referencia, permitió el análisis de variantes genéticas con parámetros aceptables de calidad, lo cual validó el análisis de genómica comparativa planteado en este estudio.

Dentro del estudio se encontraron un buen número de variantes compartidas (164,675) principalmente SNPs, lo que sugiere que podrían existir regiones potencialmente conservadas dentro del genoma de la especie. No obstante, se puede observar también que hay una pérdida de variantes ancestrales desde los individuos de la población histórica, lo que lleva a inferir un declive de la especie por una posible deriva génica o cuello de botella. Pero también se puede observar un aumento significativo de variantes exclusivas dentro de la población actual, lo que refuerza la hipótesis de que los individuos

actuales posiblemente se encuentran acumulando variantes genéticas que podrían tener un potencial efecto en la adaptación y resiliencia de esta población remanente.

Se observó también un mantenimiento de la diversidad moderado entre poblaciones, tomando en cuenta la heterocigosidad de las muestras, a pesar de que visiblemente se notaba que algunos individuos tenían una pérdida de heterocigosidad, lo cual lleva a pensar que aún existen efectos ambientales que generan presión de selección y que la población actual intenta adaptarse. Por lo cual, en este estudio se plantea que debería implementarse una estrategia de conservación donde se realicen análisis genéticos en los programas de cría en cautiverio, que ya están en marcha, donde se seleccionen individuos con una mayor variabilidad intra específica, lo que mejoraría la reproducción disminuyendo la consanguinidad.

Además, en una primera prospección se encontraron variantes potencialmente funcionales ligadas a regiones codificantes (CDS) dentro de los genes, que podrían estar relacionados a la resistencia o adaptación a ambientes adversos, como el gen 74 que genera la proteína FREM2 que contienen múltiples dominios Calx-beta y el gen 131 que codifica la proteína con dominio RRM, que está asociada a elementos transponibles, que pueden modificar la expresión génica y la regulación post-transduccional mediante inserciones o elementos transponibles. Sin embargo, al no tener una anotación funcional completa de una especie cercana a *A. ignescens* estos resultados deben ser tomados únicamente como un acercamiento a la anotación funcional de la especie, es por esta razón que se recomienda realizar más estudios de genómica funcional en esta especie y en otras especies de anuros, para poder identificar de manera precisa todos los genes que están involucrados en procesos de adaptación y resistencia en el genoma de *A. ignescens*.

Se recomienda también realizar un nuevo estudio como este, cuando ya exista un genoma de referencia de *A. ignescens*, para poder hacer una comparación de estos resultados y complementar las brechas metodológicas que se han presentado durante la realización de este trabajo. Y de esta manera poder tener la certeza de los pasos a seguir en el cuidado de la especie.

6. REFERENCIAS:

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., ... & Goecks, J. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses. *Nucleic Acids Res.*, 46(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1997). Basic local alignment search tool. *J. Mol. Biol.*, 215(3), 403-410. <https://doi.org/10.1006/jmbi.1997.9999>.
- Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: www.bioinformatics.babraham.ac.uk
- Antunes, B., Velo-Antón, G., Buckley, D., Teixeira, J., Martínez-Solano, Í., & Arntzen, J. W. (2021). Physical and ecological isolation contribute to maintain genetic differentiation between fire salamander subspecies. *Heredity*, 126, 776–789. <https://doi.org/10.1038/s41437-021-00405-0>
- Allendorf, F. W., Luikart, G., & Aitken, S. N. (2013). *Conservation and the Genetics of Populations* (2nd ed.). Wiley-Blackwell.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376. doi: 10.1371/journal.pone.0003376
- Barrio Amoroso, C. L., Costales, M., Vieira, J., Osterman, E., Kaiser, H., Arteaga, A., (2020). Back from extinction: rediscovery of the Harlequin toad *Atelopus mindoensis* Peters, (1973) in Ecuador. *Herpetol. Notes* 13, 325- 328. <https://doi.org/10.1098/rspb.2014.2881>.
- Berger, L., Speare, R., Daszak, P., Green, D. E., Cunningham, A.A., Louise Goggin, C., Parkes, H., (1998). Chytridiomycosis causes amphibian mortality associated with population declines in the rainforests of Australia and Central America. *Proc. Natl. Acad. Sci.* 95 (15), 9031–9036. <https://doi.org/10.1073/pnas.95.15.9031>.
- Bolaños, F., Meza-Lázaro, R., & Ortiz, D. (2018). Impact of climate change on amphibian populations in Ecuador: A case study on *Atelopus ignescens*. *Herpetologica*, 74(2), 234-246.
- Blum, M., Chang, H. Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., Mi, H., ... Finn, R. D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic acids res.*, 49(D1), D344–D354. <https://doi.org/10.1093/nar/gkaa977>
- Briggs, C. J., Knapp, R. A., Vredenburg, V. T. (2010). Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. *Proc. Natl. Acad. Sci.* 107 (21), 9695–9700. <https://doi.org/10.1073/pnas.0912886107>.
- Brůna, T., Lomsadze, A., Borodovsky, M. (2021). BRAKER3: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1), lqaa108. <https://doi.org/10.1093/nargab/lqaa108>.

- Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., ... & Stein, L. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biology*, 17(1), 66. <https://doi.org/10.1186/s13059-016-0924-1>
- Byrne, A. Q., Richards-Zawacki, C. L., Voyles, J., Bi, K., Ibanez, R., Rosenblum, E. B. (2021). Whole exome sequencing identifies the potential for genetic rescue in iconic and critically endangered Panamanian harlequin frogs. *Glob. Chang. Biol.* 27 (1), 50–70. <https://doi.org/10.1111/gcb.15405>.
- Catenazzi, A., Swei, A., Finkle, J., Foreyt, E., Wyman, L., Vredenburg, V. T. (2017). Epizootic to enzootic transition of a fungal disease in tropical Andean frogs: are surviving species still susceptible? *PLoS One* 12 (10), e0186478. <https://doi.org/10.1371/journal.pone.0186478>.
- Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Chen, Y., Tian, T., He, X., Li, Z., Xu, M., He, D., & Jiang, Y. (2024). VGC: an integrative and visual genomics browser for analyzing genomic variations. *BMC Bioinformatics*, 25, 59. <https://doi.org/10.1186/s12859-024-05670-y>
- Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.*, 18(2), 71–86. <https://doi.org/10.1038/nrg.2016.139>
- Collen, B., Purvis, A., Mace, G.M., 2010. When is a species really extinct? Testing extinction inference from a sighting record to inform conservation assessment. *Divers. Distrib.* 16 (5), 755–764. <https://doi.org/10.1111/j.1472-4642.2010.00689.x>.
- Coloma, L. A., Almendáriz, A., & Ron, S. R. (2004). Rediscovery of *Atelopus ignescens* and its conservation status. *Amphibian & Reptile Conservation*, 2(1), 15-22.
- Coloma, L.A., 2016. El Jambato negro del páramo, *Atelopus ignescens*, resucitó. IMCiencia. https://issuu.com/centrojambatu/docs/coloma_2016_el_jambato_negro_del_p_.
- Coloma, L. A., Duellman, W. E., Almendariz, C., Ron, S. R., Teran-Valdez, A., Guayasamin, J. M., 2010. Five new (extinct?) species of *Atelopus* (Anura:Bufonidae) from Andean Colombia, Ecuador, and Peru. *Zootaxa* 2574, 1–54. <https://doi.org/10.11646/zootaxa.2574.1.1>.
- Coloma, L. A., Ron, S. R., & Hutter, C. (2010). Declining amphibian populations in the Andes: Threats and conservation strategies. *Biotropica*, 42(5), 637-645. <https://doi.org/10.1111/j.1744-7429.2010.00629>.
- Cummins, D., Kennington, W. J., Rudin-Bitterli, T., & Mitchell, N. J. (2019). A genome-wide search for local adaptation in a terrestrial-breeding frog reveals vulnerability to climate change. *Glob. Chang. Biol.*, 25(9), 3151–3162. <https://doi.org/10.1111/gcb.14703>
- Danecek, P., Bonfield, J. K., Liddle, J. et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>.

- Davey, J. W., & Blaxter, M. L. (2011). RADseq: Next-generation population genetics. *Brief. Funct. Genomics*, 10(6), 92-101.
- Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1–15. <https://doi.org/10.1038/hdy.2010.152>
- Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, 8(3), 186–194. <https://doi.org/10.1101/gr.8.3.186>
- Ellegren H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.*, 29(1), 51–63. <https://doi.org/10.1016/j.tree.2013.09.008>
- Frankham, R. (2005). Genetics and extinction.. *Biol. Conserv.*, 126(2), 131–140. <https://doi.org/10.1016/j.biocon.2005.05.002>
- Frankham, R., Ballou, J. D., & Briscoe, D. A. (2010). *Introduction to Conservation Genetics* (2nd ed.). Cambridge University Press.
- Fritsche, L. G., Gruber, S. B., Wu, Z., Schmidt, E. M., Zawistowski, M., Moser, S. E., ... Abecasis, G. R. (2020). GCViT: Genome-wide Chromosome Visualization Tool for SNP data. *BMC Genomics*, 21, 143. <https://doi.org/10.1186/s12864-020-07217-2>
- Fujita, M. K., Sato, J., & Hasegawa, M. (2021). Molecular phylogenetics and population genetics of *Atelopus* species in the Andes. *Mol. Phylogenet. Evol.*, 161, 107215.
- Funk, C.W., Zamudio, K.R., Crawford, A.J. (2021). Advancing understanding of amphibian evolution, ecology, behavior, and conservation with massively parallel sequencing. In: Hohenlohe, P.A., Rajora, O.P. (Eds.), *Population Genomics: Wildlife*. Springer, pp. 211–254.
- Gonzalez-Maya, J. F., Belant, J. L., Wyatt, S. A., Schipper, J., Cardenal, J., Corrales, D., Fischer, A. (2013). Renewing hope: the rediscovery of *Atelopus varius* in Costa Rica. *Amphibia-Reptilia* 34, 573–578. <https://doi.org/10.1163/15685381-00002910>.
- Goodier, J. L. (2016). Restricting retrotransposons: A review. *Mobile DNA*, 7(1), 16. <https://doi.org/10.1186/s13100-016-0070-z>
- Guayasamin, J. M., Mendoza, A. M., Longo, A. V., Zamudio, K. R., Bonaccorso, E. (2014). High prevalence of *Batrachochytrium dendrobatidis* in an Andean frog community (Reserva las galarias, Ecuador). *Amphib. Reptile Conserv.* 8 (1), 33–44.
- Greenacre, M., Groenen, P. J. F., Hastie, T., Iodice d’Enza, A., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1), 100. <https://doi.org/10.1038/s43586-022-00184-w>
- Harvey, M. B., Smith, E. N., & Mendez, L. A. (2013). Amphibian chytrid fungus and its impacts on Andean amphibian populations. *J. Wildl. Dis.*, 49(4), 921-933.
- Hoban, S., et al. (2021). Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biol. Conserv.*, 253, 108906. <https://doi.org/10.1016/j.biocon.2020.108906>

- Hohenlohe, P. A., Bassham, S., Etter, P. D., et al. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS ONE*, 5(7), e12092.
- Hoelzel, A. R., Gkafas, G. A., Kang, H., Sarigol, F., Le Boeuf, B. J., Costa, D. P., ... & Li, S. (2024). Genomics of post-bottleneck recovery in the northern elephant seal. *Nature Ecology & Evolution*, 8(4), 686–694. <https://doi.org/10.1038/s41559-024-02337-4>
- Holt, C., & Yandell, M. (2011). MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, 12(1), 491. <https://doi.org/10.1186/1471-2105-12-491>
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., ... & Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods*, 12(2), 115–121. <https://doi.org/10.1038/nmeth.3252>
- IUCN. (2021). The IUCN Red List of Threatened Species. Version 2021-1. <https://www.iucnredlist.org>. Downloaded on [13 August 2021].
- James, T. Y., Toledo, L. F., Rodder, D., da Silva Leite, D., Belasen, A. M., Betancourt Roman, C. M., Longcore, J. E. (2015). Disentangling host, pathogen, and environmental determinants of a recently emerged wildlife disease: lessons from the first 15 years of amphibian chytridiomycosis research. *Ecol. Evol.* 5 (18), 4079–4097. <https://doi.org/10.1002/ece3.1672>.
- Jaynes, K. E., Páez-Vacas, M. I., Salazar-Valenzuela, D., Guayasamin, J. M., Terán-Valdez, A., Siavichay, F. R., Fitzpatrick, S. W., & Coloma, L. A. (2022). Harlequin frog rediscoveries provide insights into species persistence in the face of drastic amphibian declines. *Biological Conservation*, 276, 109784. <https://doi.org/10.1016/j.biocon.2022.109784>
- Jones, F. C., Grabherr, M. G., Chan, Y. F., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55-61.
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94. <https://doi.org/10.1186/1471-2156-11-94>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.*, 40(1), e3. <https://doi.org/10.1093/nar/gkr771>
- Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Res.*, 17(1), 44–53. <https://doi.org/10.1111/1755-0998.12549>
- Kohno, S., Kamishima, Y., & Iguchi, T. (2003). Molecular cloning of an anuran V(2) type [Arg(8)] vasotocin receptor and mesotocin receptor: functional characterization and tissue expression in

- the Japanese tree frog (*Hyla japonica*). *General and Comparative Endocrinology*, 132(3), 485–498. [https://doi.org/10.1016/S0016-6480\(03\)00140-0](https://doi.org/10.1016/S0016-6480(03)00140-0)
- Koonin, E. V., Makarova, K. S., & Aravind, L. (2006). The ancient virus world and evolution of cells. *Biology Direct*, 1, 29.
- Kozarewa, I., Ning, Z., Quail, M. A., Sanders, M. J., Berriman, M., & Turner, D. J. (2009). Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nature Methods*, 6(4), 291–295. <https://doi.org/10.1038/nmeth.1311>
- Lanciano, S., & Mirouze, M. (2018). Transposable elements: All mobile, all different, some stress responsive, some adaptive? *Current Opinion in Genetics & Development*, 49, 106–114. <https://doi.org/10.1016/j.gde.2018.04.002>
- Lawrence, M., Gentleman, R., & Carey, V. (2009). rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics*, 25(14), 1841–1842. <https://doi.org/10.1093/bioinformatics/btp328>
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., ... & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biol.*, 9(8), e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>
- Leigh, D. M., Hendry, A. P., Vázquez-Domínguez, E., & Friesen, V. L. (2019). Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications*, 12(8), 1505–1512. <https://doi.org/10.1111/eva.12810>
- Lever, J., Krzywinski, M., & Altman, N. (2017). Principal component analysis. *Nature Methods*, 14(7), 641–642. <https://doi.org/10.1038/nmeth.4346>
- Lotters, S. (2007). The fate of the harlequin toads – help through a synchronous multidisciplinary approach and the IUCN ‘amphibian conservation action plan? *Zoosyst. Evol.* 83 (S1), 69–73. <https://doi.org/10.1002/mmnz.200600028>.
- Lu, B., Jin, H., & Fu, J. (2020). Molecular convergent and parallel evolution among four high-elevation anuran species from the Tibetan region. *BMC Genomics*, 21, 839. <https://doi.org/10.1186/s12864-020-07269-4>
- McCaffery, R., Richards-Zawacki, C. L., Lips, K. R. (2015). The demography of *Atelopus* decline: harlequin frog survival and abundance in Central Panama prior to and during a disease outbreak. *Glob. Ecol. Conserv.* 4, 232–242. <https://doi.org/10.1016/j.gecco.2015.07.003>.
- Méndez, E., Silva, M. F., & Torres, A. (2022). Genetic diversity and phylogeny of *Atelopus* species in the Andes. *J. Herpetol.*, 56(2), 203-217.
- Murray, K. A., Skerratt, L. F., Speare, R., McCallum, H. (2009). Impact and dynamics of disease in species threatened by the amphibian chytrid fungus *Batrachochytrium dendrobatidis*. *Conserv. Biol.* 23 (5), 1242–1252. <https://doi.org/10.1111/j.1523-1739.2009.01211.x>.

- Nabeel, M. A. (2023). A Beginner's Guide to Genomic Data Analysis: SAMtools and Genome Mapping Visualization in IGV. Medium. Retrieved from <https://medium.com/@manabeel.vet/a-beginners-guide-to-genomic-data-analysis-samtools-and-genome-mapping-visualization-in-d2f81ed907a7>
- Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evol.*, 29(1), 1–10. <https://doi.org/10.2307/2407137>
- Ørsted, M., Hoffmann, A. A., Sverrisdóttir, E., Nielsen, K. L., Kristensen, T. N. (2019). Genomic variation predicts adaptive evolutionary responses better than population bottleneck history. *PLoS Genet*, 12;15(6):e1008205. doi: 10.1371/journal.pgen.1008205.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>.
- Premachandra, T., Cauret, C. M. S., Conradie, W., Measey, J., & Evans, B. J. (2023). Population genomics and subgenome evolution of the allotetraploid frog *Xenopus laevis* in southern Africa. *G3: Genes, Genomes, Genetics*, 13(2), jkac325. doi: 10.1093/g3journal/jkac325.
- Puschendorf, R., Hoskin, C. J., Cashins, S. D., McDonald, K., Skerratt, L. F., Vanderwal, J., Alford, R. A. (2011). Environmental refuge from disease-driven amphibian extinction. *Conserv. Biol.* 25 (5), 956–964. <https://doi.org/10.1111/j.1523-1739.2011.01728.x>.
- Ribas, A., Lasso, A., & Rivera, S. (2020). Tracking chytrid fungus in Andean amphibians using molecular techniques. *J. Applied Ecol.*, 57(1), 42-53.
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richards, S. J., McDonald, K. R., Alford, R. A. (1994). Declines in populations of Australia's endemic tropical rainforest frogs. *Pac. Conserv. Biol.* 1 (1), 66–77. <https://doi.org/10.1071/PC930066>.
- Richards-Zawacki, C. L. (2010). Thermoregulatory behaviour affects prevalence of chytrid fungal infection in a wild population of Panamanian golden frogs. *Proc. R. Soc. B* 277, 519–528. <https://doi.org/10.1098/rspb.2009.1656>.
- Roberts, D. L. (2006). Extinct or possibly extinct? *Science*, 312 (5776), 997–998. <https://doi.org/10.1126/science.312.5776.997c>.
- Rochette, N. C., Rivera-Colon, A. G., Catchen, J. M. (2019). Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* 28 (21), 4737–4754. <https://doi.org/10.1111/mec.15253>.
- Ron, S. R., Stuart, S. N., & Nieto, J. M. (2003). *Amphibians of South America*. Princeton University Press.

- Ron, S. R., Duellman, W. E., Coloma, L. A., Bustamante, M. (2003). Population decline of the Jambato toad *Atelopus ignescens* (Anura: Bufonidae) in the Andes of Ecuador. *J. Herpetol.* 37 (1), 116–126.
- Ron, S. R., Gutiérrez, R., & Ortiz, R. (2019). Distribution and habitat preferences of *Atelopus ignescens* in Ecuador. *J. Herpetol.*, 53(3), 329-340.
- Savage, A. E., Zamudio, K. R. (2011). MHC genotypes associate with resistance to a frogkilling fungus. *Proc. Natl. Acad. Sci.* 108 (40), 16705–16710. <https://doi.org/10.1073/pnas.1106893108>.
- SAMtools. (2024). samtools merge – merges multiple sorted files into a single file. Recuperado de <https://www.htslib.org/doc/samtools-merge.html>
- Sirsi, S., Rodriguez, D., & Forstner, M. R. J. (2024). Using genome-wide data to ascertain taxonomic status and assess population genetic structure for Houston toads (*Bufo* [= *Anaxyrus*] *houstonensis*). *Scientific Reports*, 14, 3306. <https://doi.org/10.1038/s41598-024-53705-w>
- Scheele, B. C., Foster, C. N., Banks, S. C., Lindenmayer, D. B. (2017a). Niche contractions in declining species: mechanisms and consequences. *Trends Ecol. Evol.* 32 (5), 346–355. <https://doi.org/10.1016/j.tree.2017.02.013>.
- Scheele, B. C., Skerratt, L. F., Grogan, L. F., Hunter, D. A., Clemann, N., McFadden, M., Berger, L. (2017b). After the epidemic: ongoing declines, stabilizations and recoveries in amphibians afflicted by chytridiomycosis. *Biol. Conserv.* 206, 37–46. <https://doi.org/10.1016/j.biocon.2016.12.010>.
- Scheele, B. C., Pasmans, F., Berger, L., Skerratt, L. F., Martel, A., Beukema, W., Canessa, S. (2019). Amphibian fungal panzootic causes catastrophic and ongoing loss of biodiversity. *Science* 363 (6434), 1459–1463. <https://doi.org/10.1126/science.aav0379>.
- Scott, J. M., Ramsey, F. L., Lammertink, M., Rosenberg, K. V., Rohrbaugh, R., Wiens, J. A., Reed, J. M. (2008). When is an ‘extinct’ species really extinct? Gauging the search efforts for Hawaiian forest birds and the Ivory-billed woodpecker. *Avian Conserv. Ecol.* 3 (2), 3.
- Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Met. Ecol. Evol.*, 8(8), 907–917. <https://doi.org/10.1111/2041-210X.12700>
- Sherlock, M. B., et al. (2024). Genomic SNPs resolve the phylogeny of an ancient amphibian island radiation from the Seychelles. *Mol. Phyl. Evo.*, 187, 107905. <https://doi.org/10.1016/j.ympev.2024.107905>
- Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., ... & Jakobsson, M. (2014). Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science*, 344(6185), 747–750. <https://doi.org/10.1126/science.1253448>

- Sparks, M. M., Schraidt, C. E., Yin, X., Seeb, L. W., & Christie, M. R. (2023). Rapid genetic adaptation to a novel ecosystem despite a large founder event. *Molecular Ecology*, 33(20), e17121. <https://doi.org/10.1111/mec.17121>
- Springborn, M. R., Weill, J. A., Lips, K. R., Ibanez, R., Ghosh, A. (2022). Amphibian collapses increased malaria incidence in Central America. *Environ. Res. Lett.* 17 (10), 104012 <https://doi.org/10.1088/1748-9326/ac8e1d>.
- Stuart, S. N., Chanson, J. S., Cox, N. A., Young, B. E., Rodrigues, A. S. L., Fischman, D. L., Waller, R. W. (2004). Status and trends of amphibian declines and extinctions worldwide. *Science* 306 (5702), 1783–1786. <https://doi.org/10.1126/science.1103538>.
- Tello-Ruiz, M. K., Naithani, S., Gupta, P., Olson, A., Wei, S., Preece, J., ... & Jaiswal, P. (2020). Gramene 2021: harnessing the power of comparative genomics and pathways for plant research. *Nucleic Acids Res.*, 49(D1), D1452–D1463. <https://doi.org/10.1093/nar/gkaa979>.
- Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the Cox model*. Springer.
- Trumbo, D. R., Hardy, B. M., Crockett, H. J., Muths, E., Forester, B. R., Cheek, R. G., Zimmerman, S. J., Corey-Rivas, S., Bailey, L. L., & Funk, W. C. (2023). Conservation genomics of an endangered montane amphibian reveals low population structure, low genomic diversity and selection pressure from disease. *Molecular Ecology*, 32(24), 6777–6795. <https://doi.org/10.1111/mec.17175>
- UniProt Consortium. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>
- Uricaru, R., et al. (2020). DiscoSnp-RAD: de novo detection of small variants for RAD-Seq population genomics. *BMC Bioinformatics*, 21, 1–14. <https://doi.org/10.1186/s12859-020-03667-6>
- Vega-Yáñez, M. A, Quezada-Riera, A. B, Rios-Touma, B., Vizcaíno-Barba, M. D. C, Millingalli, W., Ganzino, O., Coloma, L. A., Tapia, E. E., Dupérré, N., Páez-Vacas, M., Parra-Puente, D., Franco-Mena, D., Gavilanes, G., Salazar-Valenzuela, D., Valle, C. A., Guayasamin, J. M. (2024). Path for recovery: an ecological overview of the Jambato Harlequin Toad (Bufonidae: *Atelopus ignescens*) in its last known locality, Angamarca Valley, Ecuador. *PeerJ*, 21;12:e17344. doi: 10.7717/peerj.17344.
- Von May, R., Catenazzi, A., Angulo, A., Brown, J. L., Carrillo, J., Chávez, G., Twomey, E. (2008). Current state of conservation knowledge on threatened amphibian species in Peru. *Trop. Conserv. Sci.* 1 (4), 376–396. <https://doi.org/10.1177/194008290800100406>.
- Vredenburg, V. T., Knapp, R. A., Tunstall, T. S., Briggs, C. J. (2010). Dynamics of an emerging disease drive large-scale amphibian population extinctions. *Proc. Natl. Acad. Sci.* 107 (21), 9689–9694. <https://doi.org/10.1073/pnas.0914111107>.

- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
<https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the tidyverse. *J. Open Source Software*, 4(43), 1686.
<https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A grammar of data manipulation (R package version 1.1.3)* [Computer software]. <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., Vaughan, D., & Girlich, M. (2024). *tidyr: Tidy messy data (versión 1.3.1)* [Paquete R].
<https://CRAN.R-project.org/package=tidyr>
- Whitfield, S. M., Alvarado, G., Abarca, J., Zumbado-Ulate, H., Zuniga, Wainwright, M., Kerby, J. (2017). Differential patterns of *Batrachochytrium dendrobatidis* infection in relict amphibian populations following severe disease-associated declines. *Dis. Aquat. Org.* 126 (1), 33–41.
<https://doi.org/10.3354/dao03154>.
- Yu, Q., Lin, B., Xie, S., Gao, S., Li, W., Liu, Y., Wang, H., Huang, D., Xie, Z. A. (2018). Homozygous mutation p.Arg2167Trp in *FREM2* causes isolated cryptophthalmos. *Hum. Mol. Genet.* 1;27(13):2357-2366. doi:10.1093/hmg/ddy144. PMID: 29688405; PMCID: PMC6251612.