Análisis de la variación nucleotídica de regiones candidatas del cromosoma 2 en poblaciones ancestrales y derivadas de *Drosophila melanogaster*

Asier Ruiz Camara

MU Bioinf. y Bioest.
Análisis de datos Ómicos

Nombre Tutor/a de TFM
Dorcas Orengo Ferriz
Profesor/a responsable de la asignatura
David Merino Arranz

Fecha entrega
Enero 2024

Universitat Oberta de Catalunya

UNIVERSITAT DE BARCELONA

FICHA DEL TRABAJO FINAL

Título del trabajo:	Análisis de la variación nucleotídica de regiones candidatas del cromosoma 2 en poblaciones ancestrales y derivadas de *Drosophila melanogaster*
Nombre del autor:	*Asier Ruiz Camara*
Nombre del consultor/a:	*Dorcas Orengo Ferriz*
Nombre del PRA:	*David Merino Arranz*
Fecha de entrega (mm/aaaa):	*01/2024*
Titulación o programa:	*Máster en bioinformática y bioestadística*
Área del Trabajo Final:	*Análisis de datos Ómicos*
Idioma del trabajo:	*Castellano*
Palabras clave	*Drosophila melanogaster, genética poblacional, selección natural*

Resumen del Trabajo

La mosca de la fruta o *Drosophila melanogaster* se origina en África subsahariana y ha extendido su presencia por todo el mundo gracias a su relación de "comensalismo" con los humanos. Dado que esta expansión es relativamente reciente, los cambios genómicos reflejando la adaptación al nuevo clima y los efectos demográficos deberían ser evidentes en su genoma. En una población de Barcelona se habían encontrado signos de selección natural en ocho fragmentos distintos del cromosoma 2. Este estudio investiga la variación nucleotídica en estas regiones en busca de indicios de selección en seis poblaciones de *Drosophila melanogaster*, tres de América del Norte y tres de África. Conjuntamente, se realiza un análisis de diferenciación genética entre estas poblaciones. Para este propósito, se localizaron las secuencias en el genoma y se sometieron a análisis utilizando programas bioinformáticos como BLAST, DnaSp, mlcoalsim y mstatspop. A partir de los resultados obtenidos, se derivan las siguientes conclusiones: el análisis revela neutralidad en las regiones estudiadas por el estadístico *D*. En la mayoría, el estadístico *Hn* de Fay y Wu indica rechazo a la neutralidad, mostrando señales de selección positiva. La diversidad nucleotídica coincide con estudios previos, sin embargo, la región 59B puede estar sometida a un arrastre selectivo debido a su cercanía con el gen *Klp59C*. Entre las poblaciones africanas, se evidencia diferenciación genética, estando EF mayormente distanciada, posiblemente debido a su aislamiento geográfico y condiciones climáticas. En las poblaciones estadounidenses, aunque algunos fragmentos muestren diferenciación, los valores de F_{st} muestran cercanía entre ellas.

Abstract

The fruit fly, Drosophila melanogaster, originated in sub-Saharan Africa and has expanded its presence worldwide through its 'commensalism' relationship with humans. Given the relatively recent nature of this expansion, genomic changes reflecting adaptation to the new climate and demographic effects should be apparent in its genome. In a population from Barcelona, signs of natural selection had been identified in eight distinct fragments of chromosome 2. This study investigates nucleotide variation in these regions, aiming to detect indications of selection across six populations of Drosophila melanogaster, three from North America and three from Africa. Additionally, a genetic differentiation analysis is conducted among these six populations. To achieve this, sequences were mapped onto the genome and subjected to analysis using bioinformatics programs such as BLAST, DnaSp, mlcoalsim, and mstatspop. From the results obtained, the following conclusions are drawn: the analysis indicates neutrality by statistic *D*. Fay and Wu's *Hn* indicates non-neutrality, giving signs of positive selection. Nucleotide diversity aligns with previous studies; however, region 59B may experience genetic hitchhiking due to nearness to gen *Klp59C*. African populations display genetic differentiation, notably with EF geographically isolated, which possibly impacts its adaptation. Within US populations, although some segments show differentiation, F_{st} values endorse genetic propinquity.

Índice

Lista de figuras

1. Introducción

1.1. Contexto y justificación del Trabajo

Los cambios medio ambientales (ya sean bióticos o abióticos), provocan adaptaciones a través de la selección de mutaciones que confieren ventajas en el nuevo entorno[1]. Por lo tanto, el estudio de poblaciones que surgieron después de la expansión de una especie es una excelente manera de examinar cómo la selección natural afecta a la población en función de los nuevos ambientes. Es importante destacar que no solo los efectos de la selección natural afectarán al genoma, sino que también la historia demográfica resultante de la expansión tendrá un impacto en el genoma de la especie. Además, varios procesos evolutivos pueden influir en los patrones de variación genética y, en ocasiones, pueden parecerse a los efectos de la selección positiva[2], lo que dificulta la identificación de las señales de selección.

Gracias al "comensalismo" y la estrategia ecológica de *D. melanogaster* (estratega de la r[3]), se extendió desde regiones tropicales, siendo una especie originaria del África subsahariana, hacia áreas más templadas tras la última glaciación hace unos 10,000-15,000 años[4]. Esta expansión condujo a una disminución en la diversidad genética en las regiones colonizadas, debido al efecto fundador y al desafío de adaptarse a entornos distintos a su lugar de origen.

Por ello, se analizarán 3 poblaciones ancestrales (África) y otras 3 derivadas (Norte América) para comprobar si se repite o no el patrón de variación que ya ha sido observado en regiones específicas del genoma en una población catalana que se consideran regiones candidatas de haber experimentado selección natural reciente. Además, se compararán las poblaciones ancestrales entre ellas para comprobar las diferencias que puedan existir entre las poblaciones de la región geográfica ancestral.

1.2. Objetivos del Trabajo

A continuación, se exponen los objetivos a cumplimentar en el TFM.

1.2.1. Objetivos generales

I. Analizar la variación nucleotídica de regiones candidatas del cromosoma 2 en distintas poblaciones (ancestrales y derivadas) de *Drosophila melanogaster* para intentar ver si la selección natural ha actuado tras la colonización de nuevos ambientes.
II. Analizar la diferenciación genética de estas regiones entre las distintas poblaciones.

1.2.2. Objetivos específicos

I. Analizar la variación nucleotídica de regiones candidatas del cromosoma 2 en distintas poblaciones (ancestrales y derivadas) de *Drosophila melanogaster* para intentar ver si la selección natural ha actuado tras la colonización de nuevos ambientes.
   - Localizar en el genoma de referencia las secuencias a analizar.
   - Elegir las poblaciones y los individuos idóneos.
   - Obtener las secuencias correspondientes.
   - Analizar las secuencias mediante indicadores estadísticos.
   - Comparar los resultados con los obtenidos para una población de Barcelona que señaló estas regiones como candidatas de haber experimentado un barrido selectivo.

II. Analizar la diferenciación genética de estas regiones entre las distintas poblaciones.
   - Analizar la diferenciación genética de estas regiones en las poblaciones del área de origen de la especie.
   - Analizar la diferenciación genética de estas regiones en las poblaciones derivadas de la especie.

1.3. Impacto en sostenibilidad, ético-social y de diversidad

Desde una perspectiva de sostenibilidad, el análisis de la variación genética en *Drosophila melanogaster* puede proporcionar información clave sobre la adaptación de las poblaciones a cambios ambientales. Esto es relevante para la conservación de la biodiversidad y la gestión de especies en peligro de extinción, ya que nos ayuda a entender cómo las poblaciones pueden enfrentar desafíos ambientales y evolucionar para sobrevivir. Conjuntamente, los datos que se utilizarán en el estudio son públicos, no necesitando obtener nuevas muestras animales para recolectar información. A raíz de lo mencionado en la oración anterior, el impacto ambiental es nulo.

Enfocándolo en el aspecto ético-social, la investigación en genética evolutiva debe llevarse a cabo de manera ética y considerando las implicaciones sociales. Esto incluye la consideración de la equidad en el acceso a los beneficios de la investigación genética y la aplicación ética de los conocimientos derivados de estos estudios.

Finalmente, desde un punto de vista de diversidad, a la hora de la redacción de la memoria, se consultarán distintas fuentes sin tener en cuenta el género o la etnia de los autores, dándole importancia al conocimiento generado por otros investigadores que desarrollan su conocimiento en el mismo ámbito.

1.4. Enfoque y método seguido

Primero se debe localizar la región génica que ha sido analizada en el estudio realizado en la población de Barcelona en las poblaciones que se van a analizar. Para ello, se utilizará la herramienta BLAST una herramienta útil para buscar similitudes entre secuencias biológicas como ADN, ARN o proteínas[5].

Para cada una se determinará si la secuencia es intergénica o intrónica y se obtendrán las coordenadas que engloban la totalidad de región no codificadora. Aunque las secuencias analizadas en la población de Barcelona fueron fragmentos cortos de ~1kb, interesa analizar cómo es la variación nucleotídica a lo largo de toda la secuencia no codificadora.
Utilizando la base de datos de individuos de Popfly, se conseguirán las secuencias a analizar en las distintas poblaciones. Se descargarán las regiones no codificadoras completas que engloban las regiones candidatas. A continuación, se eliminará 1Kb en cada uno de los extremos de estas regiones para evitar coger parte de la secuencia reguladora de genes cercanos a dicha región o señales de *splicing* de los intrones.
Finalmente se analizarán las secuencias mediante varios indicadores estadísticos para poder proceder al análisis. Para el análisis estadístico se usará DnaSP, una herramienta popular para realizar análisis genéticos poblacionales exhaustivos en alineaciones de múltiples secuencias[6].

1.5. Planificación del Trabajo

Ahora se explicarán las tareas que se realizarán en el Trabajo de Fin de Máster. También se mostrará un calendario con las fechas de inicio y final de cada tarea y entrega y se analizarán los posibles riesgos que puedan afectar a la realización del TFM.

1.5.1. Tareas

- Localización en el genoma de referencia de las secuencias a analizar.
- Elección de los individuos más idóneos de la base de datos.
- Obtención de las secuencias nucleotídicas de cada individuo para todas las secuencias a analizar.
- Análisis de las secuencias mediante indicadores estadísticos en las poblaciones.
- Comparación de los resultados con los resultados de estudios anteriores.
- Corrección de errores y finalización de la memoria.

1.5.2. Calendario

Para el desarrollo del calendario hay que tener en cuenta la disponibilidad del autor, el cual tiene disponibilidad de dedicar mínimo 2 horas diarias compaginadas con su trabajo a jornada completa.

Tras la entrega de la primera PEC, comenzará el desarrollo del estudio propiamente dicho. Para ello, en octubre se realizarán dos hitos claves para poder tener un margen de maniobra por si ocurre algún imprevisto con las secuencias o elección de individuos.

Tras esto, comenzará el análisis de las secuencias con sus respectivos estadísticos, los cuales están planteados que se terminen en diciembre. De tal forma, después habrá 2 semanas para redactar una redacción sólida y bien contrastada con los resultados.

En la segunda quincena de diciembre se comenzará el inicio del cierre de la memoria. En esta tarea se finalizará la memoria y será entregada.

Finalmente, en enero se realizarán las dos últimas partes del trabajo: La presentación y la defensa pública.

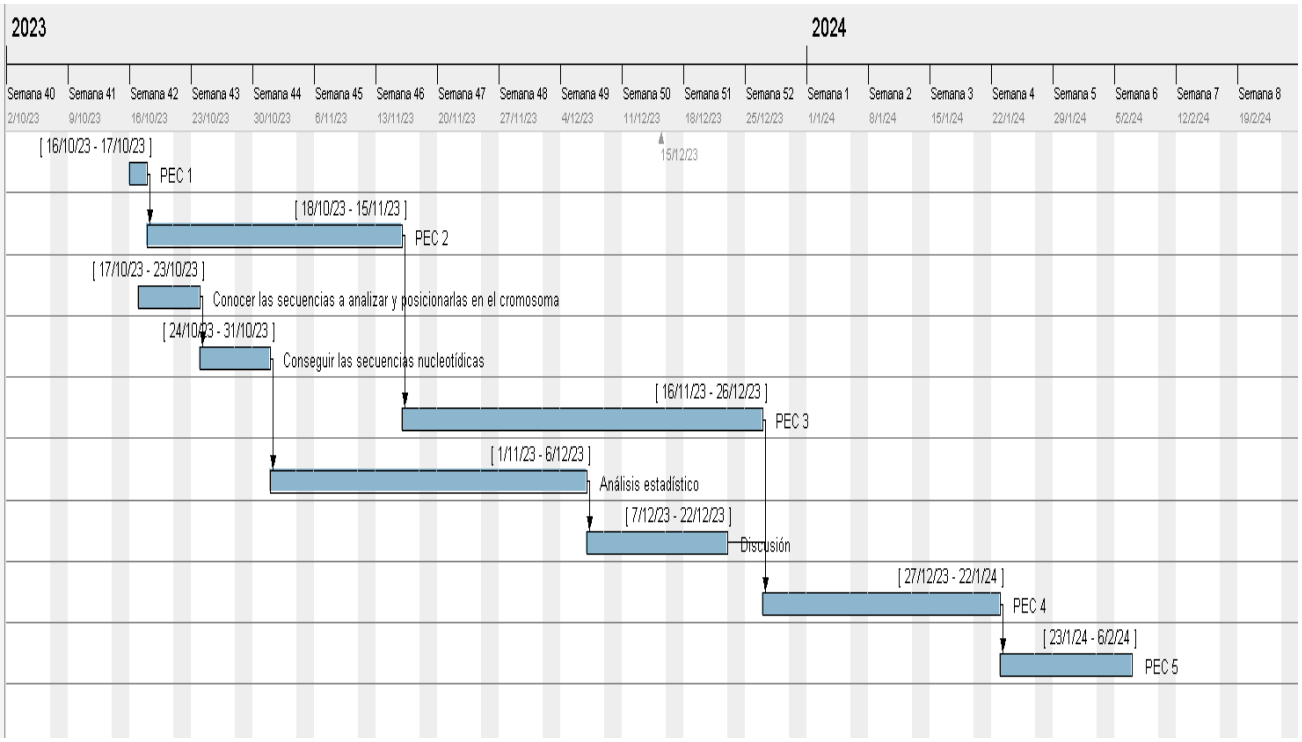


Figura 1: Diagrama de Gantt para los hitos.

1.5.3. Hitos

A continuación, se presentan los hitos con sus fechas clave para el correcto desarrollo del trabajo. Se hará un seguimiento riguroso de las actividades para completar la tarea de manera exitosa.

Tabla 1: Hitos y sus fechas de entrega

Hitos	Entrega	Fecha de entrega
Plan de trabajo	PEC 1	17/10/2023
Conocer las secuencias a analizar y posicionarlas en el cromosoma	PEC 2	23/10/2023
Conseguir las secuencias nucleotídicas	PEC 2	30/10/2023
Análisis estadístico	PEC 3	04/12/2023
Discusión	PEC 3	19/12/2023
Entregar memoria	PEC 4	16/01/2024
Elaborar presentación	PEC 4	16/01/2024
Realizar defensa pública	PEC 5	02/02/2024

1.5.4. Análisis de riesgos

Durante la realización de este trabajo podrían aparecer problemas de distinta índole que afecten negativamente a su progreso temporal.
Estos son varios de los factores a tener en cuenta:

- Imposibilidad de realizar el trabajo en el tiempo solicitado debido a un aumento de trabajo en el empleo. Se establecerán tiempos mínimos diarios a dedicar al TFM para evitar entregas fallidas.
- Problemas con la instalación y/o uso de los programas informáticos. Se realizará una instalación/prueba de uso temprana para acostumbrarse a su uso/requerimientos.
- Las secuencias que analizadas en la población de Barcelona son cortas (máximo de 2Kb) y en algunos casos a pesar de haber buscado regiones intergénicas posteriormente se han anotado genes en ellas. En caso de coincidir con parte de un gen, se comprobará que pertenece a un intrón grande para que se pueda considerar que se trata de una secuencia que evoluciona de un modo neutro. En caso de no ser así, se debería eliminar esta región del análisis.
- Posibilidad de cometer errores debido a desconocimiento del procedimiento. Para remedia este riesgo se mantendrá un contacto fluido con la tutora del máster.
- Inserciones y delecciones no mostradas en las secuencias en la base de datos. Para mitigar este error se repasarán los individuos escogidos junto con la tutora para concretar una elección correcta de estos.

1.6. Breve sumario de productos obtenidos

Tras completar los siguientes apartados se dará por finalizado el TFM:

1. Plan de trabajo:
   Es un documento entregado en la PEC 1, con el objetivo de concretar y describir el trabajo a realizar. Para el correcto desarrollo del TFM se tiene un plan de trabajo como base, usando como referencias los hitos y las fechas de entrega de las PECs.
2. Memoria:
   Este documento contendrá toda la información relativa al proyecto. Una introducción que mostrará el estado actual y conocimientos hasta ahora descubiertos de la detección de *sweeps* en *D. melanogaster*. Un apartado con la metodología utilizada para el análisis de las regiones. Los resultados obtenidos y conjuntamente la discusión de dichos resultados con estudios relacionados de otros autores.
3. Producto
   Tras el satisfactorio desarrollo de este TFM, si los resultados fuesen de cierta relevancia a nivel de publicación, sería idóneo publicar el presente TFM en una revista de impacto nacional.
4. Presentación virtual
   Se realizará una presentación que sintetizará el trabajo realizado y los resultados obtenidos en el cual se ofrecerá una perspectiva general del TFM y se recogerán los aspectos más relevantes del proyecto. Se realizará con un material de apoyo visual.

1.7. Breve descripción de los otros capítulos de la memoria

- Materiales y Métodos: Este capítulo proporciona detalles sobre los procedimientos llevados a cabo durante el estudio, así como información acerca de las herramientas bioinformáticas utilizadas.
- Resultados y Discusión: En esta sección se presentan y analizan los resultados obtenidos a través del análisis realizado.
- Conclusiones: Aquí se evalúan los resultados obtenidos y se realiza una reflexión personal sobre el trabajo realizado.
- Glosario: Incluye la definición de conceptos y la aclaración de siglas y acrónimos empleados a lo largo de la memoria.
- Bibliografía: Se enumeran todas las fuentes de información citadas durante el desarrollo del trabajo.
- Anexos: En esta sección final se incluye el listado de los individuos analizados y los archivos de entrada utilizados para las simulaciones de las poblaciones. También se muestran algunos gráficos que no se han incluido en el cuerpo del texto para facilitar la lectura.

2. Estado del arte

El concepto de "*hitchhiking* genético" se ha vuelto ampliamente utilizado en la genética de poblaciones y se refiere a cualquier situación en la que los cambios en las frecuencias alélicas causados por la selección afectan las frecuencias de variantes neutrales en sitios vinculados en el genoma. Esto incluye cualquier tipo de selección que sea lo suficientemente fuerte. Específicamente, para la selección direccional que causa "*hitchhiking* genético," se utiliza el término "barrido selectivo"[7].

Desde la publicación del artículo de Charlesworth *et al*.[8], dos modelos principales de genética de poblaciones han competido para explicar la reducción observada en la variación de nucleótidos en regiones genómicas con tasas de recombinación reducidas. Según ambos modelos, el nivel de variación neutra (o casi neutra) puede reducirse por debajo de las expectativas clásicas neutrales debido a la selección contra la entrada constante de mutaciones perjudiciales (llamada selección purificadora) o por barridos selectivos recurrentes. La observación de niveles reducidos de variación en regiones genómicas con recombinación limitada, y la subsiguiente controversia sobre su interpretación, iniciaron una fase importante en la genética de poblaciones molecular. Dado que estas observaciones no se limitaron a Drosophila, sino que se encontraron en otros organismos como humanos y plantas, se desarrollaron nuevos métodos para distinguir las contribuciones relativas de la selección purificadora y los barridos selectivos.

Los barridos selectivos han sido utilizados para inferir huellas de selección direccional positiva en los genomas de organismos con recombinación. Se han desarrollado pruebas como el "CLR test" (del inglés *Composite Likelihood Ratio* o prueba de razón de verosimilitud compuesta) para detectar reducciones locales en la diversidad de nucleótidos a lo largo de un cromosoma y predecir la intensidad y ubicación del objetivo de la selección. Sin embargo, es importante tener en cuenta que estas pruebas se basan en modelos neutros estándar y pueden no ser apropiadas cuando han ocurrido eventos demográficos significativos en la historia de una población.

Además, se ha avanzado en la inferencia conjunta de fuerzas demográficas y selectivas, lo que significa que se analizan simultáneamente todos los procesos de selección y demografía. Esto es especialmente importante para poblaciones con un gran tamaño efectivo. En el caso de la selección purificadora y los barridos selectivos, cuyos efectos son difíciles de distinguir, se han desarrollado enfoques para estudiar sus efectos conjuntos utilizando datos detallados de mapeo genético.

En resumen, el "*hitchhiking* genético" es un concepto fundamental en genética evolutiva que ha sido aplicado para entender cómo la selección afecta la variación genética en poblaciones. Los barridos selectivos son fenómenos que pueden ocurrir en el contexto de la selección natural y tienen implicaciones importantes para la evolución de los genomas. Estos conceptos se han

aplicado en diversos organismos y han contribuido significativamente a nuestra comprensión de la genética de poblaciones y la evolución molecular.

La búsqueda de barridos selectivos en dípteros en diversas partes del genoma ha experimentado un avance gracias al desarrollo de tecnologías de secuenciación de alto rendimiento y análisis de datos genómicos. Los dípteros, un orden de insectos que incluye moscas, mosquitos y tábanos, han sido objeto de numerosos estudios en genética evolutiva debido a su corto tiempo de generación, facilidad de cría en laboratorio y diversidad de especies. Estos factores han permitido a los investigadores analizar patrones de variación genética y detectar barridos selectivos en diferentes regiones del genoma[9].

Uno de los enfoques más comunes para identificar barridos selectivos en dípteros es mediante la comparación de la variación genética entre poblaciones. Esto se realiza a través de estadísticas de diversidad genética y diferenciación poblacional. La secuenciación del genoma completo de numerosas especies de dípteros ha permitido un análisis exhaustivo de las regiones genómicas bajo selección. Los estudios han revelado genes y regiones genómicas que están sujetos a barridos selectivos, lo que proporciona información valiosa sobre las adaptaciones y la evolución de estas especies. Por ejemplo, en *Drosophila melanogaster*, se han identificado genes relacionados con la resistencia a patógenos, la respuesta al estrés y la reproducción que han experimentado barridos selectivos. Estos hallazgos han arrojado luz sobre la evolución de rasgos específicos en esta especie[10].

Se ha demostrado que las poblaciones ancestrales y derivadas de esta especie de mosca de la fruta proporcionan un valioso modelo para comprender la dinámica de la variación genética en respuesta a cambios ambientales y demográficos. Además, estos estudios han revelado la influencia de la selección natural y los efectos fundadores en la estructura genética de estas poblaciones. Investigaciones previas han identificado genes candidatos relacionados con adaptaciones locales en las poblaciones derivadas, lo que ha impulsado aún más la exploración de estas regiones genómicas[11].

De acuerdo con el artículo publicado en 2007 de Stephan & Li[12], que recapitula hasta dicho año los avances conseguidos en investigación con los barridos selectivos de *D. melanogaster*. Se comenzó estudiando el genoma del díptero utilizando loci del cromosoma X, comparando genomas de especímenes norteamericanos con especímenes africanos, resultando encontrar regiones polimórficas en las mismas posiciones.

Conjuntamente, en ese mismo año (2007), se llevaron a cabo numerosos estudios sobre la adaptación en *Drosophila melanogaster*. Se utilizó el análisis de loci de rasgos cuantitativos (QTL) para estudiar la tolerancia a la temperatura y otros rasgos adaptativos. Las comparaciones moleculares de secuencias de ADN han proporcionado estimaciones de la frecuencia de cambios adaptativos a lo largo del tiempo[12].

En cuanto a la detección de selección positiva, se emplearon métodos multilocus. Al analizar los conjuntos de datos de diferentes poblaciones, se

encontró que los escenarios demográficos solos no pueden explicar los datos en algunos casos, lo que sugiere la presencia de selección positiva. Se estimó una tasa de sustitución adaptativa reciente en las poblaciones africanas y europeas, sugiriendo que la selección direccional positiva ha desempeñado un papel crítico en la adaptación de ambas poblaciones a lo largo del tiempo, probablemente en respuesta a cambios ambientales significativos. Aunque hay ciertas limitaciones en los métodos utilizados, las estimaciones indican una tasa acelerada de adaptación en ambas poblaciones debido a cambios ambientales recientes[12].

En 2010, se empezó a marcar importancia en la consanguineidad y la perdida de heterocigosidad en los individuos criados en cautiverio. Se observó que en lugar de tener su evolución "congelada", las poblaciones en cautiverio experimentaron grandes barridos selectivos en todo el genoma que afectaron no solo a los loci de aptitud, sino también a los loci neutrales vinculados[13].

De la misma forma también empezó a ser de gran relevancia conocer los barridos selectivos que afectaban a otras especies, como *Drosophila simulans* y *Drosophila sechelia*, para conocer las mutaciones adaptativas que se acumulan durante la divergencia de especies contribuyendo a las incompatibilidades reproductivas que dificultan el flujo genético; sin embargo, también puede haber un tipo de mutaciones que son generalmente ventajosas y que pueden propagarse a través de los límites de las especies. En un estudio de Schlencke y Begun de 2004[14], se observó, dentro de las dos especies arriba mencionadas, las mismas regiones presentaron un mismo barrido selectivo, solo que en el caso de *D. simulans*, se aprecia un barrido incompleto.

Para el estudio de los barridos selectivos, se destacan dos enfoques principales: los métodos basados en polimorfismo y divergencia y los métodos basados en la diferenciación de poblaciones[15].

Los métodos basados en polimorfismo y divergencia incluyen pruebas como la prueba HKA[16] y la prueba MK[17], que se utilizan para distinguir entre evolución neutral y procesos selectivos que pueden cambiar la frecuencia de alelos en las poblaciones. Estas pruebas comparan la variación de nucleótidos dentro y entre especies y analizan mutaciones sinónimas y no sinónimas para evaluar la selección natural.

Los métodos basados en el espectro de frecuencia del sitio (SFS, del inglés *Site Frequency Spectrum*) se centran en el análisis de la variación genética dentro de una población. El valor *D* de Tajima, se utiliza para evaluar diferencias entre sitios segregantes y diferencias nucleotídicas promedio. Otros métodos, como las *D* y *F* de Fu y Li, distinguen entre variantes ancestrales y derivadas y evalúan desviaciones de las expectativas neutrales en la frecuencia de alelos.

Los métodos basados en la diferenciación de poblaciones examinan la divergencia genética entre poblaciones para identificar signos de selección. F_{ST} es una medida clave que compara las frecuencias alélicas entre poblaciones. Valores elevados de F_{ST} pueden indicar selección local.

El análisis de la variación nucleotídica tiene un potencial significativo en diversas áreas. Dentro de la investigación científica, esta línea de estudio continúa siendo relevante para avanzar en nuestra comprensión de la evolución y la genética poblacional. Además, existe un interés creciente en aplicar los conocimientos derivados de estos estudios en biotecnología y agricultura. La capacidad para identificar genes específicos que pueden estar relacionados con la adaptación a diferentes entornos puede tener aplicaciones en la mejora de cultivos y en la gestión de poblaciones de plagas. Además, en el campo de la medicina, comprender cómo la variación genética influye en la adaptación de las poblaciones puede tener implicaciones en la investigación de enfermedades genéticas y en la identificación de blancos terapéuticos[11].

Cabe destacar, el estudio realizado por Kapun y colaboradores en 2020[18], donde se analizaba la variación genética de poblaciones de *D. melanogaster* de distintos países europeos usando datos genómicos *pool-seq*. En dicho estudio encontraron evidencias de barridos selectivos en las distintas poblaciones. Destacan 64 genes con evidencias de selección en todas las poblaciones europeas analizadas (13 en el cromosoma 3L; 11 en el 3R y 40 en el X). También se observó selección en al menos 7 de las 19 poblaciones analizadas para ocho genes (*wapI, HDAC6, Hen1, CR18217, mgl, phantom, Cyp18a1 y Cyp6g1*) que ya habían estado identificados anteriormente como afectados por selección reciente utilizando distintos métodos[19,20,21,22,23,24,25]. Sin embargo, los genes encontrados en dicho estudio, dentro del genoma de la mosca de la fruta, estaban prácticamente todos localizados en los cromosomas 3 y X. Es por ello, que el presente estudio presenta un valor añadido a la hora de representar un nuevo intento para encontrar barridos selectivos en una nueva región del genoma.

En resumen, el análisis de la variación nucleotídica en *Drosophila melanogaster* ofrece oportunidades significativas tanto en el ámbito de la investigación científica como en aplicaciones prácticas en diversas industrias. Además, en *D. melanogaster*, no hay demasiados estudios previos realizados en el cromosoma 2 para los *selective sweeps*, los cuales serán abordados en el TFM a desarrollar, aumentando el valor del estudio a realizar.

3. Materiales y métodos

Para la realización de los análisis se han escogido 8 fragmentos de región no codificadora, 4 del cromosoma 2R y 4 del cromosoma 2L, del genoma de *D. melanogaster*. Estos fragmentos habían mostrado un patrón de variación nucleotídica que los señalaba como candidatos de haber experimentado un barrido selectivo en una población catalana (Orengo, comunicación personal de datos no publicados).

Los fragmentos analizados se han obtenido de secuencias del cromosoma 2 de entre 12 y 15 individuos de la especie *Drosophila melanogaster* de 6 poblaciones de áreas geográficas diferentes (Estados Unidos, Etiopia, Ruanda, y Zambia). Conjuntamente, se ha utilizado la secuencia del cromosoma 2 de un individuo *Drosophila simulans*, cuya secuencia se obtuvo desde la página web Drosophila Genome Nexus (DGN)[26]. Las secuencias de *D. melanogaster* se obtuvieron haciendo uso de la herramienta BLAST de la web Flybase[27] y mediante el *Gbrowser* de la página Popfly[28]. Sin embargo, cabe mencionar que a causa de los distintos *releases* encontrados en Flybase (*release* 6) y Popfly (*release* 5), hubo que igualar la información correspondiente entre ambos *releases*. Para ello, tras hacer el BLAST de los fragmentos de interés en Flybase, se observaron los genes delimitantes presentes en la herramienta *JBrowse* de Flybase. Una vez anotados los genes delimitantes, se utilizaron como criterio de búsqueda en Popfly, asegurando así la obtención de las coordenadas correspondientes a la región no codificante entre distintos *releases*.

Se investigó en las bases de datos qué poblaciones de *D. melanogaster* disponen de suficientes individuos secuenciados. Se eligieron 3 poblaciones a lo largo de África Oriental que podrían representar el área de origen de *D. melanogaster* y que cuentan con suficientes secuencias: Etiopia (EF), Ruanda (RG) y Zambia (ZI). Las poblaciones de Estados Unidos eran de interés porque las poblaciones de *D. melanogaster* para establecerse en América experimentaron dos procesos de colonización, uno desde Europa y otro desde África[29]. Las poblaciones escogidas de EEUU corresponden a distintos estudios: la población de Raleigh (RAL)[30] la población de Raleigh, la población de Ithaca (USI)[31] y la población de Winters (USW)[32].

En cada una de estas poblaciones, se dispone de un extenso conjunto de individuos secuenciados, y fue necesario seleccionar la secuencia de 15 individuos (12 de ellos como mínimo). Para llevar a cabo esta elección, se identificaron los individuos con las secuencias más completas en las regiones sujetas a análisis, y luego se seleccionaron al azar 15 de entre esos individuos (Ver anexo 1). Se prestó especial atención también a la presencia de individuos que presentasen inversiones en ambos brazos cromosómicos (In(2L)t y In(2R)NS), ya que estos serían tomados como otra población distinta distorsionando así los resultados. Se observó en la base de datos de Popfly la presencia de estos individuos en las poblaciones de Zambia (ZI) y de Raleigh (RAL), en las cuales no se escogieron ninguno de estos individuos.

Tras ello, para el análisis de las secuencias, se utilizaron diferentes estadísticos y programas para el análisis de estos. Se calculó la diversidad nucleotídica (π[33]) y diversos indicadores estadísticos (*D*[34]; *Hn*[35,36]) utilizando el programa DnaSP[37], una herramienta ampliamente utilizada para llevar a cabo análisis genéticos poblacionales exhaustivos en alineaciones de múltiples secuencias. Tanto *D* como *H* son estadísticos empleados para identificar desviaciones de los escenarios de neutralidad. *D* destaca por su eficacia al rechazar la neutralidad en presencia de un exceso de variantes a baja frecuencia, mientras que *H* es más eficaz al rechazar la neutralidad cuando hay un exceso de variantes derivadas (distintas a las no ancestrales, a *D. simulans* en este caso) a alta frecuencia.

Durante la ejecución del análisis utilizando *sliding windows* de 2.000 bases y saltos de 1.000 *nt* para el indicador estadístico de Fay-Wu, se definieron las coordenadas de cada región. Estos fragmentos, que superan las mil bases, se emplearon posteriormente en la realización de los otros dos indicadores estadísticos (*D* de Tajima y π). Es relevante señalar que la extensión de los fragmentos no está limitada por las coordenadas del alineamiento, sino por las posiciones analizadas.

La significación estadística fue obtenida a través de simulaciones utilizando el programa mlcoalsim[38]. Estas simulaciones se llevaron a cabo teniendo en cuenta la recombinación de cada región genómica. La estimación del parámetro de recombinación de la población (R=4Nr) se realizó utilizando la frecuencia de recombinación de cada fragmento proporcionado por Orengo (Tabla 2) según datos de Hey & Kliman, 2002[39]. La población efectiva de *D. melanogaster* ($N=10^6$), según se utilizó en el estudio de Orengo y Aguadé de 2004[2]. Para contrarrestar el problema de las comparaciones múltiples se aplica la corrección de Bonferroni secuencial para cada conjunto de tests de una misma población[40].

Tabla 2: Ratio de recombinación por fragmento analizado.

Brazo del cromosoma	Fragmento	Recombinación (cM/Mb)
2L	24E	3,909
	27D	4,518
	32A	3,151
	33A	2,812
2R	55C	3,89
	55I	3,938
	57A	3,686
	59B	3,212

Para evaluar la diferenciación genética entre las poblaciones, se llevó a cabo un análisis de flujo genético y diferenciación genética usando mstatspop[41]. Este programa se utiliza para realizar análisis de variabilidad entre distintas poblaciones, con el cual se calculó el estadístico F_{st}, utilizado en genética de poblaciones para medir la diferenciación genética. La significación estadística se obtuvo por el método de las permutaciones[42]. Con los estadísticos

obtenidos, se realizó un análisis de coordenadas principales[43] (PCoA, en inglés, *Principal Coordinates Analysis*) con la ayuda de GenAlEx[44]. El PCoA es una técnica para visualizar similitudes o distancias entre muestras en un espacio multidimensional reducido.

4. Resultados y discusión

Estos son los resultados obtenidos en el estudio realizado.

Como previamente se ha indicado, se ha usado el número de posiciones segregantes y la diversidad nucleotídica. Con la intención de conocer la neutralidad, se ha usado tanto *D* como *H* para identificar desviaciones de los escenarios de neutralidad.

La tabla 3 muestra los resultados para los polimorfismos de las regiones 24E, 27D, 32A, 33A, 55C, 55I, 57A y 59B de las 6 poblaciones analizadas.

Tabla 3: Polimorfismos de las regiones analizadas por población. Las posiciones presentadas corresponden al *release* número 5 de la base de datos de Popfly.

Pob.	Loc.	*N*	Posición	*L*	*S*	*π*	*D*	p-*D*	*Hn*	p-*Hn*
EF	24E	15	4033000-4074350	29365	784	0,00821	-0,25217	n.s.	-2,051	0,002*
EF	27D	15	7545000-75681450	18629	555	0,0089	-0,33334	n.s.	-2,546	0,001*
EF	32A	15	12178000-12200900	19369	681	0,01019	-0,35189	n.s.	0,343	n.s.
EF	33A	15	12850000-12911300	52652	981	0,005	-0,58468	n.s.	n.d.	n.d.
EF	55C	15	13950700-13961700	9876	233	0,00552	-1,1282	0,046	-2,822	0,001*
EF	55I	15	15150130-15156250	5576	181	0,00918	-0,35749	n.s.	-3,258	0,001*
EF	57A	15	17480000-17490000	8646	218	0,00774	-0,00861	n.s.	-2,650	0,005*
EF	59B	15	19002000-19010000	7489	170	0,00578	-0,75745	n.s.	-2,612	0,003*
RG	24E	15	4033000-4074350	33331	1299	0,01095	-0,48755	n.s.	-2,499	0,001*
RG	27D	15	7545000-75681450	18663	684	0,01022	-0,51194	n.s.	-3,104	0,001*
RG	32A	15	12178000-12200900	19062	847	0,01277	-0,35801	n.s.	-2,923	0,001*
RG	33A	15	12850000-12911300	54067	1411	0,00726	-0,4703	n.s.	n.d.	n.d.
RG	55C	15	13950700-13961700	10139	381	0,00973	-0,70168	n.s.	-3,088	0,001*
RG	55I	15	15150130-15156250	5631	218	0,01133	-0,23408	n.s.	-3,096	0,001*
RG	57A	15	17480000-17490000	9097	316	0,00879	-0,81998	n.s.	-2,806	0,002*
RG	59B	15	19002000-19010000	7542	230	0,0084	-0,60661	n.s.	-2,128	n.s.
ZI	24E	15	4033000-4074350	33738	1608	0,0124	-0,79822	n.s.	-3,058	0,001*
ZI	27D	15	7545000-75681450	18139	803	0,01169	-0,7678	n.s.	-3,790	0,001*
ZI	32A	15	12178000-12200900	19529	1000	0,01358	-0,71888	n.s.	-2,990	0,001*
ZI	33A	15	12850000-12911300	52951	2018	0,00965	-0,86082	n.s.	n.d.	n.d.
ZI	55C	15	13950700-13961700	9883	390	0,01024	-0,8049	n.s.	-2,727	0,001*
ZI	55I	15	15150130-15156250	5546	254	0,01223	-0,5982	n.s.	-4,034	0,001*
ZI	57A	15	17480000-17490000	8985	333	0,00916	-0,9768	n.s.	-3,239	0,001*
ZI	59B	15	19002000-19010000	7493	280	0,0103	-0,59145	n.s.	-2,180	0,006*
RAL	24E	15	4033000-4074350	27808	724	0,00796	-0,04636	n.s.	-2,392	0,001*
RAL	27D	15	7545000-75681450	16433	412	0,00693	-0,5464	n.s.	-3,223	0,001*
RAL	32A	15	12178000-12200900	17840	478	0,00879	0,26582	n.s.	-2,315	0,002*
RAL	33A	15	12850000-12911300	41903	581	0,00393	-0,37872	n.s.	-2,594	0,003*

Pob.	Loc.	*N*	Posición	*L*	*S*	*π*	*D*	p-*D*	*Hn*	p-*Hn*
RAL	55C	15	13950700-13961700	7390	128	0,00513	-0,19828	n.s.	-3,545	0,001*
RAL	55I	15	15150130-15156250	4976	70	0,0035	-0,88188	n.s.	-3,157	0,001*
RAL	57A	15	17480000-17490000	8741	178	0,00505	-0,87846	n.s.	-2,949	0,003*
RAL	59B	15	19002000-19010000	6221	137	0,00614	-0,44313	n.s.	-2,694	0,001*
USI	24E	14	4033000-4074350	33188	727	0,00656	-0,2403	n.s.	-2,879	0,001*
USI	27D	13	7545000-75681450	18136	362	0,00705	0,38801	n.s.	-2,369	0,002*
USI	32A	12	12178000-12200900	19662	498	0,0083	-0,08749	n.s.	-2,391	0,001*
USI	33A	12	12850000-12911300	51496	642	0,00375	-0,44871	n.s.	n.d.	n.d.
USI	55C	15	13950700-13961700	9320	212	0,00733	0,14546	n.s.	-2,551	0,003*
USI	55I	15	15150130-15156250	5648	101	0,00552	0,0141	n.s.	-3,362	0,001*
USI	57A	15	17480000-17490000	8029	135	0,00449	-0,58125	n.s.	-2,329	0,006*
USI	59B	15	19002000-19010000	6523	112	0,00446	-0,71735	n.s.	-2,682	0,003*
USW	24E	15	4033000-4074350	23698	454	0,00491	-0,76063	n.s.	-3,243	0,001*
USW	27D	15	7545000-75681450	11203	185	0,00549	-0,11799	n.s.	-2,603	0,001*
USW	32A	15	12178000-12200900	15123	430	0,00839	-0,20743	n.s.	-3,282	0,001*
USW	33A	15	12850000-12911300	29113	364	0,00322	-0,7358	n.s.	-3,299	0,001*
USW	55C	15	13950700-13961700	5446	131	0,00733	-0,07184	n.s.	-3,250	0,001*
USW	55I	15	15150130-15156250	3834	63	0,0052	0,05312	n.s.	-3,067	0,001*
USW	57A	15	17480000-17490000	5566	100	0,00446	-0,8837	n.s.	-2,942	0,001*
USW	59B	15	19002000-19010000	5350	79	0,00422	-0,35529	n.s.	-3,203	0,001*

Nota. Pob., población a la que pertenece la secuencia; Loc., región cromosómica analizada; N, número de individuos analizados de la población; Posición, localización de la región analizada en el *release* 5 de *D. melanogaster*; L, número de nucleótidos analizados sin contar los *gaps*; *S*, número de sitios segregantes; *π*, diversidad nucleotídica; *D*, estadístico *D* de Tajima; p-*D*, significación de la *D* de Tajima; *Hn* estadístico *H* normalizado de Fay y Wu; p-*Hn* significación de *Hn* de Fay y Wu; n.s., no significativo, n.d., no determinado, se desconoce por qué el programa DNASP devolvía error como resultado para el cálculo de *H*. ***** indica que mantiene la significación al 0,05 tras realizar la corrección de Bonferroni secuencial.

Por un lado, en las regiones genómicas estudiadas, la *D* obtiene valores negativos y no llegan a ser estadísticamente significativos. Es por ello que no se tiene evidencia suficiente para afirmar que hay una desviación significativa de la neutralidad. Es posible que otros factores, como cambios demográficos o fluctuaciones aleatorias, estén contribuyendo a la variación observada en los datos. (Tabla 3).

Por otro lado, se observa una *Hn* negativa en todas las poblaciones del estudio poseyendo casi todas también una significación inferior al 0,05 incluso tras la corrección de Bonferroni. En este caso se observa un exceso de variantes derivadas a alta frecuencia en la mayoría las regiones analizados. Esto podría deberse a eventos demográficos, como una expansión poblacional después de un cuello de botella, o selección positiva reciente que ha llevado a la aparición y fijación rápida de variantes derivadas beneficiosas.

Para los resultados obtenidos de π, las poblaciones de ZI y RG muestran valores muy similares en las ocho regiones genómicas analizadas. Generalmente, no se detecta en la secuencia ninguna región que exhiba una reducción notable en la diversidad nucleotídica, lo cual sugiere la ausencia de algún efecto que altere el equilibrio genético (Tabla 3). En el caso de EF, se observa un perfil diferente, ya que esta población presenta valores de π menores que se traducen en máximos menores.

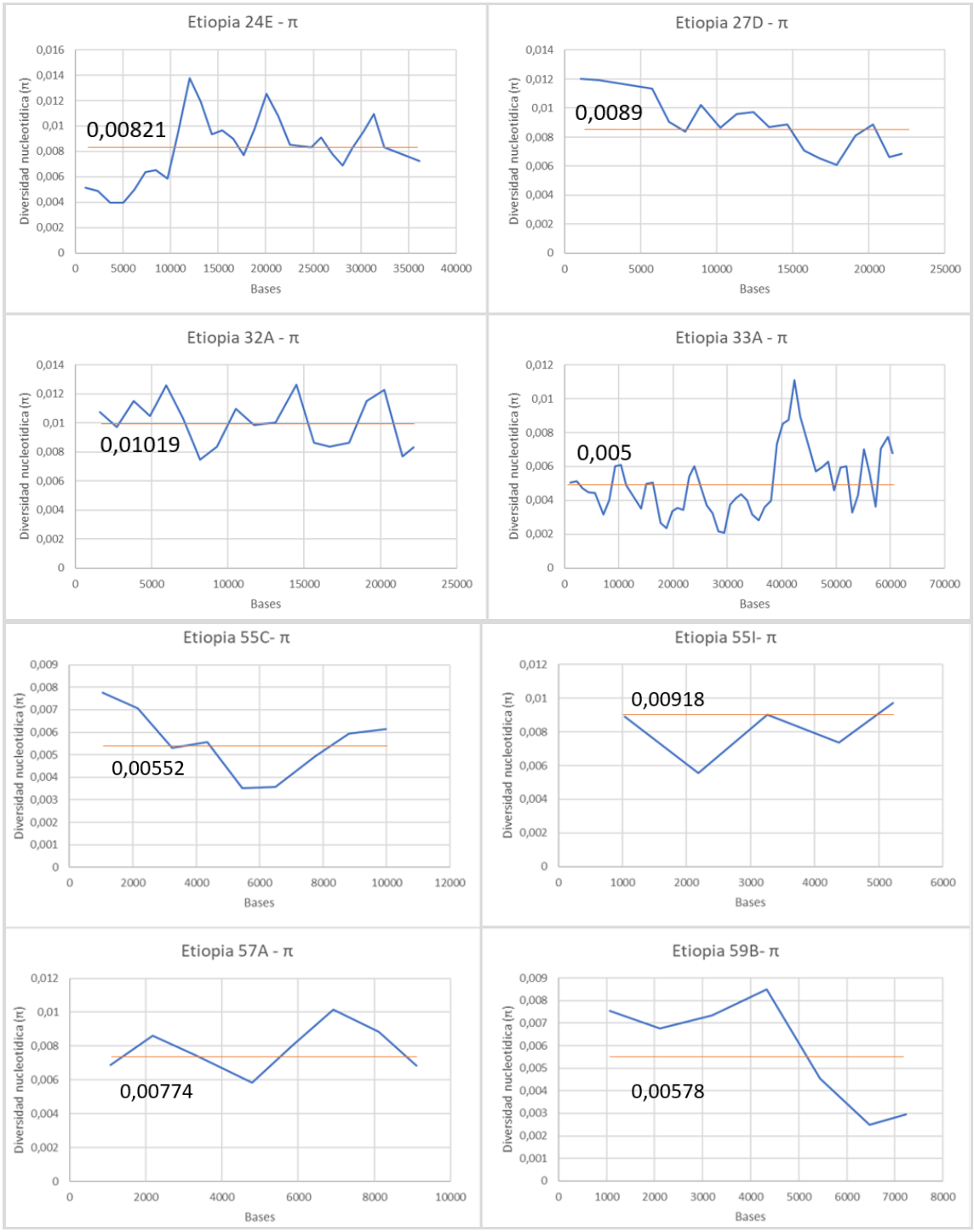


Figura 2: Gráficos de la distribución de π en las secuencias de EF. Nota: La línea naranja y el valor indican el valor de π obtenido para la región completa.

Sin embargo, al realizar un análisis por *sliding windows* se observa que la región 59B muestra un descenso en la diversidad nucleotídica en todas las poblaciones analizadas, mientras que el valor de *H* aumenta (Figura 2, Figura 3, Anexo 5). De acuerdo con la ubicación de la región 59B, los últimos fragmentos se sitúan cerca del gen *Dmel\Klp59C*. Este gen es necesario para la anafase, impulsando la separación de las cromátidas hermanas mediante la despolimerización activa de los microtúbulos cinetocóricos en sus extremos positivos asociados al cinetocoro[27]. Al estar estrictamente relacionado con la división celular, este gen debe de estar sometido a una fuerte selección purificadora. Así, la disminución de polimorfismos observada en el fragmento final de la región 59B se explicaría por un efecto de arrastre selectivo o *hitchhiking*, tanto en las poblaciones ancestrales como en las derivadas (Figura 4).

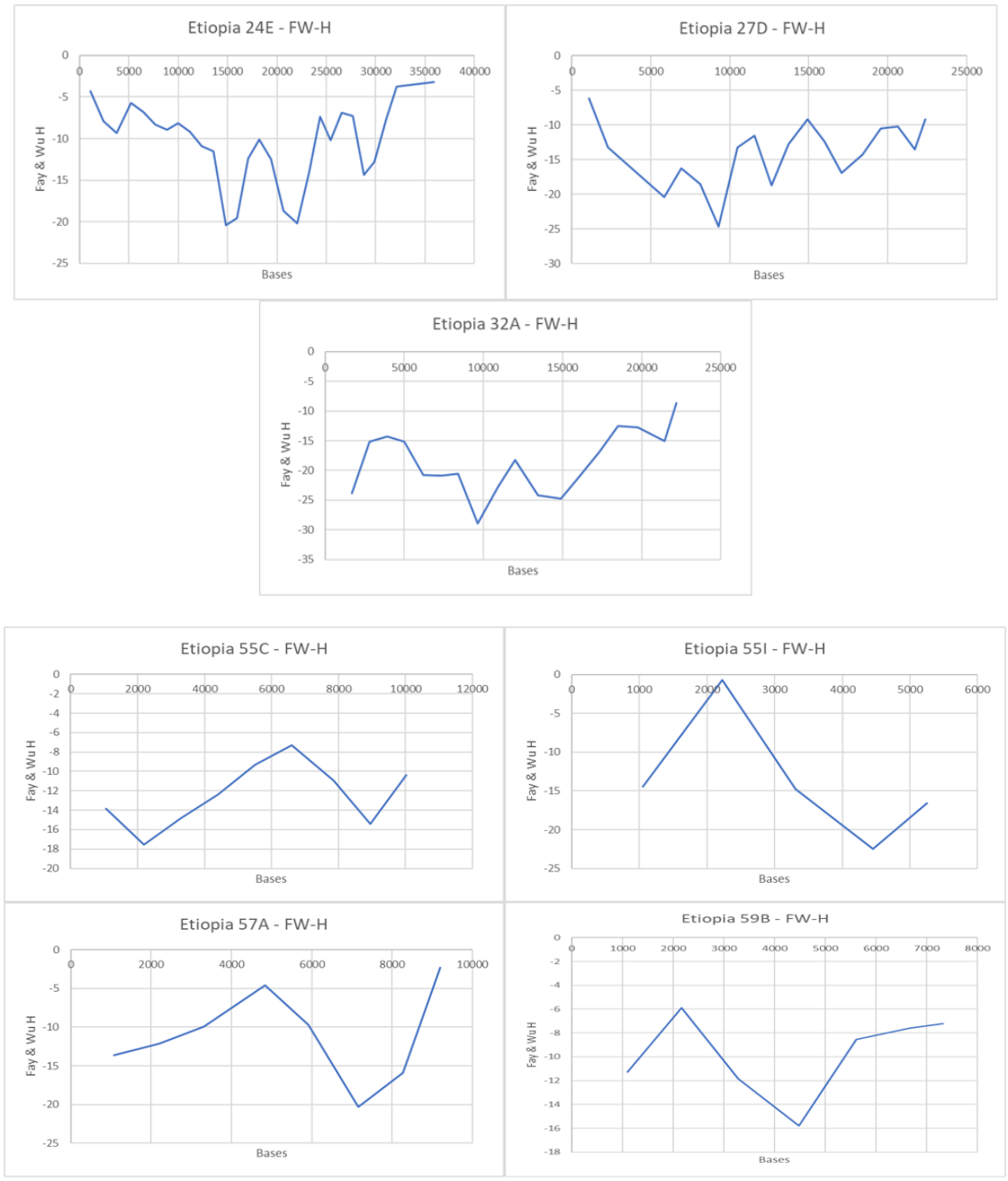


Figura 3: Gráficos de la distribución de *H* en las secuencias de EF.

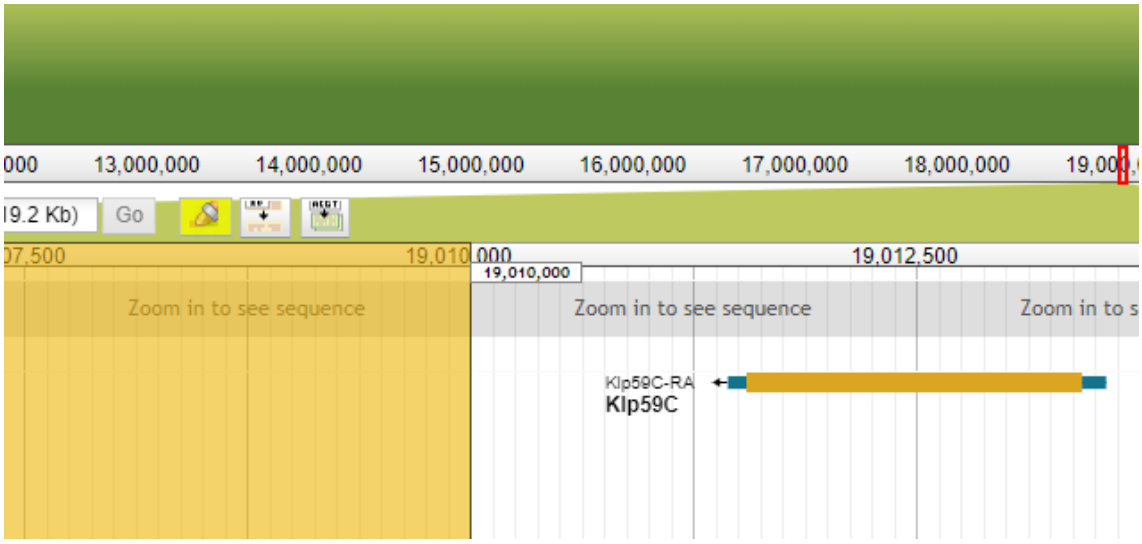
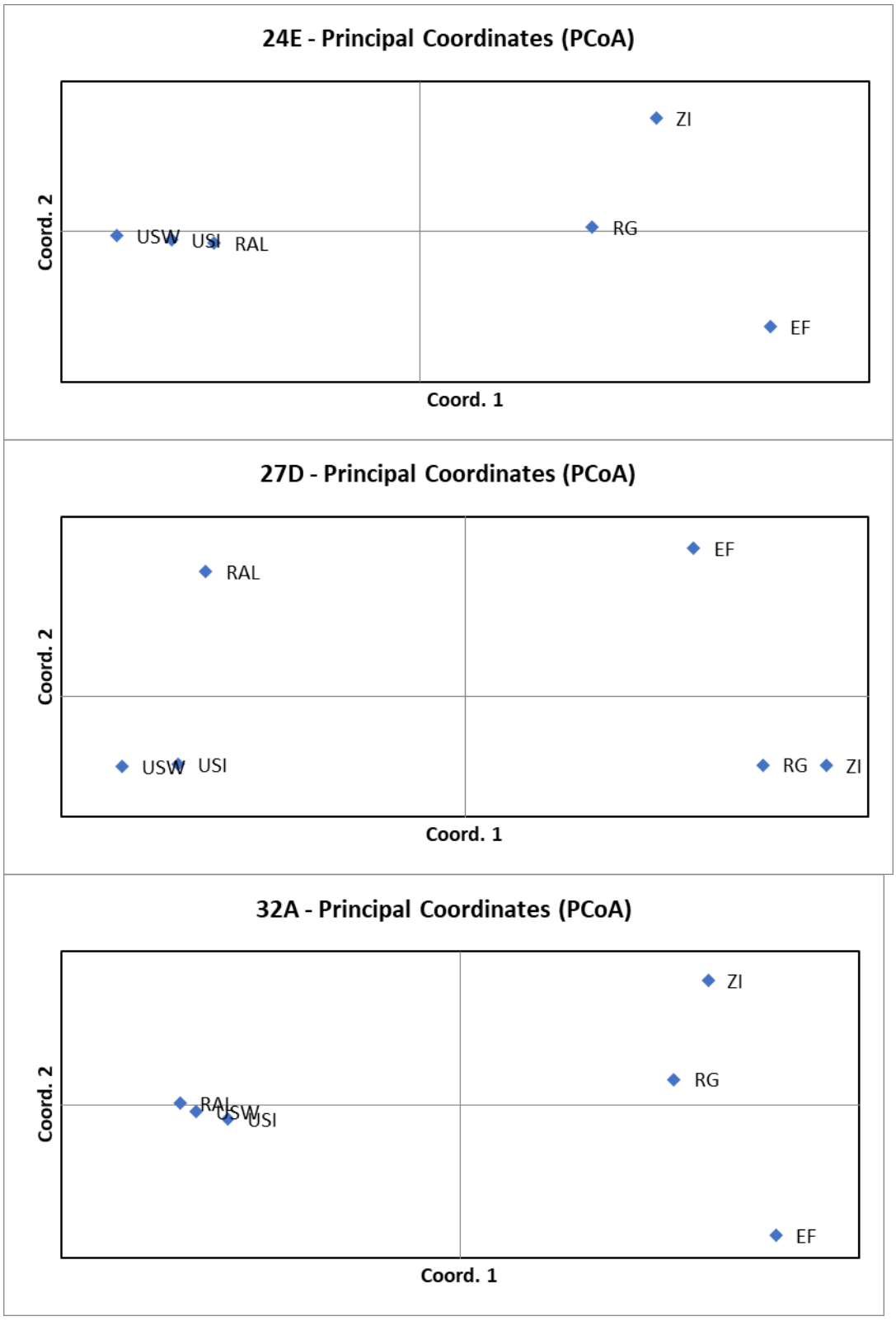
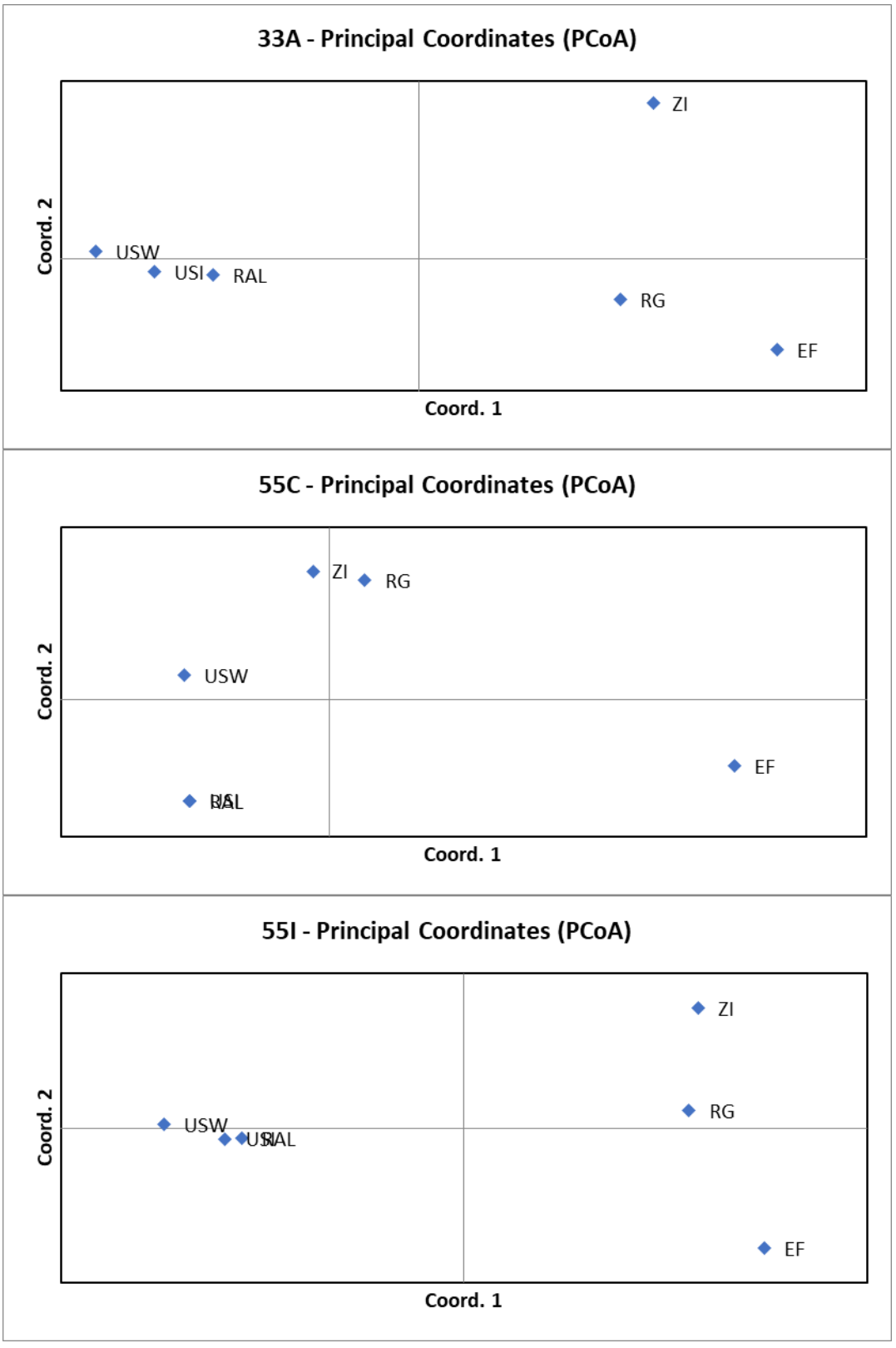


Figura 4: Localización del final de la región 59B (sombreado naranja) en el *browser* de PopFly, que muestra su cercanía al gen *Klp59C*[28].

En relación con el π observado en las poblaciones estadounidenses, los valores obtenidos son muy similares entre ellas con un resultado de bajos valores como el que presenta la población de EF (Tabla 3). La variabilidad nucleotídica es menor en las poblaciones estadounidenses (mayormente inferior a 0,009), y esto puede deberse a las expansiones de las poblaciones generadas después del cuello de botella que debió experimentar al colonizar América. De acuerdo con el estudio de Campo *et al*.[45] donde estudiaban la variabilidad genética de todos los cromosomas de las poblaciones USW y RAL, para el cromosoma 2R, se establecía un π promedio de 0,00584 para RAL y 0,00472 para USW, mientras que el 2L presentaba los siguientes valores; 0,00647 (RAL) y 0,00521 (USW). En el caso de USW, las regiones 33A (π = 0,00322) y 24E (π = 0,00497) son los más probables a mostrar un efecto en la reducción de la variación nucleotídica. Mientras que, en el caso de RAL serían; 33A (π = 0,00393) y 55I (π = 0,0035). Conjuntamente, en la gráfica de π para la región 55I de la población RAL (Anexo 5), se observa una variación nucleotídica muy baja.

La figura 5 muestra los resultados relativos a la diferenciación genética entre poblaciones.




33A - Principal Coordinates (PCoA)
ZI
USW
USI
RAL
RG
EF
Coord. 2
Coord. 1
55C - Principal Coordinates (PCoA)
ZI
RG
USW
EF
Coord. 2
Coord. 1
55I - Principal Coordinates (PCoA)
ZI
RG
USW
EF
Coord. 2
Coord. 1

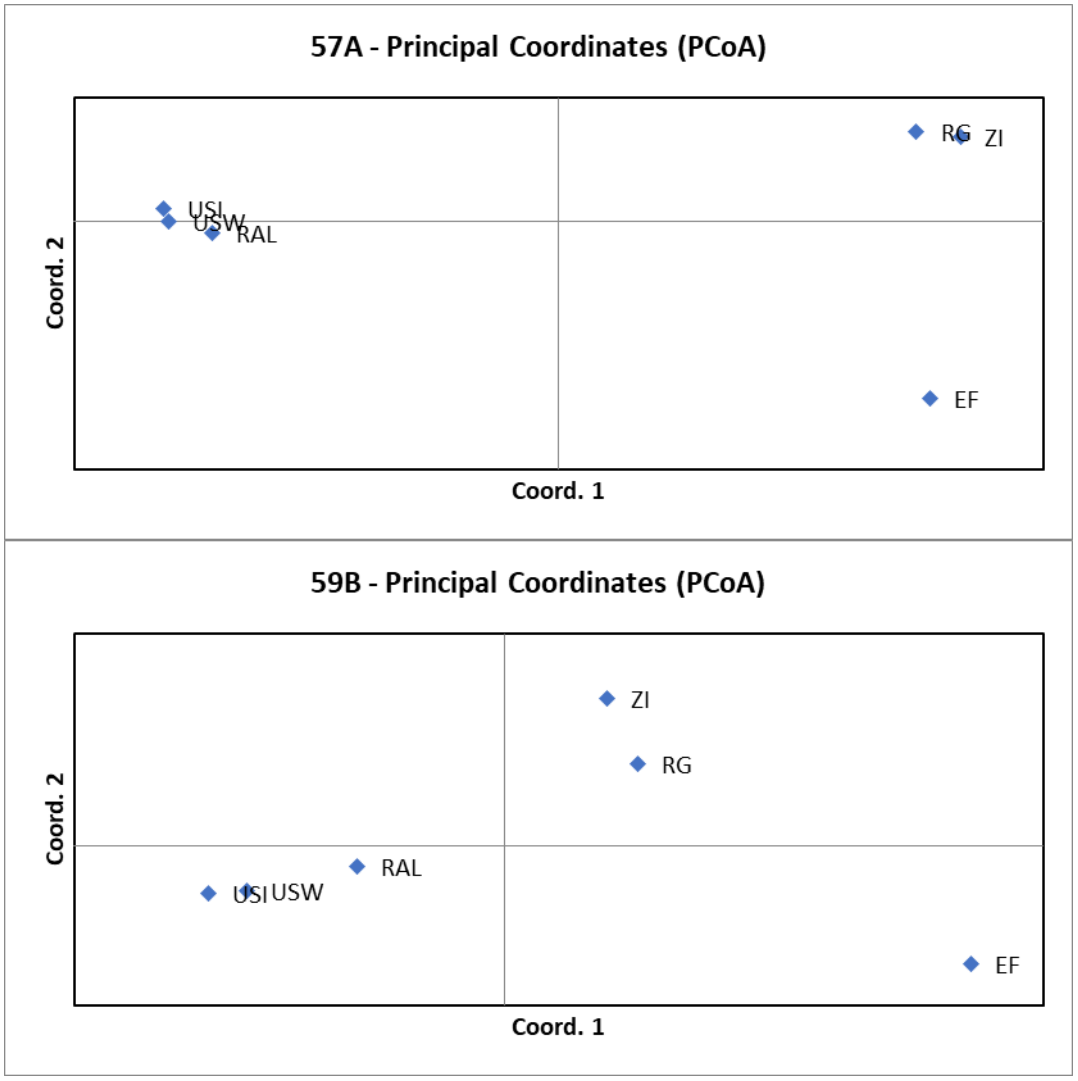


Figura 5: PCoA mostrando la diferenciación genética entre poblaciones para cada región genómica analizada.

Analizando la diferenciación genética por parejas de poblaciones (figura 4), es apreciable que no todas las poblaciones se han diferenciado entre ellas. En el caso de las tres poblaciones estadounidenses se observan valores de F_{st} negativos o inferiores a 0,05, y en su mayoría no llegan a ser estadísticamente significativos. Por lo general, estas tres poblaciones muestran cercanía genética.

En el caso de las poblaciones africanas, abordando el caso de RG-ZI, por un lado, en las regiones 55C (F_{st}: 0,028354), 55I (F_{st}: 0,011036) y 57A (F_{st}: 0,012974), se observa cercanía genética. Por otro lado, para el resto de las 5 regiones, los valores de F_{st} son algo más altos y con valores de p significativos, mostrando en estas regiones concretos distancia genética. Las poblaciones de Zambia y Ruanda se separaron hace aproximadamente 12,887 años, siendo en Zambia (ZI) la población donde se han observado mayor frecuencia de inversiones. Sin embargo, en este estudio, no se han utilizado individuos con inversiones. Es por ello por lo que el motivo de esta distanciación genética podría ser a causa de la distancia que hay entre las poblaciones o las distintas condiciones climáticas que hay entre las latitudes de las poblaciones (figura

6)[46]. Reparando a los climogramas de la figura 7, Zambia muestra un clima más seco y con temperaturas menos estables que Ruanda.

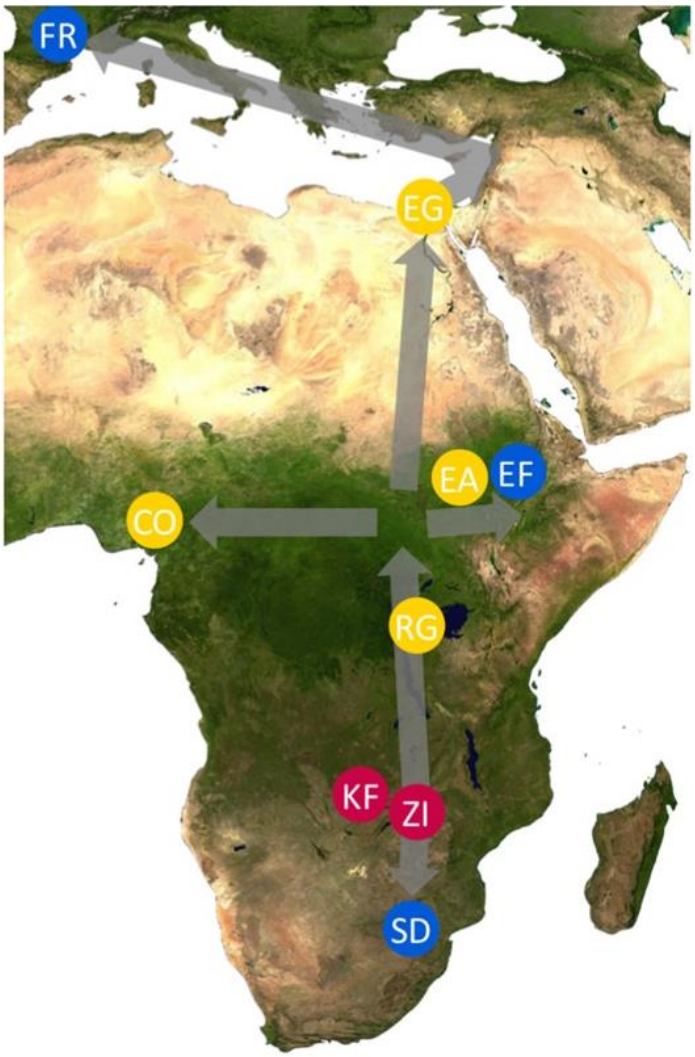


Figura 6: Mapa con las ubicaciones de diversas poblaciones de *D. melanogaster* en el área de origen[46].

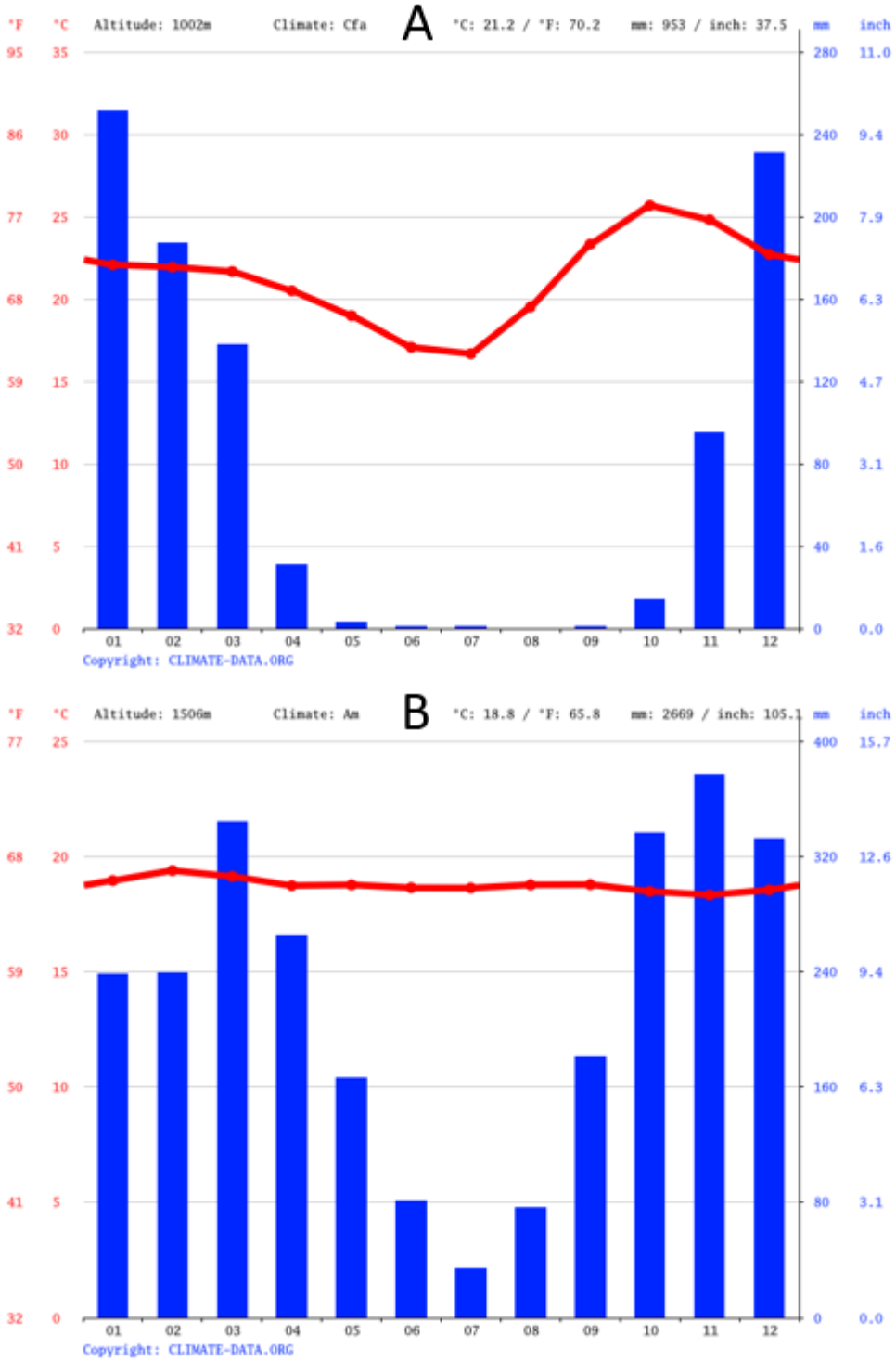


Figura 7: A; Climograma de Kafue (Zambia). B; Climograma de Gisenyi (Ruanda)[47,48].

En el caso de Etiopia (EF), la región 55I se ubica genéticamente más cerca de población RG con valores de F_{st} = 0,021822 aunque el valor de *p* denota poca significación, a pesar de ello, EF muestra una alta diferenciación al respecto de las otras dos poblaciones africanas. La diferenciación observada puede atribuirse a la ubicación geográfica de la población etíope, la cual se encuentra en el macizo etíope. Específicamente, la población etíope se localiza en la localidad de Fiche, a una altitud superior a los 3000 metros. Esta "barrera

natural" podría ser la razón de la diferenciación genética, ya que actúa como un obstáculo para el flujo genético entre ambas poblaciones[46].

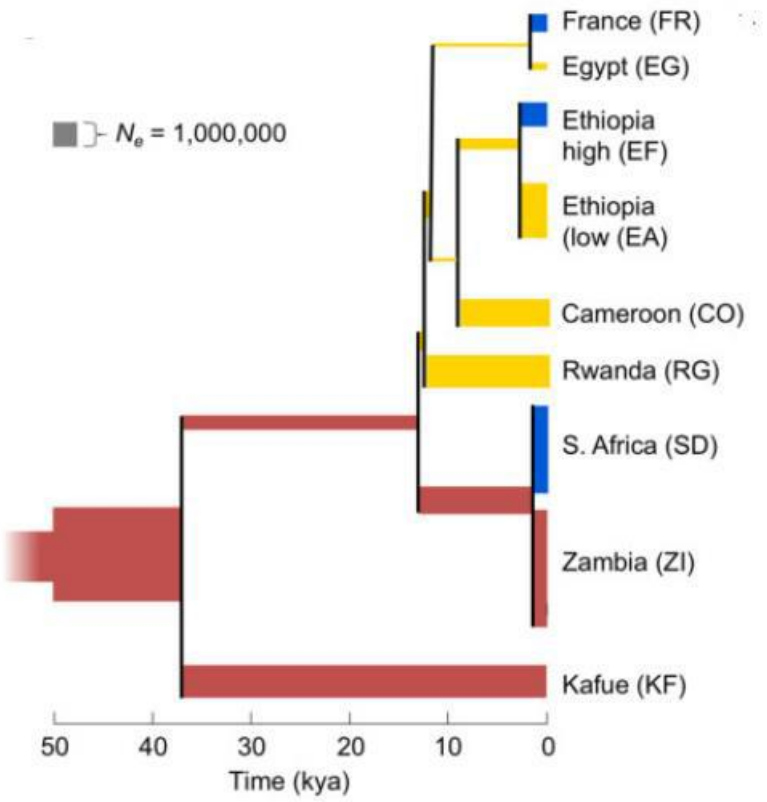


Figura 8: Historia evolutiva estimada de poblaciones de la región ancestral de *D. melanogaser*[46].

Analizando las poblaciones de manera global, era de esperar, que las poblaciones del área de origen de *D. melanogaster* respecto a las de EEUU, mostrasen una diferenciación alta en todas las regiones, mostrando valores relacionables con efecto fundador, cuello de botella y diferenciación alopátrica.

5. Conclusiones y trabajos futuros

En el presente estudio estas son las conclusiones obtenidas:

- Los valores *D* de Tajima son generalmente negativos, pero no significativos, por lo tanto, no se puede concluir de manera definitiva sobre la neutralidad en las variaciones a baja frecuencia. Es necesario considerar otros análisis y datos para obtener una comprensión más completa de los procesos evolutivos que pueden estar ocurriendo en las poblaciones estudiadas.
- En la mayoría de las regiones, de acuerdo con el estadístico normalizado *H* de Fay y Wu, donde en su mayoría se obtuvieron resultados con significación estadística, se rechaza la neutralidad en todas las poblaciones (desconociendo los resultados para 33A en las poblaciones EF, RG, ZI y USI), dando a entender que hay indicios de barrido selectivo y señales de selección positiva en estas regiones estudiadas.
- Conjuntamente, en las poblaciones no se observan valores con disminuciones notables de la diversidad nucleotídica. De hecho, si se comparan los resultados con estudios previos, se aprecian resultados similares. Sin embargo, para la región 59B al observar el *sliding windows*, se observa al final de la región una disminución en la diversidad nucleotídica. Esta disminución se asocia al gen cercano *Dmel\Klp59C* relacionado con la división celular. La importancia de este gen debe suponer que esté sometido a una selección purificadora fuerte que generaría la disminución de polimorfismos en su cercanía.
- Observando las tres poblaciones africanas en conjunto durante el análisis de diferenciación genética, se evidencia la existencia de diferenciación genética entre ellas. Mostrando una mayor distancia entre EF respecto a RG y ZI. Esto no se debe únicamente al efecto de la distancia, ya que las poblaciones de ZI y RG se encuentran a una distancia similar a las de RG y EF. Es posible que la población EF haya quedado más aislada de las otras dos debido a la orografía y que las condiciones climáticas a las que esté sometida sean algo distintas, favoreciendo que haya evolucionado independientemente. Respecto a las poblaciones estadounidenses, estas muestran resultados de cercanía entre ellas en la mayoría de las regiones estudiadas.

Aunque la confirmación completa de la presencia de señales de selección natural no se haya logrado, se han encontrado fuertes indicios que sugieren su posible existencia.

Al inicio del semestre se establecieron objetivos generales que se fueron refinando a medida que el curso progresaba. Desde el principio, se trabajó para alcanzar estos objetivos establecidos, logrando cumplir con la mayoría de todos ellos. El objetivo específico de comparar las secuencias con la población de Barcelona no se ha podido cumplimentar a causa de atrasos generados con problemas con los softwares de análisis.

La planificación y la metodología fueron establecidos al comienzo del curso. Debido al desconocimiento de las herramientas necesarias para el desarrollo del proyecto se fijaron las fechas con cierto margen para iniciarse en el correcto funcionamiento de las herramientas. Se registró un desvío cuando se inició la tarea de elección de individuos idóneos, el servicio de Popfly no se hallaba en funcionamiento, generando un atraso menor. Conjuntamente, el horario laboral del estudiante repercutió en la cumplimentación de distintos hitos a lo largo del desarrollo del TFM. Además, a la hora de compilar el programa mstatspop, surgieron diversos problemas con el sistema operativo del PC del estudiante, que acabaron solventándose mediante una máquina virtual con Linux. Gracias a los días de vacaciones planificados para la dedicación absoluta al trabajo, el estudiante pudo remediar dichos desvíos.

Hay diversas posibilidades para futuras líneas de estudio que no se han explorado. Relacionado con la no cumplimentación del objetivo específico de comparar los resultados con la población de Barcelona, sería interesante, por un lado, estudiar las regiones analizadas individualmente, dividiéndolas en subfragmentos para esclarecer con mayor precisión cuáles son los *loci* en los que se muestran rastros de evolución y conocer la procedencia del rechazo de la neutralidad. Los fragmentos de la población barcelonesa constaban de pocas pares de bases (~800bp), es por ello que realizando un estudio más específico de las regiones analizadas se podrían obtener resultados más precisos de si dichos fragmentos son candidatos a un barrido selectivo. De la misma manera, en este estudio solo se utilizaron individuos con ordenación estándar, sin embargo, cabría la posibilidad de realizar el estudio con los individuos que presentaban inversiones en ambos brazos del cromosoma 2, tratándolos como poblaciones distintas. Finalmente, realizar la fase de diferenciación genética utilizando otros estadísticos podría revelar información distinta a la obtenida en el presente estudio.

Desafortunadamente, debido a la complejidad de estudio que demandan todas las preguntas planteadas recientemente, no ha sido factible profundizar en su análisis.

6. Glosario

BLAST: Basic Local Alignment Search Tool. Herramienta informática que busca regiones parecidas entre secuencias biológicas.

D: Estadístico *D* de Tajima que se utiliza para distinguir si una secuencia evoluciona de modo neutro o no.

DnaSP: DNA Sequence Polymorphism. Herramienta bioinformática que se utiliza para varios análisis con secuencia de ADN

EF: Población de Fiche (Etiopia)

F_{st}: test estadístico que sirve para detectar diferenciación genética.

H: Estadístico *H* de Fay y Wu que sirve para distinguir entre una secuencia que evoluciona de modo neutro o que evoluciona bajo el efecto de la selección positiva.

PCoA: Análisis de Coordenadas Principales (PCoA), para visualizar similitudes o distancias entre muestras en un espacio multidimensional reducido.

RAL: Población de Raleigh (EEUU)

RG: Población de Gikongoro (Ruanda)

USI: Población de Ithaca (EEUU)

USW: Población de Winters (EEUU)

ZI: Población de Siavonga (Zambia)

π: Diversidad nucleotídica. Número promedio de diferencias por nucleótido.

7. Bibliografía

1. Gillespie, J. H. (1991). The causes of molecular evolution. Oxford University Press.

2. Orengo, D. J., & Aguadé, M. (2004). Detecting the footprint of positive selection in a European population of *Drosophila melanogaster*: multilocus pattern of variation and distance to coding regions. Genetics, 167(4), 1759-1766.

3. Sokolowski, M. B. (1985). Genetics and ecology of *Drosophila melanogaster* larval foraging and pupation behaviour. Journal of insect physiology, 31(11), 857-864.

4. Lachaise, D., Harry, M., Solignac, M., Lemeunier, F., Benassi, V., & Cariou, M. L. (1988). Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from São Tomé. Proceedings of the Royal Society of London. Series B. Biological Sciences, 239(1294), 75-84.

5. Korf, I., Yandell, M., & Bedell, J. (2003). Blast. " O'Reilly Media, Inc.".

6. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., Sánchez-Gracia, A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular biology and evolution, 34, 12, 3299-3302.

7. Stephan, W. (2019). Selective sweeps. Genetics, 211(1), 5-13.

8. Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. Genetics, 134(4), 1289-1303.

9. Hebert, P. D., Ratnasingham, S., & De Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270, 96-99.

10. Sánchez, J. A. (2016). Identificación y caracterización de genes con potencial bioinsecticida utilizando *Drosophila melanogaster* como sistema modelo.

11. Prados, A. B (2016). Estimación de la huella de la selección natural y el efecto Hill-Robertson a lo largo del genoma de *Drosophila melanogaster*.

12. Stephan, W., & Li, H. (2007). The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity, 98(2), 65-68.

13. Montgomery, M. E., Woodworth, L. M., England, P. R., Briscoe, D. A., & Frankham, R. (2010). Widespread selective sweeps affecting microsatellites in *Drosophila* populations adapting to captivity: implications for captive breeding programs. Biological Conservation, 143(8), 1842-1849.

14. Schlenke, T. A., & Begun, D. J. (2004). Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. Proceedings of the National Academy of Sciences, 101(6), 1626-1631.

15. Panigrahi, M., Rajawat, D., Nayak, S. S., Ghildiyal, K., Sharma, A., Jain, K., & Dutt, T. (2023). Landmarks in the history of selective sweeps. Animal Genetics.

16. Hudson RR, Kreitman M, Aguadé M. 1987. A Test of Neutral Molecular Evolution Based on Nucleotide Data. Genetics 116:153–159.

17. McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh *locus* in *Drosophila*. Nature 351:652–654.

18. Kapun, M., Barrón, M. G., Staubach, F., Obbard, D. J., Wiberg, R. A. W., Vieira, J., et al., (2020). Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. Molecular Biology and Evolution, 37(9), 2661-2678.

19. Beisswanger, S., Stephan, W., & De Lorenzo, D. (2006). Evidence for a selective sweep in the wapl region of *Drosophila melanogaster*. Genetics, 172(1), 265-274.

20. Svetec, N., Pavlidis, P., & Stephan, W. (2009). Recent strong positive selection on *Drosophila melanogaster HDAC6*, a gene encoding a stress surveillance factor, as revealed by population genomic analysis. Molecular biology and evolution, 26(7), 1549-1556.

21. Kolaczkowski, B., Kern, A. D., Holloway, A. K., & Begun, D. J. (2011). Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics, 187(1), 245-260.

22. Rogers, R. L., & Hartl, D. L. (2012). Chimeric genes as a source of rapid evolution in *Drosophila melanogaster*. Molecular biology and evolution, 29(2), 517-529.

23. Orengo, D. J., & Aguadé, M. (2007). Genome scans of variation and adaptive change: extended analysis of a candidate locus close to the *phantom* gene region in *Drosophila melanogaster*. Molecular biology and evolution, 24(5), 1122-1129.

24. Orengo, D. J., & Aguadé, M. (2010). Uncovering the footprint of positive selection on the X chromosome of *Drosophila melanogaster*. Molecular biology and evolution, 27(1), 153-160.

25. Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., et al. (2002). A single P450 allele associated with insecticide resistance in *Drosophila*. Science, 297(5590), 2253-2256.

26. Lack, J. B., Lange, J. D., Tang, A. D., Corbett-Detig, R. B., & Pool, J. E. (2016). A thousand fly genomes: an expanded *Drosophila* genome nexus. Molecular biology and evolution, 33(12), 3308-3313.

27. Gramates, L. S., Agapite, J., Attrill, H., Calvi, B. R., Crosby, M. A., Dos Santos, G., Goutte-Gattat D, Jenkins V, Kaufman T, Larkin A, Matthews B, Millburn G, Strelets VB, and the FlyBase Consortium (2022). FlyBase: a guided tour of highlighted features. Genetics, 220(4).

28. Hervas, S., Sanz, E., Casillas, S., Pool, J. E., & Barbadilla, A. (2017). PopFly: the *Drosophila* population genomics browser. Bioinformatics, 33(17), 2779-2780.

29. David, J. R., & Capy, P. (1988). Genetic variation of *Drosophila melanogaster* natural populations. Trends in Genetics, 4(4), 106-111.

30. Mackay, T. F., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., et al. (2012). The *Drosophila melanogaster* genetic reference panel. Nature, 482(7384), 173-178.

31. Grenier, J. K., Arguello, J. R., Moreira, M. C., Gottipati, S., Mohammed, J., Hackett, S. R., ... & Clark, A. G. (2015). Global diversity lines–a five-continent reference panel of sequenced *Drosophila melanogaster* strains. G3: Genes, Genomes, Genetics, 5(4), 593-603.

32. Campos, J. L., Zeng, K., Parker, D. J., Charlesworth, B., & Haddrill, P. R. (2013). Codon usage bias and effective population sizes on the X chromosome versus the autosomes in *Drosophila melanogaster*. Molecular biology and evolution, 30(4), 811-823.

33. Nei, M. Molecular evolutionary genetics. (New York: Columbia University Press, 1987).

34. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595 (1989).

35. Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. Genetics 155, 1405–1413 (2000).

36. Zeng K, Fu YX, Shi S, Wu CI. 2006. Statistical tests for detecting positive selection by utilizing high-frequency variants. Genetics 174:1431–1439.

37. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., Sánchez-Gracia, A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular biology and evolution, 34, 12, 3299-3302

38. Ramos-Onsins, S. E. & Mitchell-Olds, T. Mlcoalsim: Multilocus coalescent simulations. Evol. Bioinforma. 3, 41–44 (2007).

39. Hey J, Kliman RM. 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. Genetics 160:595–608.

40. Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6, 65–70.

41. Ramos-Onsins S, Ferretti L, Raineri E, Marmorini G, Burgos-Paz W, Vera G. mstatspop [Internet].

42. Wright, Sewall. "Isolation by distance." Genetics 28.2 (1943): 114.

43. Rao, C. R. 1964. The use and interpretation of principal component analysis in applied research. Sankhya A 26: 329–358.

44. Peakall, R., & Smouse, P. E. (2012). GENALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics, 28, 2537–2539.

45. Campo, D., et al. "Whole-genome sequencing of two North American *Drosophila melanogaster* populations reveals genetic differentiation and positive selection." Molecular ecology 22.20 (2013): 5084-5097.

46. Sprengelmeyer, Q. D. *et al*. Recurrent Collection of *Drosophila melanogaster* from Wild African Environments and Genomic Insights into Species History. Mol. Biol. Evol. 37, 627–638 (2020).

47. https://es.climate-data.org/africa/ruanda/iburengerazuba/gisenyi-3391/#climate-graph, 04/01/2024

48. https://es.climate-data.org/africa/zambia/lusaka-province/kafue-26142/#climate-graph, 04/01/2024

8. Anexos

Anexo 1: Listado de individuos analizados

EF	RG	ZI	RAL	USI	USW
EF101N	RG10	ZI103	RAL-105	USI02	USW_26
EF103N	RG11N	ZI104	RAL-129	USI03	USW_35
EF116N	RG13N	ZI10	RAL-136	USI04	USW_37
EF119N	RG15	ZI126	RAL-138	USI06	USW_38
EF120	RG18N	ZI129	RAL-142	USI07	USW_40
EF120N	RG19	ZI134N	RAL-149	USI13	USW_43
EF122N	RG21N	ZI138	RAL-153	USI16	USW_47
EF126N	RG22	ZI152	RAL-158	USI17	USW_49
EF12N	RG24	ZI157	RAL-176	USI22	USW_50
EF131N	RG25	ZI164	RAL-177	USI23	USW_54
EF135N	RG28	ZI165	RAL-181	USI24	USW_59
EF136N	RG2	ZI167	RAL-189	USI26	USW_60
EF15N	RG32N	ZI170	RAL-195	USI29	USW_62
EF16N	RG33	ZI172	RAL-208	USI31	USW_63
EF19N	RG34	ZI173	RAL-217	USI33	USW_66

Anexo 2: Diferenciación genética entre las poblaciones por fragmento

Poblaciones	Fragmento	F_{st}	p-F_{st}
EF-ZI	24E	0,105601	0,001
EF-RG	24E	0,082798	0,001
EF-RAL	24E	0,23472	0,001
EF-USI	24E	0,264131	0,001
EF-USW	24E	0,317209	0,001
ZI-RG	24E	0,066838	0,001
ZI-RAL	24E	0,174015	0,001
ZI-USI	24E	0,194077	0,001
ZI-USW	24E	0,234495	0,001
RG-RAL	24E	0,12037	0,001
RG-USI	24E	0,140405	0,001
RG-USW	24E	0,184133	0,001
RAL-USI	24E	-0,007614	>0,05
RAL-USW	24E	0,01421	>0,05
USW-USI	24E	-0,006942	>0,05
EF-ZI	27D	0,082816	0,001
EF-RG	27D	0,04201	>0,05
EF-RAL	27D	0,0214985	0,001
EF-USI	27D	0,150411	0,001

Poblaciones	Fragmento	F_{st}	p-F_{st}
EF-USW	27D	0,169647	0,001
ZI-RG	27D	0,037608	>0,05
ZI-RAL	27D	0,168443	0,001
ZI-USI	27D	0,13059	0,001
ZI-USW	27D	0,143294	0,001
RG-RAL	27D	0,168443	0,001
RG-USI	27D	0,102643	0,001
RG-USW	27D	0,109275	0,001
RAL-USI	27D	0,025931	>0,05
RAL-USW	27D	0,03093	>0,05
USW-USI	27D	-0,005407	>0,05
EF-ZI	32A	0,064737	0,001
EF-RG	32A	0,048805	0,001
EF-RAL	32A	0,213398	0,001
EF-USI	32A	0,181107	0,001
EF-USW	32A	0,199488	0,001
ZI-RG	32A	0,035901	0,001
ZI-RAL	32A	0,16902	0,001
ZI-USI	32A	0,151256	0,001
ZI-USW	32A	0,161579	0,001
RG-RAL	32A	0,152923	0,001
RG-USI	32A	0,115024	0,001
RG-USW	32A	0,136619	0,001
RAL-USI	32A	-0,002319	>0,05
RAL-USW	32A	-0,010079	>0,05
USW-USI	32A	-0,013282	>0,05
EF-ZI	33A	0,194682	0,001
EF-RG	33A	0,108722	0,001
EF-RAL	33A	0,278961	0,001
EF-USI	33A	0,300963	0,001
EF-USW	33A	0,361727	0,001
ZI-RG	33A	0,149812	0,001
ZI-RAL	33A	0,241589	0,001
ZI-USI	33A	0,259471	0,001
ZI-USW	33A	0,284574	0,001
RG-RAL	33A	0,171115	0,001
RG-USI	33A	0,190938	0,001
RG-USW	33A	0,243247	0,001
RAL-USI	33A	0,045815	>0,05
RAL-USW	33A	0,059081	0,019
USW-USI	33A	0,006499	>0,05
EF-ZI	55C	0,267662	0,001

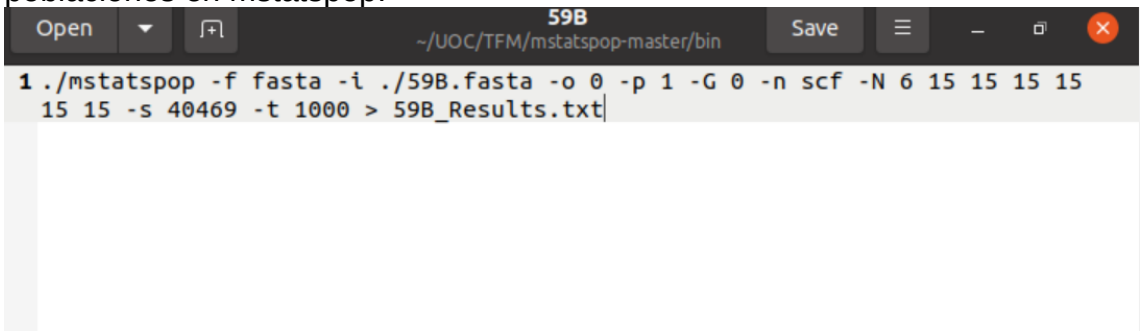
Poblaciones	Fragmento	F_{st}	p-F_{st}
EF-RG	55C	0,180391	0,001
EF-RAL	55C	0,355363	0,001
EF-USI	55C	0,356481	0,001
EF-USW	55C	0,377736	0,001
ZI-RG	55C	0,028354	>0,05
ZI-RAL	55C	0,057635	0,001
ZI-USI	55C	0,054865	0,002
ZI-USW	55C	0,068535	0,016
RG-RAL	55C	0,071349	0,001
RG-USI	55C	0,073619	0,002
RG-USW	55C	-0,017096	0,001
RAL-USI	55C	-0,035363	>0,05
RAL-USW	55C	-0,017096	>0,05
USW-USI	55C	-0,013205	>0,05
EF-ZI	55I	0,056429	0,001
EF-RG	55I	0,021822	>0,05
EF-RAL	55I	0,154135	0,001
EF-USI	55I	0,157759	0,001
EF-USW	55I	0,211018	0,001
ZI-RG	55I	0,011036	>0,05
ZI-RAL	55I	0,123343	0,001
ZI-USI	55I	0,126357	0,001
ZI-USW	55I	0,167599	0,001
RG-RAL	55I	0,101327	0,001
RG-USI	55I	0,109378	0,001
RG-USW	55I	0,149995	0,001
RAL-USI	55I	-0,037946	>0,05
RAL-USW	55I	-0,005559	>0,05
USW-USI	55I	-0,019372	>0,05
EF-ZI	57A	0,103348	0,001
EF-RG	57A	0,107064	0,001
EF-RAL	57A	0,23641	0,001
EF-USI	57A	0,268888	0,001
EF-USW	57A	0,259533	0,001
ZI-RG	57A	0,012974	>0,05
ZI-RAL	57A	0,235471	0,001
ZI-USI	57A	0,244191	0,001
ZI-USW	57A	0,243435	0,001
RG-RAL	57A	0,204162	0,001
RG-USI	57A	0,219141	0,001
RG-USW	57A	0,226944	0,001
RAL-USI	57A	-0,00892	>0,05

Poblaciones	Fragmento	F_{st}	p-F_{st}
RAL-USW	57A	0,000389	>0,05
USW-USI	57A	-0,028724	>0,05
EF-ZI	59B	0,207619	0,001
EF-RG	59B	0,152369	0,001
EF-RAL	59B	0,210473	0,001
EF-USI	59B	0,289778	0,001
EF-USW	59B	0,260492	0,001
ZI-RG	59B	0,051068	0,002
ZI-RAL	59B	0,107951	0,001
ZI-USI	59B	0,154805	0,001
ZI-USW	59B	0,140909	0,001
RG-RAL	59B	0,074113	0,001
RG-USI	59B	0,142305	0,001
RG-USW	59B	0,116765	0,001
RAL-USI	59B	0,028419	>0,05
RAL-USW	59B	0,013211	>0,05
USW-USI	59B	-0,008403	>0,05

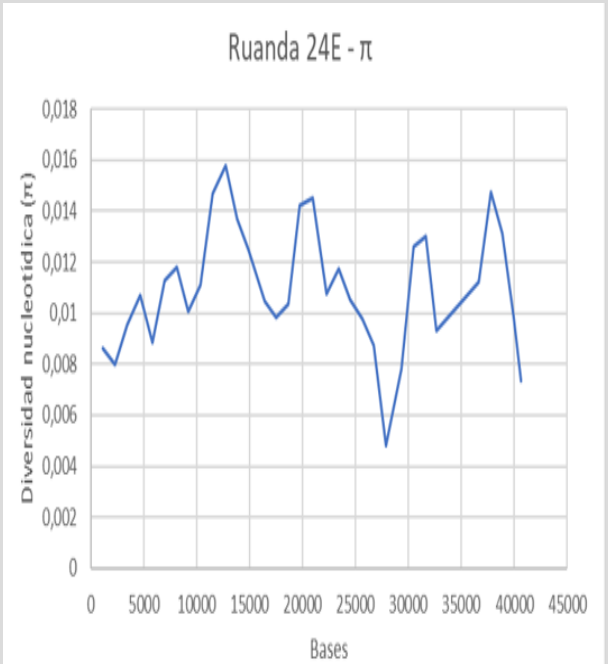
Anexo 3: Archivo de entrada para la simulación de las regiones de la población EF en mlcoalsim.

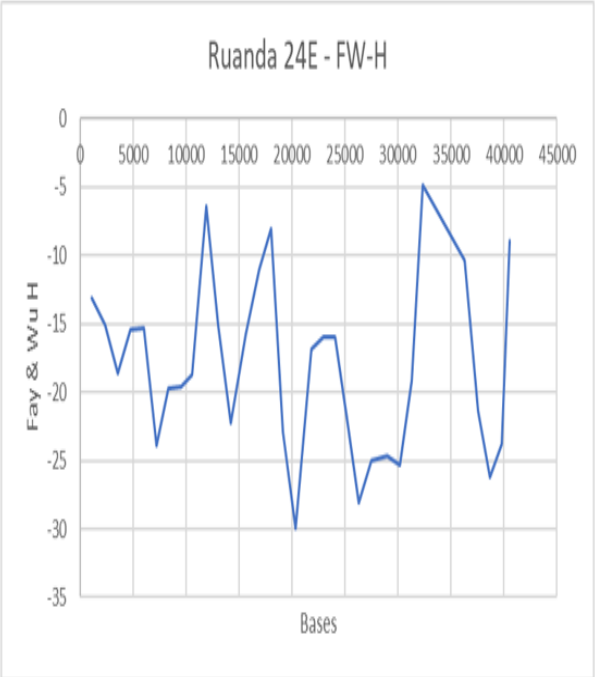
```
seed1 7354
print_matrixpol 0
print_neuttest 2

n_iterations 1000
n_loci 8
n_sites 41351 23151 22901 61301 11001 6121 10001 8001
n_samples 15 15 15 15 15 15 15 15
npop 1

recombination   15.636.000      18.072.000      12.604.000      11.248.000      15.560.000      15.752.000      14.744.000      12.848.000
mutations  918 500 651 979 247 148 212 170
factorn_chr 1 1 1 1 1 1 1 1
no_rec_males 1
likelihood_line 0

TD_obs 1 -0.25217 -0.33334 -0.35189 -0.58468 -1.1282 -0.35749 -0.00861 -0.75745
Hnorm_obs 1 -2.051 -2.546 0.343 0 -2.822 -3.258 -2.650 -2.621
```

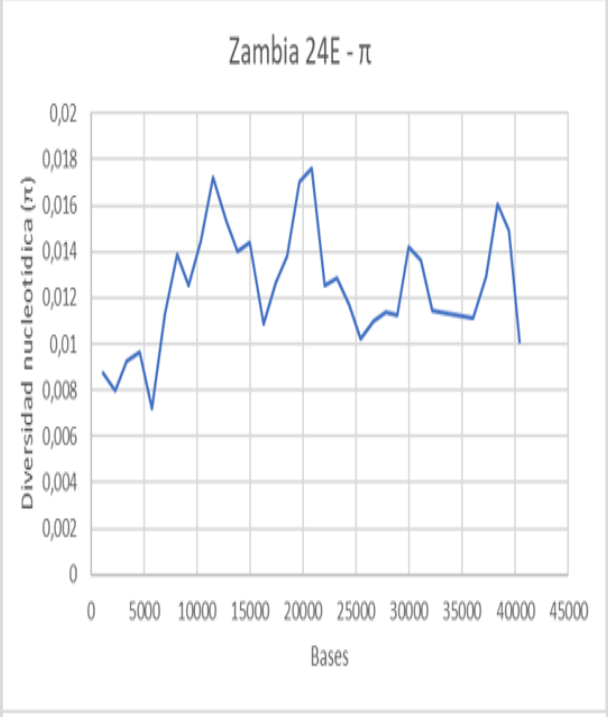
Anexo 4: Archivo de entrada para la simulación de la región 59B de todas las poblaciones en mstatspop.

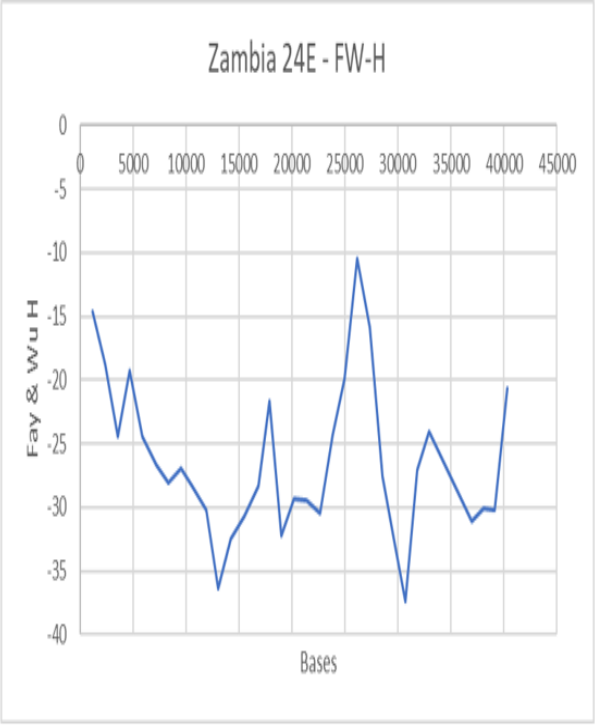


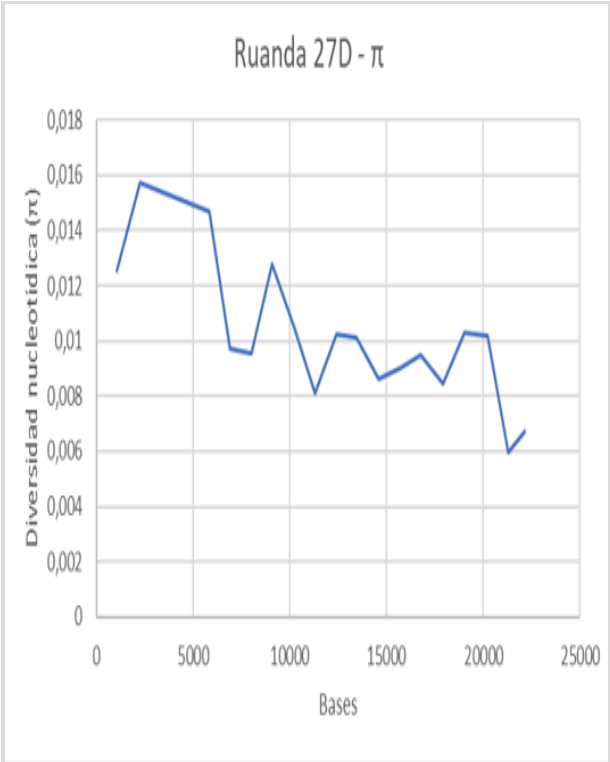
Anexo 5: Gráficos de *H* y *π* para las poblaciones de RG, ZI, RAL, USI y USW para las regiones 24E, 27D, 32A, 33A, 55C, 55I, 57A y57B.

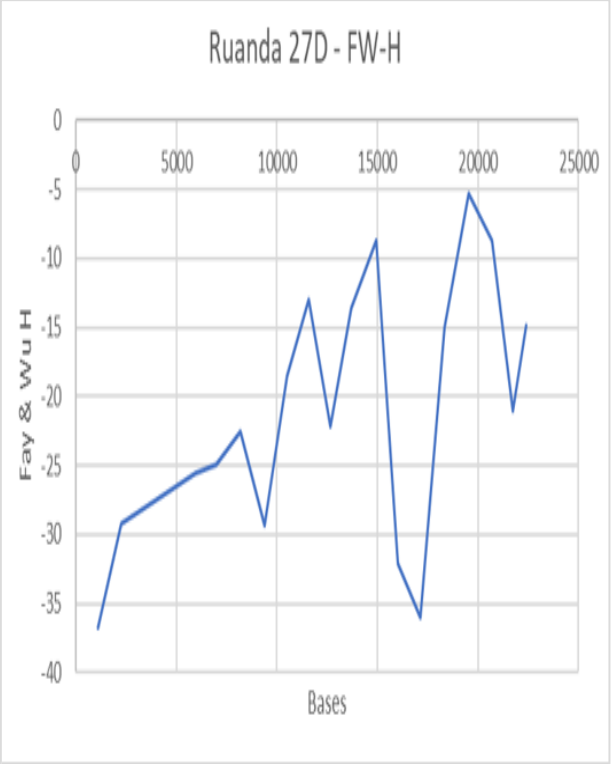


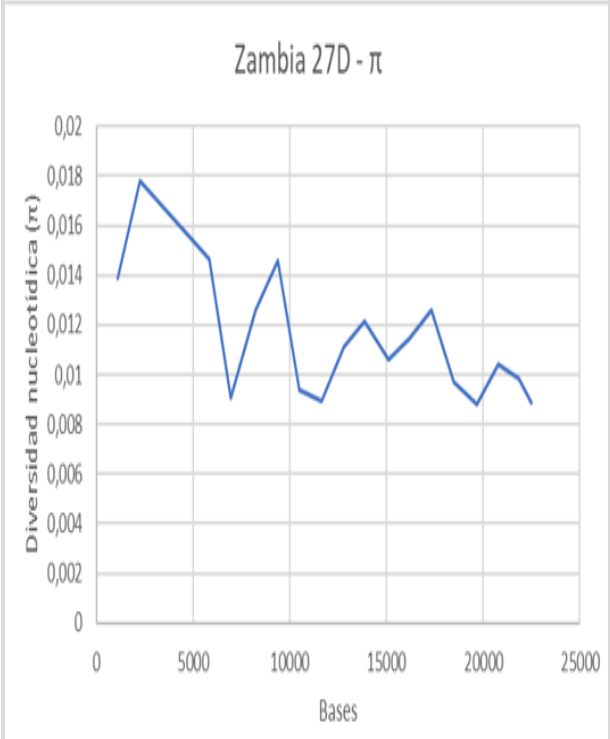


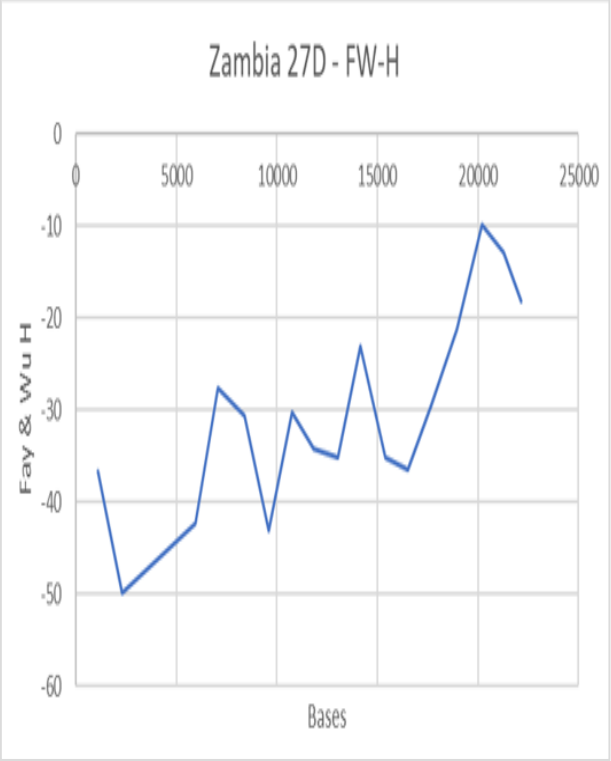


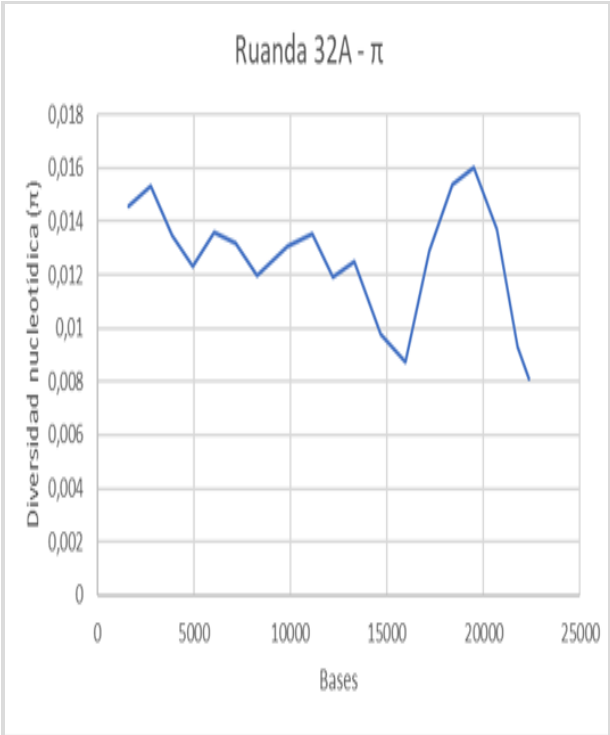


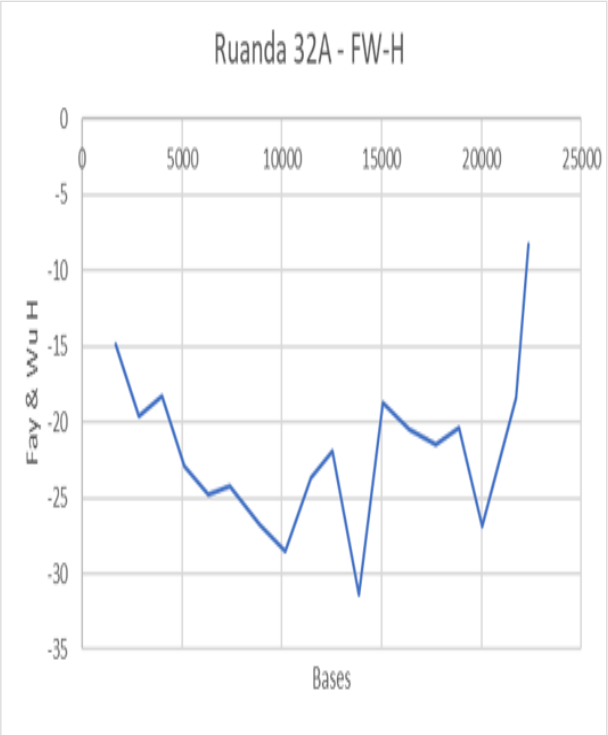

Ruanda 27D - π
Diversidad nucleotídica (π)
0,018
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
5000
10000
15000
20000
25000
Bases

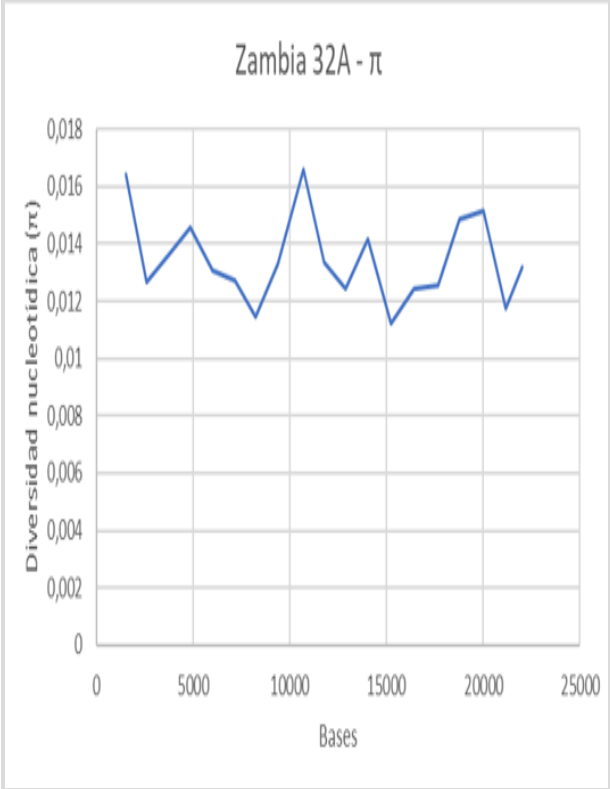

Ruanda 27D - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
-35
-40
0
5000
10000
15000
20000
25000
Bases

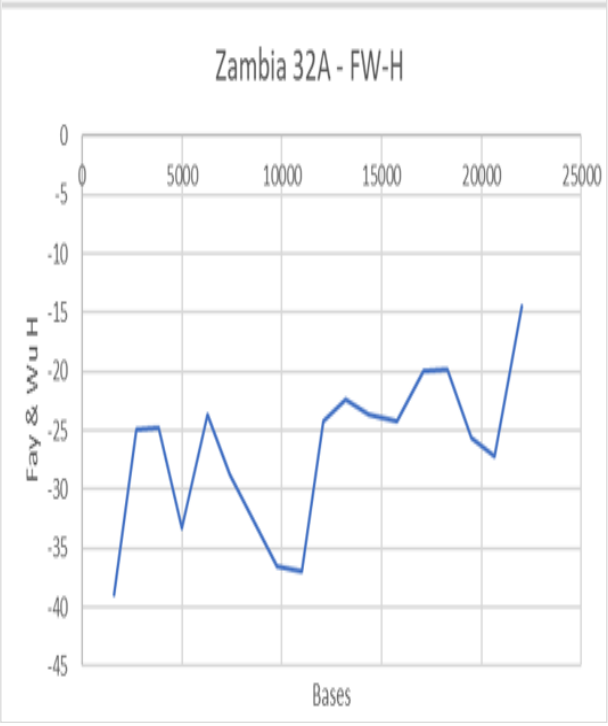

Zambia 27D - π
Diversidad nucleotídica (π)
0,02
0,018
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
5000
10000
15000
20000
25000
Bases

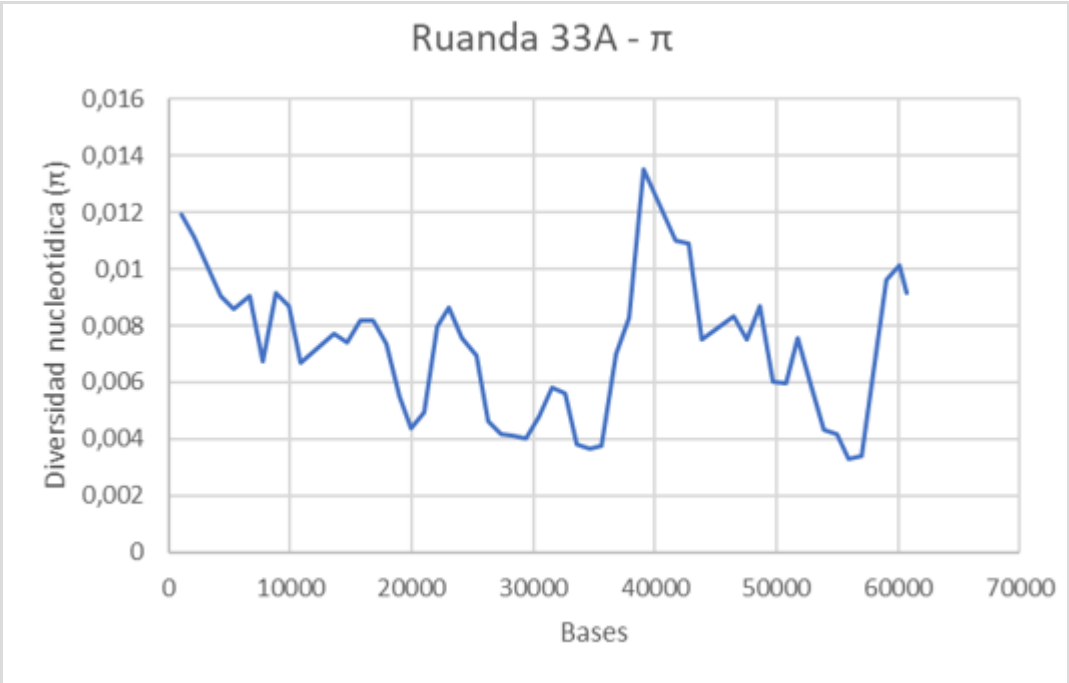

Zambia 27D - FW-H
Fay & Wu H
0
-10
-20
-30
-40
-50
-60
0
5000
10000
15000
20000
25000
Bases

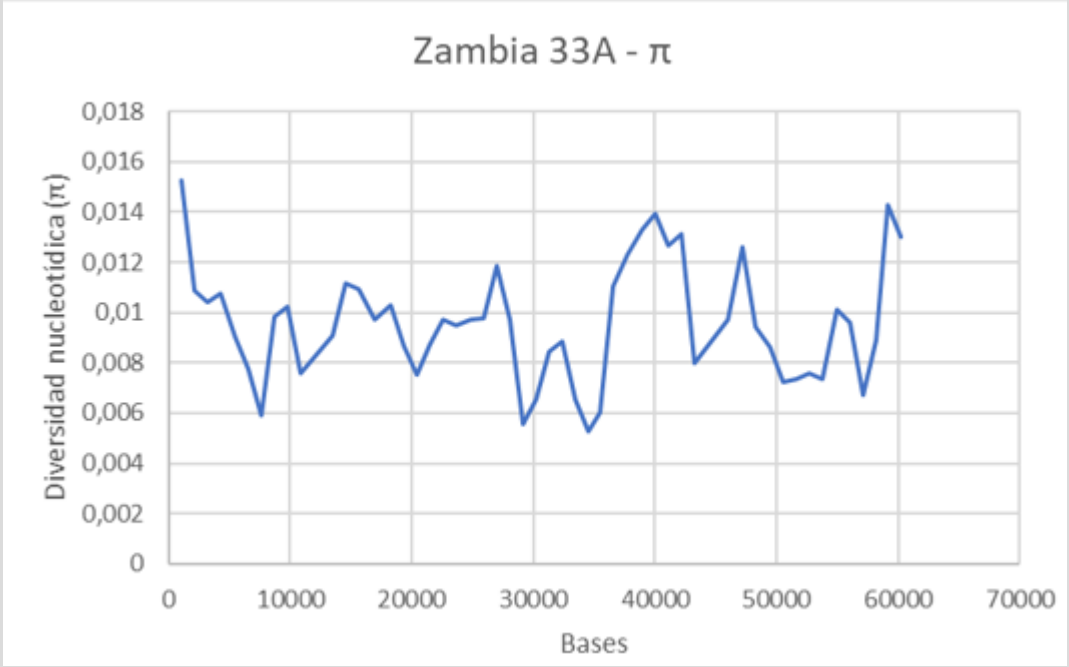

Ruanda 32A - π
Diversidad nucleotídica (π)
0,018
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
5000
10000
15000
20000
25000
Bases

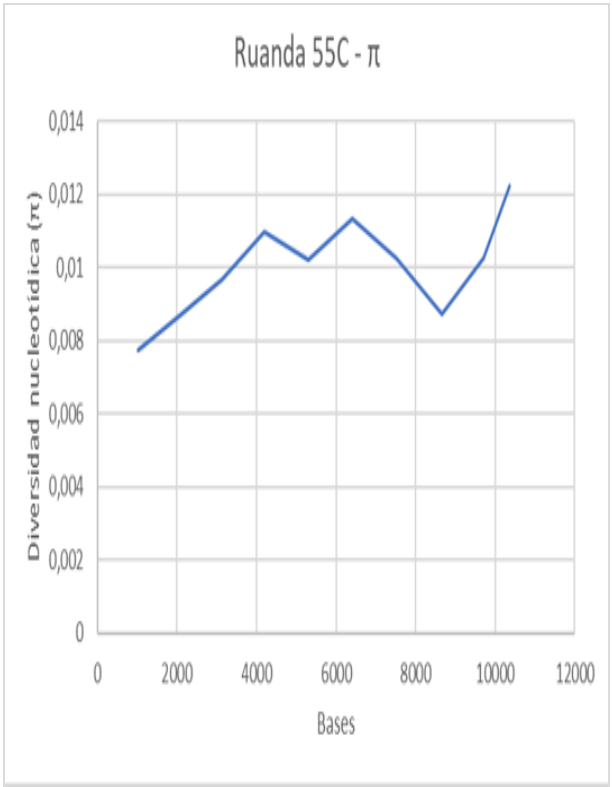

Ruanda 32A - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
-35
0
5000
10000
15000
20000
25000
Bases

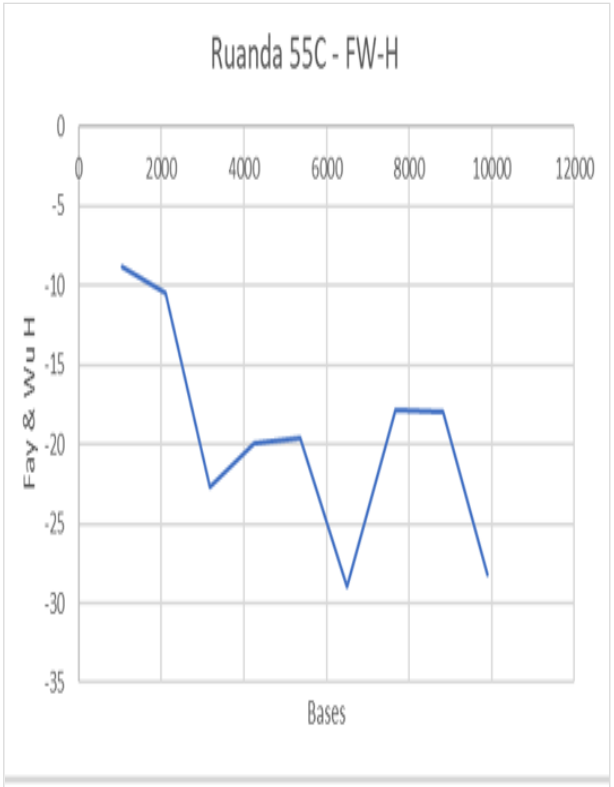

Zambia 32A - π
Diversidad nucleotídica (π)
0,018
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
5000
10000
15000
20000
25000
Bases

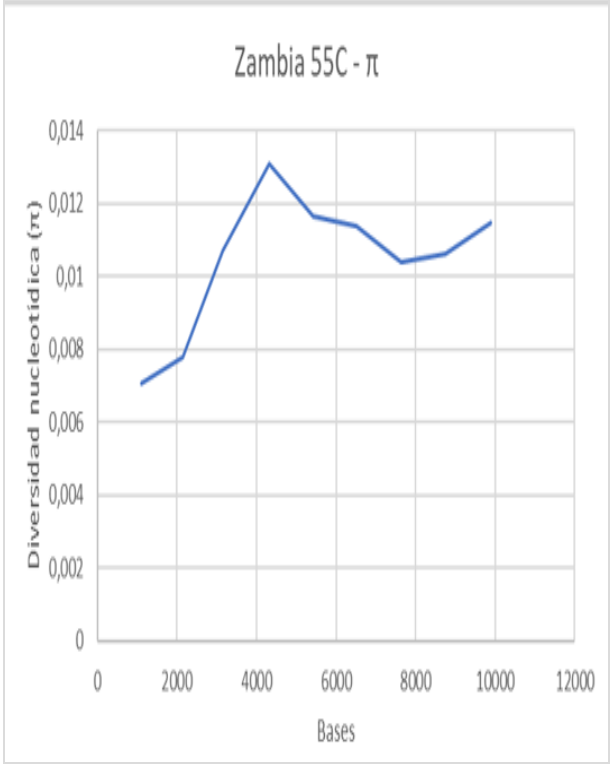

Zambia 32A - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
0
5000
10000
15000
20000
25000
Bases

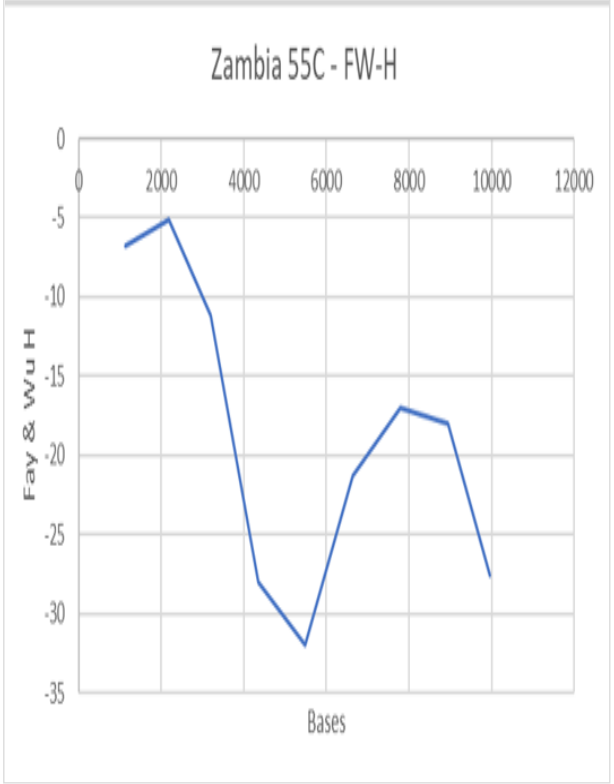

Ruanda 33A - π
Diversidad nucleotídica (π)
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
10000
20000
30000
40000
50000
60000
70000
Bases

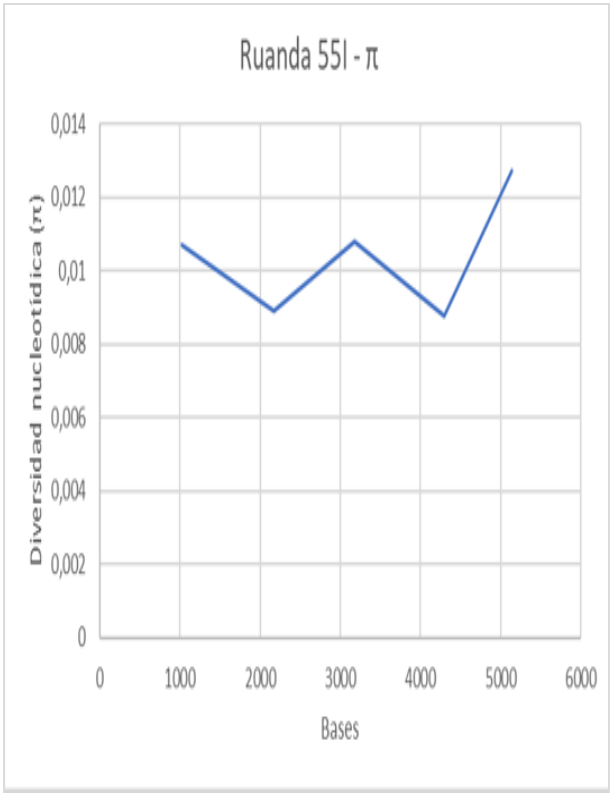

Zambia 33A - π
Diversidad nucleotídica (π)
0,018
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
10000
20000
30000
40000
50000
60000
70000
Bases

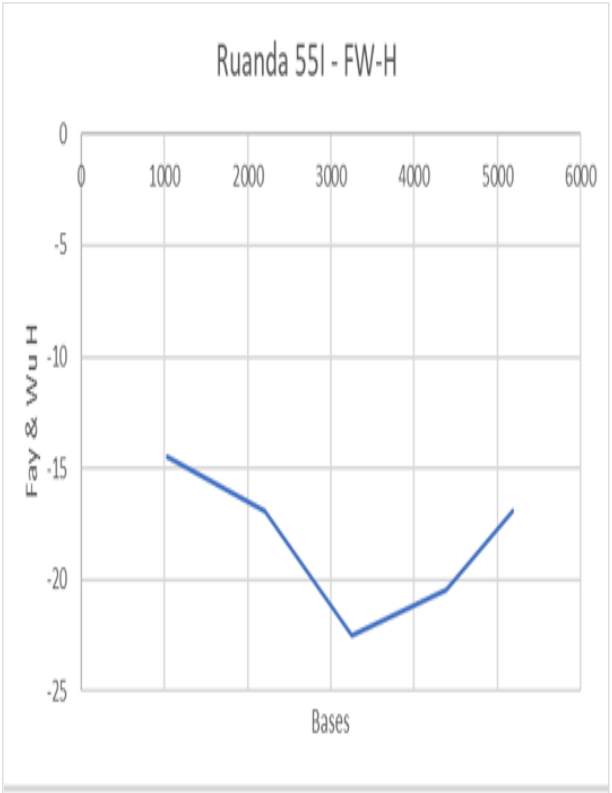

Ruanda 55C - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
2000
4000
6000
8000
10000
12000
Bases

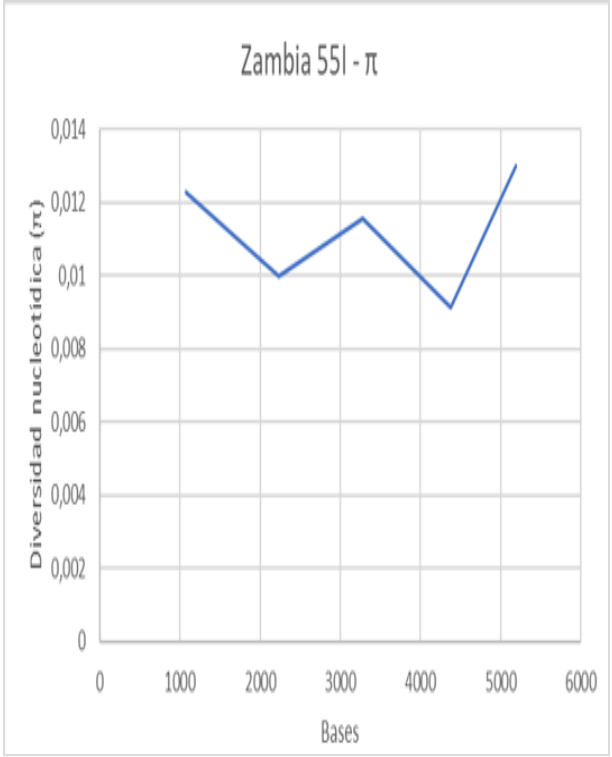

Ruanda 55C - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
-35
0
2000
4000
6000
8000
10000
12000
Bases

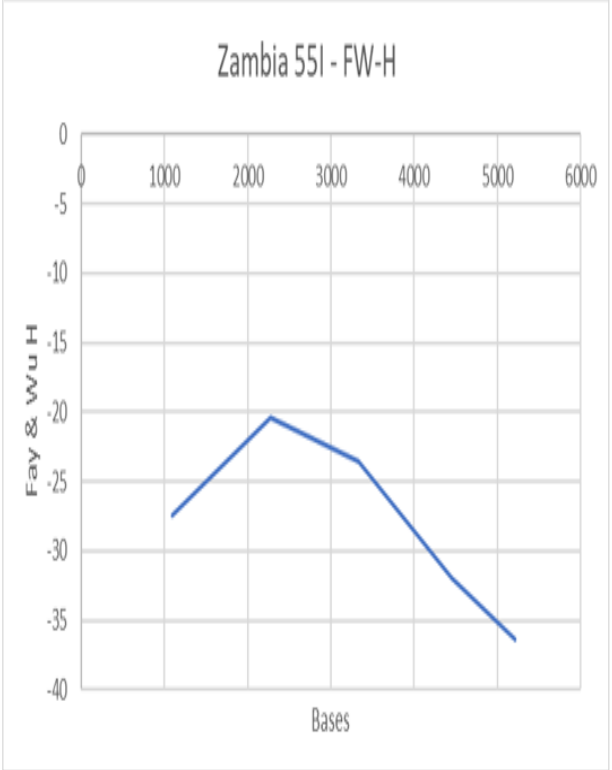

Zambia 55C - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
2000
4000
6000
8000
10000
12000
Bases

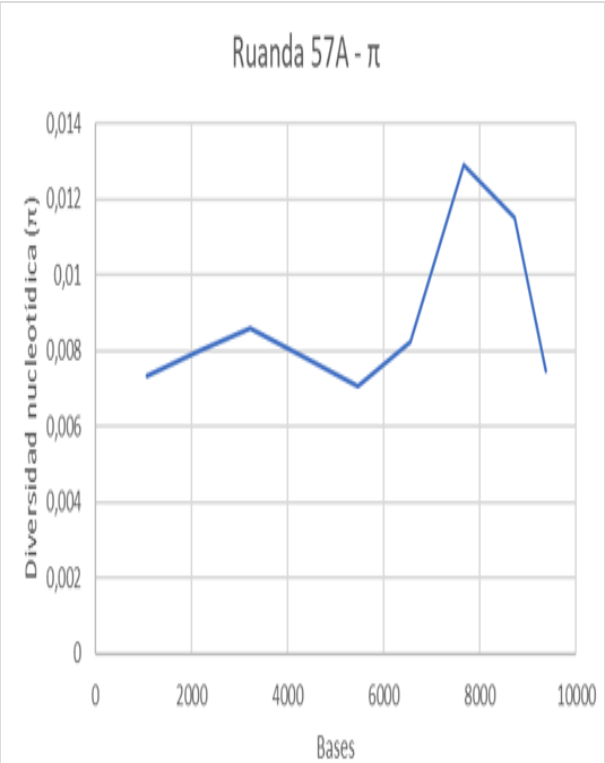

Zambia 55C - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
-35
0
2000
4000
6000
8000
10000
12000
Bases

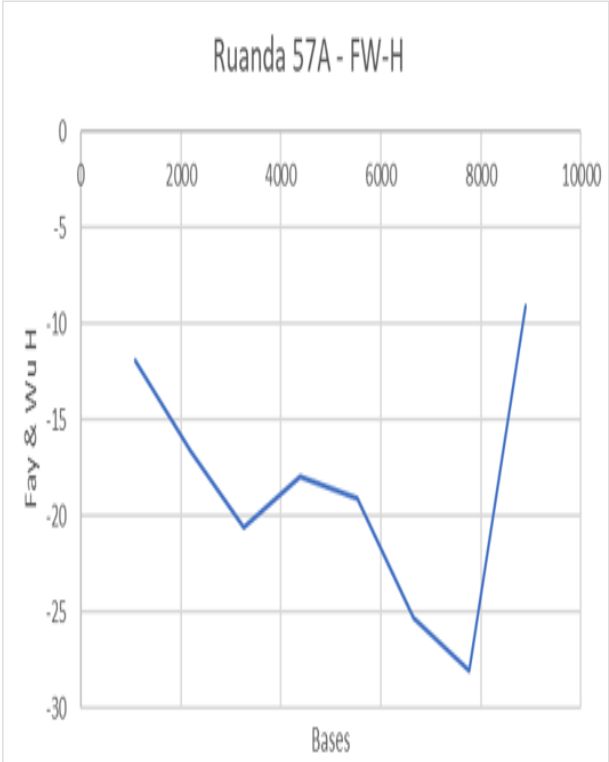

Ruanda 55I - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
1000
2000
3000
4000
5000
6000
Bases

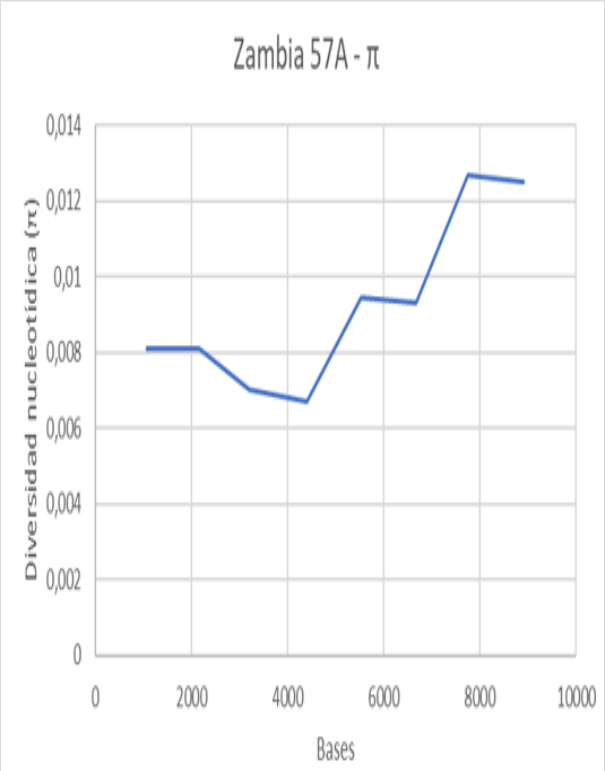

Ruanda 55I - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
0
1000
2000
3000
4000
5000
6000
Bases

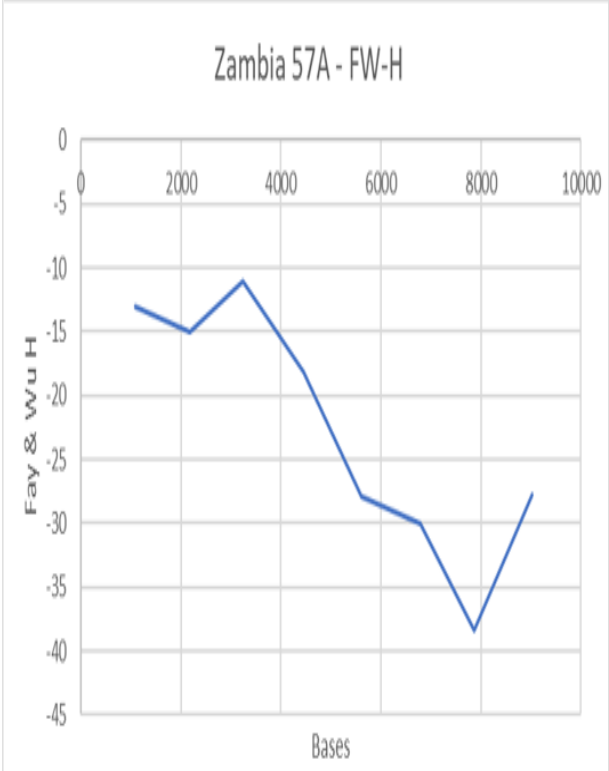

Zambia 55I - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
1000
2000
3000
4000
5000
6000
Bases

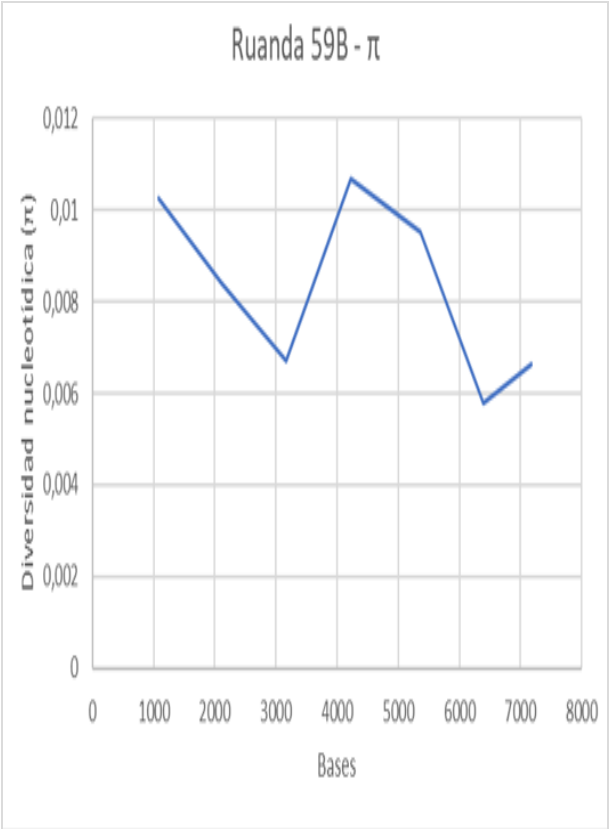

Zambia 55I - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
-35
-40
0
1000
2000
3000
4000
5000
6000
Bases

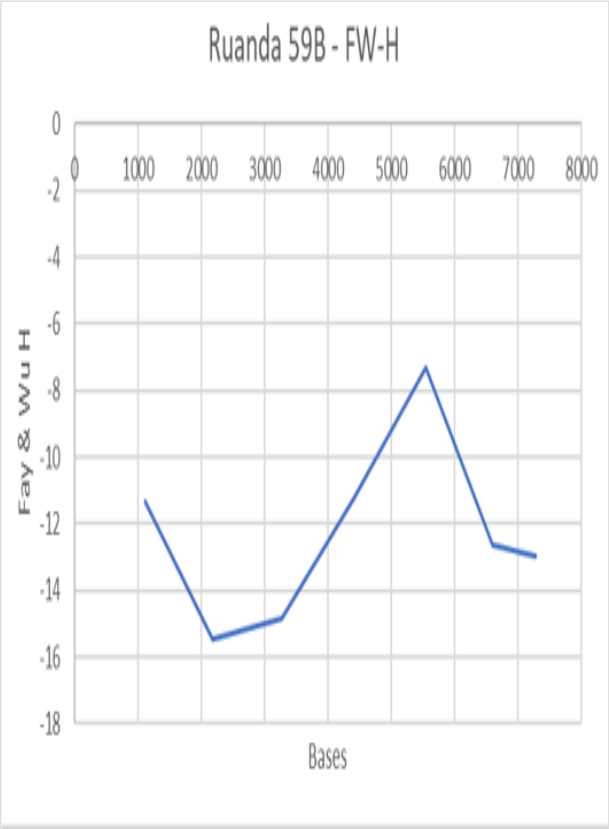

Ruanda 57A - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
2000
4000
6000
8000
10000
Bases

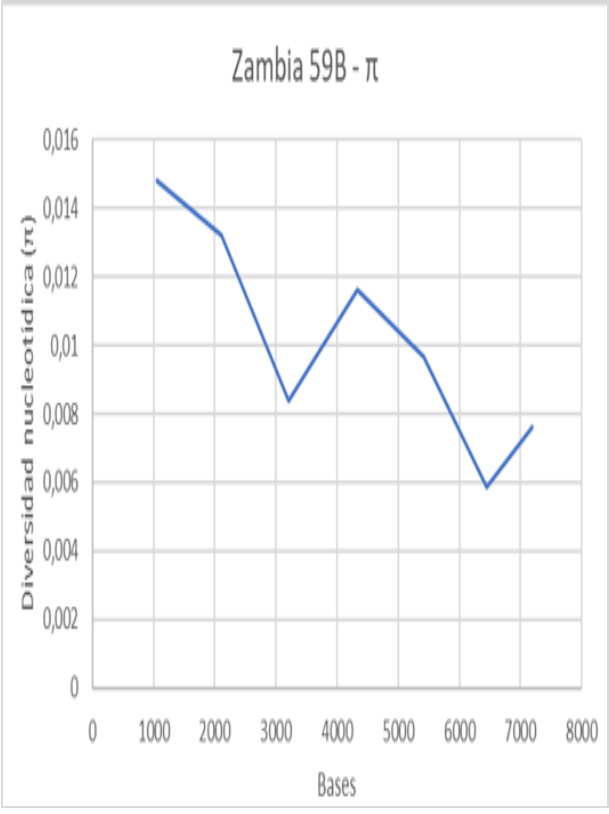

Ruanda 57A - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
0
2000
4000
6000
8000
10000
Bases

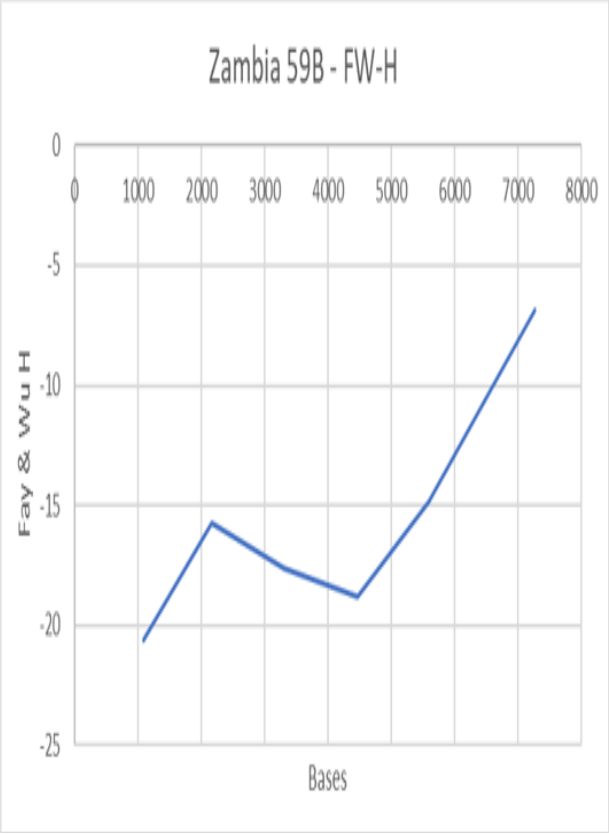

Zambia 57A - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
2000
4000
6000
8000
10000
Bases

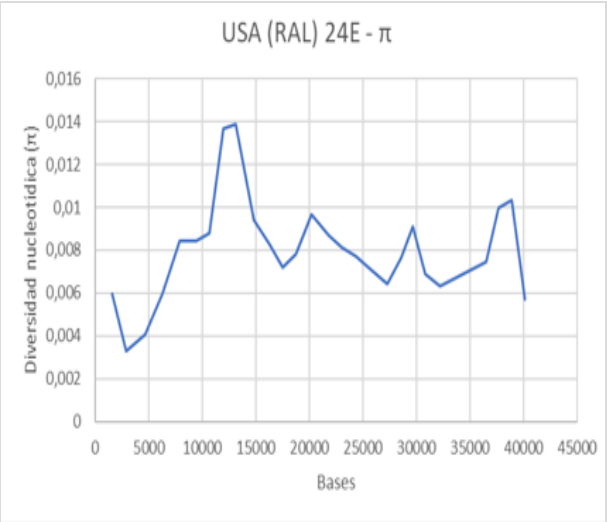

Zambia 57A - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
-35
-40
-45
0
2000
4000
6000
8000
10000
Bases

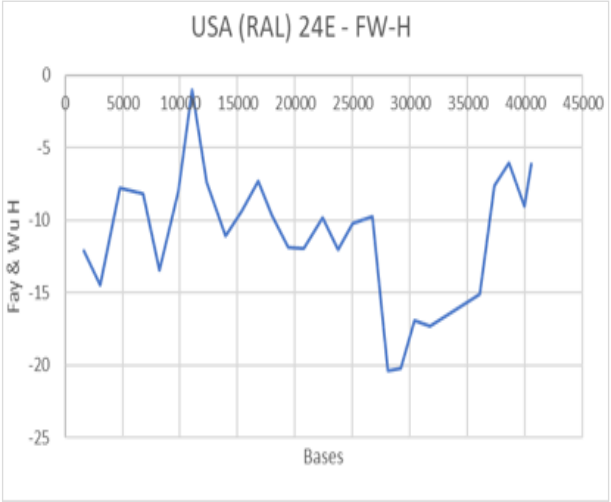

Ruanda 59B - π
Diversidad nucleotídica (π)
0,012
0,01
0,008
0,006
0,004
0,002
0
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases

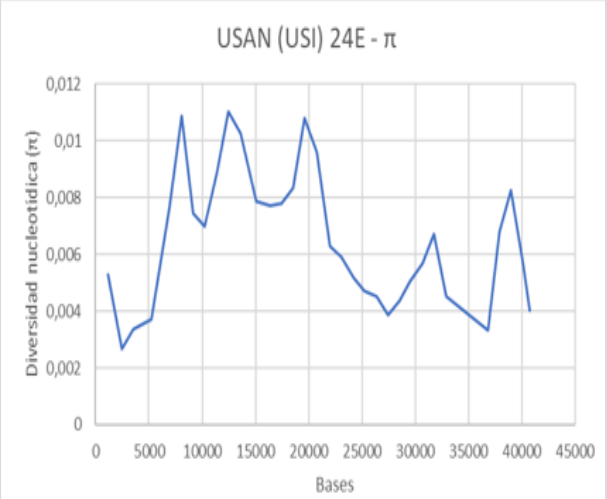

Ruanda 59B - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases

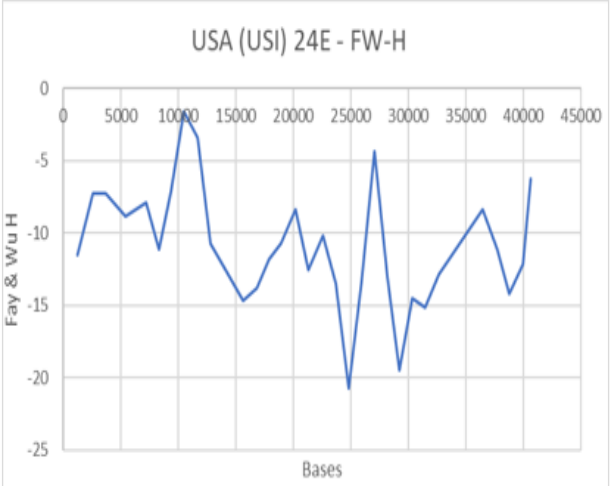

Zambia 59B - π
Diversidad nucleotídica (π)
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases

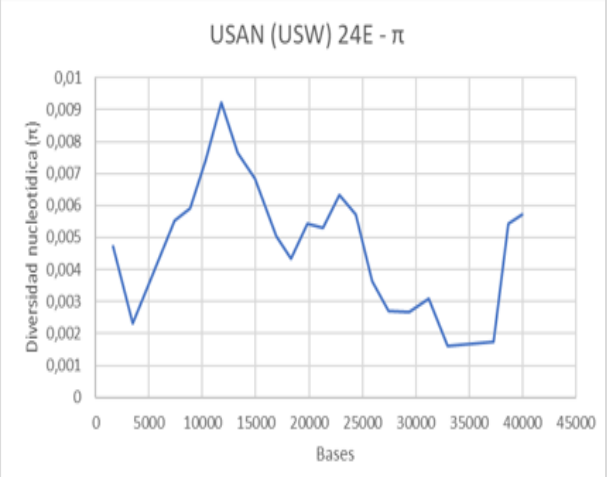

Zambia 59B - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases

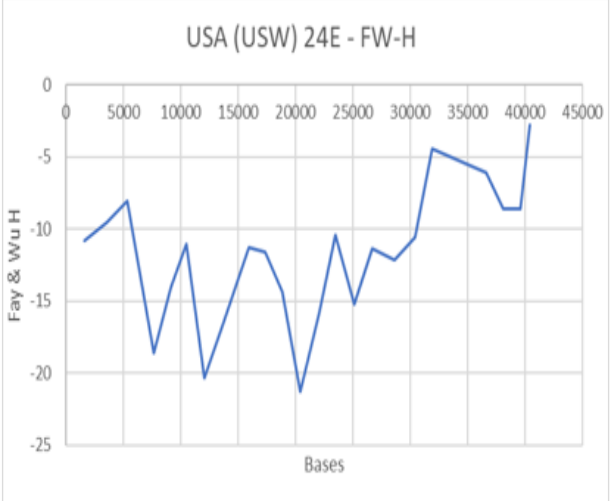

USA (RAL) 24E - π
Diversidad nucleotidica (π)
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0 5000 10000 15000 20000 25000 30000 35000 40000 45000
Bases

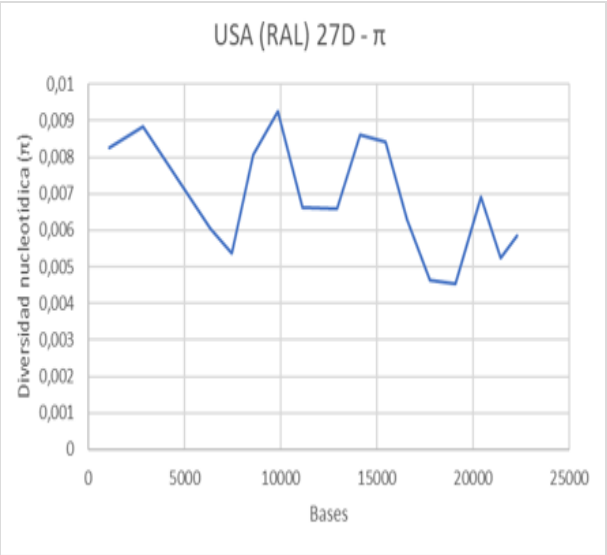

USA (RAL) 24E - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
0 5000 10000 15000 20000 25000 30000 35000 40000 45000
Bases

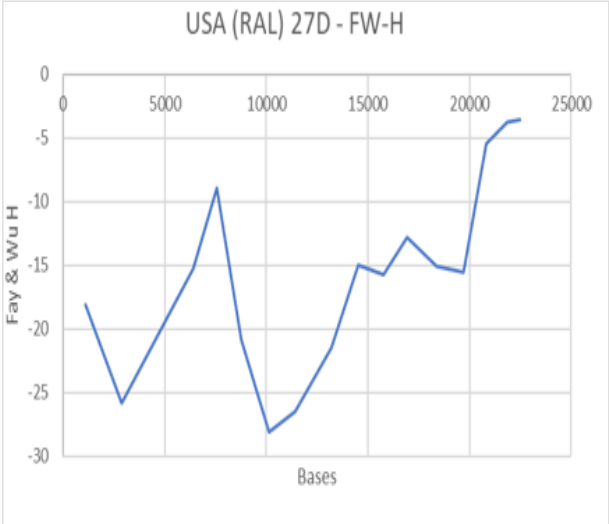

USAN (USI) 24E - π
Diversidad nucleotidica (π)
0,012
0,01
0,008
0,006
0,004
0,002
0
0 5000 10000 15000 20000 25000 30000 35000 40000 45000
Bases

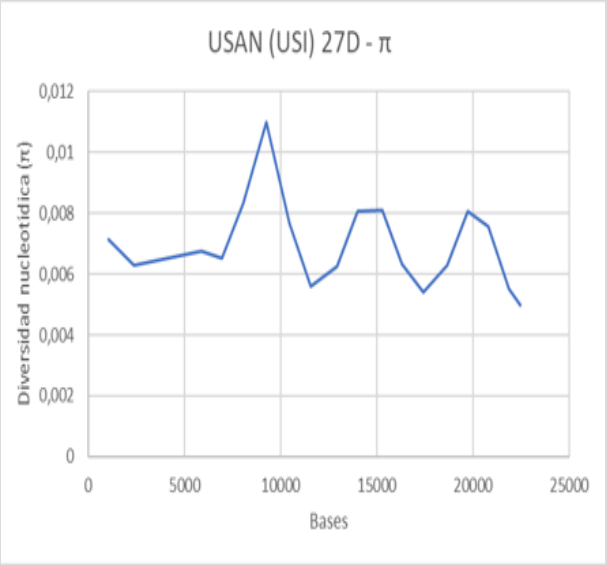

USA (USI) 24E - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
0 5000 10000 15000 20000 25000 30000 35000 40000 45000
Bases

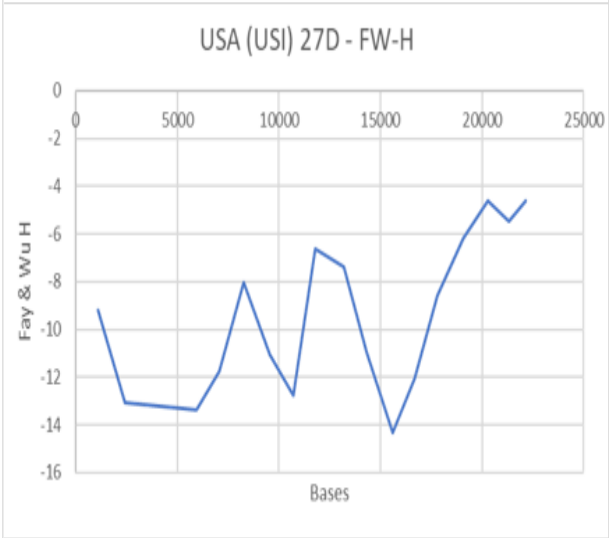

USAN (USW) 24E - π
Diversidad nucleotidica (π)
0,01
0,009
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0 5000 10000 15000 20000 25000 30000 35000 40000 45000
Bases

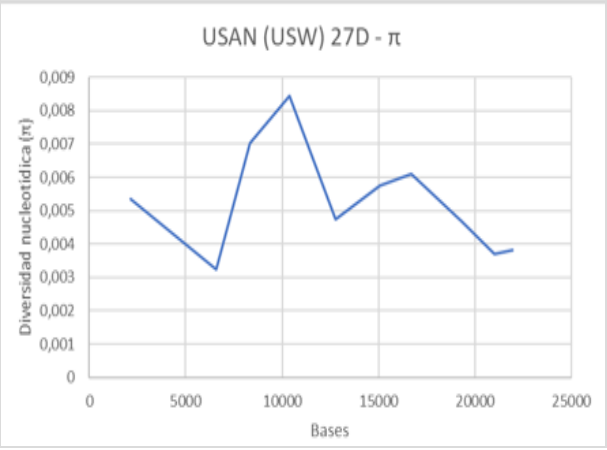

USA (USW) 24E - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
0 5000 10000 15000 20000 25000 30000 35000 40000 45000
Bases

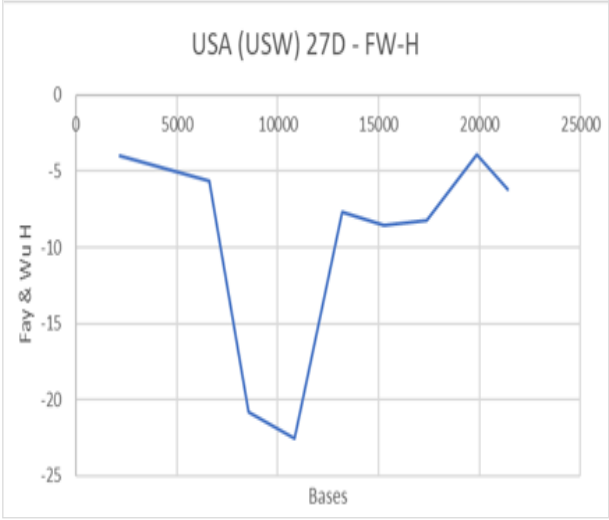

USA (RAL) 27D - π
Diversidad nucleotidica (π)
0,01
0,009
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
5000
10000
15000
20000
25000
Bases

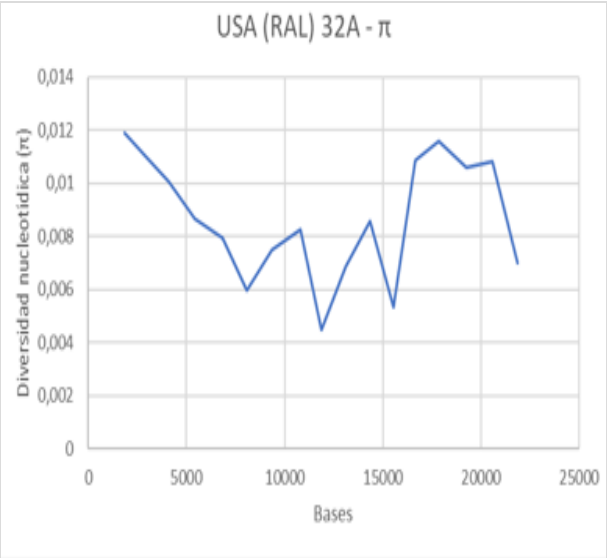

USA (RAL) 27D - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
-30
0
5000
10000
15000
20000
25000
Bases

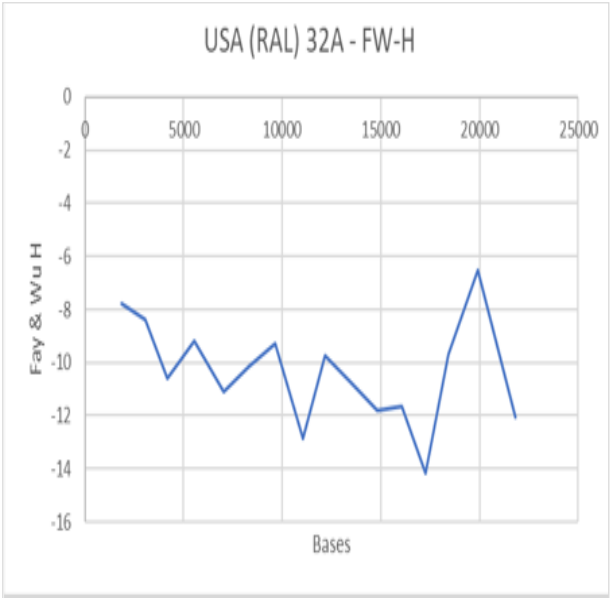

USAN (USI) 27D - π
Diversidad nucleotidica (π)
0,012
0,01
0,008
0,006
0,004
0,002
0
0
5000
10000
15000
20000
25000
Bases

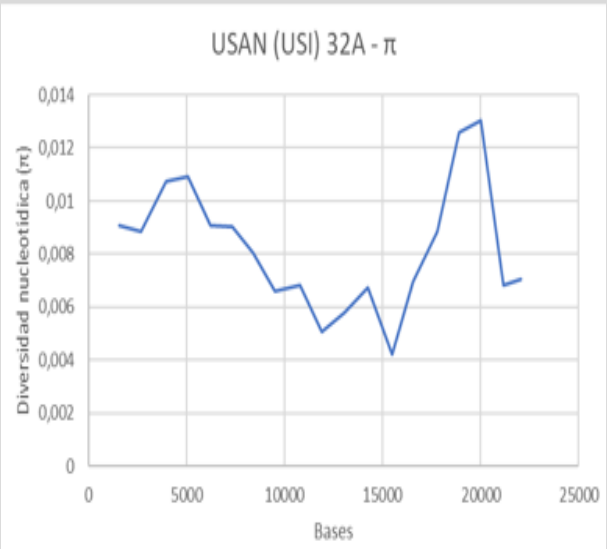

USA (USI) 27D - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
0
5000
10000
15000
20000
25000
Bases

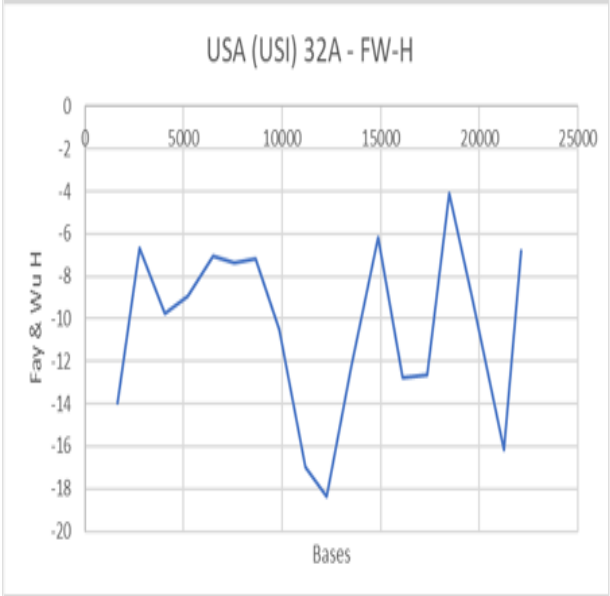

USAN (USW) 27D - π
Diversidad nucleotidica (π)
0,009
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
5000
10000
15000
20000
25000
Bases

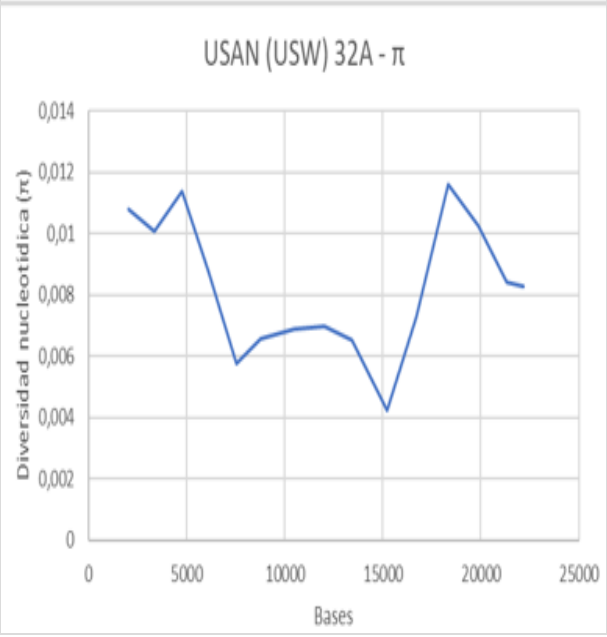

USA (USW) 27D - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
0
5000
10000
15000
20000
25000
Bases

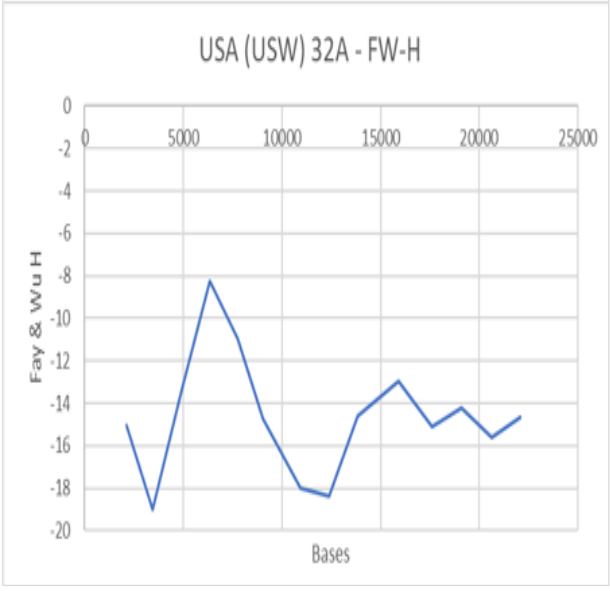

USA (RAL) 32A - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
5000
10000
15000
20000
25000
Bases

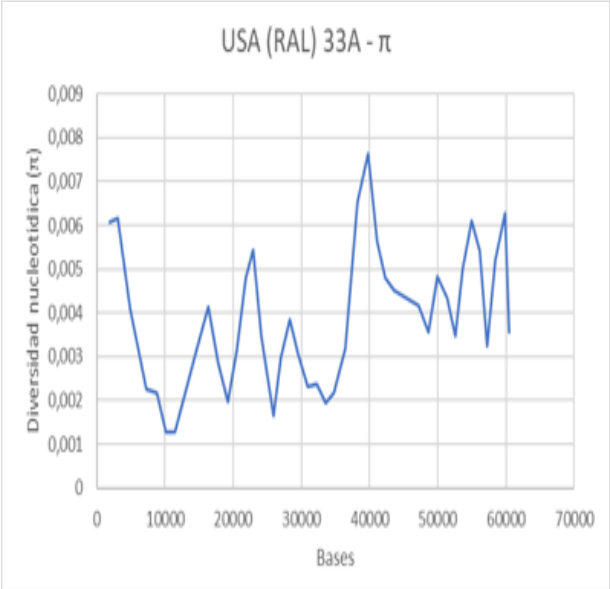

USA (RAL) 32A - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
0
5000
10000
15000
20000
25000
Bases

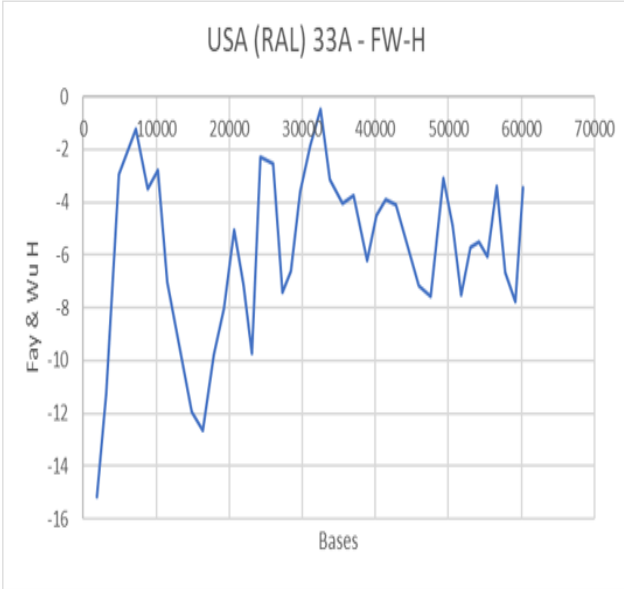

USAN (USI) 32A - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
5000
10000
15000
20000
25000
Bases

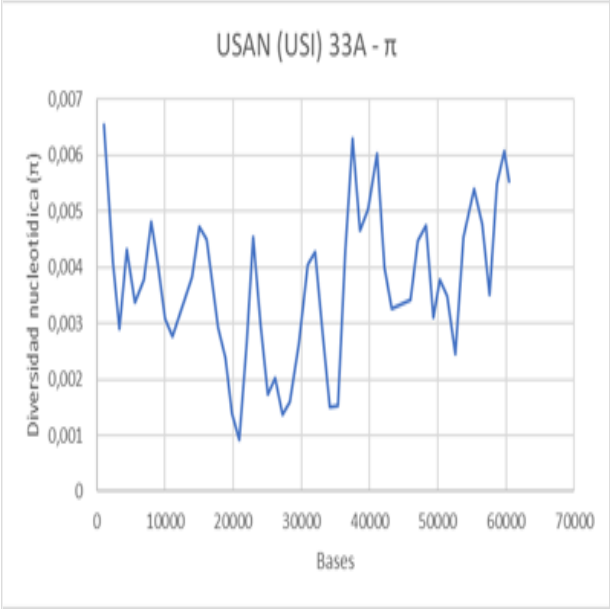

USA (USI) 32A - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
0
5000
10000
15000
20000
25000
Bases

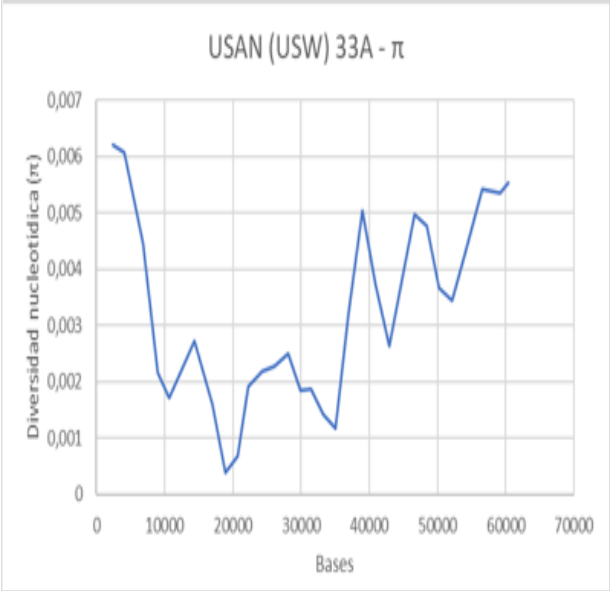

USAN (USW) 32A - π
Diversidad nucleotídica (π)
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0
5000
10000
15000
20000
25000
Bases

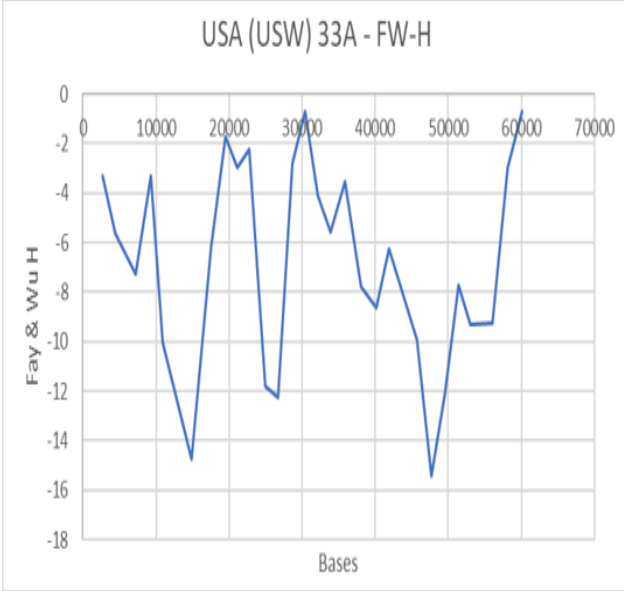

USA (USW) 32A - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
0
5000
10000
15000
20000
25000
Bases

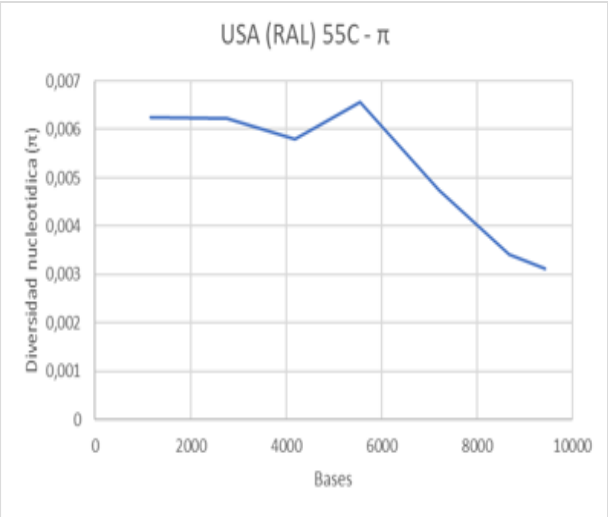

USA (RAL) 33A - π
Diversidad nucleotídica (π)
0,009
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
10000
20000
30000
40000
50000
60000
70000
Bases

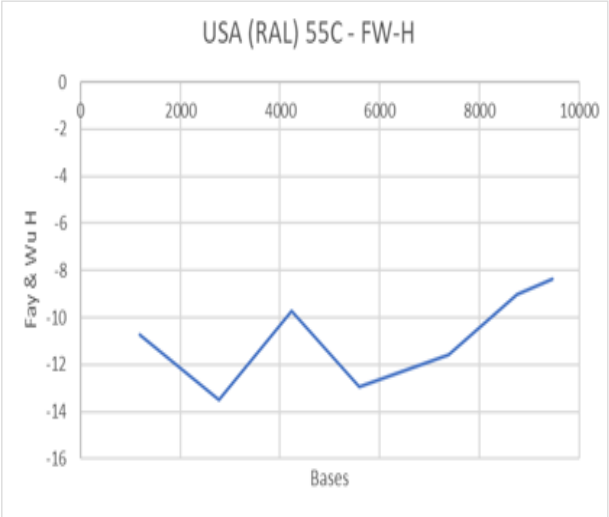

USA (RAL) 33A - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
0
10000
20000
30000
40000
50000
60000
70000
Bases

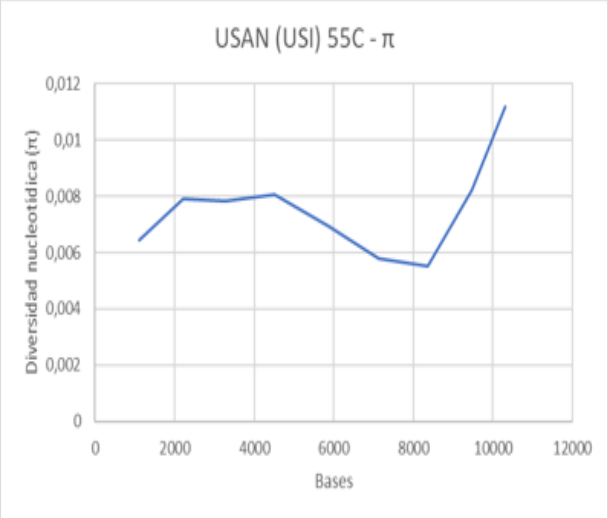

USAN (USI) 33A - π
Diversidad nucleotídica (π)
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
10000
20000
30000
40000
50000
60000
70000
Bases

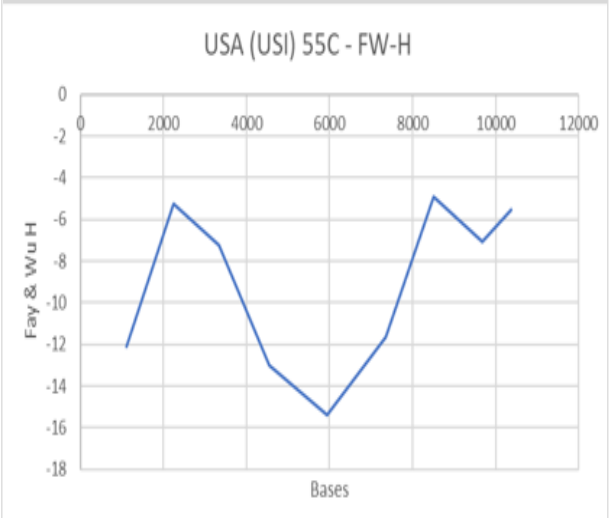

USAN (USW) 33A - π
Diversidad nucleotídica (π)
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
10000
20000
30000
40000
50000
60000
70000
Bases

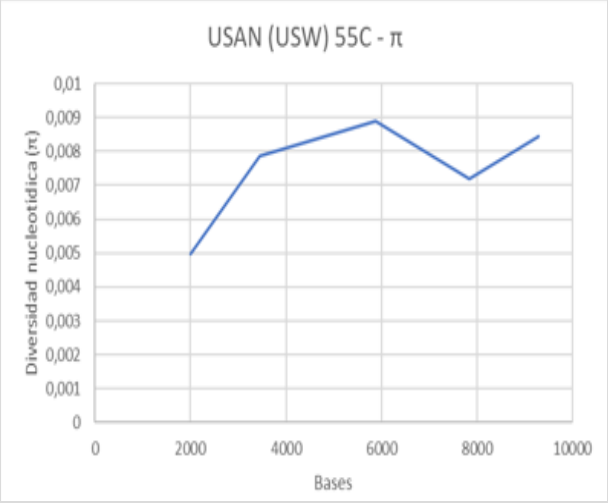

USA (USW) 33A - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
0
10000
20000
30000
40000
50000
60000
70000
Bases

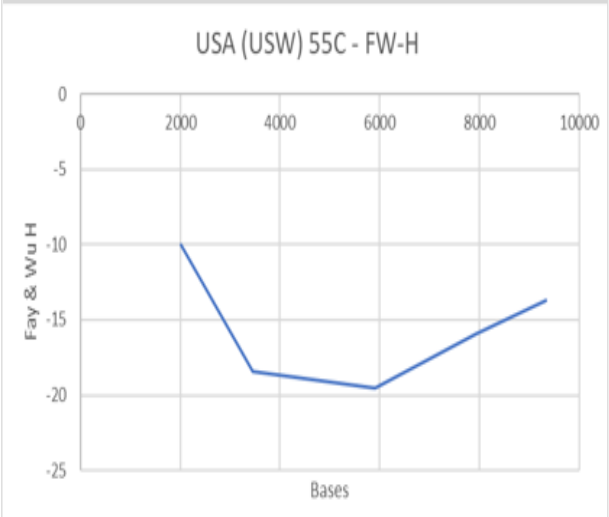

USA (RAL) 55C - π
Diversidad nucleotidica (π)
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
2000
4000
6000
8000
10000
Bases

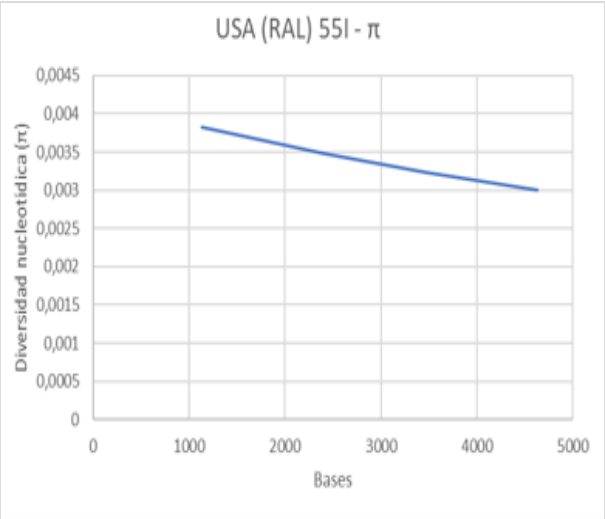

USA (RAL) 55C - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
0
2000
4000
6000
8000
10000
Bases

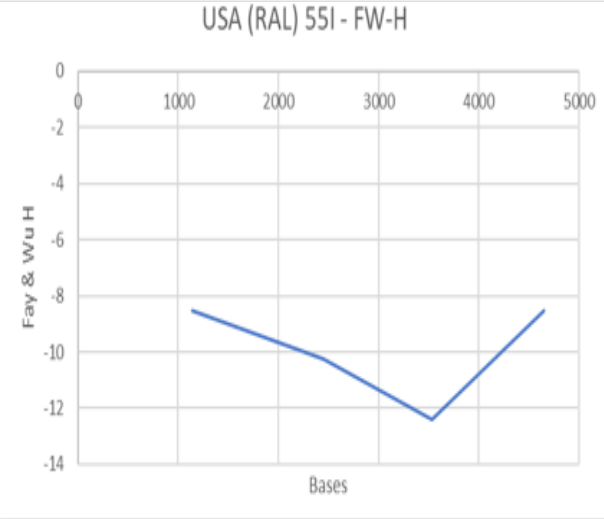

USAN (USI) 55C - π
Diversidad nucleotidica (π)
0,012
0,01
0,008
0,006
0,004
0,002
0
0
2000
4000
6000
8000
10000
12000
Bases

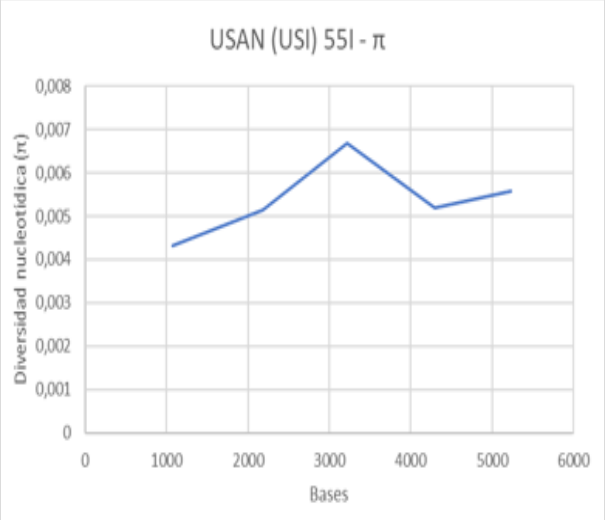

USA (USI) 55C - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
0
2000
4000
6000
8000
10000
12000
Bases

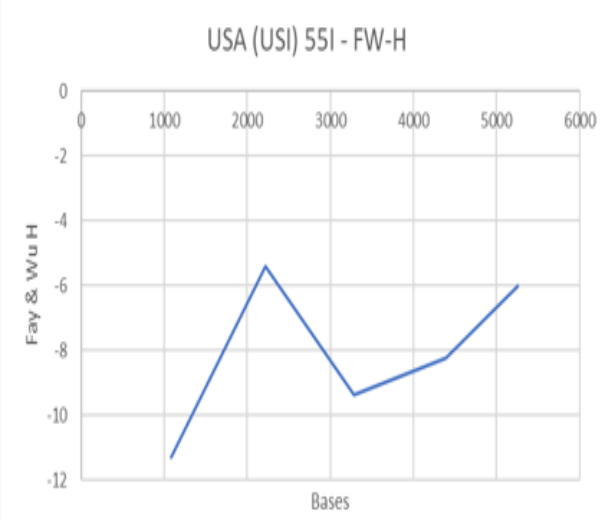

USAN (USW) 55C - π
Diversidad nucleotidica (π)
0,01
0,009
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
2000
4000
6000
8000
10000
Bases

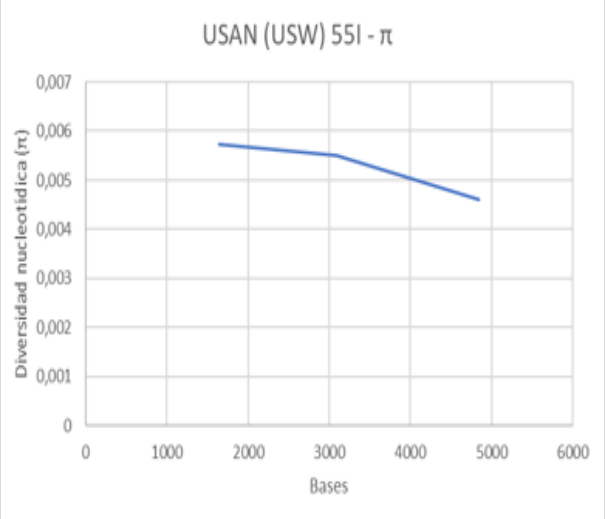

USA (USW) 55C - FW-H
Fay & Wu H
0
-5
-10
-15
-20
-25
0
2000
4000
6000
8000
10000
Bases

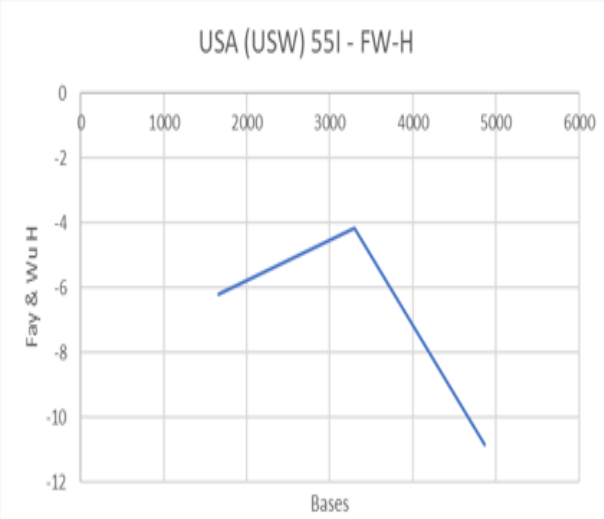

USA (RAL) 55I - π
Diversidad nucleotidica (π)
0,0045
0,004
0,0035
0,003
0,0025
0,002
0,0015
0,001
0,0005
0
0
1000
2000
3000
4000
5000
Bases

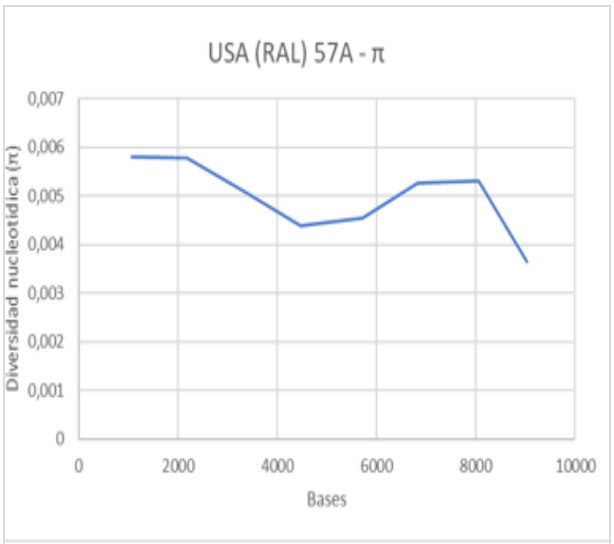

USA (RAL) 55I - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
0
1000
2000
3000
4000
5000
Bases

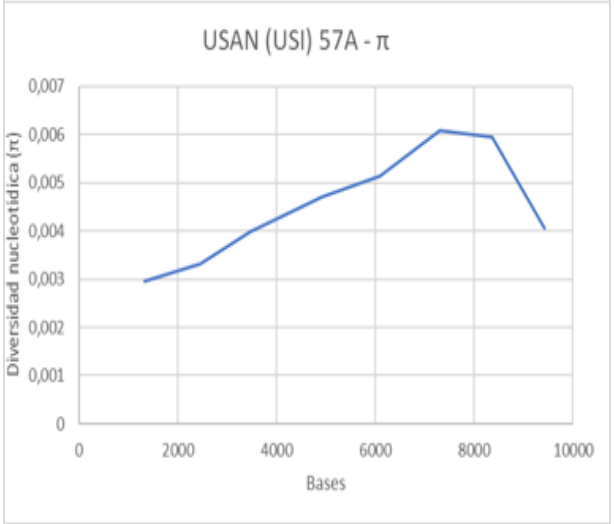

USAN (USI) 55I - π
Diversidad nucleotidica (π)
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
1000
2000
3000
4000
5000
6000
Bases

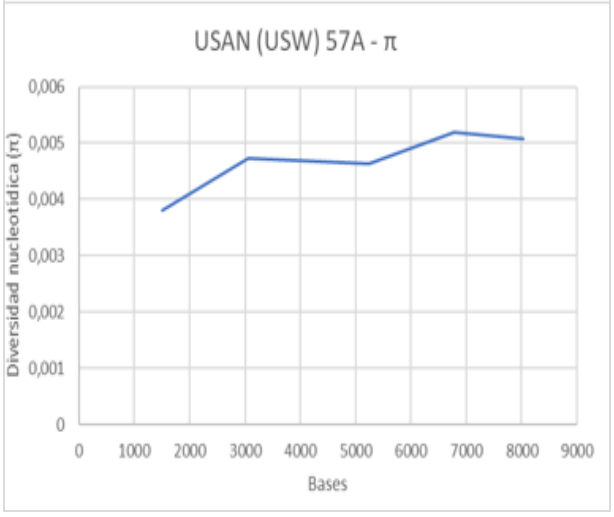

USA (USI) 55I - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
0
1000
2000
3000
4000
5000
6000
Bases

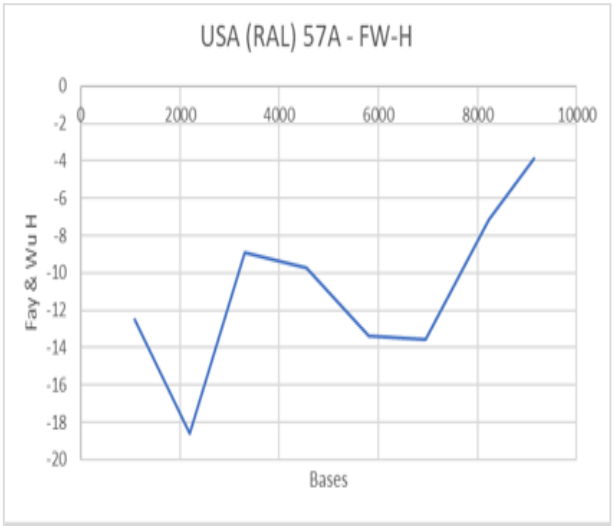

USAN (USW) 55I - π
Diversidad nucleotidica (π)
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
1000
2000
3000
4000
5000
6000
Bases

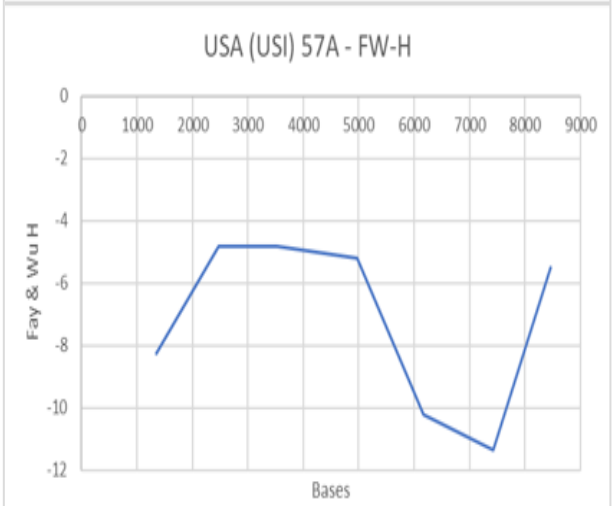

USA (USW) 55I - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
0
1000
2000
3000
4000
5000
6000
Bases

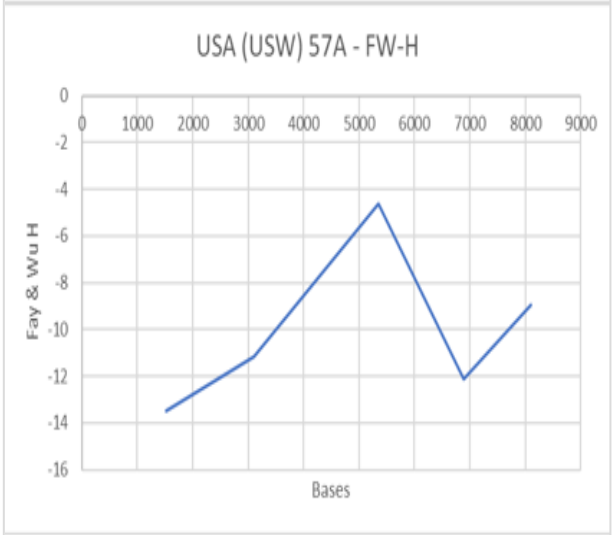

USA (RAL) 57A - π
Diversidad nucleotidica (π)
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
2000
4000
6000
8000
10000
Bases

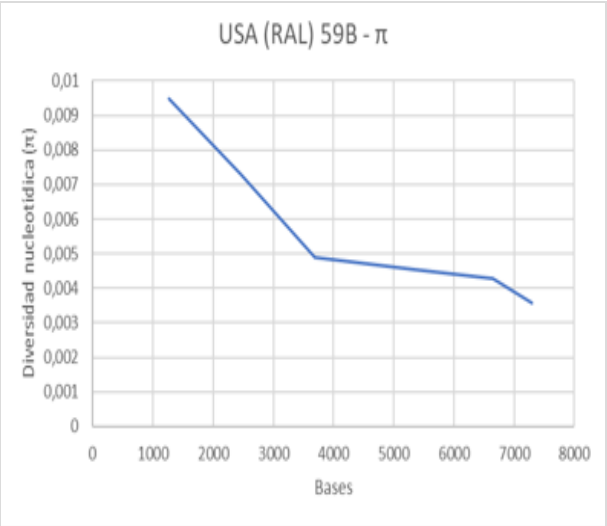

USAN (USI) 57A - π
Diversidad nucleotidica (π)
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0
2000
4000
6000
8000
10000
Bases

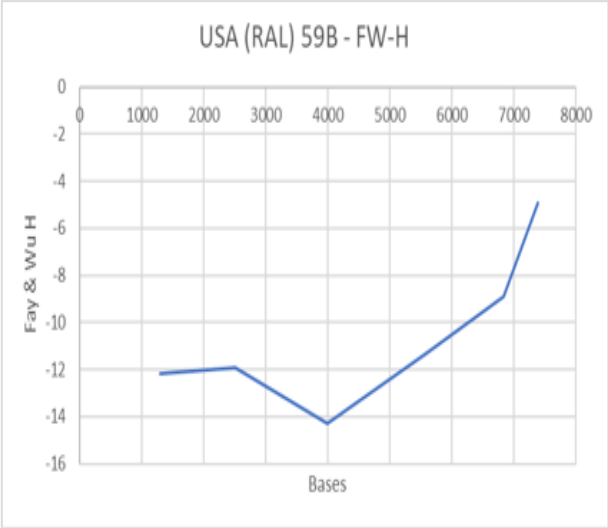

USAN (USW) 57A - π
Diversidad nucleotidica (π)
0,006
0,005
0,004
0,003
0,002
0,001
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Bases

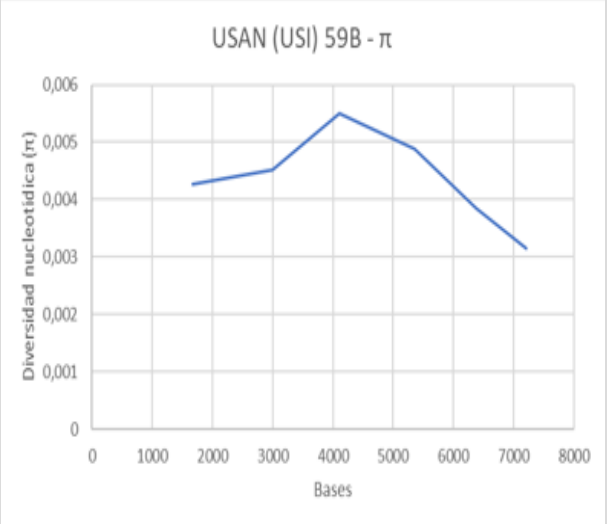

USA (RAL) 57A - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
0
2000
4000
6000
8000
10000
Bases

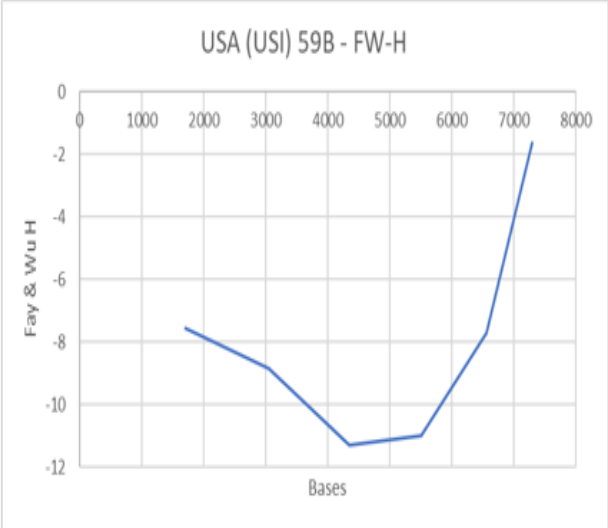

USA (USI) 57A - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Bases

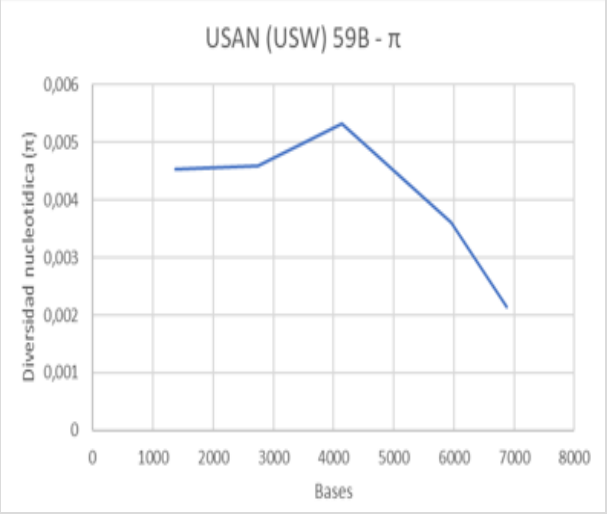

USA (USW) 57A - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Bases

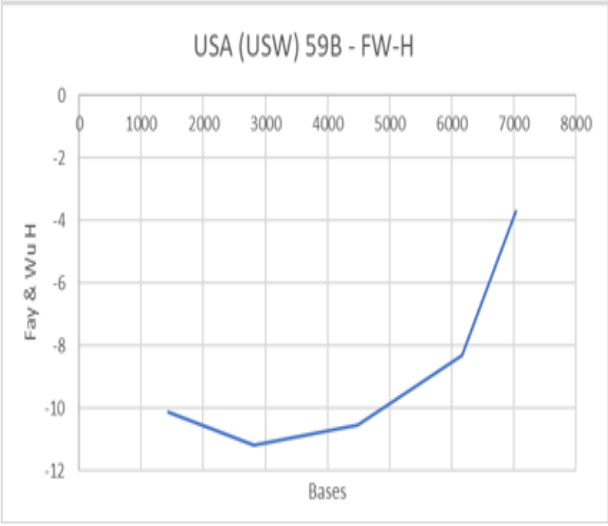

USA (RAL) 59B - π
Diversidad nucleotídica (π)
0,01
0,009
0,008
0,007
0,006
0,005
0,004
0,003
0,002
0,001
0
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases


USA (RAL) 59B - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
-14
-16
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases


USAN (USI) 59B - π
Diversidad nucleotídica (π)
0,006
0,005
0,004
0,003
0,002
0,001
0
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases


USA (USI) 59B - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases


USAN (USW) 59B - π
Diversidad nucleotídica (π)
0,006
0,005
0,004
0,003
0,002
0,001
0
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases


USA (USW) 59B - FW-H
Fay & Wu H
0
-2
-4
-6
-8
-10
-12
0 1000 2000 3000 4000 5000 6000 7000 8000
Bases