Universitat Oberta de Catalunya

uoc.edu

Análisis de la variabilidad nucleotídica en regiones no codificantes del cromosoma 2L de poblaciones naturales de *Drosophila melanogaster*.

Jose Antonio Arco Martín de Rosales
Máster en Bioinformática y Bioestadística
Área 3 – Evolución molecular

Nombre Consultor/a
Dorcas Orengo Ferriz

Nombre Profesor/a responsable de la asignatura
Laura Calvet Liñan

06/2022

FICHA DEL TRABAJO FINAL

Título del trabajo:	Análisis de la variabilidad nucleotídica en regiones no codificantes del cromosoma 2L de poblaciones naturales de *Drosophila melanogaster.*
Nombre del autor:	Jose Antonio Arco Martín de Rosales
Nombre del consultor/a:	Dorcas Orengo Ferriz
Nombre del PRA:	Laura Calvet Liñan
Fecha de entrega (mm/aaaa):	06/2022
Titulación:	Máster en Bioinformática y Bioestadística
Área del Trabajo Final:	Área 3 – Evolución molecular
Idioma del trabajo:	Castellano
Número de créditos:	15
Palabras clave	*Drosophila melanogaster,* selección natural, variabilidad nucleotídica

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

Drosophila melanogaster es un organismo perfecto para el estudio de los modelos de adaptación local debido fundamentalmente a ser una especie cosmopolita, que se ha expandido fuera de África Subsahariana en épocas relativamente recientes. Esto hace que, en el caso de las poblaciones derivadas, se espera poder detectar la huella de su adaptación a nuevos hábitats.

Este trabajo ha estudiado la variabilidad nucleotídica en el cromosoma 2L en regiones intergénicas de tres poblaciones de África, una de Europa y una de Estados Unidos. Además, para dos de estas poblaciones geográficas se han considerado dos muestras distintas según su ordenación cromosómica.

Para la selección de las regiones a estudiar, se ha desarrollado un script en R que extrae información del archivo de la anotación del genoma de *D. melanogaster*. Para cada región se han calculado estadísticos de neutralidad y de diferenciación genética entre poblaciones, utilizando diversos programas.

El patrón de variabilidad nucleotídica de todas las poblaciones es muy similar, pero se ha detectado diferenciación genética entre seis de las poblaciones, comportándose de igual manera cuando se considera cada región

independientemente que cuando se consideran todas las regiones conjuntamente, o agrupadas según se encuentran dentro o fuera de la inversión.

Destaca que mientras las dos poblaciones “cromosómicas” de Zambia, están muy diferenciadas, las dos de EEUU, no lo están en absoluto. Este hecho parece indicar que la barrera a la recombinación entre los dos tipos de cromosomas, que parece persistir en África, se hubiera roto en las poblaciones derivadas.

Abstract (in English, 250 words or less):

Drosophila melanogaster is a perfect organism for the study of local adaptation patterns mainly because it is a cosmopolitan species, which has expanded out of sub-Saharan Africa in relatively recent times. This means that, in the case of derived populations, we hope to be able to detect the imprint of its adaptation to new habitats.

This work has studied nucleotide variability on chromosome 2L in intergenic regions of three populations from Africa, one from Europe and one from the United States. In addition, for two of these geographic populations, two different samples have been considered according to their chromosomal arrangement.

For the selection of the regions to be studied, an R script has been developed that extracts information from the *D. melanogaster* genome annotation file. For each region, statistics of neutrality and genetic differentiation between populations were calculated using different programs.

The pattern of nucleotide variability of all the populations is very similar, but genetic differentiation has been detected among six of the populations, behaving in the same way when each region is considered independently as when all the regions are considered together, or grouped according to whether they are inside or outside the inversion.

It is noteworthy that while the two "chromosomal" Zambian populations are highly differentiated, the two US populations are not differentiated at all. This fact seems to indicate that the barrier to recombination between the two types of chromosomes, which seems to persist in Africa, has been broken in the derived populations.

Índice

Lista de tablas

Lista de figuras

1 Resumen

Drosophila melanogaster es un organismo perfecto para el estudio de los modelos de adaptación local debido fundamentalmente a haberse convertido en una especie cosmopolita repartida por todo el mundo, y que su expansión fuera de África Subsahariana se ha producido en épocas relativamente recientes.

En el caso de las poblaciones derivadas de *D. melanogaster*, dado el corto tiempo transcurrido desde su salida de África en relación con el tamaño de la población de la especie, se espera detectar la huella de su adaptación a nuevos hábitats.

En este trabajo se ha estudiado la variabilidad nucleotídica en el cromosoma 2L en regiones intergénicas de tres poblaciones de África, una de Europa y una de Estados Unidos. Para dos de estas poblaciones geográficas (una de África y la de Estados Unidos) se han considerado dos muestras distintas según su ordenación cromosómica.

Se ha desarrollado un *script* en R para procesar el archivo que contenía toda la información del genoma de referencia de *D.* melanogaster que selecciona las regiones a partir de las cuales se han obtenido las secuencias para el estudio.

A partir de las secuencias elegidas se han calculado estadísticos de neutralidad y de diferenciación genética de las poblaciones, utilizando diversos programas bioinformáticos.

Se ha observado un patrón de variabilidad nucleotídica muy similar entre todas las poblaciones. Se ha detectado diferenciación genética entre seis de las poblaciones en estudio comportándose de igual manera cuando se consideran una a una que cuando se consideran todas las regiones conjuntamente, o agrupadas según se encuentran dentro o fuera de la inversión.

Se ha observado que mientras las dos poblaciones “cromosómicas” de Zambia (ZI_ST y ZI_INV), están muy diferenciadas, las dos poblaciones de EEUU, (RAL_ST y RAL_INV) no lo están en absoluto. Este hecho parece indicar que la barrea a la recombinación que parece persistir en África, se hubiera roto en las poblaciones derivadas.

2 Introducción

2.1 Contexto y justificación del Trabajo

La evolución molecular estudia cómo cambia la secuencia del ADN a lo largo del tiempo y entre poblaciones. El estudio de la variación en el ADN es importante porque es la base de cualquier cambio morfológico o funcional que observemos en los organismos. La variación nucleotídica que se observa entre los individuos de una población se origina por las mutaciones en el ADN y es necesaria para que tenga lugar cualquier proceso evolutivo. A su vez, los distintos procesos evolutivos como la deriva genética, las migraciones, los cuellos de botella y los procesos de selección natural son factores que influyen en la variabilidad nucleotídica de una población. Así, el estudio de la variación genética que podemos observar entre individuos de distintas poblaciones es una excelente fuente de información de sus orígenes y de los eventos de adaptación experimentados por las poblaciones a lo largo del tiempo.

Drosophila melanogaster es un organismo perfecto para el estudio de los modelos de adaptación local debido fundamentalmente a dos factores, haberse convertido en una especie cosmopolita repartida prácticamente por todo el mundo, y que su expansión fuera de África central y meridional, de donde es originaria[1], se ha producido en épocas relativamente recientes [2].

A nivel teórico se espera que el efecto de los procesos selectivos sobre el patrón de variación nucleotídica se ciña bastante a la región más cercana a la localización de la diana de selección debido a que la recombinación rompe la asociación entre variantes. En cambio, factores demográficos como un cuello de botella, provocarán una disminución en la variabilidad que será bastante más homogénea a lo largo de todo el genoma.

El patrón de la variabilidad nucleotídica se ve afectado tanto por procesos adaptativos como por procesos demográficos. En el caso de las poblaciones derivadas de *D. melanogaster*, dado el corto tiempo transcurrido en relación con el tamaño de la población de la especie, se espera detectar la huella de su adaptación a hábitats templados sobre un efecto más general debido al cuello de botella que supuso su salida de África. En este sentido, ya se realizó un estudio sobre la variabilidad nucleotídica a lo largo del cromosoma X en una población europea de *D. melanogaster.* [3]

Ante la disponibilidad de datos genómicos completos de diversas poblaciones naturales de todo el mundo de *D. melanogaster*, se puede intentar responder a nuevas preguntas: ¿Es también posible detectar

alteraciones en el patrón de variabilidad nucleotídica a lo largo de un autosoma? ¿Cómo afecta el hecho de la existencia de una inversión polimórfica en este patrón de variación? ¿Todas las poblaciones derivadas se comportan igual? ¿Algunas regiones bajo selección ya lo estaban en su área de origen?

Para intentar responder a algunas de estas preguntas, se recuperarán secuencias de diversas poblaciones africanas (ancestrales) y europeas y americanas (derivadas). En ellas se estudiará el patrón de variación nucleotídica a lo largo del cromosoma 2L para regiones no codificantes. También se estudiará si las poblaciones están diferenciadas genéticamente.

2.2 Objetivos del Trabajo

Se plantean los objetivos generales y específicos detallados a continuación:

2.2.1 Objetivos generales

1. Estudiar la variabilidad nucleotídica de diversos fragmentos del cromosoma 2L en regiones no codificantes en poblaciones naturales de *D. melanogaster* de distintas áreas geográficas (África, Europa y América) en busca de posibles huellas de la selección natural.

2. Comparar el patrón de variación nucleotídica observado en las poblaciones naturales de África (ancestrales), Europa y América (derivadas) de *D. melanogaster.*

2.2.2 Objetivos específicos

Se desglosa cada objetivo general en los siguientes objetivos específicos:

1. Estudiar la variabilidad nucleotídica de diversos fragmentos del cromosoma 2L en regiones no codificantes en poblaciones naturales de *Drosophila melanogaster* de distintas áreas geográficas (África, Europa y América) en busca de posibles huellas de la selección natural.

    1.1. Calcular la distancia entre pares de genes contiguos a lo largo del cromosoma 2L.

1.2. Elegir las regiones intergénicas a analizar de manera que estén distribuidas homogéneamente a lo largo de todo el cromosoma.

1.3. Obtener una tabla con las características de estas regiones, posición en el cromosoma, el nombre de los genes que las flanquean, etc…

1.4. Comprobar que las regiones elegidas no se corresponden con regiones repetitivas.

1.5. Obtener las secuencias correspondientes para las distintas poblaciones elegidas.

1.6. Seleccionar las secuencias individuales que se analizarán para cada población y ordenación.

1.7. Alinear la secuencia de *D. melanogaster* con la secuencia ortóloga de *D. simulans.*

1.8. Calcular los indicadores de variabilidad genética y los estadísticos de neutralidad.

1.9. Relacionar la variabilidad nucleotídica con la distancia a los puntos de rotura de la inversión polimórfica, la distancia al gen más cercano y la frecuencia de recombinación, entre otras características.

2. Comparar el patrón de variación nucleotídica observado en las poblaciones naturales de África (ancestrales), Europa y América (derivadas) de *D. melanogaster.*

2.1. Calcular el estadístico de diferenciación genética F_{ST}.

2.2. Comparar las distancias genéticas y geográficas entre poblaciones.

2.3. Comparar las posibles diferencias entre ordenaciones cromosómicas dentro de cada población natural.

2.3 Enfoque y método seguido

Con el fin de llevar a cabo el estudio de la variabilidad y la diferenciación genética se realizará el calculo de diversos indicadores y test estadísticos que asignen un valor de significación entre los valores observados y los valores esperados de acuerdo siempre a las teorías neutralistas.

Para realizar estos cálculos, se seleccionarán muestras de poblaciones de tres áreas geográficas distintas, tres de África, de donde es originaría la especie, una de Europa y una de América. Así se analizará y comparará la variabilidad nucleotídica de estas tres poblaciones para evaluar si las posibles diferencias encontradas son efecto de la adaptación a climas menos cálidos o también existen estas diferencias en poblaciones africanas.

Se dispone de gran cantidad de información genómica de acceso libre sobre *D. melanogaster* debido a que es un organismo ampliamente utilizado para fines científicos. Esta información es relativa tanto a su genoma de referencia como del genoma particular de numerosos individuos de diferentes poblaciones naturales. Por tanto, se obtendrá la información de bases de datos públicas como PopFly [4] y FlyBase [5] . En concreto se decide extraer el genoma de referencia de FlyBase donde se encuentra la última versión, FB2022_01, de 8 de febrero de 2022, en lugar de utilizar Ensembl [6], donde no está la citada versión.

Junto a la secuencia de *D. melanogaster* se necesita tener una secuencia *outgroup.* Para este objeto se obtendrá la secuencia de *D. simulans,* una de las especies más cercanas a *D. melanogaster* y de la que se dispone de la secuencia completa de su genoma.

Tras seleccionar los individuos de las tres poblaciones, se obtendrán sus secuencias y se alinearán con la secuencia *outgroup*. Asimismo, se realizará el calculo de los indicadores de polimorfismo y estadísticos de neutralidad (*D* de Tajima [7] y *H* de Fay y Wu [8]. Además, se calculará el estadístico de diferenciación genética F_{ST} [9] Para asignar un valor de significación estadística entre los valores observados y esperados se realizarán diferentes simulaciones de acuerdo a las teorías neutralistas[10]. Para la realización de estos cálculos se utilizarán diversos programas bioinformáticos entre los que destaca DnaSP6 [11], que es uno de los más ampliamente utilizados en análisis poblacionales de secuencias. También se usará el programa mlcoalsim[12] para realizar simulaciones que permitan obtener la significación estadística de la *D* de Tajima y *H* de Fay y Wu y el programa mstastpop[13] que permite obtener valores de F_{ST} y su significación estadística.

2.4 Planificación del Trabajo

Basándonos en los objetivos marcados previamente se realiza una planificación del trabajo definiendo las tareas a realizar. Por otro lado, se evalúa los posibles riesgos que se pueden presentar durante la realización del trabajo y las acciones posibles para mitigarlos.

2.4.1 Basándonos en los objetivos, tanto generales como específicos, se definen las tareas a realizar:

1. Cálculo de la distancia entre pares de genes contiguos a lo largo del cromosoma 2L.

2. Clasificación de las distancias obtenidas atendiendo al número de pares de bases que contienen.

3. Elección del tamaño mínimo de la región intergénica que se quiere analizar y obtención de las coordenadas de inicio y fin de las regiones seleccionadas.

4. Identificación de los nombres de los genes que flanquean a estas regiones intergénicas, así como la cadena en la que están (+,-).

5. Revisión de la naturaleza de las regiones intergénicas seleccionadas.

6. Transformación de las coordenadas de las secuencias seleccionadas a la estructura de coordenadas de Popfly.

7. Obtención de los fragmentos de las secuencias genómicas de las poblaciones a comparar desde Popfly.

8. Selección de las secuencias a analizar en cada población y ordenación.

9. Comprobación de la posible existencia de una secuencia similar alineada con nuestra secuencia de *D. melanogaster* para *D. simulans*

10. Obtención de la secuencia correspondiente de *D. simulans* y alineación a los alineamientos de *D. melanogaster* en el caso de que no existiera dicha secuencia ya alineada entre ambas especies.

11. Análisis de la variabilidad genética, polimorfismo y divergencia mediante el calculo de los indicadores de polimorfismo y estadísticos de neutralidad.

12. Realización de simulaciones para la obtención de la significación estadística.

13. Cálculo del estadístico de diferenciación genética $F_{ST.}$

14. Comparación de las distancias genéticas y geográficas entre secuencias.

15. Análisis de los resultados de la variabilidad y de la diferenciación genética.

16. Ajustes y corrección de errores.

17. Redacción de la memoria.

Aunque se propone una revisión constante de las tareas a realizar con el objetivo de la detección y corrección de posibles errores lo mas tempranamente posible, también se propone una tarea previa a la redacción de la memoria para ajuste y corrección de potenciales errores no detectados

2.4.2 Calendario

Se estima una dedicación de 24 h semanales para la realización del calendario. Basado en esta dedicación se planifican las distintas tareas de acuerdo con la siguiente tabla y diagrama de Gantt asociado:

Tarea	F. Inicio	F. Fin
PEC0 – Definición de los contenidos	**16/02/22**	**23/02/22**
PEC1 – Plan de trabajo	**24/02/22**	**07/03/22**
PEC2 – Desarrollo del trabajo Fase 1	**08/03/22**	**11/04/22**
Calculo de la distancia entre pares de genes a lo largo del cromosoma 2L	08/03/22	14/03/22
Clasificación de las distancias obtenidas atendiendo al número de pares de bases que contienen.	15/03/22	16/03/22
Elección del tamaño mínimo de la región intergénica que queremos analizar y obtención de las coordenadas de inicio y fin de las regiones seleccionadas.	17/03/22	17/03/22
Identificación de los nombres de los genes que flanquean a estas regiones así como la cadena en la que están (+,-)	18/03/22	18/03/22
Revisión de la naturaleza de las regiones intergénicas seleccionadas.	19/03/22	20/03/22
Transformación de las coordenadas de las secuencias seleccionadas a la estructura de coordenadas de Popfly.	21/03/22	21/03/22
Obtención de los fragmentos de las secuencias genómicas de las poblaciones a comparar desde Popfly.	22/03/22	25/03/22
Selección de las secuencias por cada ordenación.	26/03/22	07/04/22
Comprobación de la posible existencia de una secuencia similar alineada con nuestra secuencia de *D. melanogaster* para *D. simulans.*	08/04/22	09/04/22
Obtención de la secuencia correspondiente de *D. simulans* y alineación a los alineamientos de *D. melanogaster* en el caso de que no existiera dicha secuencia similar entre ambas variedades.	N/A	N/A
Cálculo de los indicadores de polimorfismo y estadísticos de neutralidad.	10/04/22	11/04/22
PEC3 – Desarrollo del trabajo Fase 2	**12/04/22**	**16/05/22**
Cálculo de los indicadores de polimorfismo y estadísticos de neutralidad. (cont.)	12/04/22	26/04/22
Realización de simulaciones para la obtención de la significación estadística.	27/04/22	10/05/22
Cálculo del estadístico de diferenciación genética F_{ST}	11/05/22	16/05/22
PEC4 – Cierre de la memoria	**17/05/22**	**02/06/22**
Comparación de las distancias genéticas y geográficas entre secuencias.	17/05/22	18/05/22
Análisis de los resultados de la variabilidad y de la diferenciación genética.	19/05/22	24/05/22
Redacción de la memoria	25/05/22	29/05/22
Ajuste y corrección de errores	30/05/22	02/06/22
PEC5a – Elaboración de la presentación	**30/05/22**	**06/06/22**

Tabla 1: *Planificación de tareas*

Figura 1: *Diagrama de Gantt con la planificación del proyecto*

2.4.3 Hitos

En base a la planificación se establecen los siguientes hitos. El retraso en el cumplimiento de las tareas asociadas a estos hitos podría poner en riesgo los objetivos planteados. Estas tareas están resaltadas en amarillo en el diagrama de Gantt.

1. Cálculo de los indicadores de polimorfismo y estadísticos de neutralidad.

2. Realización de simulaciones para la obtención de la significación estadística.

3. Cálculo de la diferenciación genética.

4. Análisis de los resultados de la variabilidad y de la diferenciación genética.

5. Redacción de la memoria.

2.4.4 Análisis de riesgos

Se identifican los factores de riesgo que podrían afectar al cumplimiento de los objetivos planteados:

- Problemas en la gestión del tiempo debido a imprevistos de tipo profesional como personal. Se han acotado los compromisos profesionales y se tiene una planificación de la dedicación a los mismos. No obstante, podrían surgir imprevistos que impacten en el trabajo. Para mitigar estos casos se está dedicando semanalmente más de las 24 horas estipuladas para la realización del trabajo durante las semanas que no hay carga profesional y se dedicarán fines de semana y festivos en caso de necesidad.

- Detección tardía de problemas o retrasos. Para mitigar este impacto se ha establecido una comunicación fluida entre el autor y el consultor del proyecto con continuo cruce de información de seguimiento.

- Existencia de secuencias con un porcentaje elevado sin identificar con la consiguiente dificultad en su análisis. Para mitigar este problema se seleccionarán los individuos que contengan las secuencias más completas para la región de estudio.

2.5 Breve sumario de contribuciones y productos obtenidos

A la finalización del proyecto se generarán los siguientes entregables:

- Plan de Trabajo. Documento en el que se determinan los objetivos y alcance del trabajo. Además, se establece el calendario, la planificación de las distintas tareas y el análisis de riesgos.

- Memoria. Documento donde se informa del trabajo realizado durante el proyecto detallando el contexto y justificación del trabajo, los métodos utilizados, los resultados obtenidos y las conclusiones alcanzadas tras el análisis de los resultados.

- Presentación virtual. Presentación resumen del desarrollo del trabajo y de los resultados obtenidos junto con los aspectos más significativos.

- Informe de autoevaluación del proyecto.

2.6 Breve descripción de los otros capítulos de la memoria

La memoria incluye los siguientes capítulos, además de la introducción:

- Estado del arte: Se recoge el estado actual de los estudios relativos a la variación nucleotídica observada en el genoma que se focalizan en aclarar si el origen de la disminución de la variabilidad es originada por procesos demográficos o bien por eventos de selección.

- Metodología: Se detallan los procedimientos llevados a cabo durante el proyecto para analizar la variabilidad nucleotídica y la diferenciación genética de las secuencias objeto del trabajo, así como de los programas bioinformáticos utilizados.

- Resultados: Se exponen los resultados obtenidos a partir de la metodología aplicada al análisis de las secuencias en las distintas poblaciones en estudio.

- Discusión: Se discuten los resultados obtenidos relativos a la variabilidad nucleotídica y a la diferenciación genética entre las poblaciones en estudio.

- Conclusiones: Se describen las conclusiones extraídas del trabajo realizado, la valoración del seguimiento de la planificación propuesta al inicio del trabajo y las líneas de trabajo futuro.

- Glosario: Se definen los términos y acrónimos utilizados a lo largo de esta memoria.

- Bibliografía: Se listan las referencias bibliográficas utilizadas como fuente de información para la redacción de esta memoria.
- Anexos: Se incluyen tablas y scripts utilizados para generar la información de este trabajo.

3 Estado del arte

D. melanogaster es un organismo perfecto para el estudio de los modelos de adaptación local debido fundamentalmente a dos factores, haberse convertido en una especie cosmopolita repartida prácticamente por todo el mundo, y que su expansión fuera de África central y meridional, de donde es originaria [1], se ha producido en épocas relativamente recientes. Así su expansión a territorios del África occidental se data hace 70.000-72.000 años, coincidiendo con la migración humana y con cambios climáticos [14], mientras que su expansión a territorios fuera de África se data hace 10.000-15.000 años.

El patrón de la variabilidad nucleotídica se ve afectado tanto por procesos adaptativos como por procesos demográficos. En el caso de las poblaciones derivadas de *D. melanogaster*, dado el corto tiempo transcurrido en relación con el tamaño de la población de la especie, se espera detectar la huella de su adaptación a hábitats templados sobre un efecto más general debido al cuello de botella que supuso su salida de África. [15,16]

Diversos estudios multilocus realizados en distintas poblaciones naturales de *D*. *melanogaster*, dónde se analizaron diferentes regiones de los cromosomas X [17,18] , 3 [19] y 2 (Orengo y Aguadé, datos sin publicar), han mostrado que los efectos demográficos no explicarían por completo los patrones de variabilidad observada. En este sentido ya se realizó un análisis de componentes principales (PoCA) para explorar si 14 genomas haploides completamente secuenciados de una muestra de individuos de *D*. *melanogaster* del norte de Suecia eran consistentes con la hipótesis citada. [20] Los resultados mostraron que el primer componente principal (PC1) separó consistentemente las muestras europeas y africanas y que las líneas suecas estaban menos dispersas, lo que refleja una diversidad más baja en comparación con una muestra africana, de Zambia concretamente. Una importante excepción a este patrón se observó en el cromosoma 2L. Además de la separación específica del continente en PC1, identificaron un agrupamiento igualmente fuerte en PC2, causado por la inversión cromosómica cosmopolita común In(2L)t, para la cual ya se informó una alta diferenciación genética entre haplotipos estándar e invertidos.

En otro estudio utilizando pool-seq para secuenciar 48 muestras de poblaciones de 32 localidades también encontraron indicios de diferentes genes que han experimentado selección en poblaciones europeas. [21] Estos son solo algunos ejemplos en los que se ha encontrado indicios de eventos de selección natural en poblaciones derivadas de *D. melanogaster*.

En las dos últimas décadas, a medida que las facilidades de secuenciación aumentaban, el número de genomas individuales secuenciados de *D. melanogaster* ha aumentado considerablemente. Distintos consorcios como, por ejemplo, el *Drosophila melanogaster* Genetic Reference Panel [22] han registrado la variación molecular y fenotípica en 168 cepas de moscas de la fruta completamente secuenciadas derivadas de una única población natural exogámica. El primer conjunto de análisis de datos de DGRP proporcionó información sobre el panorama genómico de la variación genética, la selección positiva y negativa y la rápida evolución del cromosoma X. Los resultados también revelaron muchas variantes de baja frecuencia en nuevos loci que estaban asociadas con rasgos cuantitativos y explicaban una gran fracción de la variación fenotípica.

Otros, como el Drosophila Genome Nexus [23], un recurso genómico de población de 623 genomas de *Drosophila melanogaster*, incluidos 197 de una única población de rango ancestral, también han asumido la secuenciación de numerosos genomas. Al mismo tiempo, se han desarrollado herramientas que permiten el fácil acceso a estos datos, como puede ser PopFly [24] que recogen genomas de muestras poblacionales. Ante la disponibilidad de datos genómicos completos de diversas poblaciones naturales de todo el mundo de *D. melanogaster*, se puede intentar responder a nuevas preguntas: ¿Todas las poblaciones derivadas se comportan igual? ¿Algunas regiones bajo selección ya lo estaban en su área de origen? En este sentido parece que estudios como el que llevaron a cabo Hutter y colaboradores [19] donde se concluía que el tamaño de la población femenina en África era mayor o igual al tamaño de la población masculina en contraste con la población europea, donde se mostraba un enorme exceso de varones, puede indicar lo contrario.

4 Metodología

4.1 Elección de las regiones genómicas a analizar.

El primer paso para el desarrollo del proyecto fue la elección de las regiones genómicas a analizar. Para ello, se accedió a Flybase donde se obtuvo la última *release* (6.44) del genoma de referencia de *Drosophila melanogaster* en formato GFF3. Tras revisar este archivo y comprender sus características, se diseñó un script en R para procesarlo de modo que realizaba las siguientes tareas:

- Se extrajeron los registros que hacían referencia al cromosoma 2L.
- Se ordenó el registro resultante por la posición inicial del cada gen.
- Se seleccionaron los registros que hacían referencia al tipo "gene", excluyendo, entre otros, exón, codón, etc.
- Se identificaron las áreas intergénicas y las distancias entre los genes colindantes.

En base al estudio de la información que arrojó el script se decidió trabajar con fragmentos de 5001 nucleótidos localizados en la parte central de cada región intergénica. Una vez se tuvieron localizados todos estos fragmentos se escogieron 25 regiones de 5001 nucleótidos distribuidos homogéneamente a lo largo de todo el cromosoma 2L y se anotaron las coordenadas de inicio y fin de cada región. También se identificó el nombre de los genes colindantes a esas regiones intergénicas, así como la cadena en la que estaban localizados (+, -). *(ver Anexo 2)*

4.2 Obtención de las secuencias a analizar.

Se comprobó que las coordenadas de las regiones intergénicas seleccionadas, basadas en Flybase, coincidían con las coordenadas de Popfly para el cromosoma 2L pese a que la versión del genoma era distinta. Esto facilitó la descarga desde Popfly de los fragmentos de las secuencias genómicas de los individuos en estudio.

Para seleccionar las poblaciones a analizar y, dentro de cada una de ellas, los individuos a considerar, se contó con los archivos "TableS1_individuals" y "TableS2_populations", que se descargaron desde *The Drosophila genoma nexus* [23] *.* Estos archivos almacenan un registro de individuos de diversas poblaciones de *D. melanogaster* de todo el mundo cuyo genoma ha sido secuenciado. Además, incluyen información sobre la ordenación de sus cromosomas, los cromosomas secuenciados parcial o totalmente, la localización geográfica de las poblaciones, etc.

A partir de esta información y teniendo en cuenta la cantidad y calidad de las secuencias de las diversas poblaciones y los objetivos del estudio, se

decidió trabajar con poblaciones de individuos de Zambia en sus dos ordenaciones cromosómicas, estándar (ST) e invertida (ln(2L)t), Etiopia, Ruanda, dentro del continente africano, Francia, en Europa y Estados Unidos en América, también en las dos ordenaciones cromosómicas del cromosoma 2L (ST y ln(2L)t).

Con esta información y con las coordenadas de inicio y fin de las regiones que se había decidido estudiar, se descargaron las poblaciones citadas desde Popfly con todos los individuos que incluían.

Inicialmente se consideró óptimo generar una muestra de 15 individuos por cada ordenación (INV y ST) distribuidos en 2 poblaciones africanas, 2 europeas y 1 americana. Para ello se desarrolló un *script* por cada población, *(ver Anexo 3),* para seleccionar los individuos de acuerdo a estos criterios. Sin embargo, los resultados obtenidos a partir de este *script* mostraron la imposibilidad de mantener este criterio basándonos en las poblaciones disponibles en PopFly por la falta de información suficiente. Por tanto, y en base a las poblaciones e individuos disponibles, se limitó la muestra a 10 individuos por población y se escogieron tres poblaciones africanas, una de ellas con las dos ordenaciones cromosómicas, una europea y una americana, también con las dos ordenaciones cromosómicas. Así se minimizó el número de nucleótidos faltantes y al mismo tiempo se obtuvo una muestra lo suficientemente significativa para que el estudio fuera representativo.

Con el mismo objetivo se eliminaron 2 regiones de las 25 seleccionadas inicialmente ya que para la mayoría de los individuos no tenían información para estas regiones, es decir, toda su secuencia eran Ns. Además, se permitió que algunos individuos tuvieran, para algunas regiones, su secuencia incompleta, con muchas Ns, o incluso totalmente incompleta, es decir todo Ns. *(ver Anexo 3 y 5)*

En base a todo lo expuesto se seleccionó la siguiente muestra de individuos con sus ordenaciones y su tipo de genoma, por cada una de las regiones que se detallan:

Continente	N.º individuos	Ordenación	Tipo Genoma	Población
África	10	ST	Haploid embryo	ZI, Zambia
África	10	INV	Haploid embryo	ZI, Zambia
África	10	ST	Inbred line	EF, Ethiopia
África	10	ST	Haploid embryo	RG, Rwanda
Europa	10	ST	Inbred line	FR, France
América	10	ST	Inbred line	RAL, USA
América	10	INV	Inbred line	RAL, USA

Tabla 2: *Selección de poblaciones en estudio*

Además, se obtuvo la secuencia *outgroup* de *Drosophila simulans* que se utilizó para el cálculo de los estadísticos que lo requerían. Esta secuencia se obtuvo de Drosophila Genome Nexus [23]

Se concatenaron todas las secuencias de las distintas regiones genómicas por individuo, usando el programa bioinformático DnaSP6[11]. Además, se añadió al archivo resultante la secuencia *outgroup* mediante un script en Python.

4.3 Estudio de la variabilidad nucleotídica y estadísticos de selección.

Los estadísticos de polimorfismo (*S*, *π*) y de divergencia (*K*) y los estadísticos de neutralidad (*D* de Tajima y *H* de Fay & Wu) se obtuvieron utilizando el programa bioinformático DnaSP6.

Para el cálculo de S, *π y D de Tajima* se utilizó la opción de *Multidomain Analysis* que devuelve los estadísticos para cada región previamente concatenada con este programa.

En el caso del estadístico de divergencia *K* se utilizó la opción *Polymorphism and Divergence* y para el estadístico de neutralidad *H* se utilizó la opción *Fu and Li´s test with an outgroup*. En ambos casos se recurrió a un análisis de *sliding windows* donde las secuencias se analizaron en ventanas de 5.001 nucleótidos con un desplazamiento de 5.001 nucleótidos entre ellas, lo que corresponde a obtener los resultados para cada región independiente.

La significación estadística de los estadísticos *D* de Tajima y *H* de Fay & Wu se obtuvo a partir de la realización de simulaciones de cada una de las poblaciones de *D. melanogaster* de 1.000 iteraciones. Se consideró cada ventana definida en el cálculo de los distintos indicadores y estadísticos como loci independientes.

Para realizar las simulaciones se utilizó el programa mlcoalsim.[12] En los ficheros de entrada al programa, se debe utilizar el número de nucleótidos por loci y la recombinación asociada a cada uno de ellos.

El cálculo del parámetro de recombinación de la población (R= 4Nr) se calculó considerando una tasa de recombinación r (HR recomb cM/Mb) [25] obtenida en Popfly mediante el *track* correspondiente y para cada región en estudio.

A partir de las simulaciones y las probabilidades se identificaron los fragmentos que eran significativos con un nivel de significación α igual a 0,025 al estar evaluando una distribución de dos colas.

Tanto para *D* de Tajima como para *H* de Fay & Wu se realizaron múltiples test para regiones independientes. Con ello se incrementa el riesgo de cometer un error de tipo I (rechazar la hipótesis nula cuando en realidad es correcta). Por ello se aplicó una corrección de Bonferroni secuencial con un nivel de significación alfa = 0,025 para los análisis de cada estadístico y población.

4.4 Estudio de la diferenciación genética.

Para analizar la diferenciación genética entre las poblaciones se calculó el estadístico F_{ST} a través del programa bioinformático mstastpop.[13]

Con este fin se realizaron comparaciones dos a dos entre las distintas poblaciones, una comparación global entre poblaciones teniendo en cuenta todas las regiones analizadas conjuntamente, otra solo con las regiones que se encontraban dentro de la inversión y otra con las regiones que se encontraban fueran de la inversión.

Se realizó una representación gráfica de los resultados mediante un análisis de componentes principales *PCoA* a través del complemento de Excel GenAIEx 6.5 [26] el cual arrojó el porcentaje de variabilidad explicada a partir de cada eje, así como la localización de cada población con respecto a los ejes definidos mediante este análisis.

Además, para descubrir un posible efecto de la distancia geográfica sobre la diferenciación genética, se compararon las dos matrices de distancias mediante un test de Mantel. Este test se realizó también a través del complemento de Excel GenAIEx 6.5, que devuelve la representación gráfica de la relación entre estas distancias, así como su grado de significación, obtenido a partir de un test de permutaciones.

5 Resultados

5.1 Diferenciación genética entre poblaciones.

A través del *PCoA* se observa para el estadístico F_{ST} tras comparar las siete poblaciones entre ellas, la tendencia general en todas las regiones de que las dos poblaciones de EE. UU. (RAL_ST y RAL_INV) están muy poco diferenciadas. En muchos de estos casos, además, se observan para F_{ST} valores <0 en la matriz de población por pares.

Se observa que la población de Zambia con inversión (ZI_INV) generalmente se diferencia más del resto de poblaciones. En las regiones donde la población más diferenciada no es ZI_INV, la población de Etiopia, (EF), es, normalmente, la población más diferenciada.

Por otro lado, el primer eje del *PCoA* diferencia bien entre las poblaciones ancestrales (africanas) y las poblaciones derivadas (europea y americana) Asimismo, el *PCoA* para F_{ST} explica el 74,78 % de la variabilidad observada con los dos primeros ejes.

Axis	**1**	**2**	**3**
%	48,94	25,85	12,17
Cum %	48,94	74,78	86,95

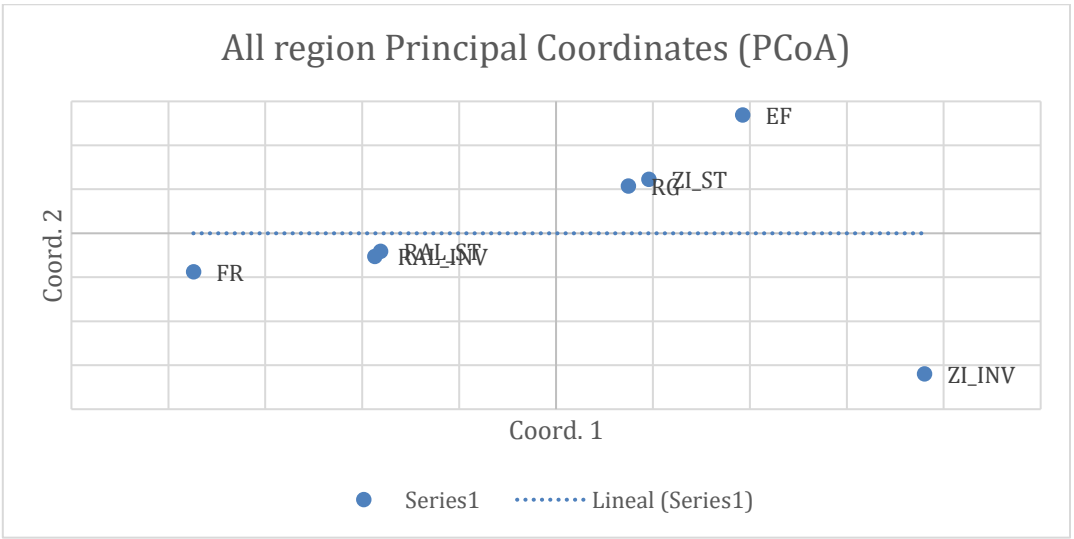

Figura 2: *Porcentaje de variación del estadístico F_{ST} explicado con los tres primeros ejes del PCoA y representación gráfica del mismo usando los dos primeros ejes.*

Tras evaluar las correlaciones entre el estadístico F_{ST} y la distancia geográfica (en Km) a través del *test de Mantel* para todas las regiones en estudio concatenadas, se ha comprobado que no se observa una diferenciación genética paralela a la distancia geográfica. Esto se puede aprecia en la recta de regresión que es prácticamente plana y además comprobamos que no existe correlación (p>0,05) *(figura 3).*

SSx	SSy	SPxy	Rxy	P(rxy-rand >= rxy-data)
436131327,671	0,182	1003,088	0,112	0,300

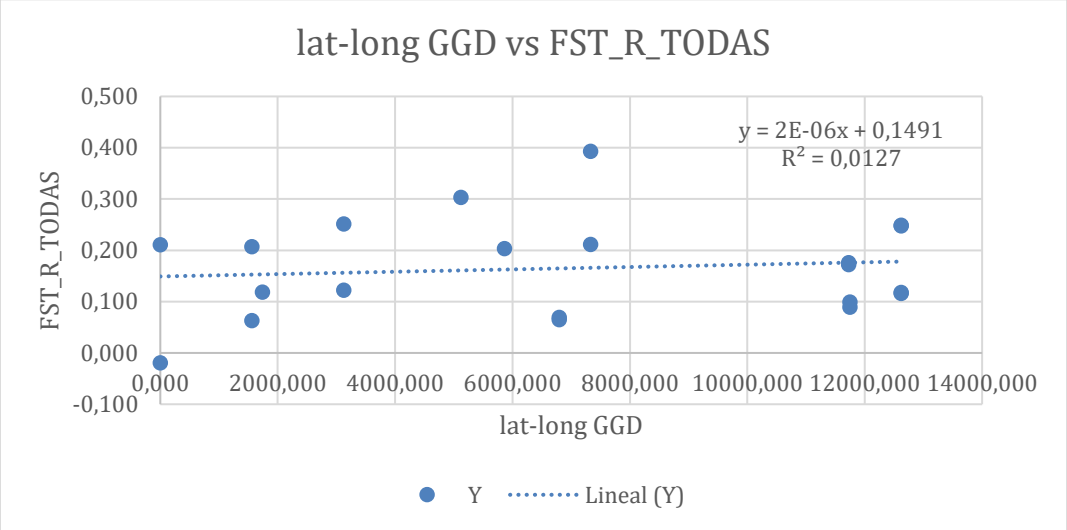

Figura 3: *Correlación de F_{ST} con la distancia geográfica en Km*

También se ha evaluado, a través del test de Mantel, la correlación entre el estadístico F_{ST} y la distancia geográfica, específicamente para las poblaciones con ordenación estándar (ST). De este modo, se mantienen todos los puntos geográficos analizados, pero se elimina cualquier posible efecto distorsionador de la ordenación cromosómica sobre la diferenciación genética. En este caso, en el gráfico obtenido (fig.4) no se observa tampoco una diferenciación genética paralela a la distancia geográfica lo que queda confirmado por el bajo valor de la correlación obtenida (Rxy = 0,018 *P*(Rxy-rand ≥ Rxy-data) = 0.49, obtenido a partir de 99 permutaciones).

SSx	SSy	SPxy	Rxy	P(rxy-rand >= rxy-data)
153175058,951	0,051	49,169	0,018	0,490

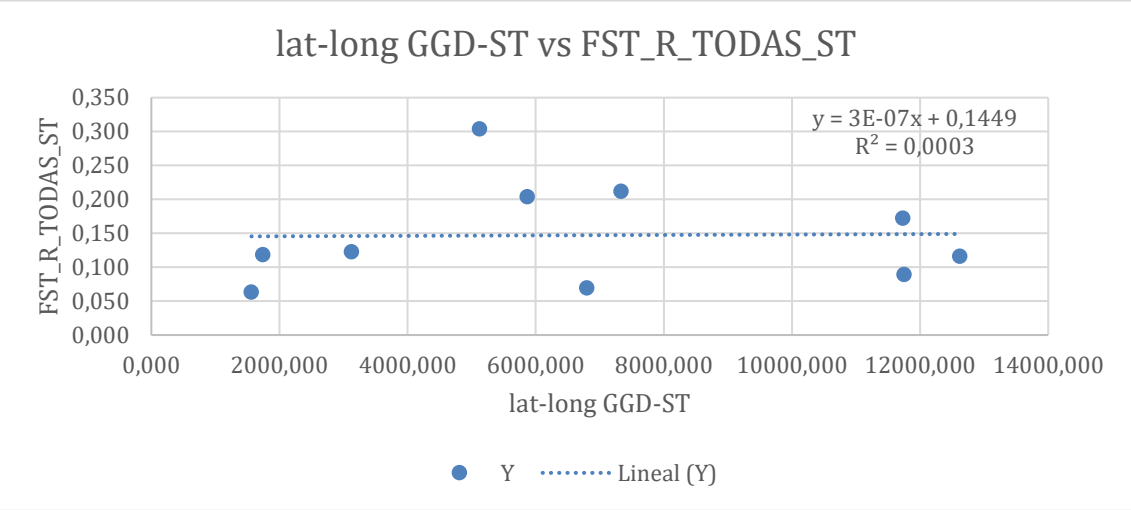

Figura 4: *Correlación de FST con la distancia geográfica para las poblaciones estándar, ST.*

5.2 Resultados de los indicadores de polimorfismo y divergencia genética.

A partir del análisis de los indicadores de polimorfismos *S* y π observamos dos patrones diferentes. Por un lado, las poblaciones de Zambia, (ZI_INV), Etiopía y Francia presenta unos valores más bajos de *S* y de π que las poblaciones de Zambia (ZI_ST), Ruanda y EEUU (RAL_ST y RAL_INV). Estos datos no se correlacionan con la distribución geográfica de las poblaciones.

En cambio, para el indicador de divergencia genética *K*, las siete poblaciones presentan valores muy similares.

Población	Valor Medio *S*	Valor medio π	Valor medio *K*
Zambia (ZI_ST)	155	0,0107	0,047
Zambia (ZI_INV)	78,6	0,0061	0,048
Etiopia (EF)	96	0,0074	0,045
Ruanda (RG)	124,6	0,0092	0,047
Francia (FR)	66,6	0,0052	0,047
EEUU (RAL_ST)	106,7	0,0081	0,047
EEUU (RAL_INV)	100,9	0,0079	0,045

Tabla 3: *Valores medios para los indicadores de polimorfismo y divergencia genética.*

Gráficamente, se observa en general que a partir de la región 20 los valores observados para π se sitúan por debajo de la media, acentuándose la distancia a la media cuanto más al extremo del cromosoma 2L nos encontramos. *(Fig. 5)*

Sin embargo, para los valores observados para el estadístico *K*, se observa que a partir de la región 20 se sitúan por encima de la media, alejándose de la media cuanto más al extremo del cromosoma 2L nos encontramos. *(Fig. 6)*

Se observan discontinuidades en las gráficas por dos razones, la primera entre las regiones R2 y R3 y entre las regiones R14 y R15 para hacer visible el inicio y el fin de la inversión. Por otro lado, para Etiopia y EEUU (INV) se observan dos discontinuidades adicionales que se corresponden con regiones sin información nucleotídica para esas poblaciones y por tanto sin posibilidad de calcular los estadísticos en estudio.

Si atendemos a la relación π/K se observa una zona por encima de la media entre las regiones 4 y 11 para las poblaciones de Zambia (ST), Ruanda, Francia y EEUU. *(Fig. 7)*

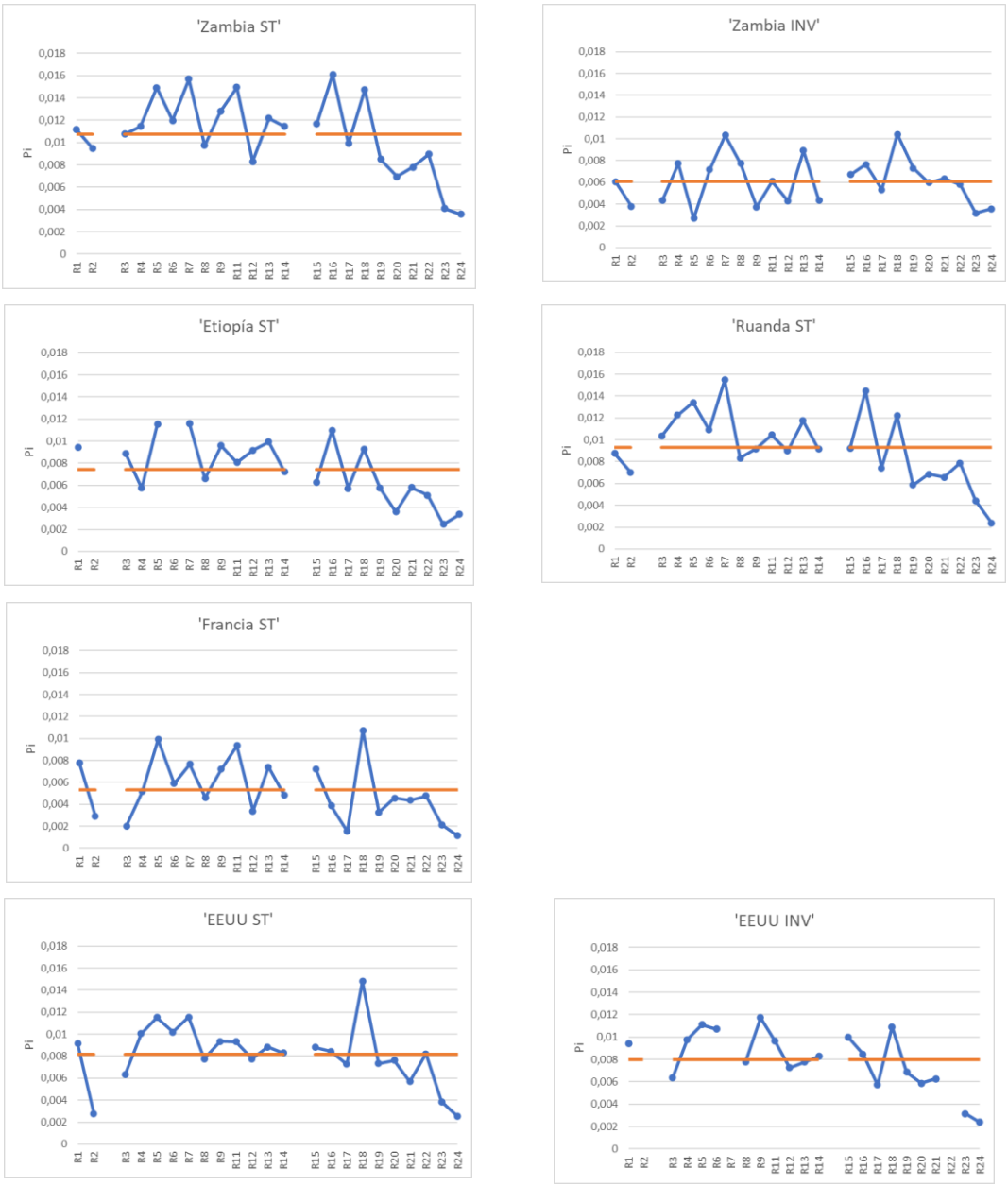

Figura 5: *Distribución de la diversidad nucleotídica (π).*

Los espacios entre las regiones 2 y 3 y entre 14 y 15 corresponden al inicio y final de la inversión respectivamente.

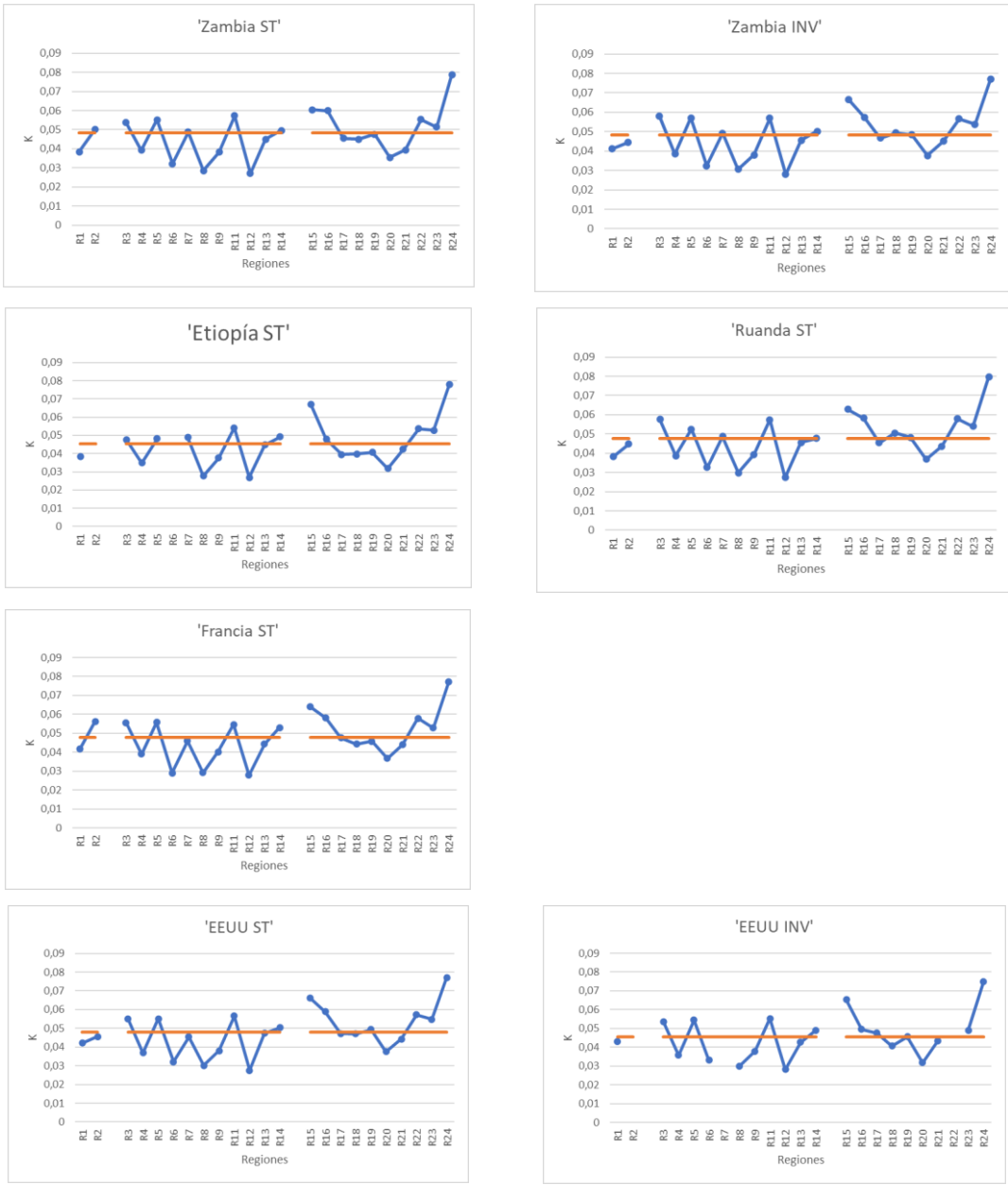

Figura 6: *Distribución de la divergencia genética (K).*

Los espacios entre las regiones 2 y 3 y entre 14 y 15 corresponden al inicio y final de la inversión respectivamente.

Figura 7: *Distribución de la relación π /K).*

Los espacios entre las regiones 2 y 3 y entre 14 y 15 corresponden al inicio y final de la inversión respectivamente

En cuanto al test *D* de Tajima, la población que obtiene un mayor número de regiones estadísticamente significativas es Zambia (ZI_ST) aunque solo representan el 34% de todas las regiones. A parte de esta población solo Zambia (ZI_INV) y EEUU (RAL_ST) tienen alguna región donde el valor de *D* de Tajima sea significativo. No ocurre lo mismo para Etiopia, Ruanda, Francia y EEUU (RAL_INV). *(Tabla 4).*

En cuanto al estadístico *H* de Fay & Wu, se observa que el número de poblaciones estadísticamente significativas es menor que los obtenidos para el estadístico *D* de Tajima.

Las poblaciones que han obtenido más fragmentos significativos son las que tienen la ordenación cromosómica invertida, Zambia (ZI_INV) y EEUU (RAL_INV) con tres regiones, aunque no coinciden.

Destaca la falta de regiones donde el estadístico *H* sea estadísticamente significativo en las poblaciones africanas con ordenación cromosómica estándar.

D de Tajima

	ZI_ST	ZI_INV	EF_ST	RG_ST	FR_ST	RAL_ST	RAL_INV
R1							
R2	*						
R3		*					
R4	*						
R5							
R6	*						
R7	*						
R8							
R9		*					
R11	*						
R12							
R13	*	*					
R14							
R15							
R16							
R17							
R18	*						
R19						*	
R20							
R21						*	
R22		*					
R23							

H de Fay&Wu

	ZI_ST	ZI_INV	EF_ST	RG_ST	FR_ST	RAL_ST	RAL_INV
R1							
R2							
R3							*
R4							*
R5		*					
R6							
R7							
R8					*		
R9		*					
R11							
R12							
R13							*
R14							
R15							
R16							
R17		*			*		
R18							
R19							
R20							
R21							
R22							
R23							

Tabla 4: *Regiones estadísticamente significativas para el estadístico D y H. Solo se indica para las que siguen siendo significativas tras la corrección de Bonferroni.*

Los valores de *H* son mayoritariamente negativos, excepto para la población con ordenación estándar de Zambia (ZI_ST) y la población con ordenación invertida de EEUU (RAL_INV). *(Tabla 5)*

Destaca que las dos poblaciones con ordenación cromosómica invertida, ZI_INV y RAL_INV, muestren una tendencia inversa a la que muestran las poblaciones estándar en relación a los valores positivos y negativos del estadístico *H*.

Los valores de *H* para los que han mostrado significación estadística son para todas las poblaciones, excepto para una región en RAL_INV, negativos, lo que implica un exceso de variantes derivadas a frecuencias elevadas.

Población	N.º H positivo	N.º H negativo	N.º total
Zambia (ZI_ST)	18	5	23
Zambia (ZI_INV)	5	18	23
Etiopia (EF_ST)	8	13	21
Ruanda (RG_ST)	9	14	23
Francia (FR_ST)	3	20	23
EEUU(RAL_ST)	9	14	23
EEUU(RAL_INV)	10	7	17

Tabla 5: *Número de regiones con valores positivos y negativos del estadístico H*

La representación gráfica de los valores del estadístico *D* de Tajima no muestra diferencias apreciables entre poblaciones más allá de algunos valores más extremos, comparados con el resto de las poblaciones, en las poblaciones de Zambia INV y Francia ST. *(Figura 8)*

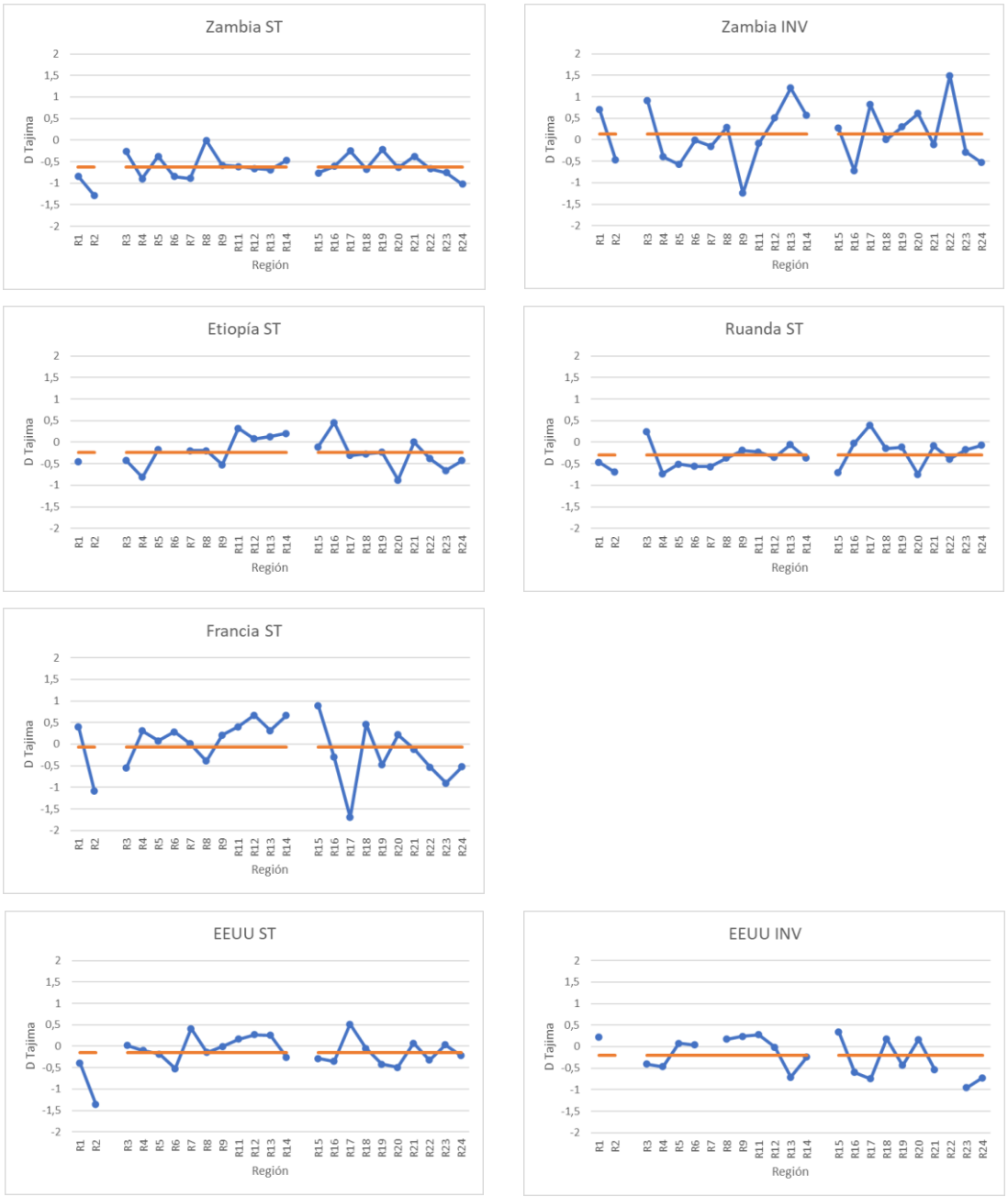

Figura 8: *Distribución del estadístico D de Tajima*

Los espacios entre las regiones 2 y 3 y entre 14 y 15 corresponden al inicio y final de la inversión respectivamente

La representación gráfica de los valores del estadístico *H* de Fay & Wu no presenta diferencias o tendencias apreciables entre las distintas poblaciones más allá de la tendencia en la región 3 a valores similares y en torno a -15 en las poblaciones de Zambia ST, Etiopía ST, Ruanda ST y EEUU ST e INV. *(Figura 9)*

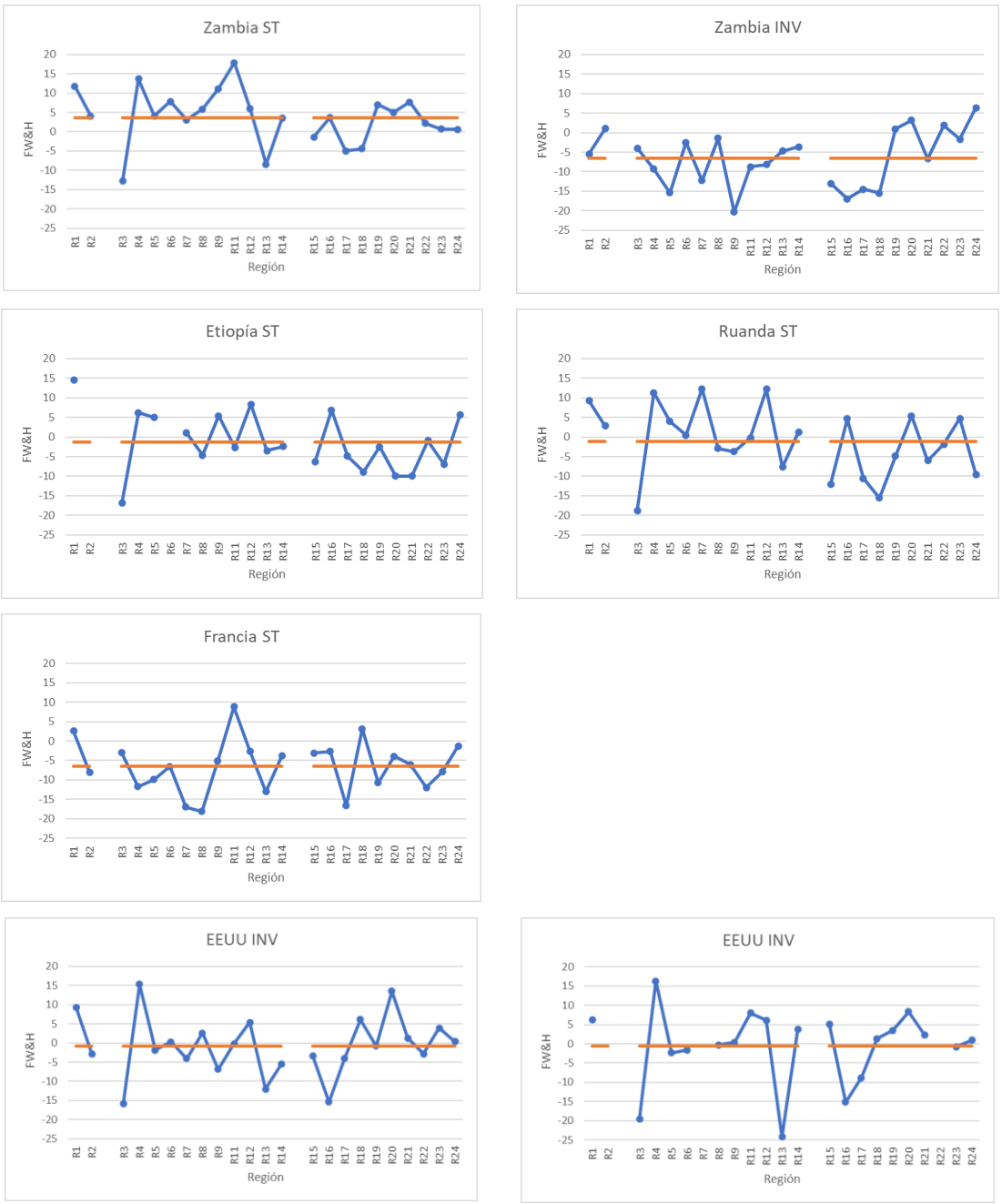

Figura 9: *Distribución del estadístico H de FW&H*

Los espacios entre las regiones 2 y 3 y entre 14 y 15 corresponden al inicio y final de la inversión respectivamente

6 Discusión

6.1 Análisis de resultados de diferenciación genética entre poblaciones.

Cuando se realiza el análisis de coordenadas principales (*PCoA*), se observa que la gran mayoría de regiones se comportan de igual manera cuando se consideran una a una que cuando se consideran todas las regiones conjuntamente, o las regiones que se encuentran dentro o fuera de la inversión.

Se observa que, generalmente, la población más diferenciada es la de ZI_INV y las más cercanas las dos poblaciones de EEUU, RAL_ST y RAL_INV. En muchos de los análisis entre las dos poblaciones americanas, además, si comprobamos los valores del estadístico F_{ST} se observan valores <0 que son indicativos de que las muestras pueden pertenecer a una única población panmítica. Esto nos puede llevar a pensar que al encontrarse en un ambiente nuevo (América) y con un menor número de individuos, al menos inicialmente, se dieran procesos especiales de mezcla entre los dos tipos de cromosomas.

En contraposición tenemos las dos poblaciones de Zambia, ZI_ST y ZI_INV, las cuales también coexisten como las americanas, pero que al tratarse de poblaciones ancestrales están mucho más diferenciadas como podemos observar a partir del análisis de coordenadas principales *(PCoA)* y a partir de los valores observados del *estadístico* F_{ST}*,* siempre positivos*.*

También se comprueba que las poblaciones derivadas más cercanas a las ancestrales son las de EEUU, RAL_ST y RAL_INV, lo que se podría explicar por la doble colonización de *D. melanogaster* en América. [27] Parece que la colonización de esta especie en América tuvo lugar en dos fases, una hace 500 años, con la introducción de la especie desde África tropical, de donde es originaria, a América Tropical, y otra a finales del siglo XIX, con ejemplares procedentes de Europa a Norte América. Esta última sufrió un importante cuello de botella debido al reducido número de ejemplares y se sugiere que esto llevo a la mezcla de las poblaciones de ambas regiones del continente americano con el resultado de una proporción en torno al 15% de ancestros de las poblaciones originarias africanas.

A partir de la evaluación de las correlaciones entre el estadístico F_{ST} y la distancia geográfica (en Km) para todas las regiones en estudio concatenadas, se ha comprobado que no se observa una diferenciación genética paralela a la distancia geográfica. Lo mismo ocurre cuando solamente tenemos en cuenta las poblaciones con ordenaciones cromosómicas estándar.

Los resultados obtenidos del *PCoA* y test de *Mantel ponen* de manifiesto el importante papel que debe haber jugado la diferenciación genética entre estas poblaciones tanto para procesos demográficos como selectivos. En este sentido cabe destacar la semejanza entre las poblaciones americanas y las africanas, lo que, como se ha comentado, se podría explicar por la doble colonización de *D. melanogaster* en América.

Asimismo, la gran diferenciación entre las dos poblaciones de Zambia, una con la ordenación cromosómica estándar y la otra con la ordenación cromosómica invertida, indican que una estructura poblacional puede mantenerse por selección. Esta afirmación se refuerza cuando se comprueba que ZI_INV es la más diferente, lo cual se explica por la aparición de la inversión que atrapa una variante inicial determinada.

Por otro lado, se ha encontrado que la siguiente población más diferenciada es la de Etiopia, EF, lo cual entra en concordancia con lo que ya se ha visto para algunas regiones del cromosoma X (Orengo, comunicación personal) donde se indica que esta población puede estar sometida a una selección distinta debido a un ambiente muy distinto determinado por la altitud.

6.2 Análisis de los resultados de los indicadores de polimorfismo y divergencia genética.

La comparación entre la ordenación cromosómica estándar e invertida en la población de Zambia, ZI, a partir de los valores de *S* y de *Pi*, nos indica que ZI_ST es mucho más variable que ZI_INV. Se podría realizar un test de signos (no paramétrico) que pudiera confirmar esta tendencia de mayor variabilidad.

Se ha observado que la gran mayoría de los valores obtenidos para el estadístico *D* de Tajima son negativos, lo cual indica un exceso de polimorfismo de baja frecuencia. En particular en las poblaciones con ordenación estándar, *D* de Tajima siempre es negativo, mientras que, en las poblaciones con ordenación invertida, existen algunos valores positivos. Podría ser interesante comprobar cómo se comportan las regiones donde la diferencia es más acusada.

Destaca que las dos poblaciones con ordenación cromosómica inversa, ZI_INV y RAL_INV, muestren una tendencia inversa a la que muestran las poblaciones estándar en relación a los valores positivos y negativos del estadístico *H*.

Los valores de *H* para los que se ha obtenido significación estadística son siempre negativos, excepto para una región en RAL_INV. Los valores, negativos implican un exceso de variantes derivadas a frecuencias elevadas, que es un indicativo de selección positiva.

7 Conclusiones

7.1 Conclusiones

En base a los resultados obtenidos a lo largo de la realización de este trabajo se pueden extraer las siguientes conclusiones:

- Se ha detectado diferenciación genética entre seis de las poblaciones en estudio comportándose de igual manera cuando se considera cada región genómica una a una que cuando se consideran todas las regiones conjuntamente, o agrupando las regiones que se encuentran dentro o fuera de la inversión. La única excepción es entre las dos poblaciones de EEUU, en sus dos ordenaciones cromosómicas, RAL_ST y RAL_INV, las cuales no han presentado diferenciación genética independientemente del escenario elegido.

- La diferenciación genética entre poblaciones no es paralela a la distancia geográfica a partir de los resultados obtenidos con el test de Mantel para las regiones en estudio concatenadas. Se ha obtenido el mismo resultado cuando solo se han tenido en cuanta las poblaciones con ordenación estándar. Esto indica que, en la diferenciación de las poblaciones, intervienen otros factores más importantes como puede ser la adaptación a distintos ambientes, pero también la obstrucción a la recombinación entre ordenaciones cromosómicas distintas.

- La gran diferenciación entre las dos poblaciones de Zambia, una con la ordenación cromosómica estándar y la otra con la ordenación cromosómica invertida, indican una estructura poblacional que puede mantenerse por una barrera en la recombinación entre cromosomas de distintas ordenaciones. Esta afirmación se refuerza cuando se comprueba que ZI_INV es la más diferente del resto, lo cual se explica por la aparición de la inversión que atrapa una variante inicial determinada.

- Se ha encontrado que la población de Etiopía, EF, se diferencia genéticamente del resto de poblaciones africanas, lo cual indica que esta población puede estar sometida a una selección distinta debido a un ambiente muy distinto determinado posiblemente por la altitud.

- Las dos poblaciones americanas RAL_ST y RAL_INV, siempre quedan agrupadas en el análisis de PCoA y, en base a la presencia de valores negativos para el estadístico F_{ST}, parecen pertenecer a una única población panmictica.

- Se ha observado un exceso de polimorfismo de baja frecuencia basado en que la gran mayoría de los valores obtenidos para el estadístico *D* de Tajima son negativos.

- Se ha detectado indicios de selección positiva a través del exceso de variantes derivadas a frecuencias elevadas apoyado en que los valores de *H* para los que se ha obtenido significación estadística son siempre negativos

7.2 Líneas de futuro

- Analizar en mayor profundidad con el fin de buscar posibles dianas de selección aquellas regiones cromosómicas que presentan una diferencia más acusada entre los valores del estadístico *D* de Tajima al comparar las dos poblaciones cromosómicas de una misma localidad.

- Relacionar la variabilidad nucleotídica con la distancia a los puntos de rotura de la inversión polimórfica, la distancia al gen más cercano y la frecuencia de recombinación, entre otras características.

- Añadir más regiones estratégicamente localizadas a lo largo del cromosoma 2L para obtener un muestreo de más precisión.

7.3 Seguimiento de la planificación

La dificultad para seleccionar una muestra suficiente de poblaciones e individuos para que el estudio fuera representativo ha tenido un impacto considerable sobre la planificación inicial que ha sido compensado con el aumento de la dedicación al trabajo, cumpliéndose en líneas generales, los objetivos planteados inicialmente.

La falta de experiencia con los programas bioinformáticos utilizados ha tenido impacto en la planificación inicial no tanto por la dificultad de uso sino por problemas de configuración e instalación que finalmente se han sorteado y solucionado.

8 Glosario

- *D*: Estadístico *D* de Tajima: Es una estadística de prueba genética poblacional. Se calcula como la diferencia entre el número medio de diferencias por pares y el número de sitios segregantes.

- DnaSP: DNA *Sequence Polymorphism. Software* para el análisis de polimorfismos de ADN utilizando datos de un solo locus o de varios loci.

- F_{ST}: Este estadístico se utiliza para analizar la diferenciación genética entre las poblaciones comparadas dos a dos.

- GenAIEx: Aplicación para el análisis genético de poblaciones que funciona como complemento de Excel.

- *H:* Estadístico H de Fay & Wu cuyo propósito es distinguir entre una secuencia de ADN que evoluciona aleatoriamente ("neutralmente") y una que evoluciona bajo selección positiva.

- *K:* Estadístico que muestra la divergencia genética observada como la media de las diferencias nucleotídicas observadas al comparar secuencias dos a dos.

- mlcoalsim*: Multilocus coalescent simulations. Software* que realiza simulaciones y realiza análisis multilocus.

- mstatpop: Software que realiza análisis de polimorfismos de secuencias de ADN.

- π*:* Estadístico que muestra la diversidad nucleotídica. Número medio de diferencias entre pares que se esperan observar por nucleótido.

- PCoA: Análisis de coordenadas principales analiza una matriz de distancia o disimilitud. El objetivo del PCoA es representar estas distancias en un espacio euclidiano de pocas dimensiones (usualmente de 2 o 3 ejes) con la menor pérdida de información posible.

- *S*: Número de posiciones segregantes.

- Población panmíctica: Población en la que el apareamiento es libre y al azar y las frecuencias genotípicas se pueden calcular a partir de las gaméticas.

- Test de Mantel: Test estadístico de la correlación entre dos matrices del mismo rango. Una matriz puede contener las estimaciones de las distancias genéticas mientras que la otra puede contener estimaciones

de la distancia geográfica entre los rangos de cada especie y todas las demás especies.

9 Bibliografía

1. Pool, J. E. *et al.* Population Genomics of Sub-Saharan Drosophila melanogaster: African Diversity and Non-African Admixture. *PLoS Genetics* **8**, (2012).
2. Sprengelmeyer, Q. D. *et al.* Recurrent Collection of Drosophila melanogaster from Wild African Environments and Genomic Insights into Species History. *Molecular Biology and Evolution* **37**, 627–638 (2020).
3. Orengo, D. J. & Aguadé, M. Detecting the footprint of positive selection in a European population of Drosophila melanogaster: Multilocus pattern of variation and distance to coding regions. *Genetics* **167**, 1759–1766 (2004).
4. Hervas S, Sanz E, Casillas S, Pool JE & Barbadilla A. PopFly: the Drosophila population genomics browser. Bioinformatics, 33, 2779-2780; https://doi.org/10.1093/bioinformatics/btx301. (2017).
5. Larkin A *et al.* FlyBase: updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res. 49(D1) D899–D907. https://flybase.org/ (2021).
6. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Research* **49**, D884–D891 (2021).
7. Tajima, F. *Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism*. https://academic.oup.com/genetics/article/123/3/585/5998755 (1989).
8. Fay, J. C. & Wu, C.-I. *Hitchhiking Under Positive Darwinian Selection*. https://academic.oup.com/genetics/article/155/3/1405/6050858 (2000).
9. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
10. Kimura, M. & others. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
11. Rozas, J. *et al.* DnaSP, DNA Sequence Polymorphism. (2019).
12. Ramos-Onsins, S. E. & Mitchell-Olds, T. Mlcoalsim: multilocus coalescent simulations. *Evol Bioinform Online* **3**, 41–44 (2007).
13. Ramos-Onsins, S. E. *et al.* mstatspop: Statistical Analysis using Multiple Populations for Genomic Data. (2022).
14. Kapopoulou, A., Pfeifer, S. P., Jensen, J. D. & Laurent, S. The demographic history of african drosophila melanogaster. *Genome Biology and Evolution* vol. 10 2338–2342 (2018).
15. Thornton, K. & Andolfatto, P. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster. *Genetics* **172**, 1607–1619 (2006).
16. Li, H. & Stephan, W. Inferring the Demographic History and Rate of Adaptive Substitution in Drosophila. *Plos Genetics* (2016) doi:10.1371/journal.pgen.
17. Glinka, S., Ometto, L., Mousset, S., Stephan, W. & de Lorenzo, D. Demography and Natural Selection Have Shaped Genetic Variation in Drosophila melanogaster: A Multi-locus Approach. *Genetics* **165**, 1269–1278 (2003).

18. Orengo, D. J. & Aguadé, M. Detecting the footprint of positive selection in a European population of Drosophila melanogaster: Multilocus pattern of variation and distance to coding regions. *Genetics* **167**, 1759–1766 (2004).
19. Hutter, S., Li, H., Beisswanger, S., de Lorenzo, D. & Stephan, W. Distinctly different sex ratios in African and European populations of Drosophila melanogaster inferred from chromosomewide single nucleotide polymorphism data. *Genetics* **177**, 469–480 (2007).
20. Kapopoulou, A. *et al.* Demographic analyses of a new sample of haploid genomes from a Swedish population of Drosophila melanogaster. *Scientific Reports* **10**, (2020).
21. Kapun, M. *et al.* Drosophila Evolution over Space and Time (DEST): A New Population Genomics Resource . *Molecular Biology and Evolution* **38**, 5782–5805 (2021).
22. MacKay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
23. Lack, J. B. *et al.* The Drosophila genome nexus: A population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics* **199**, 1229–1241 (2015).
24. Hervas S, Sanz E, Casillas S, Pool JE & Barbadilla A. PopFly: the Drosophila population genomics browser. Bioinformatics, 33, 2779-2780; https://doi.org/10.1093/bioinformatics/btx301. (2017).
25. Comeron, J. M., Ratnappan, R. & Bailin, S. The Many Landscapes of Recombination in *Drosophila melanogaster*. *PLoS Genetics* **8**, 33–35 (2012).
26. PEAKALL, R. O. D. & SMOUSE, P. E. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288–295 (2006).
27. Campo, D. *et al.* Whole-genome sequencing of two North American *Drosophila melanogaster* populations reveals genetic differentiation and positive selection. *Molecular Ecology* **22**, 5084–5097 (2013).

10 Anexos

Anexo 1: Coordenadas de inicio y fin de cada región en estudio, genes colindantes y hebra en la que se encuentran.

Region	Inicio Reg.	Fin Reg.	Lon. Reg	Gen Inicio Reg.	Gen Fin Reg.	Hebra
1	965017	970017	5.001	CG4341	lncRNA:CR44801	-
2	1811524	1816524	5.001	Gr22a	CG31933	-
Inicio Inversión	**2225744**					
3	2478088	2483088	5.001	lncRNA:CR45286	Slh	-
4	3228853	3233853	5.001	lncRNA:CR45336	lncRNA:CR45337	-
5	4017302	4022302	5.001	CG43774	ed	+
6	5447958	5452958	5.001	lncRNA:CR31647	mid	+
7	6240142	6245142	5.001	lncRNA:CR44823	Ddr	+
8	7292426	7297426	5.001	Wnt4	wg	+
9	8801573	8806573	5.001	CG42710	SoxN	+
11	10714968	10719968	5.001	lncRNA:CR44407	CG6431	+
12	11411693	11416693	5.001	lncRNA:CR43681	salm	-
13	12325582	12330582	5.001	lncRNA:CR45283	CG31862	+
14	13080000	13085000	5.001	CG9932	lncRNA:CR44197	+
Fin Inversión	**13154180**					
15	13336948	13341948	5.001	hbt	CG16826	+
16	14249880	14254880	5.001	lncRNA:CR44216	wb	+
17	15344755	15349755	5.001	esg	CG15258	-
18	16382196	16387196	5.001	jhamt	CG5888	-
19	17342719	17347719	5.001	CG45691	lncRNA:CR43239	-
20	18420629	18425629	5.001	lncRNA:CR44487	lncRNA:CR44488	+
21	19236285	19241285	5.001	lncRNA:CR44189	CG31797	-
22	20124598	20129598	5.001	CG13970	CG10651	-
23	20954346	20959346	5.001	lncRNA:CR44606	CG42238	-
24	21805841	21810841	5.001	lncRNA:CR44269	lncRNA:CR44786	+

Nota: Gen Inicio Reg: Gen que encontramos justo antes del inicio de la región. Gen Fin Reg: Gen que encontramos justo después del fin de la región.

Anexo 2: Script en R selección de individuos. Ejemplo para la población ZI, existe un script por cada una de las siete poblaciones

```
#Función que genera un dataframe con tres columnas, una con el individuo, otra
#con el número de R y otra con la región
ind_ns_r <- function(f,r){

  library(stringr)

  #Añadimos una columna con el número de Ns en la secuencia de ese individuo en
  #esa región
  for (i in 1:(nrow(f)-1)){
    f$Ns[i] <- str_count(f$X1[i] , "N")
  }
  #Añadimos esa información a la misma fila donde está el codigo del individuo
  for (i in 1:(nrow(f)-1)){
    if(i%%2!=0)
      f$Ns[i] <- f$Ns[i+1]
```

```
  }

  #Eliminamos el caracter '>' y la
  #cadena '_Chr2L', dejando así solo el código del individuo.
  for (i in 1:(nrow(f))){
    if(i%%2!=0)
      f$X1[i] <- gsub('.{6}$', '', substring(f$X1[i], 2))
  }

  #Asignamos el valor 0 a las filas de las secuencias para eliminarlas después
  for (i in 1:(nrow(f))){
    if(i%%2==0)
      f$Ns[i] <- 0
  }

  #Eliminamos las filas con las secuencias pues ya no las necesitamos
  f <- f[f$Ns != 0, ]

  #Renombramos la columna X1 a ind
  library(dplyr)
  f <- rename(f, ind = X1)

  #Añadimos una columna con la región
  f <- cbind(f, region = r)

  #Se elimina los individuos con Ns>4999
  f <- f[f$Ns < 4999, ]

  return(f)
}


#Cambiamos el directorio de trabajo a la carpeta donde tenemos los ficheros
setwd("~/JOSE/UOC/R/INDIVIDUOS")

#Cargamos fichero
library(readr)

ZI1 <- read_csv("ZI_Chr2L_965017_970017.fasta", col_names = FALSE)
ZI2 <- read_csv("ZI_Chr2L_1811524_1816524.fasta", col_names = FALSE)
ZI3 <- read_csv("ZI_Chr2L_2478088_2483088.fasta", col_names = FALSE)
ZI4 <- read_csv("ZI_Chr2L_3228853_3233853.fasta", col_names = FALSE)
ZI5 <- read_csv("ZI_Chr2L_4017302_4022302.fasta", col_names = FALSE)
ZI6 <- read_csv("ZI_Chr2L_5447958_5452958.fasta", col_names = FALSE)
ZI7 <- read_csv("ZI_Chr2L_6240142_6245142.fasta", col_names = FALSE)
ZI8 <- read_csv("ZI_Chr2L_7292426_7297426.fasta", col_names = FALSE)
ZI9 <- read_csv("ZI_Chr2L_8801573_8806573.fasta", col_names = FALSE)
ZI10 <- read_csv("ZI_Chr2L_9478520_9483520.fasta", col_names = FALSE)
ZI11 <- read_csv("ZI_Chr2L_10714968_10719968.fasta", col_names = FALSE)
ZI12 <- read_csv("ZI_Chr2L_11411693_11416693.fasta", col_names = FALSE)
ZI13 <- read_csv("ZI_Chr2L_12325582_12330582.fasta", col_names = FALSE)
```

```
ZI14 <- read_csv("ZI_Chr2L_13080000_13085000.fasta", col_names = FALSE)
ZI15 <- read_csv("ZI_Chr2L_13336948_13341948.fasta", col_names = FALSE)
ZI16 <- read_csv("ZI_Chr2L_14249880_14254880.fasta", col_names = FALSE)
ZI17 <- read_csv("ZI_Chr2L_15344755_15349755.fasta", col_names = FALSE)
ZI18 <- read_csv("ZI_Chr2L_16382196_16387196.fasta", col_names = FALSE)
ZI19 <- read_csv("ZI_Chr2L_17342719_17347719.fasta", col_names = FALSE)
ZI20 <- read_csv("ZI_Chr2L_18420629_18425629.fasta", col_names = FALSE)
ZI21 <- read_csv("ZI_Chr2L_19236285_19241285.fasta", col_names = FALSE)
ZI22 <- read_csv("ZI_Chr2L_20124598_20129598.fasta", col_names = FALSE)
ZI23 <- read_csv("ZI_Chr2L_20954346_20959346.fasta", col_names = FALSE)
ZI24 <- read_csv("ZI_Chr2L_21805841_21810841.fasta", col_names = FALSE)
ZI25 <- read_csv("ZI_Chr2L_22422241_22427241.fasta", col_names = FALSE)


#Llamos a la función que cuenta el número de Ns y añade la región como parámetro
ZI1N <- ind_ns_r(ZI1,"R1")
ZI2N <- ind_ns_r(ZI2,"R2")
ZI3N <- ind_ns_r(ZI3,"R3")
ZI4N <- ind_ns_r(ZI4,"R4")
ZI5N <- ind_ns_r(ZI5,"R5")
ZI6N <- ind_ns_r(ZI6,"R6")
ZI7N <- ind_ns_r(ZI7,"R7")
ZI8N <- ind_ns_r(ZI8,"R8")
ZI9N <- ind_ns_r(ZI9,"R9")
ZI10N <- ind_ns_r(ZI10,"R10")
ZI11N <- ind_ns_r(ZI11,"R11")
ZI12N <- ind_ns_r(ZI12,"R12")
ZI13N <- ind_ns_r(ZI13,"R13")
ZI14N <- ind_ns_r(ZI14,"R14")
ZI15N <- ind_ns_r(ZI15,"R15")
ZI16N <- ind_ns_r(ZI16,"R16")
ZI17N <- ind_ns_r(ZI17,"R17")
ZI18N <- ind_ns_r(ZI18,"R18")
ZI19N <- ind_ns_r(ZI19,"R19")
ZI20N <- ind_ns_r(ZI20,"R20")
ZI21N <- ind_ns_r(ZI21,"R21")
ZI22N <- ind_ns_r(ZI22,"R22")
ZI23N <- ind_ns_r(ZI23,"R23")
ZI24N <- ind_ns_r(ZI24,"R24")
ZI25N <- ind_ns_r(ZI25,"R25")


#Unimos todos los dataframe solo con los individuos que están en ambos

ZI <- merge(x = ZI1N, y = ZI2N,by = c("ind"))
ZI <- rename(ZI, Ns1 = Ns.x, region1 = region.x, Ns2 = Ns.y, region2 = region.y)
ZI <- merge(x = ZI, y = ZI3N,by = c("ind"))
ZI <- rename(ZI, Ns3 = Ns, region3 = region)
ZI <- merge(x = ZI, y = ZI4N,by = c("ind"))
ZI <- rename(ZI, Ns4 = Ns, region4 = region)
ZI <- merge(x = ZI, y = ZI5N,by = c("ind"))
```

```
ZI <- rename(ZI, Ns5 = Ns, region5 = region)
ZI <- merge(x = ZI, y = ZI6N,by = c("ind"))
ZI <- rename(ZI, Ns6 = Ns, region6 = region)
ZI <- merge(x = ZI, y = ZI7N,by = c("ind"))
ZI <- rename(ZI, Ns7 = Ns, region7 = region)
ZI <- merge(x = ZI, y = ZI8N,by = c("ind"))
ZI <- rename(ZI, Ns8 = Ns, region8 = region)
ZI <- merge(x = ZI, y = ZI9N,by = c("ind"))
ZI <- rename(ZI, Ns9 = Ns, region9 = region)
ZI <- merge(x = ZI, y = ZI10N,by = c("ind"))
ZI <- rename(ZI, Ns10 = Ns, region10 = region)
ZI <- merge(x = ZI, y = ZI11N,by = c("ind"))
ZI <- rename(ZI, Ns11 = Ns, region11 = region)
ZI <- merge(x = ZI, y = ZI12N,by = c("ind"))
ZI <- rename(ZI, Ns12 = Ns, region12 = region)
ZI <- merge(x = ZI, y = ZI13N,by = c("ind"))
ZI <- rename(ZI, Ns13 = Ns, region13 = region)
ZI <- merge(x = ZI, y = ZI14N,by = c("ind"))
ZI <- rename(ZI, Ns14 = Ns, region14 = region)
ZI <- merge(x = ZI, y = ZI15N,by = c("ind"))
ZI <- rename(ZI, Ns15 = Ns, region15 = region)
ZI <- merge(x = ZI, y = ZI16N,by = c("ind"))
ZI <- rename(ZI, Ns16 = Ns, region16 = region)
ZI <- merge(x = ZI, y = ZI17N,by = c("ind"))
ZI <- rename(ZI, Ns17 = Ns, region17 = region)
ZI <- merge(x = ZI, y = ZI18N,by = c("ind"))
ZI <- rename(ZI, Ns18 = Ns, region18 = region)
ZI <- merge(x = ZI, y = ZI19N,by = c("ind"))
ZI <- rename(ZI, Ns19 = Ns, region19 = region)
ZI <- merge(x = ZI, y = ZI20N,by = c("ind"))
ZI <- rename(ZI, Ns20 = Ns, region20 = region)
ZI <- merge(x = ZI, y = ZI21N,by = c("ind"))
ZI <- rename(ZI, Ns21 = Ns, region21 = region)
ZI <- merge(x = ZI, y = ZI22N,by = c("ind"))
ZI <- rename(ZI, Ns22 = Ns, region22 = region)
ZI <- merge(x = ZI, y = ZI23N,by = c("ind"))
ZI <- rename(ZI, Ns23 = Ns, region23 = region)
ZI <- merge(x = ZI, y = ZI24N,by = c("ind"))
ZI <- rename(ZI, Ns24 = Ns, region24 = region)
ZI <- merge(x = ZI, y = ZI25N,by = c("ind"))
ZI <- rename(ZI, Ns25 = Ns, region25 = region)

#Calculamos la media de todos los Ns por fila y lo añadimos a una nueva columna
#med_Ns a nuestro dataframe
ZI <- ZI %>%
  dplyr::rowwise() %>%
  dplyr::mutate(med_Ns = mean(c(Ns1, Ns2, Ns3, Ns4, Ns5, Ns6, Ns7, Ns8,
                                Ns9, Ns10, Ns11, Ns12, Ns13, Ns14, Ns15, Ns16,
                                Ns17, Ns18, Ns19, Ns20, Ns21, Ns22, Ns23, Ns24,
                                Ns25), na.rm = TRUE))
```

```
#Restablecemos el directorio de trabajo
setwd("~/JOSE/UOC/R/")

#Cargamos el fichero con la información relativa a los individuos
library(readxl)
individuals <- read_excel("individuals.xlsx")

#Realizamos una leftjoin, es decir todos los elementos de ZI y solo los que
#están incluido en ZI de individuals.
ZIm <- merge(x = ZI, y = individuals, by = c("ind"),all.x = TRUE)

#Nos quedamos solo con los individuos que tienen el cromosoma 2l
ZIm <- ZIm[grep( "2L",ZIm$chrom), ]

#Nos quedamos solo con los individuos que son INV o ST
ZIm <- ZIm[grep( "INV|ST",ZIm$ln2lt ), ]

#Ordenamos por el número medio de Ns
ZIm <- ZIm[order(ZIm$med_Ns),]
View(ZIm)
```

Anexo 3: Script en Python que genera el fichero. fasta con los individuos seleccionados para cada región:

```
#Función que abre el fichero que contiene todos los individuos, antes de la selección, por cada región y escribe en el fichero .fasta solo los seleccionados
def sel(fil,fil1,ind):
    f = open(fil, 'r')
    f1 = open(fil1, 'a')
    i = 0
    for line in f:
        if i == 1:
            f1.write(line)
            i = 0
        elif ind in line:
            f1.write(line)
            i = 1
    f.close()
    f1.close()

#Se crea un vector con los nombres de los individuos seleccionados a partir del script anterior en R.
ind = ['ZI291','ZI402','ZI211','ZI233','ZI228','ZI477','ZI197N','ZI458','ZI56','ZI317',
      'ZI276','ZI486','ZI342','ZI357N','ZI505','ZI405','ZI161','ZI235','ZI269','ZI316',
      'EF19N','EF101N','EF126N','EF6N','EF95N','EF24N','EF93N','EF35N','EF31N','EF135N',
      'RG10','RG33','RG22','RG13N','RG8','RG19','RG15','RG6N','RG4N','RG21N',
      'FR320N','FR225N','FR370N','FR240N','FR16N','FR348N','FR293N','FR37N','FR213N','FR263N',
      'RAL-177','RAL-589','RAL-819','RAL-787','RAL-361','RAL-348','RAL-855','RAL-491','RAL-897','RAL-189',
      'RAL-440','RAL-850','RAL-705','RAL-354','RAL-383','RAL-129','RAL-642','RAL-439','RAL-49','RAL-375']
```

```
#Se llama a la función una vez por cada región para que escriba el fichero .fasta con los individuos seleccionados.
for i in ind:
    sel('t1.txt','R1_SEL_IND.fasta',i)
    sel('t2.txt','R2_SEL_IND.fasta',i)
    sel('t3.txt','R3_SEL_IND.fasta',i)
    sel('t4.txt','R4_SEL_IND.fasta',i)
    sel('t5.txt','R5_SEL_IND.fasta',i)
    sel('t6.txt','R6_SEL_IND.fasta',i)
    sel('t7.txt','R7_SEL_IND.fasta',i)
    sel('t8.txt','R8_SEL_IND.fasta',i)
    sel('t9.txt','R9_SEL_IND.fasta',i)
    sel('t11.txt','R11_SEL_IND.fasta',i)
    sel('t12.txt','R12_SEL_IND.fasta',i)
    sel('t13.txt','R13_SEL_IND.fasta',i)
    sel('t14.txt','R14_SEL_IND.fasta',i)
    sel('t15.txt','R15_SEL_IND.fasta',i)
    sel('t16.txt','R16_SEL_IND.fasta',i)
    sel('t17.txt','R17_SEL_IND.fasta',i)
    sel('t18.txt','R18_SEL_IND.fasta',i)
    sel('t19.txt','R19_SEL_IND.fasta',i)
    sel('t20.txt','R20_SEL_IND.fasta',i)
    sel('t21.txt','R21_SEL_IND.fasta',i)
    sel('t22.txt','R22_SEL_IND.fasta',i)
    sel('t23.txt','R23_SEL_IND.fasta',i)
    sel('t24.txt','R24_SEL_IND.fasta',i)
```

Anexo 4: Individuos seleccionados:

ind	med_Ns	gen_type	ln2lt
ZI291	101,00	haploid_embryo	ST
ZI402	101,65	haploid_embryo	ST
ZI211	102,09	haploid_embryo	ST
ZI233	102,22	haploid_embryo	ST
ZI228	102,26	haploid_embryo	ST
ZI477	102,35	haploid_embryo	ST
ZI197N	103,04	haploid_embryo	ST
ZI458	103,22	haploid_embryo	ST
ZI56	103,61	haploid_embryo	ST
ZI317	104,39	haploid_embryo	ST
ZI276	87,48	haploid_embryo	INV
ZI486	89,22	haploid_embryo	INV
ZI342	93,91	haploid_embryo	INV
ZI357N	94,61	haploid_embryo	INV
ZI505	96,48	haploid_embryo	INV
ZI405	97,26	haploid_embryo	INV
ZI161	97,57	haploid_embryo	INV
ZI235	97,70	haploid_embryo	INV
ZI269	98,35	haploid_embryo	INV
ZI316	98,91	haploid_embryo	INV
EF19N	188,26	inbred_line	ST
EF101N	197,96	inbred_line	ST
EF126N	203,09	inbred_line	ST
EF6N	217,30	inbred_line	ST
EF95N	247,17	inbred_line	ST
EF24N	325,52	inbred_line	ST
EF93N	340,09	inbred_line	ST
EF35N	359,83	inbred_line	ST
EF31N	374,91	inbred_line	ST
EF135N	432,43	inbred_line	ST
RG10	106,17	haploid_embryo	ST
RG33	107,04	haploid_embryo	ST
RG22	113,74	haploid_embryo	ST
RG13N	118,09	haploid_embryo	ST
RG8	119,04	haploid_embryo	ST
RG19	120,91	haploid_embryo	ST
RG15	121,87	haploid_embryo	ST
RG6N	121,96	haploid_embryo	ST
RG4N	122,52	haploid_embryo	ST
RG21N	124,57	haploid_embryo	ST
FR320N	111,17	inbred_line	ST
FR225N	116,30	inbred_line	ST
FR370N	120,57	inbred_line	ST
FR240N	123,48	inbred_line	ST
FR16N	127,91	inbred_line	ST
FR348N	134,70	inbred_line	ST
FR293N	150,78	inbred_line	ST
FR37N	156,78	inbred_line	ST
FR213N	159,78	inbred_line	ST
FR263N	163,26	inbred_line	ST
RAL-177	100,43	inbred_line	ST
RAL-589	105,13	inbred_line	ST
RAL-819	108,91	inbred_line	ST
RAL-787	112,48	inbred_line	ST
RAL-361	117,30	inbred_line	ST
RAL-348	117,43	inbred_line	ST
RAL-855	119,70	inbred_line	ST
RAL-491	121,74	inbred_line	ST
RAL-897	124,78	inbred_line	ST
RAL-189	134,78	inbred_line	ST
RAL-440	301,17	inbred_line	INV
RAL-850	302,96	inbred_line	INV
RAL-705	309,61	inbred_line	INV
RAL-354	311,04	inbred_line	INV
RAL-383	315,09	inbred_line	INV
RAL-129	369,65	inbred_line	INV
RAL-642	377,43	inbred_line	INV
RAL-439	453,74	inbred_line	INV
RAL-49	492,17	inbred_line	INV
RAL-375	509,30	inbred_line	INV

Nota: med_Ns: Número medio de valores faltantes, Ns.

gen_type: tipo de genoma. ln2lt : Ordenación cromosómica.

Anexo 5: Indicadores de polimorfismo y estadísticos de neutralidad para cada población.

ZI_ST

Región	N	Net sites	S	H	Pi	K	Pi/K	T-D	Sig. D	B	H	Sig.H	B
R1	10	4536	168	10	0,0112	0,038	0,2904	-0,84	0,00200		11,73	0,06800	
R2	10	3835	136	10	0,0095	0,050	0,1899	-1,28	0,00100	*	4,09	0,57600	
R3	10	4671	148	10	0,0108	0,054	0,2006	-0,27	0,20200		-12,80	0,05000	
R4	10	4565	172	10	0,0114	0,039	0,2910	-0,90	0,00100	*	13,69	0,04600	
R5	10	4424	197	10	0,0149	0,055	0,2702	-0,38	0,05800		4,09	0,62000	
R6	10	4390	175	10	0,0120	0,032	0,3717	-0,85	0,00100	*	7,82	0,34200	
R7	10	4515	232	10	0,0157	0,049	0,3220	-0,89	0,00100	*	3,02	0,75600	
R8	10	4493	121	10	0,0097	0,029	0,3414	-0,01	0,94600		5,87	0,30600	
R9	10	4657	189	10	0,0128	0,038	0,3349	-0,58	0,00400		11,11	0,15400	
R11	10	4630	217	10	0,0149	0,057	0,2607	-0,61	0,00100	*	17,87	0,02000	
R12	10	4796	128	10	0,0083	0,027	0,3070	-0,66	0,03000		5,96	0,42000	
R13	10	4685	185	9	0,0122	0,045	0,2704	-0,69	0,00100	*	-8,44	0,29600	
R14	10	4669	163	10	0,0114	0,050	0,2309	-0,48	0,02800		3,47	0,65400	
R15	10	4576	175	10	0,0117	0,060	0,1935	-0,77	0,01000		-1,33	0,78800	
R16	10	4509	228	10	0,0161	0,060	0,2688	-0,61	0,00400		3,64	0,75200	
R17	10	4634	136	10	0,0099	0,045	0,2184	-0,25	0,31200		-4,98	0,45400	
R18	10	4433	209	10	0,0148	0,045	0,3284	-0,67	0,00100	*	-4,44	0,55400	
R19	10	4722	117	10	0,0085	0,048	0,1791	-0,22	0,47200		7,02	0,29400	
R20	10	4708	103	8	0,0069	0,035	0,1960	-0,64	0,22600		5,07	0,58200	
R21	10	4704	111	10	0,0078	0,039	0,1976	-0,37	0,37200		7,64	0,33200	
R22	10	4687	135	10	0,0090	0,055	0,1617	-0,66	0,30000		2,13	0,96000	
R23	10	4808	66	10	0,0041	0,051	0,0799	-0,76	0,00800		0,71	0,97400	
R24	10	4778	61	9	0,0036	0,079	0,0453	-1,02	0,34000		0,62	0,70400	
Media			**155**		**0,0107**	**0,047**	**0,2413**	**-0,63**			**3,63**		
FULL_SEQ	10	105425	3572		0,0104	0,047	0,2217	-0,64			83,56		
IN_INV	10	50495	1927		0,0116	0,043	0,2704	-0,61			51,64		
OUT_INV	10	54930	1645		0,0091	0,050	0,1808	-0,67			31,91		

NOTAS: Líneas verdes marcan inicio y fin de inversión. Línea gris separa análisis para la secuencia completa, dentro y fuera de la inversión. Región: Región analizada. *N*: Número de individuos. Net Site: Número de nucleótidos considerados después de excluir sitios con *gaps*. S: Número de sitios segregados. *H*: Número de Haploides. *Pi*: Diversidad Nucleotídica. *K*: Divergencia nucleotídica a *D. Simulans*. Pi/K: Relación entre diversidad y divergencia nucleotídicas. T-D: Estadístico *D* de Tajima. Sig.D: Significación estadístico D de Tajima. B: Valores que mantienen su significación estadística para α = 0.025 después de aplicar la corrección de Bonferroni.

ZI_INV

Región	N	Net sites	S	H	Pi	K	Pi/K	T-D	Sig. D	B	H	Sig.H	B
R1	10	4672	69	9	0,0061	0,041	0,1472	0,70	0,01800		-5,42	0,22200	
R2	10	4791	55	10	0,0038	0,044	0,0858	-0,46	0,14600		1,07	0,78600	
R3	10	4928	51	9	0,0043	0,058	0,0750	0,91	0,00100	*	-4,00	0,26000	
R4	10	4801	113	8	0,0077	0,039	0,2005	-0,40	0,08600		-9,33	0,11200	
R5	10	4738	41	7	0,0027	0,057	0,0473	-0,58	0,06400		-15,38	0,00100	*
R6	10	4668	94	9	0,0072	0,033	0,2206	-0,01	0,91600		-2,49	0,54800	
R7	10	4643	139	9	0,0103	0,049	0,2098	-0,15	0,47800		-12,27	0,05600	
R8	10	4651	94	9	0,0077	0,031	0,2520	0,29	0,23800		-1,42	0,69400	
R9	10	4742	66	8	0,0037	0,038	0,0980	-1,25	0,00100	*	-20,36	0,00100	*
R11	10	4723	83	9	0,0061	0,057	0,1071	-0,09	0,79200		-8,71	0,06000	
R12	10	4818	53	9	0,0043	0,028	0,1532	0,50	0,14200		-8,18	0,03600	
R13	10	4779	97	7	0,0089	0,046	0,1960	1,20	0,00100	*	-4,71	0,37000	
R14	10	4642	51	9	0,0043	0,050	0,0862	0,57	0,05000		-3,64	0,33000	
R15	10	4831	87	8	0,0067	0,066	0,1010	0,27	0,39000		-12,98	0,05800	
R16	10	4653	118	5	0,0077	0,057	0,1334	-0,72	0,00600		-16,98	0,01000	
R17	10	4654	60	8	0,0053	0,047	0,1137	0,82	0,01000		-14,49	0,00100	*
R18	10	4550	133	8	0,0104	0,049	0,2114	0,01	0,99200		-15,47	0,01400	
R19	10	4623	89	8	0,0073	0,048	0,1508	0,30	0,36200		0,98	0,92400	
R20	10	4723	69	9	0,0060	0,038	0,1589	0,61	0,22800		3,20	0,69400	
R21	10	4733	86	7	0,0063	0,045	0,1406	-0,12	0,80000		-6,67	0,26600	
R22	10	4833	61	5	0,0058	0,057	0,1027	1,50	0,00200	*	1,87	0,93400	
R23	10	4929	47	6	0,0032	0,054	0,0590	-0,28	0,37800		-1,69	0,59200	
R24	10	4674	53	6	0,0036	0,077	0,0464	-0,53	0,59600		6,40	0,58000	
Media			**78,65**		**0,0061**	**0,048**	**0,1346**	**0,13**			**-6,55**		
FULL_SEQ	10	108799	1809		0,0059	0,048	0,1220	0,10			-150,67		
IN_INV	10	52133	882		0,0059	0,044	0,1333	0,07			-90,49		
OUT_INV	10	56666	927		0,0059	0,052	0,1128	0,13			-60,18		

EF_ST

Región	N	Net sites	S	H	Pi	K	Pi/K	T-D	Sig. D	B	H	Sig.H	B
R1	10	4561	130	10	0,0094	0,038	0,2462	-0,46	0,03200		14,49	0,00400	
R2	10												
R3	10	4359	118	10	0,0089	0,048	0,1865	-0,43	0,04800		-16,89	0,00800	
R4	10	4715	90	10	0,0058	0,035	0,1649	-0,81	0,00200		6,22	0,17000	
R5	10	4047	135	10	0,0116	0,048	0,2399	-0,17	0,43000		4,98	0,42800	
R6	10												
R7	10	4291	140	10	0,0116	0,049	0,2375	-0,20	0,40800		1,07	0,86200	
R8	10	4529	88	10	0,0066	0,028	0,2383	-0,20	0,38800		-4,62	0,34400	
R9	10	4514	135	10	0,0096	0,038	0,2556	-0,52	0,02400		5,42	0,39200	
R11	10	4613	98	10	0,0081	0,054	0,1491	0,32	0,23200		-2,67	0,59600	
R12	10	4585	116	9	0,0092	0,027	0,3418	0,08	0,79600		8,36	0,15400	
R13	10	4637	126	10	0,0099	0,045	0,2215	0,13	0,60400		-3,47	0,55400	
R14	10	4562	89	10	0,0073	0,049	0,1475	0,20	0,40800		-2,40	0,60000	
R15	10	4837	88	10	0,0063	0,067	0,0937	-0,11	0,76000		-6,40	0,28400	
R16	10	3862	109	8	0,0110	0,048	0,2291	0,45	0,12600		6,84	0,20000	
R17	10	4417	76	9	0,0057	0,039	0,1449	-0,31	0,31400		-4,89	0,32600	
R18	10	4243	117	10	0,0093	0,040	0,2337	-0,27	0,25400		-8,89	0,16400	
R19	10	4380	75	10	0,0058	0,041	0,1415	-0,24	0,46400		-2,49	0,56400	
R20	10	3938	48	9	0,0036	0,032	0,1133	-0,88	0,12000		-9,96	0,07000	
R21	10	4748	78	10	0,0058	0,042	0,1369	0,00	0,92800		-9,96	0,15400	
R22	10	4544	70	10	0,0051	0,054	0,0951	-0,38	0,56400		-0,98	0,73800	
R23	10	4927	40	7	0,0025	0,053	0,0470	-0,66	0,05800		-6,93	0,05200	
R24	10	4771	50	9	0,0034	0,078	0,0434	-0,43	0,65400		5,60	0,75000	
Media			**96,00**		**0,0074**	**0,045**	**0,1765**	**-0,23**			**-1,31**		
FULL_SEQ	10	94080	2016		0,0072	0,045	0,1589	-0,20			-27,56		
IN_INV	10	44852	1135		0,0085	0,042	0,2043	-0,17			-4,00		
OUT_INV	10	49228	881		0,0060	0,048	0,1228	-0,24			-23,56		

RG_ST

Región	N	Net sites	S	H	Pi	K	Pi/K	T-D	Sig. D	B	H	Sig.H	B
R1	10	4572	123	9	0,0088	0,038	0,2291	-0,47	0,04200		9,24	0,11400	
R2	10	2588	59	10	0,0070	0,045	0,1572	-0,70	0,04800		2,84	0,53000	
R3	10	4771	131	10	0,0103	0,058	0,1795	0,25	0,22000		-18,76	0,00600	
R4	10	4649	183	10	0,0123	0,039	0,3168	-0,73	0,00200		11,29	0,13400	
R5	10	4424	177	10	0,0134	0,052	0,2561	-0,51	0,01200		4,00	0,67400	
R6	10	4370	149	10	0,0109	0,033	0,3346	-0,56	0,00800		0,44	0,96000	
R7	10	4493	209	10	0,0155	0,049	0,3177	-0,57	0,00200		12,27	0,09600	
R8	10	4553	113	9	0,0083	0,030	0,2797	-0,37	0,11200		-2,93	0,59000	
R9	10	4641	123	7	0,0092	0,039	0,2328	-0,19	0,46000		-3,73	0,52200	
R11	10	4486	138	9	0,0105	0,057	0,1820	-0,23	0,33400		-0,18	0,98800	
R12	10	4746	129	10	0,0090	0,027	0,3294	-0,35	0,25400		12,18	0,05000	
R13	10	4730	155	10	0,0118	0,046	0,2580	-0,06	0,80200		-7,64	0,29600	
R14	10	4439	122	10	0,0091	0,048	0,1917	-0,37	0,15400		1,24	0,87200	
R15	10	4656	140	10	0,0092	0,063	0,1471	-0,71	0,01600		-12,00	0,13000	
R16	10	4337	176	9	0,0145	0,058	0,2491	-0,03	0,87800		4,62	0,61200	
R17	10	4585	89	9	0,0074	0,045	0,1633	0,40	0,20000		-10,58	0,04400	
R18	10	4630	164	10	0,0122	0,050	0,2420	-0,15	0,50200		-15,56	0,05000	
R19	10	4714	80	9	0,0059	0,048	0,1215	-0,11	0,75000		-4,89	0,35800	
R20	10	4696	107	10	0,0068	0,037	0,1855	-0,75	0,15400		5,42	0,59200	
R21	10	4703	88	10	0,0066	0,044	0,1507	-0,09	0,82800		-6,04	0,35400	
R22	10	4767	114	10	0,0079	0,058	0,1357	-0,39	0,55400		-1,87	0,70800	
R23	10	4859	63	9	0,0044	0,054	0,0819	-0,17	0,55200		4,71	0,23400	
R24	10	4842	33	8	0,0024	0,080	0,0297	-0,07	0,99800		-9,60	0,10800	
Media			**124,57**		**0,0093**	**0,048**	**0,2074**	**-0,30**			**-1,11**		
FULL_SEQ	10	104251	2865		0,0090	0,048	0,1893	-0,33			-25,51		
IN_INV	10	50302	1629		0,0106	0,043	0,2438	-0,37			8,18		
OUT_INV	10	53949	1236		0,0075	0,052	0,1449	-0,28			-33,69		

FR_ST

Región	N	Net sites	S	H	Pi	K	Pi/K	T-D	Sig. D	B	H	Sig.H	B
R1	10	4654	95	8	0,0078	0,042	0,1863	0,40	0,06000		2,58	0,65400	
R2	10	2138	23	8	0,0029	0,056	0,0521	-1,09	0,13200		-8,00	0,00800	
R3	10	4653	30	10	0,0020	0,056	0,0363	-0,55	0,48200		-2,93	0,25000	
R4	10	4771	66	10	0,0052	0,039	0,1333	0,31	0,00400		-11,73	0,00600	
R5	10	4388	119	10	0,0099	0,056	0,1776	0,08	0,03000		-9,87	0,09600	
R6	10	4573	70	8	0,0059	0,029	0,2030	0,29	0,05400		-6,49	0,14200	
R7	10	4487	96	9	0,0077	0,046	0,1671	0,01	0,02000		-16,98	0,01000	
R8	10	4651	66	8	0,0046	0,029	0,1584	-0,39	0,18800		-18,13	0,00100	*
R9	10	4667	90	7	0,0072	0,040	0,1793	0,21	0,50600		-5,07	0,25600	
R11	10	4629	112	10	0,0093	0,055	0,1708	0,40	0,31200		8,80	0,08000	
R12	10	4876	41	9	0,0034	0,028	0,1216	0,67	0,32400		-2,58	0,40800	
R13	10	4682	92	9	0,0074	0,044	0,1670	0,31	0,90000		-12,98	0,03400	
R14	10	4721	56	8	0,0048	0,053	0,0914	0,66	0,22600		-3,82	0,27400	
R15	10	4801	83	8	0,0072	0,064	0,1127	0,89	0,03800		-3,11	0,53800	
R16	10	4562	53	9	0,0039	0,058	0,0664	-0,30	0,95400		-2,67	0,44000	
R17	10	4757	32	7	0,0015	0,048	0,0324	-1,69	0,27400		-16,62	0,00100	*
R18	10	4389	121	10	0,0107	0,044	0,2431	0,46	0,49200		3,20	0,60400	
R19	10	4697	48	10	0,0033	0,046	0,0712	-0,48	0,73000		-10,76	0,01200	
R20	10	4380	54	9	0,0046	0,037	0,1239	0,22	0,18000		-3,91	0,43600	
R21	10	4827	61	9	0,0044	0,044	0,0993	-0,11	0,84800		-6,04	0,23400	
R22	10	4775	72	7	0,0048	0,058	0,0823	-0,53	0,59400		-11,91	0,14800	
R23	10	4840	36	7	0,0021	0,053	0,0405	-0,90	0,62400		-7,91	0,02000	
R24	10	4714	17	7	0,0011	0,077	0,0146	-0,52	0,94200		-1,24	0,45200	
Media			**66,65**		**0,0053**	**0,048**	**0,1187**	**-0,07**			**-6,44**		
FULL_SEQ	10	104632	1533		0,0052	0,047	0,1107	0,09			-148,18		
IN_INV	10	51098	838		0,0060	0,043	0,1402	0,20			-81,78		
OUT_INV	10	53534	695		0,0045	0,052	0,0867	-0,04			-66,40		

RAL_ST

Región	N	Net sites	S	H	Pi	K	Pi/K	T-D	Sig. D	B	H	Sig.H	B
R1	10	4591	128	9	0,0092	0,042	0,2175	-0,39	0,08600		9,33	0,07800	
R2	10	3093	34	10	0,0028	0,046	0,0613	-1,36	0,00200		-2,84	0,31800	
R3	10	4766	85	10	0,0063	0,055	0,1150	0,01	0,94000		-15,91	0,00200	
R4	10	4681	134	10	0,0101	0,037	0,2727	-0,10	0,67800		15,29	0,00400	
R5	10	4438	147	10	0,0115	0,055	0,2087	-0,18	0,40400		-1,87	0,74600	
R6	10	4436	137	10	0,0102	0,032	0,3169	-0,52	0,01800		0,27	0,98400	
R7	10	4408	129	10	0,0116	0,045	0,2546	0,41	0,07400		-4,09	0,44400	
R8	10	4545	100	10	0,0078	0,030	0,2573	-0,15	0,54000		2,58	0,61000	
R9	10	4642	122	7	0,0094	0,038	0,2457	-0,01	0,98200		-6,84	0,25000	
R11	10	4708	118	9	0,0093	0,057	0,1648	0,16	0,50000		-0,18	0,99800	
R12	10	4704	98	10	0,0078	0,027	0,2828	0,27	0,35200		5,42	0,34000	
R13	10	4628	110	10	0,0088	0,047	0,1867	0,26	0,31200		-12,00	0,07200	
R14	10	4728	115	10	0,0083	0,050	0,1646	-0,25	0,31200		-5,51	0,34200	
R15	10	4690	121	9	0,0088	0,066	0,1333	-0,28	0,34400		-3,29	0,63400	
R16	10	4623	118	9	0,0085	0,059	0,1435	-0,35	0,19600		-15,38	0,02600	
R17	10	4664	86	10	0,0073	0,047	0,1541	0,51	0,11200		-4,00	0,46200	
R18	10	4447	185	10	0,0148	0,047	0,3140	-0,05	0,85400		6,13	0,41800	
R19	10	4627	105	10	0,0073	0,049	0,1485	-0,42	0,00100	*	-0,80	0,87200	
R20	10	4597	109	9	0,0076	0,038	0,2031	-0,49	0,42800		13,51	0,07400	
R21	10	4800	76	8	0,0057	0,044	0,1284	0,07	0,00100	*	1,24	0,91200	
R22	10	4552	112	9	0,0082	0,057	0,1431	-0,32	0,61800		-2,84	0,65800	
R23	10	4810	52	9	0,0039	0,055	0,0703	0,03	0,91000		3,91	0,29400	
R24	10	4667	34	9	0,0025	0,077	0,0328	-0,22	0,88200		0,36	0,72400	
Media			**106,74**		**0,0082**	**0,048**	**0,1835**	**-0,15**			**-0,76**		
FULL_SEQ	10	104845	2455		0,0080	0,048	0,1678	-0,12			17,51		
IN_INV	10	50684	1295		0,0089	0,043	0,2059	-0,02			-22,84		
OUT_INV	10	54161	1160		0,0072	0,052	0,1369	-0,24			5,33		

RAL_INV

Región	N	Net sites	S	H	Pi	K	Pi/K	T-D	Sig. D	B	H	Sig.H	B
R1	10	4466	113	9	0,0094	0,043	0,2191	0,22	0,38000		6,22	0,24600	
R2	10												
R3	10	4473	88	10	0,0064	0,054	0,1188	-0,41	0,07800		-19,47	0,00100	*
R4	10	4487	136	9	0,0098	0,036	0,2735	-0,47	0,03600		16,27	0,00100	*
R5	10	4008	122	10	0,0111	0,054	0,2042	0,08	0,67400		-2,31	0,66200	
R6	10	4398	129	7	0,0107	0,033	0,3237	0,04	0,84600		-1,60	0,84400	
R7	10												
R8	10	4574	97	8	0,0078	0,030	0,2582	0,17	0,54600		-0,27	0,88600	
R9	10	4306	136	8	0,0117	0,038	0,3088	0,24	0,26600		0,36	0,98000	
R11	10	4614	117	10	0,0096	0,055	0,1740	0,28	0,23000		8,00	0,15600	
R12	10	4813	99	10	0,0073	0,028	0,2580	-0,01	0,94400		6,13	0,28200	
R13	10	4496	115	10	0,0078	0,043	0,1818	-0,71	0,00600		-24,18	0,00100	*
R14	10	4426	108	10	0,0083	0,049	0,1689	-0,25	0,34200		3,73	0,52800	
R15	10	4353	113	10	0,0100	0,065	0,1527	0,34	0,28600		5,16	0,49600	
R16	10	3926	106	9	0,0085	0,050	0,1710	-0,60	0,02600		-15,11	0,02600	
R17	10	4681	89	10	0,0057	0,048	0,1200	-0,75	0,00200		-8,80	0,10000	
R18	10	4030	119	10	0,0109	0,041	0,2683	0,18	0,47600		1,33	0,84600	
R19	10	4321	91	9	0,0069	0,046	0,1506	-0,44	0,14600		3,47	0,58000	
R20	10	4236	68	10	0,0059	0,032	0,1851	0,17	0,72600		8,36	0,11400	
R21	10	4586	91	10	0,0063	0,043	0,1444	-0,54	0,18800		2,31	0,83000	
R22	10												
R23	10	4163	46	9	0,0031	0,049	0,0642	-0,95	0,00600		-0,80	0,79000	
R24	10	4535	35	9	0,0024	0,075	0,0319	-0,72	0,44000		0,98	0,74200	
Media			**100,90**		**0,0080**	**0,046**	**0,1889**	**-0,21**			**-0,51**		
FULL_SEQ	10	87892	2018		0,0077	0,045	0,1696	-0,15			-10,22		
IN_INV	10	44595	1147		0,0086	0,042	0,2058	-0,10			-13,33		
OUT_INV	10	43297	871		0,0067	0,049	0,1372	-0,22			3,11		