



**PONTIFICIA UNIVERSIDAD CATÓLICA DEL
ECUADOR**

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

TÍTULO DE MÁSTER EN BIOLOGÍA COMPUTACIONAL

**Comparación de genomas de dos especies de café *Coffea arabica* L. y
Coffea canephora Pierre.**

Trabajo de Titulación

Autor: Blgo. Pardo Jaramillo, Sebastián Paúl Mgtr.

Directora: cPh.D Gonzalez García, Laura Natalia Mgtr.

QUITO – ECUADOR

2024

DEDICATORIA

Le dedico este trabajo a mis padres y hermana, por haberme apoyado en esta iniciativa de estudiar este programa. Nada de lo que sucede en mi vida sería posible de no ser por ustedes.

A mi abuelita Eliza y a mi hermanito Christopher que, aunque ya no estén en este mundo, en cada momento he sentido su apoyo y presencia.

Al amor de mi vida... Sin todo tu amor y paciencia no hubiera podido salir adelante en todo esto.

A mi hermano Felipe, espero que este trabajo te motive y te recuerde que siempre se puede conseguir las cosas que queremos, sé que serás el mejor biólogo marino. Gracias por ser uno de mis mayores apoyos en esta vida.

AGRADECIMIENTOS

Le agradezco inmensamente a Dios por toda fuerza que me ha dado en todos estos procesos incluso aún con mis problemas de salud.

A mi padre, por siempre invertir en mi educación, ser mi apoyo y creer que siempre puedo.

Al amor de mi vida por toda su paciencia y comprensión con todo lo que mis estudios implican, siempre estar apoyándome y nunca quejarte ¡Te amo!

A la doctora Laura Gonzalez, por su inagotable paciencia y predisposición para ayudarme a desarrollar este proyecto. Gracias por ser mi ejemplo de como ser un gran maestro. ¡Siempre le estaré eternamente agradecido!

INDICE

DEDICATORIA	2
AGRADECIMIENTOS	3
ÍNDICE DE FIGURAS	7
RESUMEN	8
INTRODUCCIÓN	9
MARCO TEÓRICO	11
1.1. El Café	11
1.2. Genómica del café.....	11
1.3. Comparación de genomas en plantas.....	13
1.4. Gen SH3.....	14
1.5. Cafeína	14
METODOLOGÍA.....	15
2.1. Obtención de datos	15
2.2. Análisis de calidad del ensamblado.....	16
2.2.1. Quast	16
2.2.2. Busco	16
2.2. Alineamiento de genomas	16
2.3. Visualización de bloques de sintenia.....	16
2.4. Comparación de genomas	17
RESULTADOS	18
3.1. Quast	18
3.2. BUSCO	19
3.3. Comparación genómica de <i>C. canephora</i> , subgenoma de <i>C. arabica</i> (sgCC) y tres subgenomas CC de <i>C. arabica</i>	20
3.4. Comparación genómica entre <i>C. eugenioides</i> y subgenomas de <i>C. arabica</i> y tres variedades.....	22

3.4. Gen SH3.....	24
3.5. Presencia/Ausencia de familia de genes.....	26
3.6. Familias de genes <i>Core</i> y <i>Accesory</i>	27
DISCUSIÓN	27
CONCLUSIONES	32
ANEXOS.....	33
6.1. Script para instalación y corrida de QUAST	33
6.2. Script para instalación y corrida de BUSCO	34
6.3. Tabla de información de la calidad del ensamblado de los genomas obtenida por QUAST	36
6.4. Dotplots de genomas y subgenomas	36
6.4.1. <i>C. canephora</i> (CC1.8_V2) vs <i>C. eugenoides</i> _BUA (CEUG_B)...	36
6.4.2. <i>C. canephora</i> (CC1.8_V2) vs <i>C. eugenoides</i> (CEUG_C).....	37
6.4.3. <i>C. canephora</i> (CC1.8_V) vs <i>C. arabica</i> var bourbon (sgCC_bourbon) 37	
6.4.4. <i>C. canephora</i> (CC1.8_V) vs <i>C. arabica</i> var caturra (sgCC_caturra)	38
6.4.5. <i>C. canephora</i> (CC1.8_V) vs <i>C. arabica</i> var geisha (sgCC_geisha)	38
6.4.6. <i>C. eugenoides</i> (<i>C. eugenoides</i> _BUA) vs sgEE <i>C. arabica</i> var. bourbon (sgEE_bourbon)	39
6.4.7. <i>C. eugenoides</i> (<i>C. eugenoides</i> _BUA) vs sgEE <i>C. arabica</i> var. caturra	39
6.4.8. <i>C. eugenoides</i> (<i>C. eugenoides</i> _BUA) vs sgEE <i>C. arabica</i> var. geisha	40
6.5. Genes pertenecientes a las familias con mayor número de repeticiones en los genomas evaluados.	40
BIBLIOGRAFÍA.....	50

ÍNDICE DE FIGURAS

Figura 1. Tamaño total de los genomas ensamblado	18
Figura 2. Porcentaje total de GC en los genomas ensamblados.....	19
Figura 3. Valores de N50 para cada genoma ensamblado.....	19
Figura 4. Resultado de la evaluación del ensamblado de genomas y subgenomas de Coffea mediante el BUSCO.	20
Figura 5. . Sintenia entre el genoma de C. canephora (CC1.8) y sus genomas Canephora de C. arabica y sus variedades: C. arabica ((CARA_EC), C. arabica var bourbon (CARA_BC), C. arabica var caturra (CARA_CC), C. arabica var geisha (CARA_GC).....	22
Figura 6 Sintenia construida entre dos genomas de C. eugenoides (CEUG_B, CEUG_C) y subgenomas sgEE de C. arabica (CARA_EE), C. arabica var. caturra (CARA_CE), C. arabica var. bourbon (CARA_BE), C. arabica var. geisha (CARA_GE).	23
Figura 7 Microsintenia de cromosoma 1 de C. canephora, C. eugenoides y C humblotiana donde se evidencia la ausencia del gen DXMT en C. humblotiana.....	24
Figura 8. Microsintenia del gen SH3	25
Figura 9 Heatmap de presencia/ausencia de familia de genes multicopia >5	26
Figura 10 . Dotplot que compara los genomas de C. canephora y C. eugenoides	36
Figura 11 Dotplot que compara los genomas de C. canephora y C. eugenoides.....	37
Figura 12. Dotplot que compara C. canephora y sgCC C. arabica var. bourbon	37
Figura 13. Dotplot que compara C. canephora y sgCC C. arabica var. bourbon	38
Figura 14. Dotplot que compara C. canephora y sgCC C. arabica var. geisha.....	38
Figura 15. Dotplot C. eugenoides_BUA y sgEE_bourbon.....	39
Figura 16. Dotplot C.eugenoides y sgEE_caturra.....	39
Figura 17. Dotplot C.eugenoides y sgEE_geisha	40

RESUMEN

El género *Coffea* es uno de los grupos botánicos más importantes en el mundo, pues de este género se derivan plantas de gran impacto económico como lo son: *Coffea canephora*, *Coffea eugeniodes*, *Coffea arabica*. Sin embargo, no todas estas especies son aprovechadas por igual o consideradas para producción de café dado que, los procesos evolutivos han generado en sus genomas ciertos cambios, confiriéndoles diferentes características como, por ejemplo: resistencia a enfermedades o niveles de cafeína. En dicho contexto, recientes estudios han generado notoriedad en una nueva especie de café, cuya carencia del gen DXMT la ha convertido en una especie de *Coffea* naturalmente descafeinado; por otra parte, el SH3 que confiere resistencia a la roya de café, presenta variaciones en su estructura lo que ha hecho a algunas especies más propensas a padecer la enfermedad. Es así que, al comparar sus genomas, se evidencia que aunque pertenecen a la misma familia existen eventos evolutivos como las inversiones, transversiones o las translocaciones las cuales, han generado nuevas variantes en sus genes o modificaciones completas.

INTRODUCCIÓN

El género *Coffea* es un grupo botánico perteneciente a la familia Rubiaceae, el cual está conformado por 125 especies que están distribuidas mayoritariamente en África, en donde se puede encontrar incluso individuos silvestres (Mishra et al., 2012). A nivel genético, todas las especies de *Coffea* son diploides ($2n = 2x = 22$) y autoestériles con contadas excepciones, como la de *C. arabica* la cual, es una especie tetraploide ($2n = 4x = 44$) y cuyo origen proviene de una hibridación natural que involucra a las especies *C. canephora* y *C. eugenioides*. Siendo *C. eugenioides* el parental femenino, de acuerdo con estudios del genoma del cloroplasto lo que ha generado interés por indagar, a partir de esta especie, cual es el verdadero origen de *C. arabica*. (Cenci et al., 2012; Lashermes et al., 1999; Pearl et al., 2004).

El café, nombre común de estas especies, es utilizado preferentemente como bebida, la cual posee mayor demanda en el mundo y que, históricamente, ha sido el pilar de una industria multimillonaria, que posee una cadena de valor que va desde el agricultor hasta el consumidor (International Coffee Organization (ICO), 2018). Su obtención, se da a partir de los frutos de las especies *C. arabica* o *C. canephora* las cuales, son cultivadas en un 70% y 30% respectivamente (Davis, 2011; Davis et al., 2019). Cabe destacar, que estas especies son conocidas como café Arábica y café Robusta (Merot-L'anthoene et al., 2019).

Sin embargo, la industria cafetalera se ha expandido a otras industrias como: médica, farmacológica, cosmética y alimenticia pues, los metabolitos secundarios que poseen estas especies han aportado a la creación de nuevos productos (Villamar et al., 2022). Entre los principales metabolitos encontramos a la cafeína, la cual es usada en el campo de salud humana dado que, desempeña un papel importante en medicamentos y tratamientos contra enfermedades; llegando a mitigar el estrés actuando como antioxidante y antiinflamatorio (Baroni et al., 2021).

Sin embargo, estudios previos, como el realizado por Perrois et al., 2015 han demostrado que existe una variación en la producción de la cafeína entre las especies de *Coffea*; ya sea por una combinación de la regulación transcripcional o por la evolución del genoma (Perrois et al., 2015). Un ejemplo de este fenómeno es la especie *Coffea hublotiana* Baill. la cual carece del gen DXMT, responsable de la síntesis de cafeína, debido a procesos evolutivos del genoma orquestados por elementos móviles. Sin embargo, este fenómeno no está investigado para otras especies de *Coffea* o no se ha realizado comparaciones genómicas con las especies más representativas de este género, a excepción de *C. canephora* (Raharimalala et al., 2021). En el mismo contexto,

el gen SH3, responsable de resistencia a la roya en café, ha sido evaluado únicamente de forma individual entre las especies de *Coffea*, más no se conoce como ha evolucionado este gen al comparar las especies ya mencionadas mediante una microsintenia (Cenci et al., 2012; Raharimalala et al., 2021).

En este contexto, surge la necesidad de realizar una comparación de genomas pues, los estudios enmarcados en la comparativa de genomas han ayudado a desvelar similitudes, diferencias y variaciones genéticas entre diferentes especies generalmente relacionadas. Además, se puede estudiar su evolución y determinar características particulares de los genomas, como el reordenamiento de secuencias y elementos transponibles (Cenci et al., 2012). Es así como, la comparativa de genomas en plantas ha aportado con información que ha servido en programas de selección y mejoramiento. Cabe destacar que, las plantas cuyos datos genómicos son similares, predicen caracteres compartidos como la producción de los mismos metabolitos o similares, tolerancia a estrés hídrico y distribución ecológica (Mawalagedera et al., 2019; Polturak et al., 2022).

Bajo estos antecedentes, el presente proyecto se plantea como objetivo general:

- Comparar genomas completos de *Coffea arabica* L. y *Coffea canephora* Pierre. publicados en NCBI

Y como objetivos específicos:

- Identificar diferencias y similitudes entre los genomas de *C. arabica* y *C. canephora*.
- Analizar inserciones, deleciones y repeticiones de los genes DXMT el cual interviene en la síntesis de la cafeína y el gen SH3 el cual confiere resistencia a la enfermedad de la roya en el café

MARCO TEÓRICO

1.1. El Café

El café es el producto con mayor comercialización a nivel mundial pues, se ha llegado a cultivar en más de 50 países, debido a su demanda al generar la bebida denominada café. La especie de *Coffea* con mayor preferencia de cultivo es *Coffea arabica* la cual, representa un aproximado del 60% de la producción mundial (Pilozo, 2021). Su origen se remonta al suroeste de Etiopía a los bosques tropicales de África, pudiendo encontrar especies silvestres nativas de esta región. Incluso, nuevas investigaciones han reportado nuevas especies diploides de *Coffea* provenientes de la India, tales como: *C. bengalensis*, *C. jenkinsii*, *C. wightiana*, *C. khasiana* y *C. travancorensis*. (Mishra et al., 2020).

En Ecuador, gracias a la ubicación geográfica privilegiada y la extensa variedad de microclimas, han aportado al desarrollo de la biodiversidad del cultivo de café. Pues, el país produce aproximadamente 700.000 sacos anuales de 60 kg (Rivas & Herrera, 2000). Entre las principales provincias cafetaleras encontramos a Manabí, Galápagos, El Oro, Loja y Zamora Chinchipe, siendo Manabí la principal productora de café robusta (*C. canephora*) y Loja de café arábico (*C. arabica*) (International Coffee Association, 2016).

La práctica cafetalera en el país es una actividad tradicionalmente familiar, lo que ha fomentado el desconocimiento de material cultivado, la mezcla de variedades, la diferente tecnificación (mal uso del agua, fertilizantes, manejo y control de plagas, podas sanitarias y de formación) y la renovación de cafetales (Alarcó López, 2011). Por lo que, subsanar estos desconocimientos es de vital importancia pues, la conservación de recursos fitogenéticos es necesaria para la seguridad alimentaria e implica que se hagan manejos tanto *in situ* o *ex situ* (Cardenas, 2007). Se debe mencionar que, el desconocimiento de la variabilidad genética del café y las características agronómicas de poblaciones es limitante para su conservación por lo que, estudios genómicos permitirá mejorar el manejo de los recursos genéticos locales y su producción (Espinosa Guzmán, 2009).

1.2. Genómica del café

En la familia Rubiaceae, la familia a la que pertenece el género *Coffea*, el número estándar de cromosomas es $x=11$, y muchas de las especies leñosas son diploides. Cabe destacar, que estas características están presentes en el género *Coffea*, siendo

todas sus especies diploides $2n=2x=22$ cromosomas, con la excepción de *C. arabica* la cual presenta una tetraploidía $2n=4x=44$ cromosomas (Clarindo & Carvalho, 2009).

La genómica del café ha nacido como un campo de estudio trascendental para comprender y mejorar una de las bebidas más consumidas en el mundo. Entre las especies más cultivadas se encuentra *Coffea canephora*, la cual, posee un genoma diploide y además cuenta con dos variedades, las cuales son: Robusta y Canilon (Maurin et al., 2007). Estas variedades, presentan rasgos diferentes pero complementarias entre sí, por ejemplo: Robusta posee crecimiento vertical, hojas largas, maduración tardía pero una gran resistencia a las plagas y enfermedades; por otro lado, Canilon presenta un crecimiento desordenado, hojas largas, rápida maduración y resistencia a la sequía, pero es mayormente propensa a las plagas y enfermedades (Oliveira et al., 2018).

El genoma de *C. canephora* fue publicado por Denoeud et al., (2014) mostrando en esta, que el 80% de 700 Mb de *C. canephora* ha sido secuenciado y publicado. Sin embargo, actualmente se cuenta con una secuencia más completa y refinada de este genoma publicado por Salojärvi et al., (2024) en donde se indica que el nuevo genoma tiene un total de 672 Mb. Este genoma, está comprendido por 25.570 genes que codifican para proteínas y una alta proporción de secuencias repetitivas, denominadas elementos transponibles; estas secuencias representan aproximadamente el 50% del genoma secuenciado y son denominadas como: Long Terminal Repeat retrotransposones (LTR-RT) (Chaparro et al., 2015). Desde su publicación este genoma se ha convertido en una herramienta de gran apoyo y referencia para comprender la composición del genoma del café, ayudando a mejoradores genéticos a la creación de nuevas variedades resistentes a plagas y adaptadas al cambio climático (Denoeud et al., 2014).

La especie *Coffea eugenioides*, es el parental de mayor importancia dada su implicación en la evolución de *C. arabica*. La publicación del genoma de esta especie ha revelado que posee un genoma de menor tamaño pero complejo. Su genoma fue publicado en el 2018, y posee un tamaño de 665 Mbp (Hamon et al., 2015). Al igual que *C. canephora*, su genoma es altamente repetitivo pues, de acuerdo con Salojärvi et al., (2024), este genoma está representado por un 59.7% de Elementos Transponibles. Se debe mencionar que, esta especie no es comercialmente producida debido a que posee una baja producción natural de granos de café pero ha sido incluida en programas de mejoramiento donde se busca reducir los niveles de cafeína y mejorar la calidad de la taza pues los granos de esta especie son pequeños con baja cantidad de cafeína en comparación con otras especies de *Coffea* (Mazzafera & Carvalho, 1991).

Un estudio desarrollado por Villamar et al. (2022), determinó, al comparar genomas de café y cacao que: la contribución genética de *C. eugenioides* y *C. canephora* en *C. arabica* no es igual. Se asume que, probablemente en el proceso de especiación se presentaron eventos evolutivos como: duplicaciones y deleciones, lo que pudo haber causado un cambio en la proporción génica de los dos ancestros de *C. arabica*. Sin embargo, dada la cantidad de proteínas compartidas con *C. eugenioides* y *C. arabica*, se sugiere que este puede desempeñar un papel importante en la selección de café y programas de mejoras, pero, se necesitan mayores estudios entre *C. arabica* y *C. canephora* para determinar la influencia de este parental en la formación del híbrido.

1.3. Comparación de genomas en plantas

La comparativa de genomas es la determinación y estudio de las similitudes y diferencias en la estructura y función de genes entre diferentes taxa. Se vale del uso de herramientas moleculares para investigar muchos eventos evolutivos que precedieron durante mucho tiempo a la identificación del ADN como molécula hereditaria. (Zewdu & Mastewal, 2022). La ley de variación de series homólogas de Vavilov fue una sugerencia temprana de las similitudes en los modelos genéticos de muchas especies de plantas (Bulatova, 2020).

Durante las últimas dos décadas, múltiples investigaciones de muchos taxa vegetales han arrojado evidencia de que en la mayoría de las plantas, la evolución de una pequeña pero vital porción del genoma, que codifica proteínas esenciales del organismo, ha avanzado relativamente lento; dando como resultado que los taxa que han estado aislados reproductivamente durante millones de años hayan conservado secuencias de ADN, así como disposiciones similares de genes a lo largo de cromosomas (Meyers et al., 2004). Además, un gran compendio de sucesos, como duplicaciones cromosómicas o segmentarias antiguas, movilidad de secuencias de ADN, deleción de genes y reordenamientos localizados, se han superpuesto al ritmo lento de evolución cromosómica y causan muchas desviaciones de la colinealidad (Paterson et al., 2000).

De modo que, la comparación de secuencias completas del genoma revela macroeventos evolutivos como poliploidía o re-arreglos cromosómicos entre diferentes linajes, permitiendo así la reconstrucción de la evolución del genoma de grandes grupos de especies (Salse et al., 2008). Sin embargo, la disponibilidad de secuencias completas del genoma está disponible únicamente para un limitado número de especies. Por otro lado, es necesario establecer que es posible realizar análisis comparativos a escalas más pequeñas en largos tramos de secuencias genómicas, lo que también nos permitirá inferir la historia evolutiva de un genoma (Cenci et al., 2010).

A pesar de la importancia mundial del café, hasta la actualidad se han desarrollado pocos esfuerzos para entender la genética y genómica del *C. arabica* y *C. canephora* (Anthony et al., 2010). Pues, sabiendo que *C. arabica* tiene un genoma tetraploide; Wolfe (2001) determinó que, luego del establecimiento de un genoma poliploide, estos tienen a evolucionar hacia una estructura diploide mediante reordenamientos cromosómicos y pérdida de genes redundantes en regiones homólogas. Este fenómeno, puede ser observado mediante análisis completos de secuencias de genoma, esto se demostró analizando poliploides antiguos como las uvas quienes conservan un número muy reducido de genes duplicados (The French–Italian Public Consortium for Grapevine Genome Characterization, 2007).

1.4. Gen SH3

El gen SH3 se cree que puede ayudar a combatir la roya del café, una enfermedad mortal causada por el hongo *Hemileia vastatrix*. Este gen se encontró en *Coffea liberica* y se transfirió a *Coffea arabica* a través de programas de mejoramiento genético para que sea más resistente a la roya. Su capacidad para provocar respuestas defensivas en la planta huésped es lo que lo hace efectivo (Lashermes et al., 2008). El SH3 produce una serie de señales de tipo defensiva cuando el patógeno invade la planta. Esta cascada de señales conduce a la producción de proteínas de defensa y a la creación de barreras físicas que impiden que el hongo se propague. La activación de la ruta de ácido salicílico, esencial para la resistencia, es parte de esta respuesta (Prakash et al., 2004).

La introducción y expresión del gen SH3 en *Coffea arabica* mejoran la resistencia a la roya y mejoran el rendimiento en una variedad de condiciones ambientales. Se ha demostrado que el locus del SH3 está estrechamente vinculado a una variedad de marcadores utilizados en procesos de selección asistida (Mahé et al., 2007). Estos marcadores acortan el tiempo y los recursos necesarios para desarrollar nuevas variedades resistentes al permitir a los mejoradores identificar y seleccionar plantas que proporcionen el gen de resistencia de manera más eficiente. Es importante destacar que, los mapeos genéticos y físicos del gen SH3 han proporcionado una descripción detallada de su organización y función (Noirot et al., 2002).

1.5. Cafeína

La cafeína es un alcaloide sintetizado por varias plantas del grupo eudicotiledóneas, estas son café, el cacao (*Theobroma cacao*) y el té (*Camellia sinensis*). En el café, este alcaloide es sintetizado en varias partes de su estructura vegetal como en las hojas, en donde actúa como insecticida, y en los frutos y semillas en donde actúa como inhibidor de la germinación de semillas de otras especies (Nathanson, 1984; Pacheco et al.,

2008). Cabe destacar, que *C. canephora* es la especie de *Coffea* que presenta mayores niveles de cafeína (Campa et al., 2005).

La ruta metabólica que sintetiza la cafeína involucra tres pasos de metilación catalizada por diferentes genes *N*-metiltransferasa (NMT) estos son: el gen XMT (xanthosine 7 N-methyltransferase), el gen MXMT (7-methylxanthine methyltransferase) y el gen DXMT (3,7-dimethylxanthine methyltransferase or caffeine synthase) (Raharimalala et al., 2021). Estos tres genes, están localizados en dos distintas regiones en *C. canephora*; en el cromosoma 1 para DXMT y en el cromosoma 9 para XMT y MXMT. En el contexto industrial, es vital conocer cómo se forma este compuesto pues, en el ser humano, participa en la formación de metabolitos los cuales, son capaces de activar los cinco receptores de sabor, además juega rol destacable en la calidad final de la taza de café (Denoëud et al., 2014).

Es así como, para limitar la cantidad de cafeína en algunas bebidas, se ha realizado diferentes enfoques investigativos, tales como: interferencia de RNA o cruza interespecíficas; sin embargo, no ha habido éxito. Por otro lado, investigaciones basadas en conocer la razón de la variación en el contenido de cafeína de algunas especies lograron determinar la existencia de variedades de *C. arabica* que casi no contenían cafeína; y se determinó, además, que esto se debía a una combinación de regulación transcripcional y a una mutación en el gen DXMT (Maluf et al., 2009).

METODOLOGÍA

2.1. Obtención de datos

Los genomas completos de *C. canephora*, *C. eugenioides* y *C. arabica*, fueron obtenidos de la base de datos Comparative Genomic (COGE) (<http://https://genomeevolution.org/cogel/>), los cuales, fueron identificados y luego descargados de acuerdo con su Genome ID: *C. canephora*: 50947; *C. eugenioides* BUA: 67315; y *C. arabica* ET39: 66663. Mientras que, el genoma de *C. humblotiana* fue obtenido de la base de datos ORCAE (<https://bioinformatics.psb.ugent.be/orcae/>) buscándolo y posteriormente descargándolo, usando el nombre científico de la especie. El genoma de *C. arabica* var. Geisha se obtuvo del repositorio de NCBI bajo el número de accesión PRJNA1032043. El genoma de *C. arabica* var Bourbon se obtuvo del repositorio de NCBI mediante el número de accesión PRJNA94414. Mientras que, el genoma de *C. arabica* var. Caturra, se obtuvo del repositorio del NCBI bajo el número de accesión PRJNA1001613.

2.2. Análisis de calidad del ensamblado

Para el análisis de calidad del ensamblado, se utilizaron los softwares Quast y BUSCO dentro de la plataforma Google Colab. Los scripts utilizados fueron los siguientes:

2.2.1. Quast

Para el análisis por QUAST, se usó el software en Google Colab, mediante el siguiente script:

```
!python quast.py /Directorio con archivos en formato fasta
```

2.2.2. Busco

El software BUSCO fue ejecutado en Google Colab, mediante el siguiente script

```
!busco -i /content/archivo en formato fasta -o nombre del archivo de salida -l eudicod_odb10 -m genome -c 2 -f
```

Para generar el plot de BUSCO, se utilizó el siguiente script

```
!python3 generate_plot.py -wd /content/drive/MyDrive/Tesis/Busco_summaries
```

2.2. Alineamiento de genomas

Los genomas fueron alineados usando la herramienta NGSEP, mediante el siguiente script de LINUX

```
java -jar NGSEPcore.jar GenomesAligner <OPTIONS> <GENOME1>  
<TRANSCRIPTOME1> <GENOME2> <TRANSCRIPTOME2>
```

2.3. Visualización de bloques de sintenia

Para la visualización de los archivos de salida del software NGSEP, se usó el software SynVisio, siguiendo los siguientes parámetros:

1. Dentro de la sección UPLOAD OWN DATA TO DASHBOARD, cargaremos los archivos de salida de NGSEP.
 - 1.1. En MCScanX Collinearity File se cargó el archivo `_SynvisioCollinearity`
 - 1.2. En GFF File se cargó el archivo `_SynvisioAnnot`
2. Una vez cargados los archivos, se cambió a la sección Synteny Dashboard donde de acuerdo con los filtros, se seleccionó los cromosomas de los distintos genomas cargados para observar los bloques de sintenia.

2.4. Comparación de genomas

Para conocer los eventos evolutivos suscitados en los genomas, las sintenias fueron realizadas agrupando los genomas basándonos en que *C. arabica* es un híbrido natural por lo tanto se analizaron sus subgenomas por separado.

Para el análisis de los genes de la cafeína, se consideró, de acuerdo con la teoría, a *C. canephora*, y *C. eugenioides* versus *C. humblotiana*.

Tabla 1. Listado de genomas y subgenomas evaluados.

Especie	Genoma	Código SynVisio
<i>Coffea canephora</i>	C.C1.8_V2	CC1.8
<i>Coffea arabica</i> subgenoma <i>Canephora</i>	sgCC_ET39	CARA_EC
<i>Coffea arabica</i> var. <i>bourbon</i>	sgCC_bourbon	CARA_BC
<i>Coffea arabica</i> var. <i>caturra</i>	sgCC_caturra	CARA_CC
<i>Coffea arabica</i> var. <i>geisha</i>	sgCC_geisha	CARA_GC
Especie	Genoma	Código SynVisio
<i>Coffea eugenioides</i>	C.eugenioides_BUA	CEUG_B
<i>Coffea eugenioides</i>	C.eugenioides_CCC68	CEUG_C
<i>Coffea arabica</i> subgenoma <i>Eugenioides</i>	sgEE_ET39	CARA_EE
<i>Coffea arabica</i> var. <i>bourbon</i>	sgEE_bourbon	CARA_BE
<i>Coffea arabica</i> var. <i>caturra</i>	sgEE_caturra	CARA_CE
<i>Coffea arabica</i> var. <i>geisha</i>	sgEE_geisha	CARA_GE
Especie	Genoma	Código SynVisio
<i>Coffea canephora</i>	C.C1.8_V2	CC1.8
<i>Coffea eugenioides</i>	C.eugenioides_BUA	CEUG_B
<i>Coffea humblotiana</i>	C.humblotiana	CHUM

RESULTADOS

3.1. Quast

Se elaboró un análisis de la calidad del ensamblado de todos los genomas y subgenomas bajo estudio usando diferentes métricas (Tabla 2). El análisis reveló que todos los ensamblados presentaron 11 cromosomas en todas las categorías; señalando que el ensamblado está constituido por un número limitado de fragmentos pero que son de gran tamaño; lo que indica que el genoma no está fragmentado pudiendo encontrar una gran cantidad de genes completos (Figura 1).

Por otro lado, el contenido de GC, entre los diferentes genomas y subgenomas, están entre 37.77% (*Coffea arabica* sgEE_ET39) y 35.79% (*Coffea humblotiana*). Estos valores están enmarcados en rangos normales para el tipo de genoma vegetal (Figura 2)

También se consideraron las métricas de calidad como el N50, estos valores se encuentran entre 61.286.361 (C_eugenioides_BUA) y 29.861.901 (C_humblotiana). Por tanto, considerando que el número de cromosomas es 11 y de acuerdo con el rango de tamaño de N50 se puede establecer nuevamente que los contigs de gran tamaño representan los cromosomas completos (Figura 3).

La tabla con la información completa de estadísticos QUASt está en el [Anexo 6.3](#)

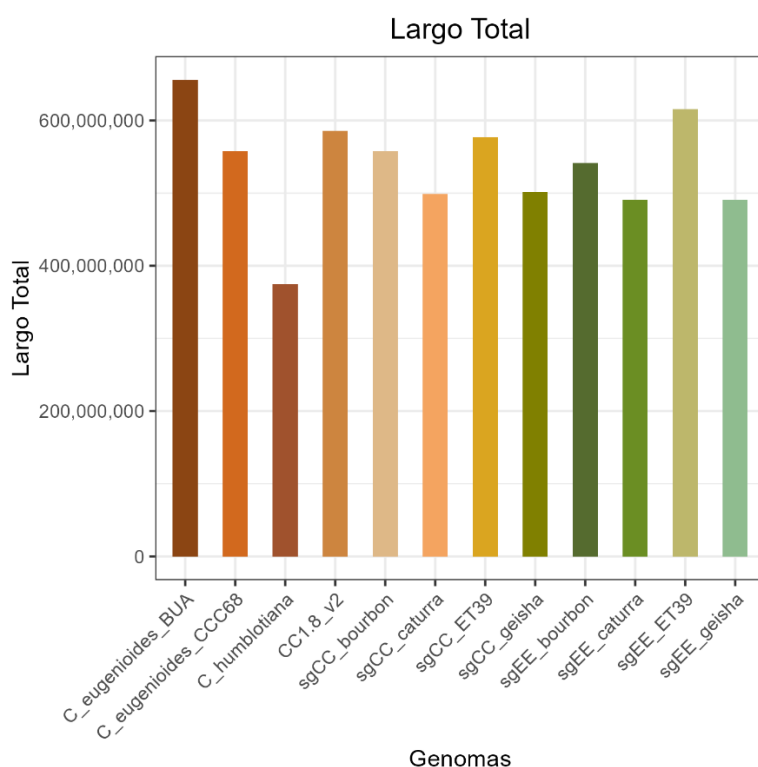


Figura 1. Tamaño total de los genomas ensamblado

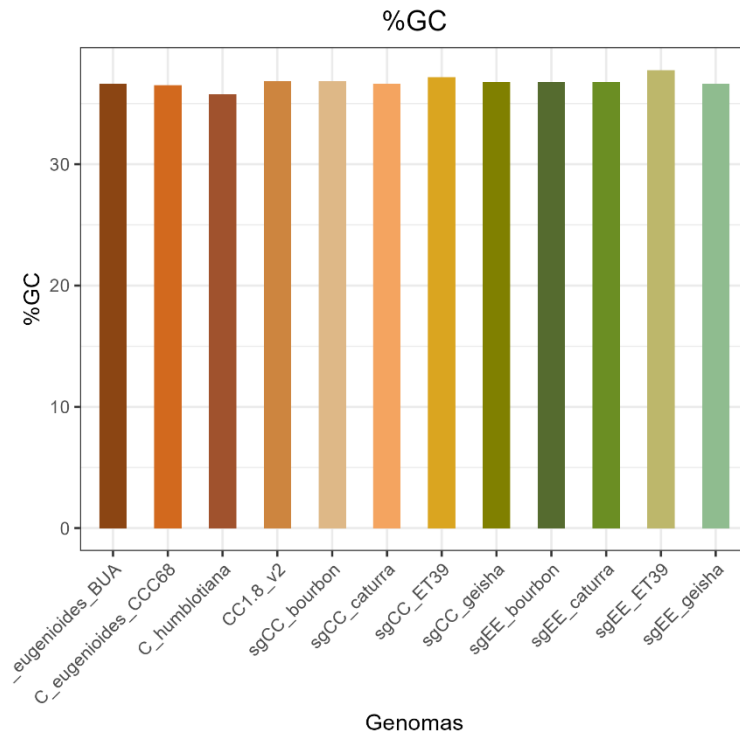


Figura 2. Porcentaje total de GC en los genomas ensamblados

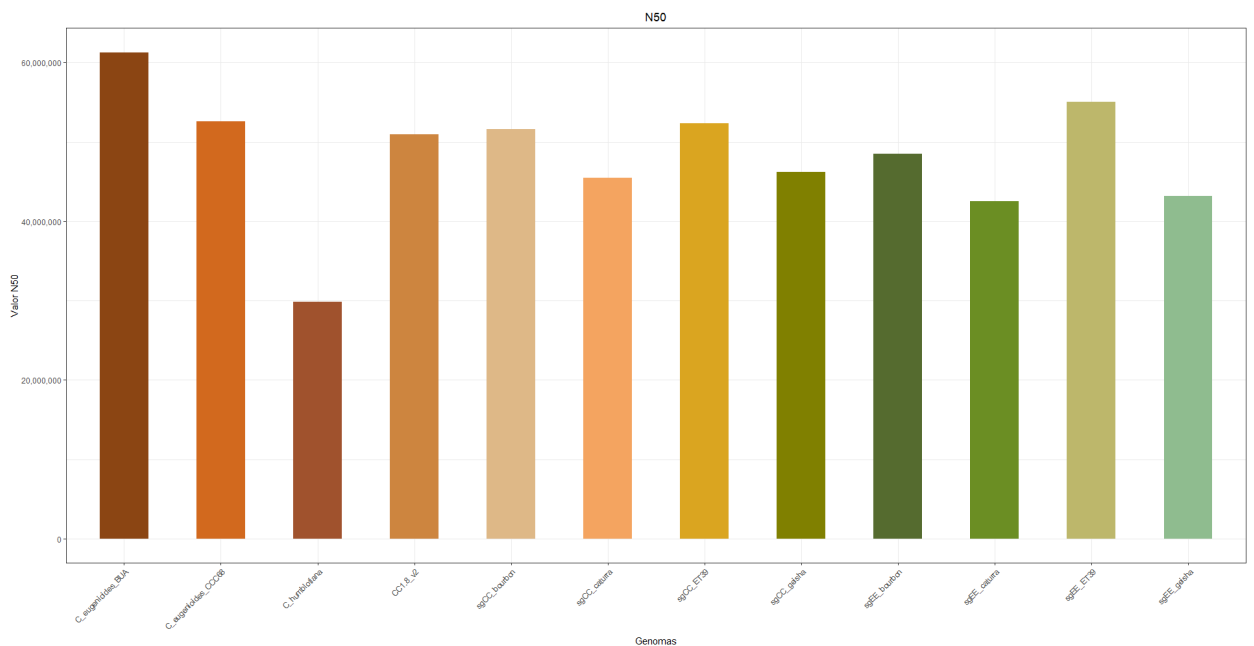


Figura 3. Valores de N50 para cada genoma ensamblado.

3.2. BUSCO

Los resultados obtenidos mediante BUSCO (Figura 4) muestran que en los genomas y subgenomas existe una mayor cantidad de genes ortólogos completos (color azul claro) en comparación con la cantidad de genes fragmentados (amarillo) o ausentes (rojos).

La gráfica (Figura 4) muestra que los sugenomas sgEE y sgCC de Bourbon son los genomas con mayor cantidad de genes ortólogos completos y de copia única denotando así, una alta calidad de ensamblado. Por el contrario, los subgenomas sgCC y sgEE de caturra, en comparación con el resto de genomas, presentan una mayor proporción de genes ortólogos duplicados ausentes.

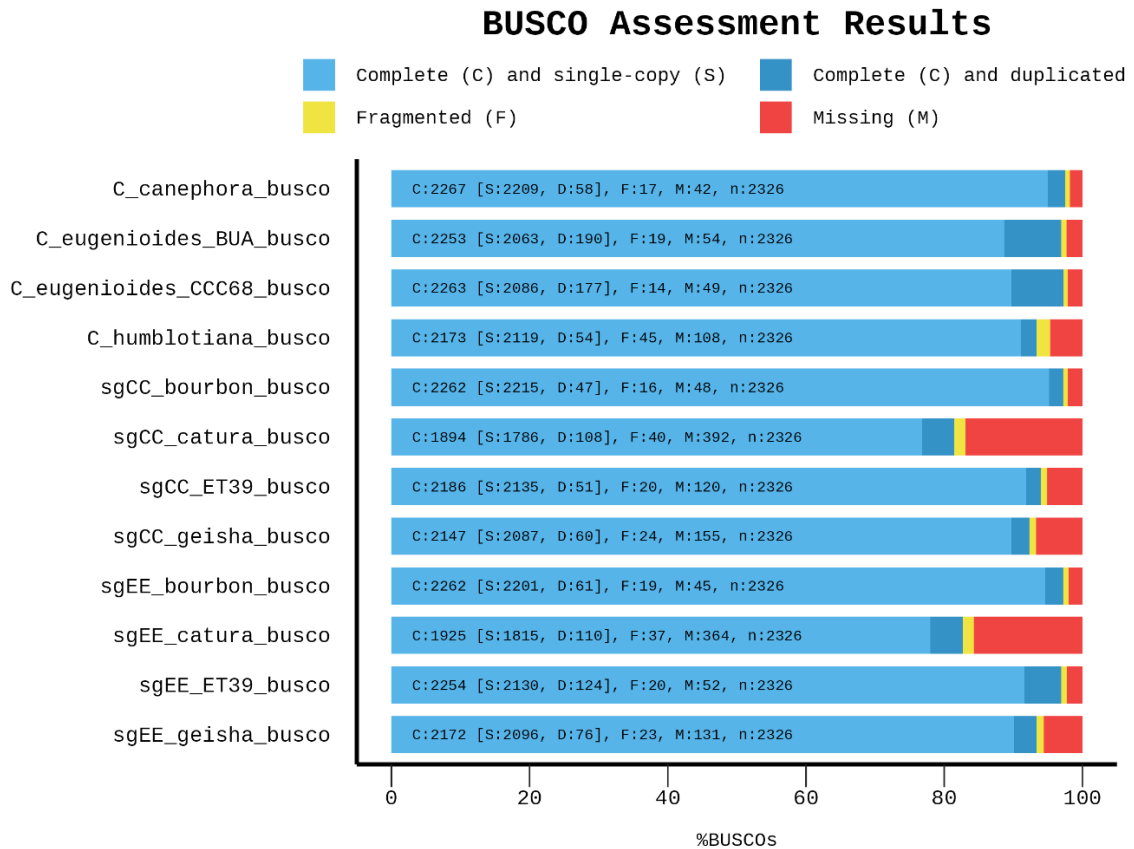


Figura 4. Resultado de la evaluación del ensamblado de genomas y subgenomas de *Coffea* mediante BUSCO.

3.3. Comparación genómica de *C. canephora*, subgenoma de *C. arabica* (sgCC) y tres subgenomas CC de *C. arabica*.

Al comparar los genomas y subgenomas y variedades de *C. arabica*, *C. canephora*, se pudo observar bloques de genes muy conservados a lo largo de todos los genomas evaluados (Figura 1), los cuales pueden ser distinguidos por las cintas azules que se conectan entre cromosomas. Estas similitudes indican que dichos genes se conservaron relativamente durante la evolución. Un ejemplo de esto se encuentra en los cromosomas 4 (Chr04) del genoma *C. canephora* (CC1.8) versus el subgenoma sgCC de *C. arabica* (CARA_EC) o el cromosoma 11 de todos los genomas, observando que escasamente se han suscitado eventos evolutivos.

Sin embargo, existen bloques de genes que si se han sometido a eventos evolutivos de reversión y traslocación. En total se observaron en esta sintenia: 2526 eventos de traslocación y 320 eventos de inversión de genes. Estos eventos son evidentes en las cintas que conectan un bloque de genes de un cromosoma con un bloque de genes de otro cromosoma, como es el caso de un bloque genes pertenecientes al cromosoma 7 del genoma de CC1.8 que están presentes en el cromosoma 5 del subgenoma CARA_EC. Por otro lado, las cintas de color rojo mostraron la existencia de fenómenos de inversión de genes entre cromosomas ortólogos o en el mismo cromosoma. Dentro de este fenómeno, es notable que ha sucedido con mayor frecuencia entre los subgenomas de las variedades de *C. arabica* (CARA_BC, CARA_CC, CARA_GC) en donde es remarcable lo sucedido el cromosoma 3 del subgenoma *Coffea arabica* var. *geisha* (CARA_GC) pues, la gran mayoría de sus genes se encuentran invertidos si tomamos como referencia su ortólogo, el cromosoma 3 del subgenoma de *C. arabica* var *caturra* (CARA_CC).

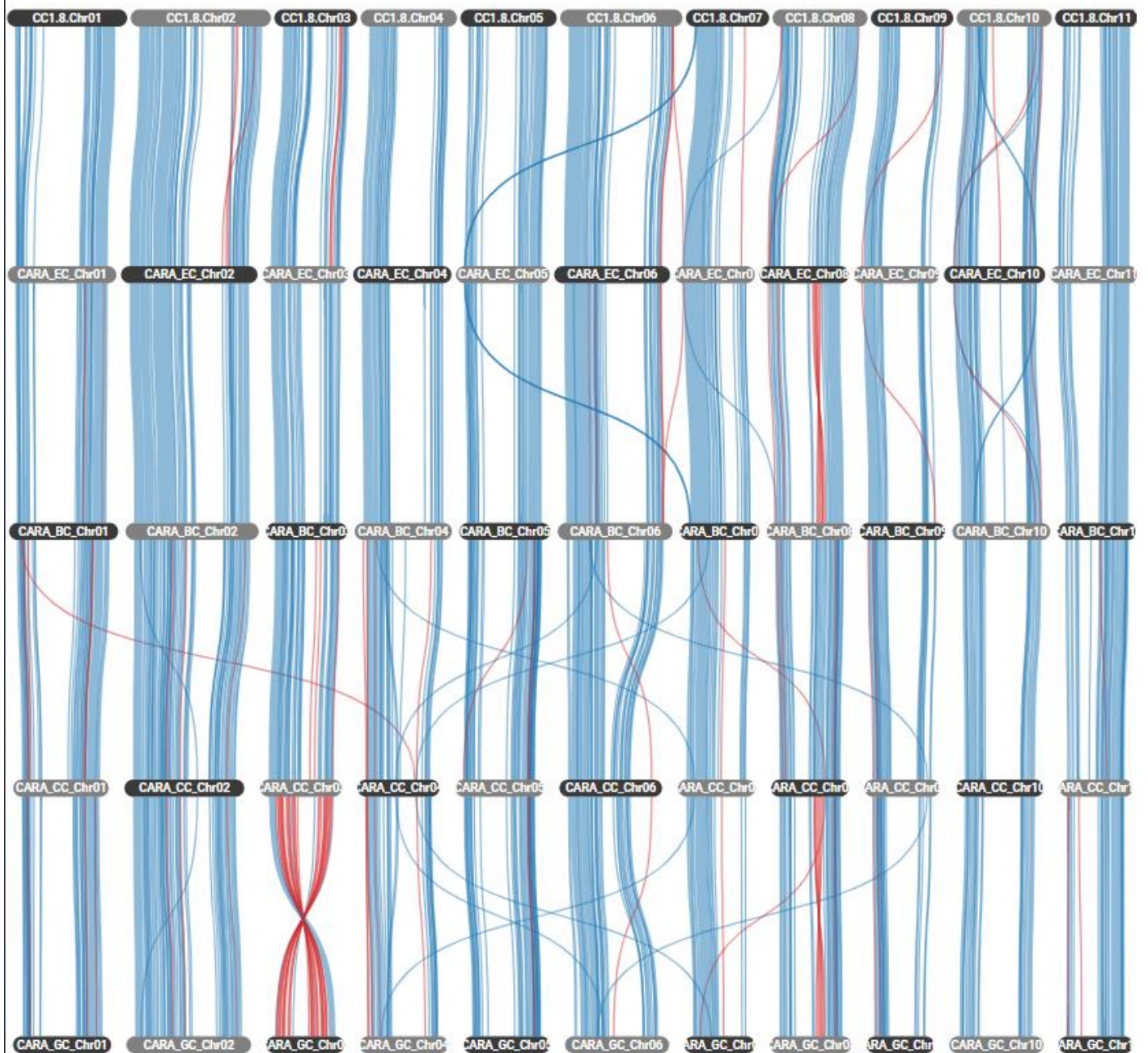


Figura 5. . Sintenia entre el genoma de *C. canephora* (CC1.8) y sus genomas *Canephora* de *C. arabica* y sus variedades: *C. arabica* ((CARA_EC), *C. arabica* var *bourbon* (CARA_BC), *C. arabica* var *caturra* (CARA_CC), *C. arabica* var *geisha* (CARA_GC)

3.4. Comparación genómica entre *C. eugenioides* y subgenomas de *C. arabica* y tres variedades.

Por otro lado, la sintenia obtenida al comparar genomas de *C. eugenioides* de dos individuos diferentes y subgenomas sgEE de *C. arabica*, nos muestra que ha existido un total de 4136 translocaciones y 998 inversiones. Se destaca, que entre los dos individuos de *C. eugenioides* (CEUG_B, CEUG_C) han existido 147 eventos de inversión de genes y 304 eventos de traslocación. Se debe señalar que los bloques

sinténicos en el cromosoma 1 de CEUG_B son muy pequeños si los comparamos con otros bloques sinténicos del mismo cromosoma 1. Por el contrario, los bloques de sintenia de los subgenomas sgEE de las variedades de *C. arabica* se muestran mayormente conservados evolutivamente presentando escasamente eventos de traslocación o inversión.

Es necesario mencionar, que en ambas sintenias es notorio que los bloques en cada cromosoma se muestran separados, distribuyéndose en ambos extremos del cromosoma y siendo los de un extremo más grandes que los del otro extremo.



Figura 6 Sintenia construida entre dos genomas de *C. eugenioides* (CEUG_B, CEUG_C) y subgenomas sgEE de *C. arabica* (CARA_EE), *C. arabica* var. *caturrea* (CARA_CE), *C. arabica* var. *bourbon* (CARA_BE), *C. arabica* var. *geisha* (CARA_GE).

3.3. Gen DXMT

El gen DXMT, se encuentra ubicado entre 1.1 y 1.3 Mb del cromosoma 1 en *Coffea*. En dicho sentido, de acuerdo con la sintenia elaborada entre *C. canephora* (CC1.8), *C.*

eugenioides (CEUG_B) y *C. humblotiana* (CHUM), se evidenció la ausencia del gen DXMT en *C. humblotiana*.

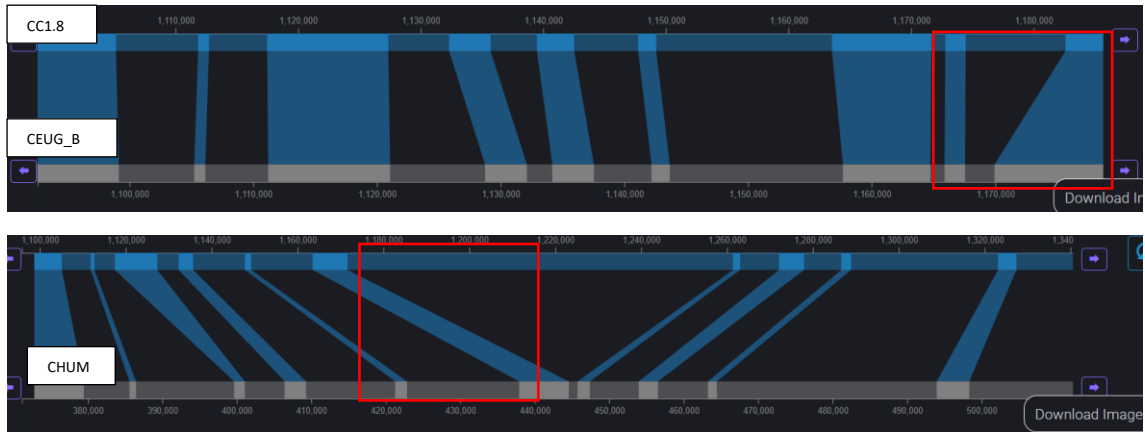
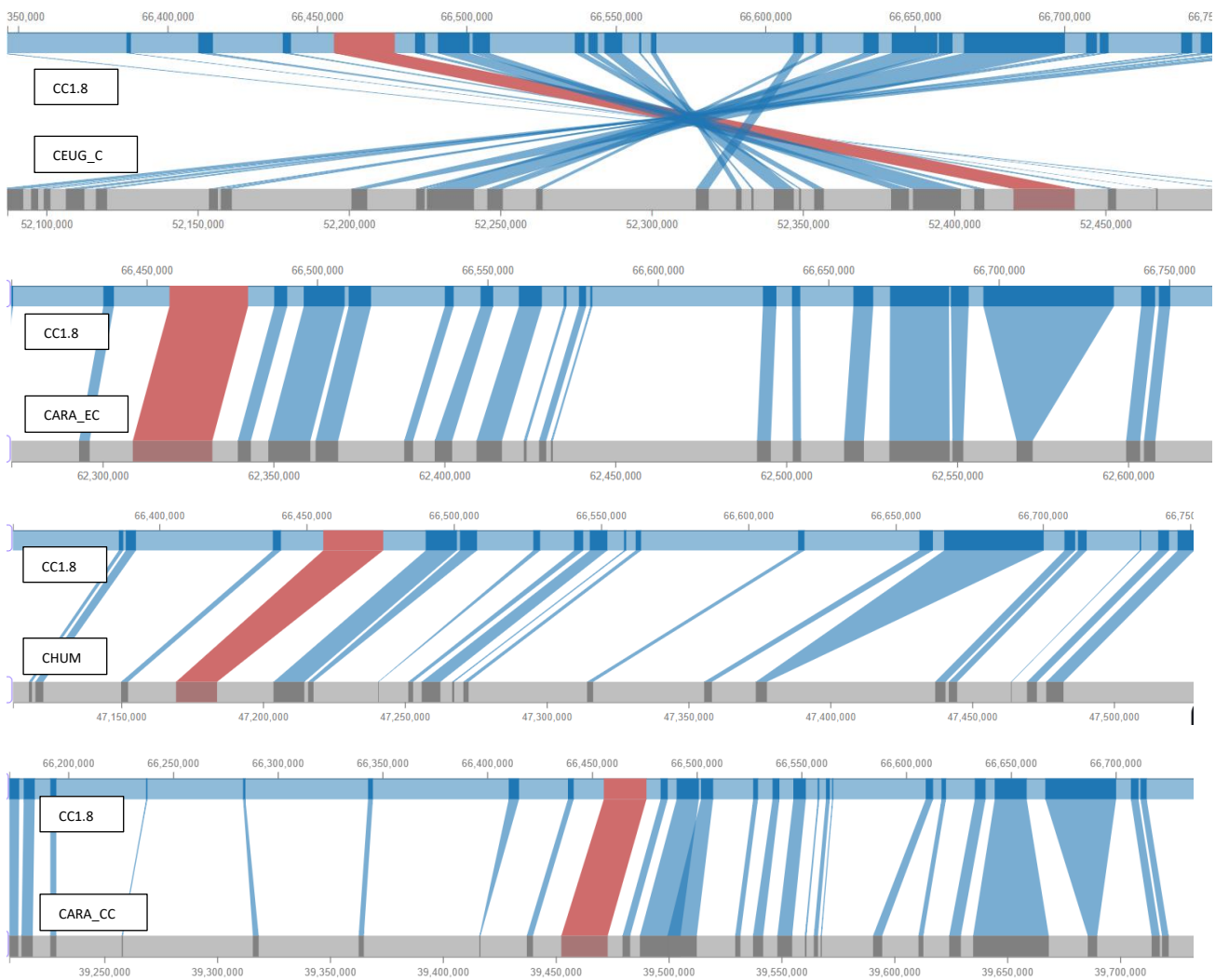


Figura 7 Microsintenia de cromosoma 1 de *C. canephora*, *C. eugenioides* y *C. humblotiana* donde se evidencia la ausencia del gen DXMT en *C. humblotiana*.

3.4. Gen SH3



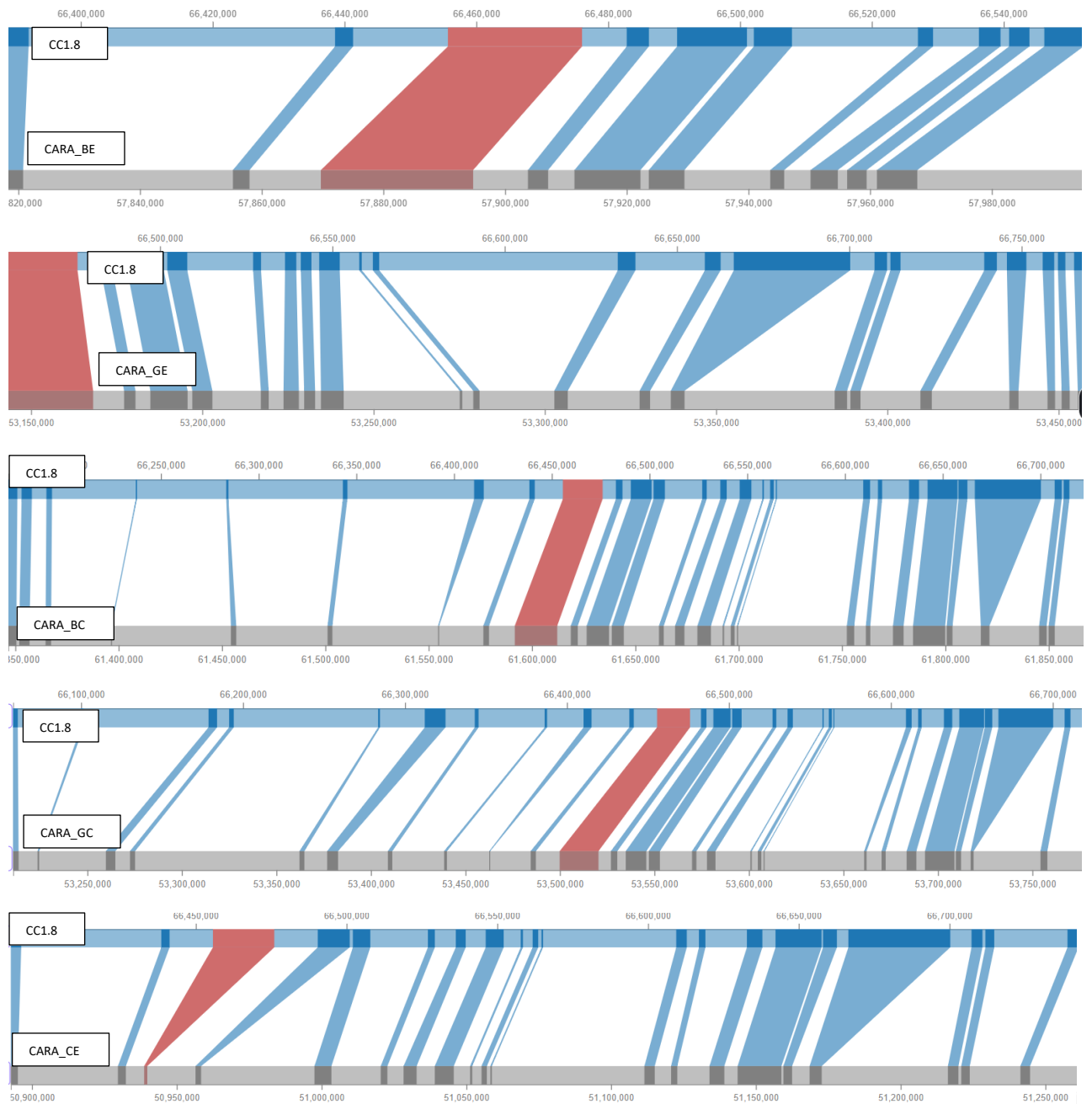


Figura 8. Microsintenia del gen SH3

De acuerdo con las micro-sintenas obtenidas (Figura 8) para analizar el gen SH3, se pudo observar que el gen no ha sido eliminado por completo en ningún genoma ni subgenoma pero si ha sufrido cambios en su tamaño y posición en el mismo cromosoma. El caso de CARA_BE se muestra como un bloque de mayor tamaño al igual que CARA_GE. Por el contrario, en el subgenoma CARA_CE evidente una reducción drástica del tamaño del gen si comparamos con el resto de los genomas evaluados. Otro evento suscitado en este gen, es la inversión completa en la región en el genoma CEUG_C.

3.5. Presencia/Ausencia de familia de genes.

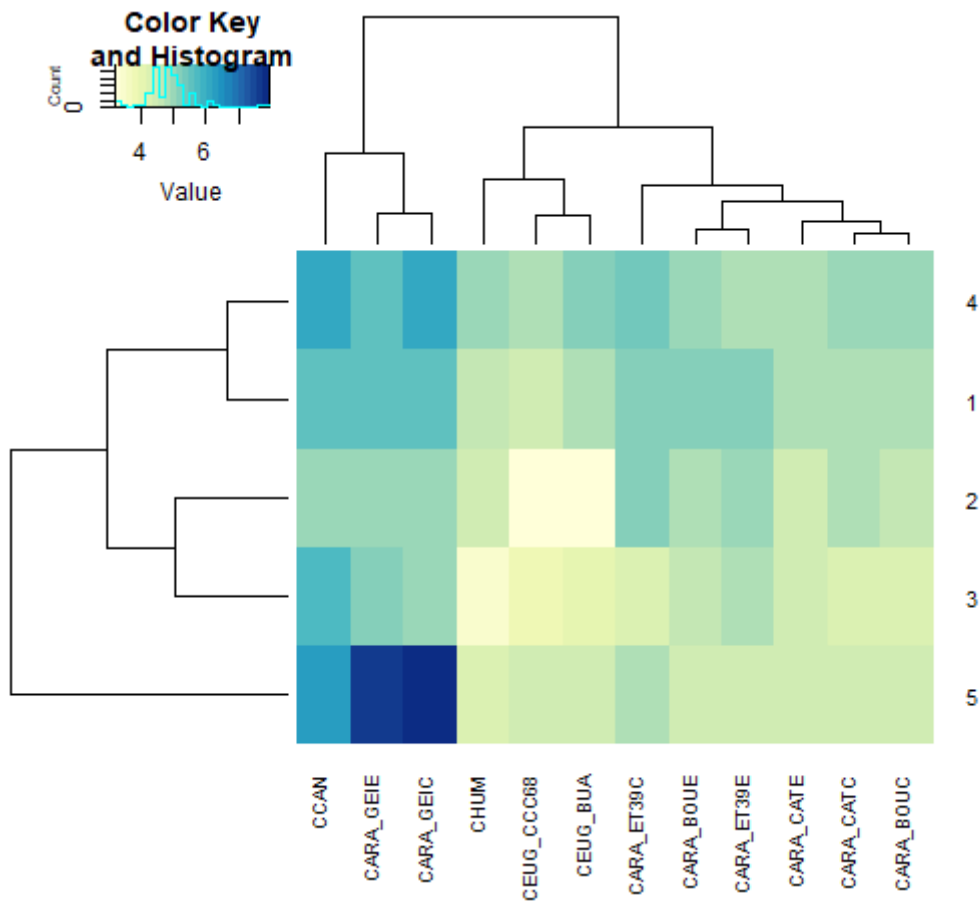


Figura 9 Heatmap de presencia/ausencia de familia de genes multicopia >20

De acuerdo con la figura 9, el dendrograma jerárquico muestra que CCAN muestra un patrón de abundancia de genes distinto lo que lo coloca en un clado aislado. Por otro lado, los genomas CARA_GEIC Y CARA_GEIE tiene un patrón de abundancia de genes similar. Cabe señalar que las familias de genes con mayor abundancia en todos los genomas son las familias 5,3,1,2,4; siendo la familia 5 mayormente abundante en CARA_GEIC y CARA_GEIE.

Los genes pertenecientes a estas familias y su respectiva descripción se encuentran en la tabla del [Anexo 6.5](#).

3.6. Familias de genes Core y Accesory

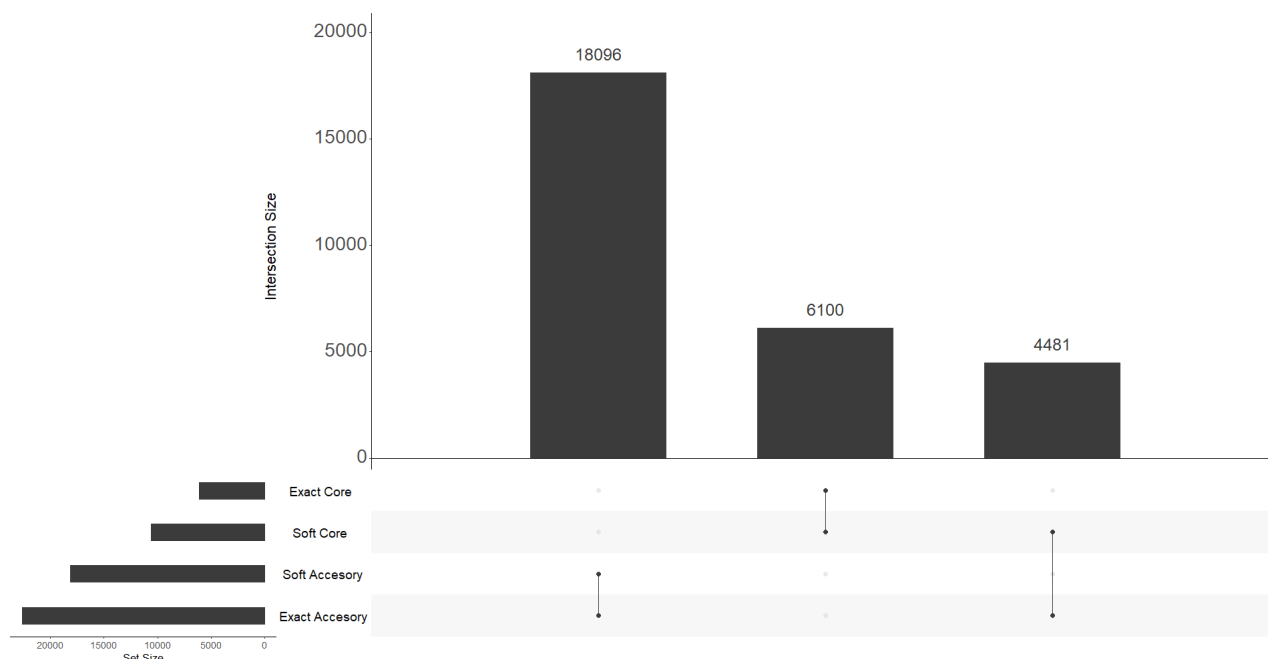


Figura 10. UpSet de las familias de genes catalogadas como Exact core, Soft Core, Soft Accesory y Exact Accesory

Al analizar las diferentes familias de genes existentes en los distintos genomas bajo estudio (Figura 10) se encontró, que 22.577 (78.75%) elementos forman parte de la categoría *Exact Accesory* es decir que, gran parte de las familias genéticas solo están presentes en algunos genomas. Por otro lado, 6.100 (21.24%) elementos pertenecen a la categoría *Exact Core* indicando, que 6.100 familias de genes están presentes en todos los genomas. En cuanto a las relaciones existentes entre las categorías *Exact Accesory* y *Soft Accesory* representan ser el grupo de familias de genes más grande con 18.906 familias lo que nos informa que, aunque estas familias de genes no están conservadas entre todos los genomas, representan adaptaciones o especializaciones en ciertos genomas. Mientras que la interacción de *Exact Core* y *Soft Core* muestra que hay familias de genes, los cuales son fundamentales, y que están en la mayoría o todos los genomas; se encuentran en una cantidad de 6.100 familias.

DISCUSIÓN

En el presente trabajo, se evidenciaron las similitudes y diferencias existentes entre cuatro especies de café y sus variedades, notándose que en los genomas han existido eventos evolutivos notables como son las deleciones, translocaciones e incluso inversiones de regiones del genoma. Esto fue posible gracias a herramientas bioinformáticas de genómica comparativa para el análisis de sintenia. En dicho contexto, se puede concordar con Gordon et al., (2016) y Cotton et al., (2016) quienes establecen,

que la parte central de la genómica comparativa está enmarcada en como comparamos los genomas, pues esto nos permitirá develar las relaciones evolutivas que existen entre las especies.

Por otro lado, el alineamiento de tan solo dos genomas puede resultar en un reto, pues a nivel de pares de bases existen millones o billones de diferencias incluso entre especies estrechamente cercanas. Por lo que, resulta computacionalmente demandante si se trata de genomas de plantas dado el gran contenido de múltiples copias idénticas de ADN, especialmente los elementos transponibles (X. Chen & Tompa, 2010; Uricaru et al., 2015). Por lo tanto, la aplicación de bloques de sintenia es una alternativa más usada y confiable considerando que son definidos como la comparación entre regiones cromosómicas de genomas evaluados que comparten un orden de genes ortólogos que provienen de un ancestro en común (Vergara & Chen, 2010).

Como paso previo a la elaboración de bloques de sintenia, se evaluó la calidad del ensamblaje de los genomas utilizados mediante métricas como: el número de contigs, el N50 y el largo total del ensamblaje. Es importante determinar la calidad del ensamblado según lo determinado por Liu et al., (2018) quienes establecieron la importancia de verificar la calidad del ensamblado previo a la elaboración de un análisis de sintenia, pues un genoma fragmentado puede guiar a descubrimientos erróneos y pérdida de información valiosa como la presencia de genes. Nuestro análisis de calidad mostró un número de contigs pequeño (11) que corresponde a los cromosomas del café, valores de N50 altos (entré 61.286.361 y 29.861.901 y valores del largo del ensamblado grandes (entre 655.786.469 y 374.304.561) lo que permite establecer que el ensamblaje realizado en los genomas evaluados es de buena calidad. Estudios realizados por Purayil et al., (2024) y Wang & Xu, (2024), obtuvieron métricas cuyos valores se asemejan a las del presente estudio, determinando que obtener una pequeña cantidad de contigs de gran tamaño, ayudan a generar una versión completa y continua de un genoma, lo cual es la base para pasos posteriores de análisis como lo es la predicción de genes y un análisis funcional de los mismos.

Bajo el mismo contexto, se realizó un análisis con el software BUSCO con la finalidad de conocer la cantidad de genes ortólogos presentes en los genomas evaluados, pues las medidas propuestas anteriormente no reportan información complementaria del ensamblado como el contenido de genes, lo cual es importante pues afecta a la interpretación de los resultados y ayuda a mejorar las estrategias de ensamblado y anotación (Mende et al., 2013). El resultado de este análisis determinó la presencia de un >90% de genes ortólogos presentes en los genomas a excepción de los subgenomas

de *C. arabica* var. *caturrea* el cual presenta 80 % de genes identificados. Nuestros resultados coinciden con los obtenidos por Salojärvi et al., (2024), quienes analizando los mismos genomas lograron obtener valores mayores a 95% de genes ortólogos encontrados con BUSCO.

Las sintenias generadas al comparar, por un lado *C. canephora* con los subgenomas de sgCC de *C. arabica*, sgCC de *C. arabica* var. *bourbon*, sgCC de *C. arabica* var. *caturrea*, sgCC *C. arabica* var *geisha* y por otro *C. eugenioides* con los subgenomas de sgEE de *C. arabica*, sgEE de *C. arabica* var. *bourbon*, sgEE de *C. arabica* var *caturrea*, sgEE de *C. arabica* var *geisha*; han mostrado la conservación de bloques sinténicos entre todos los genomas, concentrándose más en los brazos de los cromosomas. La conservación de estos bloques de genes es esencial en la especie pues tienen una importancia funcional crítica debiendo ser conservadas para poder mantener la estabilidad genética (Y.-L. Chen et al., 2013). Además, Guyot et al., (2012) en su comparación genómica de especies pertenecientes a los clados Asterids y Rosids demostró, que existen genes que se conservan a lo largo de la historia evolutiva entre individuos pertenecientes a la misma familia botánica.

Por otro lado, también se evidenció la presencia de traslocaciones e inversiones entre los cromosomas de los genomas analizados. En la sintenia que abarca *C. canephora* y sgCC de var. de *C. arabica*; estos fenómenos son más notorios entre los subgenomas de las variedades de *C. arabica*; mientras que en la sintenia que abarca *C. eugenioides* y sgEE de var. de *C. arabica* estos eventos son en extremo evidentes entre los genomas de los dos individuos de la especie *C. eugenioides*. Cenci et al., (2010). Estos fenómenos ocurren desde que un especie o antepasado diverge en dos nuevas especies, por lo tanto, son fenómenos implicados en eventos evolutivos. Es necesario señalar que dentro del genoma de especies de plantas, las estructuras más abundantes son los elementos móviles, los cuales son entidades genéticas con la potencialidad de insertarse de una región a otra (Kleckner, 1977). Su actividad se ve enmarcada en alterar genes, mediar reordenamientos genómicos que producen translocaciones, duplicaciones o eliminaciones del material genético; incluso puede llegar a influir en la expresión genética, por lo tanto, llegan a tener un impacto sustancial en la estructura y función del genoma (Feschotte & Pritham, 2007).

En cuanto a los eventos relevantes observados en las sintenias, como lo ocurrido en el cromosoma 3 de sgCC *C. arabica* var. *geisha*, en donde se evidencia una inversión completa (Anexo 6.4.5), se puede establecer este resultado como un mal ensamblado de dicho cromosoma, pues si lo comparamos con los otros genomas dentro de la misma

sintenia, los bloques sinténicos se conservan, dado que son taxas cercanas filogenéticamente. Guyot et al., (2012) recalca que las relaciones de sintenia son mayormente conservadas entre especies estrechamente relacionadas, y llegan a perderse cuando incrementa la distancia filogenética. Por otro lado, Catacchio et al., (2018), en su estudio comparativo de primates y humano, encontró que muchos de los genomas poseían inversiones erróneas, lo cual era asociado con la tecnología de secuenciación, cantidad de gaps existentes y el genoma de referencia usado para lograr ensamblar los genomas.

La microsintenia realizada para analizar el gen DXMT, develó que entre una distancia de entre 1.1 y 1.3 Mb del cromosoma 1, *C. humblotiana* carece del gen DXMT responsable de transformar la teobromina en cafeína. De acuerdo con Raharimalala et al., (2021) este gen esta flanqueado por una duplicación genética de aproximadamente 10 kb en *C. canephora*, pero en *C. humblotiana* está solo presente una sola copia. Por lo tanto, se establece que la ausencia del gen DXMT puede haberse originado por un proceso de delección en *C. humblotiana* o una inserción de tamaño considerable en *C. canephora*. De ese modo, *Coffea humblotiana* es considerada una especie naturalmente descafeinada.

En dicho contexto, la microsintenia realizada para analizar el gen SH3 en todos los genomas, nos mostró que el gen sí está presente, pero se presenta diferente en cada uno de ellos, como es el caso de *C. eugenioides* (CEUG_C) donde esa región génica presenta una inversión o como sgEE *C. arabica* var. *caturra* (Cara_CE) que en comparación con el genoma de *C. arabica*, es notorio que el gen SH3 ha sufrido un proceso de delección. Noël et al., (1999) y Parniske & Jones, (1999) establecen en sus investigaciones comparando genes de resistencia de diferentes especies, que dichos genes están sometidos a un proceso dinámico que involucra duplicaciones, delecciones, intercambio de secuencias, puntos de mutación, recombinación, conversión de genes e inserción de elementos transponibles. Es así como, al ser este gen un punto de diversidad genética puede llegar a considerarse un recurso de variación genética que permita la creación de nuevas especificidades de resistencia, a través de conversión genética o entrecruzamiento (Hulbert et al., 2001).

Para analizar la presencia/ausencia de las familias de genes presentes en los genomas evaluados se elaboró un heatmap, que nos mostró que las familias 1,2,3,4,5 son las más abundantes especialmente en los genomas de *C. canephora* y *C. arabica* var *geisha* (2 subgenomas). En este contexto la familia 5 tiene mayor presencia en los subgenomas de *C. arabica* var. *geisha*, de acuerdo con el [anexo 6.5](#) los genes que

conforman esta familia son: Ccan008g019490, Ccan008g019630, Ccan008g019540 y Ccan008g019570. Sin embargo, la anotación de estos genes los reporta como “unnamed product”; esta determinación está ligada a que, a pesar de los avances en secuenciación, mucha información genética se mantiene anotados con denominaciones no específicas. Esto se debe a que se realizan identificaciones iniciales de secuencias genéticas en ausencia de suficiente información, que permita determinar una función biológica específica (Leale et al., 2018). Por otro lado, la abundancia de esta familia se debe también a la duplicación de su genoma. Medrano et al., (2024) determinó que el contenido repetitivo de la variedad Geisha abarca un 60.6% de su genoma, resultados que son consistentes con otros genomas de *Coffea* como: *C. eugenioides* 61.5%, *C. arabica* var. caturra 62.5%, *C. arabica* var. red bourbon 59.2%.

Por otro lado, *C. canephora* también presenta una incidencia alta de genes pertenecientes a la familia 5 y 4. Los genes que conforman esta familia (4) son Ccan008g019870 y Ccan008g019710, cuya descripción menciona que son genes que codifican para “Leucine-rich repeat domain superfamily” y para “Serine/threonine-protein kinase”, active site respectivamente ([Anexo 6.5](#)). Por un lado, “Leucine-rich repeat” es uno de los mayores receptores de inmunidad reportados en plantas, conocidos también como LRR, poseen una amplia superficie de interacción con la capacidad de soportar altos niveles de variabilidad o estresores (Padmanabhan et al., 2009). Además, actúan dentro de una línea de defensa basal que se activa cuando falla la primera línea defensiva PRR-*mediated* por lo que, su alta presencia representa un rasgo evolutivo de adaptación y buena respuesta a ataques de patógenos (Jones & Dangl, 2006). Por otro lado, “Serine/threonine-protein kinase” en plantas, es considerada como una red de proteínas que actúa de manera similar a una unidad de procesador central (CPU), pues recibe todas las señales captadas por los receptores con respecto a las condiciones ambientales, fitohormonas y otros factores externos y los transforma en una respuesta adecuada como, por ejemplo: cambios en el metabolismo, expresión genética e incluso en el crecimiento y división celular (Hardie, 1999).

Adicionalmente, se analizó el pangenoma a partir de los genomas evaluados. Un pangenoma clasifica los genes de una especie en las categorías de *core* y *accessory*, por un lado, los *core* genes (*exact core*) son los genes que están presentes en todos los genomas, mientras que los genes *accessory* (*exact accessory*) son genes que están presentes solo en ciertos genomas (Morgante et al., 2007). En dicho contexto, nuestro análisis arrojó que un 70.75% de las familias de genes son consideradas *exact accessory* mientras que un 24.25% son consideradas *exact core*. Estos valores discrepan del pangenoma publicado por Salojärvi et al., (2024) pues, en esta publicación únicamente

se usa un genoma para establecer dicha afirmación. La presente investigación abarca el análisis de cuatro especies de *Coffea* y tres variedades de *C. arabica* por lo tanto, se realiza un análisis de un mayor número de familias genéticas. De ese modo, se ha tenido concordancia con lo manifestado por Tettelin et al., (2005) quienes establecen el concepto de pangenoma al comparar 8 cepas de *Streptococcus agalactiae*.

La gran cantidad de exact accessory, está ligada a la existencia de los elementos transponibles (TE) del tipo LTR-retrotransposones. Los resultados obtenidos por Salojärvi et al., (2024) en cuanto a la cantidad de TEs existentes en los genomas muestran que, el contenido de estos elementos en *C. canephora* y *C. eugenioides* es de 67.5% y 59.7% respectivamente y en *C. arabica* sgCC 63.1% y en sgEE 63.8%. Morgante et al., (2007), menciona que el movimiento de los TEs es por mucho responsable de la variación, tanto en las regiones intergénicas como en el contenido del gen. Además, sus resultados obtenidos, se asemejan a los obtenidos en este ensayo pues, al comparar 8 especies de maíz se evidenció que, aunque el core genoma no se redujo drásticamente, la cantidad de exact accessories incrementaba dibujando nuevas líneas en las sintenias realizadas, demostrando inserciones de secuencias en otras regiones intergénicas.

CONCLUSIONES

- La comparación de genomas es una herramienta informativa, cuando se cuenta con varios genomas completos, y correctamente ensamblados de especies similares pues, nos permite debelar información evolutiva como: inversiones o traslocaciones, que han suscitado en las especies a lo largo de su historia evolutiva.
- Contar con una buena calidad de ensamblado de los datos, es crucial para análisis posteriores pues, de esto depende que no se asuman por correctos eventos de inversión que son producto de un mal ensamblado genómico.
- Se debe evaluar la calidad del ensamblaje de los genomas usando más de una herramienta ya que, dependiendo del algoritmo del software, los estadísticos del ensamblado pueden variar. Hay que considerar también que, el usar herramientas gráficas para conocer la existencia de genes funcionales es una opción correcta
- Aunque la literatura disponible mencione que la *Coffea* ha vivido eventos de cuellos de botella en su diversidad genética, no se puede dejar de lado el papel crucial que juegan los elementos móviles para generar diversidad genética en

esta especie, permitiendo en algunos casos su supervivencia en otros ecosistemas, así como también volviéndolo más vulnerable.

ANEXOS

6.1. Script para instalación y corrida de QUAST

```
## Quast
```

1. Clonamos en este repositorio a QUAST que se encuentra en GitHub

```
!!!!
```

```
!git clone https://github.com/ablab/quast #El comando git clone sirve para clonar el repositorio, sintáxis: !git clone [url del sitio donde se desea clonar]
```

```
!pip install matplotlib numpy scipy pysam reportlab #Instalamos dependencias que sirven a quast para realizar su análisis
```

2. Ingresamos a la carpeta quast

```
pwd
```

```
# Commented out IPython magic to ensure Python compatibility.
```

```
# %cd quast
```

```
ls
```

3. Vinculamos a nuestra cuenta drive

```
from google.colab import drive
drive.mount('/content/drive')
```

```
!python quast.py
```

```
ls
```

```
!python quast.py /content/drive/MyDrive/Tesis/Datos/CC1.8_v2.fasta\
/content/drive/MyDrive/Tesis/Datos/C_eugenioides_BUA.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/C_eugenioides_CCC68.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/C_humblotiana.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/sgCC_ET39.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/sgCC_bourbon.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/sgCC_catura.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/sgCC_geisha.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/sgEE_ET39.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/sgEE_bourbon.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/sgEE_catura.renamed.fasta\
/content/drive/MyDrive/Tesis/Datos/sgEE_geisha.renamed.fasta
```

6.2. Script para instalación y corrida de BUSCO

```
##### BUSCO
```

```
1. Clonamos el repositorio desde ezlab
```

```
#####
```

```
pwd
```

```
# Commented out IPython magic to ensure Python compatibility.  
# %cd /content
```

```
# Commented out IPython magic to ensure Python compatibility.  
!git clone https://gitlab.com/ezlab/busco.git  
# %cd busco/
```

```
pwd
```

```
#####Instalamos BUSCO#####
```

```
!python -m pip install .
```

```
#####Instalamos las dependencias de BUSCO#####
```

```
# Instalar dependencias  
!apt-get update  
!apt-get install -y git cmake build-essential python3-dev zlib1g-dev libbz2-dev liblzma-  
dev libcurl4-gnutls-dev wget
```

```
# Descargar e instalar Miniprot Index desde el repositorio de GitHub  
!git clone https://github.com/valv/Miniprot_index.git  
!cd Miniprot_index && make
```

```
# Install BBTools  
!wget https://sourceforge.net/projects/bbmap/files/latest/download -O bbmap.tar.gz  
!tar -xvzf bbmap.tar.gz  
!mv bbmap /usr/local/bbmap  
!ln -s /usr/local/bbmap/*.sh /usr/local/bin/  
!ln -s /usr/local/bbmap/*.pl /usr/local/bin/
```

```
# Verify BBTools installation  
!bbmap.sh --version
```

```
# Install HMMER  
!wget http://eddylib.org/ware/hmmer/hmmer-3.3.2.tar.gz  
!tar -xvzf hmmer-3.3.2.tar.gz  
!cd hmmer-3.3.2 && ./configure && make && make install
```

```
# Install AUGUSTUS  
!apt-get install -y augustus
```

```
# Install BUSCO  
!pip install busco
```

```
# Install additional python packages
```

```

!pip install biopython

# Commented out IPython magic to ensure Python compatibility.
# Clone the Miniprot repository from GitHub
!git clone https://github.com/lh3/miniprot.git

# Navigate to the Miniprot directory
# %cd miniprot

# Compile Miniprot
!make

# Move the Miniprot executable to a directory in the PATH
!mv miniprot /usr/local/bin/

# Verify the installation by checking the Miniprot version
!miniprot --version

pwd

# Commented out IPython magic to ensure Python compatibility.
# %cd ../

""""Corremos BUSCO""""

!busco -i /content/Coffea_arabica.faa -o BUSCO_C.arabica -l eudicots_odb10 -m
genome -c 2 -f

pwd

""""GRAFICO DE RESULTADOS BUSCO

##### CONOCEMOS LA FUNCIÓN GENERATE_PLOT
""""

!python3 scripts/generate_plot.py -h

# Commented out IPython magic to ensure Python compatibility.
# %cd scripts

ls

!python3 generate_plot.py -wd /content/drive/MyDrive/Tesis/Busco_summaries2

```

6.3. Tabla de información de la calidad del ensamblado de los genomas obtenida por QUAST

Report

	CC1.8_v2	C_eugenioides_BUA.renamed	C_eugenioides_CCC08.renamed	C_humboldtiana.renamed	sgCC_ET39.renamed	sgCC_bourbon.renamed	sgCC_catura.renamed	sgCC_geisha.renamed	sgEE_ET39.renamed	sgEE_bourbon.renamed	sgEE_catura.renamed	sgEE_geisha.renamed
# contigs (>= 0 bp)	11	11	11	11	11	11	11	11	11	11	11	11
# contigs (>= 1000 bp)	11	11	11	11	11	11	11	11	11	11	11	11
# contigs (>= 5000 bp)	11	11	11	11	11	11	11	11	11	11	11	11
# contigs (>= 10000 bp)	11	11	11	11	11	11	11	11	11	11	11	11
# contigs (>= 25000 bp)	11	11	11	11	11	11	11	11	11	11	11	11
# contigs (>= 50000 bp)	11	11	11	11	11	11	11	11	11	11	11	11
Total length (>= 0 bp)	385604637	655786469	558110372	374304561	577113884	557634504	498797991	501416724	615481186	541154740	491123899	490229983
Total length (>= 1000 bp)	385604637	655786469	558110372	374304561	577113884	557634504	498797991	501416724	615481186	541154740	491123899	490229983
Total length (>= 5000 bp)	385604637	655786469	558110372	374304561	577113884	557634504	498797991	501416724	615481186	541154740	491123899	490229983
Total length (>= 10000 bp)	385604637	655786469	558110372	374304561	577113884	557634504	498797991	501416724	615481186	541154740	491123899	490229983
Total length (>= 25000 bp)	385604637	655786469	558110372	374304561	577113884	557634504	498797991	501416724	615481186	541154740	491123899	490229983
Total length (>= 50000 bp)	385604637	655786469	558110372	374304561	577113884	557634504	498797991	501416724	615481186	541154740	491123899	490229983
# contigs	11	11	11	11	11	11	11	11	11	11	11	11
Largest contig	79363681	83462375	79780123	57522413	76906396	74718367	66155350	69315366	77261657	72705077	71633312	68027359
Total length	385604637	655786469	558110372	374304561	577113884	557634504	498797991	501416724	615481186	541154740	491123899	490229983
GC (%)	36.87	36.64	36.52	35.79	37.18	36.85	36.78	36.65	37.77	36.75	36.75	36.64
NSI	50946795	61296361	52590270	29861901	53223663	51611264	45429025	46182948	55030746	48524476	42507420	43221077
N90	42304562	46417428	40525418	28010604	44254144	39995608	38064651	34437611	48754064	39730949	35839895	37227601
auM	55904987.2	62118921.8	52972072.7	36594671.9	54425886.3	52876495.5	46956584.6	47917692.6	57719549.7	51015232.1	46835762.7	46512790.7
LS0	5	5	5	5	5	5	5	5	5	5	5	5
LS9	10	10	10	10	10	10	10	10	10	10	10	10
# N's per 100 kbp	10.62	85.09	2209.63	11.06	2.85	5.65	227.22	20.48	5.93	8.32	164.54	19.08

6.4. Dotplots de genomas y subgenomas

6.4.1. C. canephora (CC1.8_V2) vs C. eugenioides_BUA (CEUG_B)

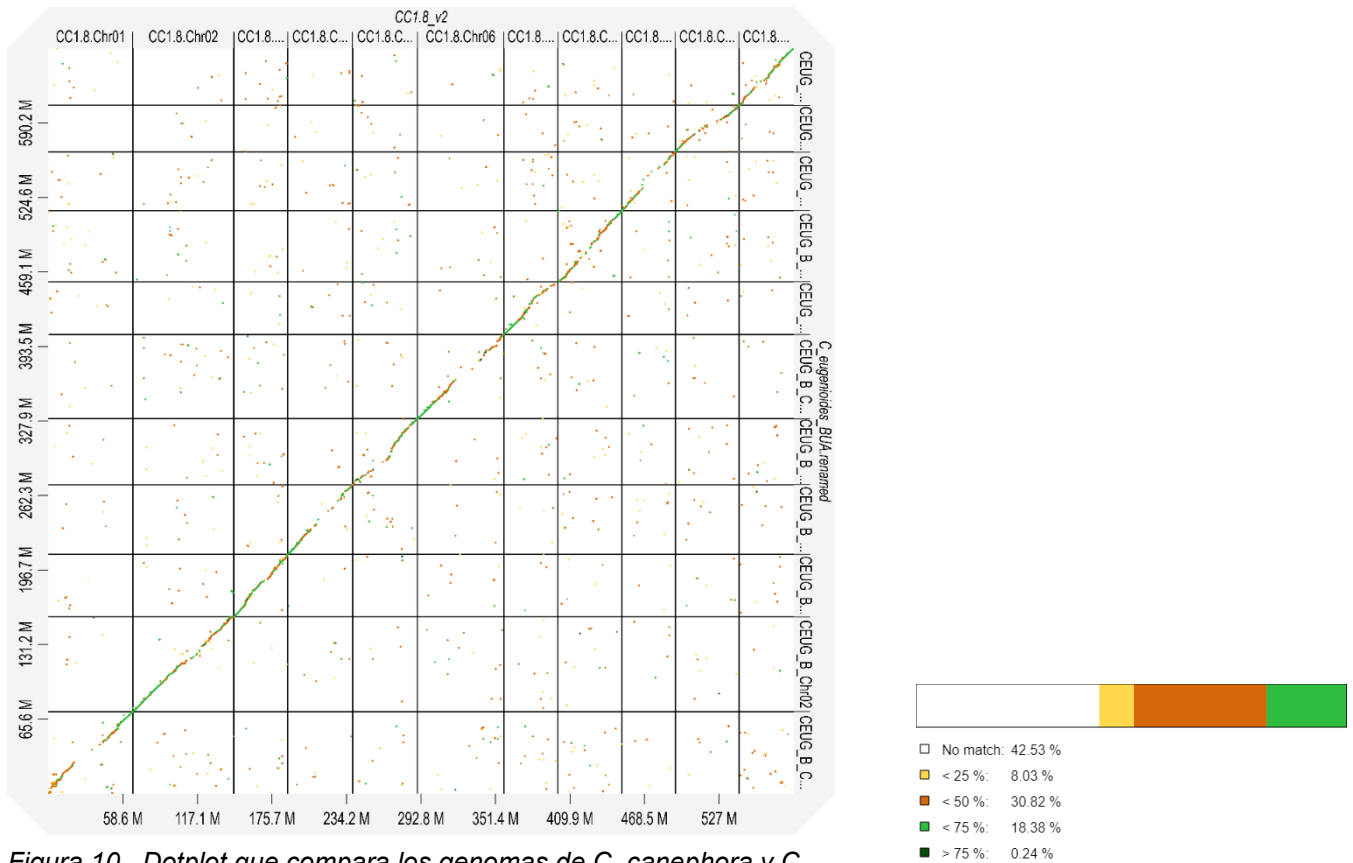


Figura 10. Dotplot que compara los genomas de *C. canephora* y *C. eugenioides*

6.4.2. *C. canephora* (CC1.8_V2) vs *C. eugenioides* (CEUG_C)

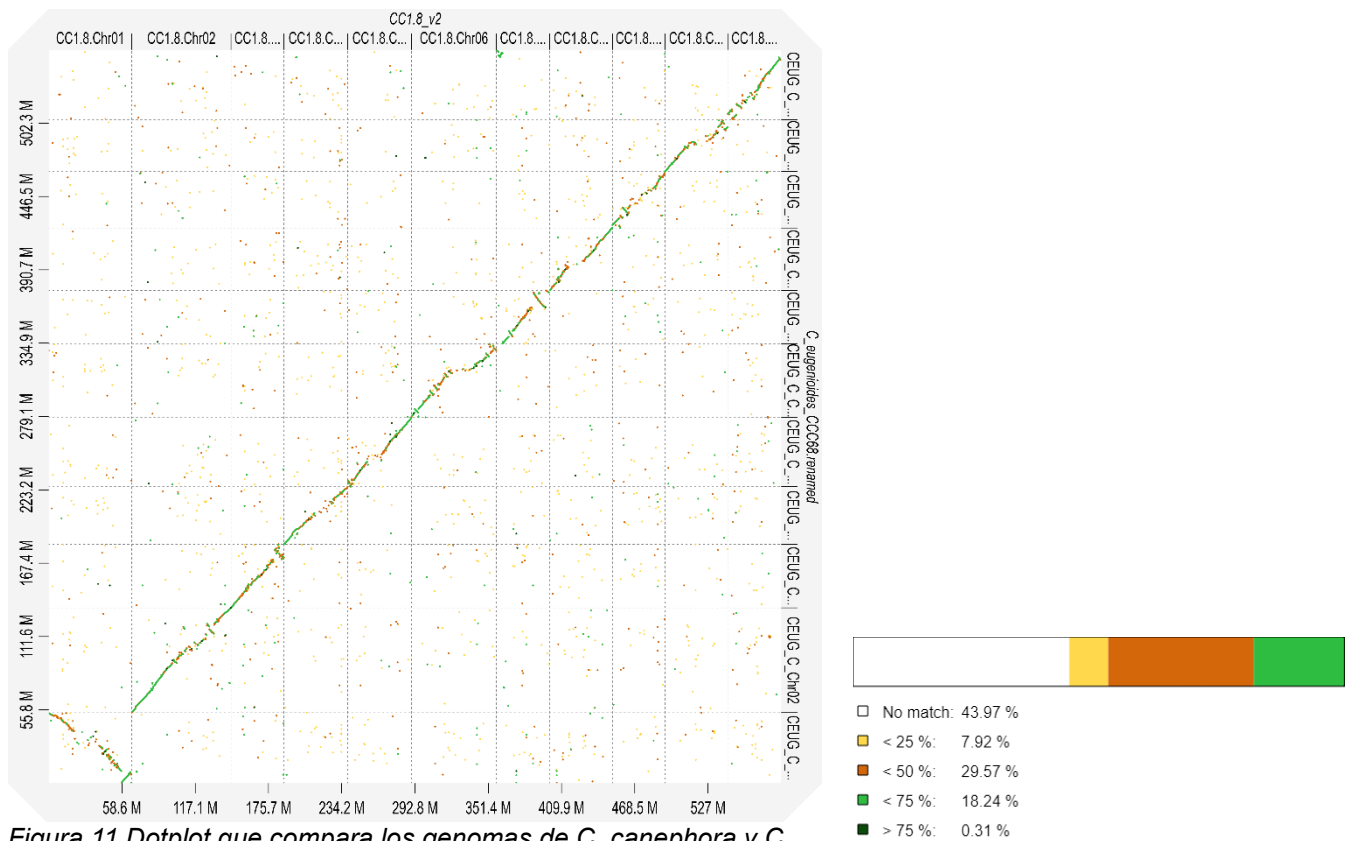


Figura 11 Dotplot que compara los genomas de *C. canephora* y *C. eugenioides*

6.4.3. *C. canephora* (CC1.8_V) vs *C. arabica* var bourbon (sgCC_bourbon)

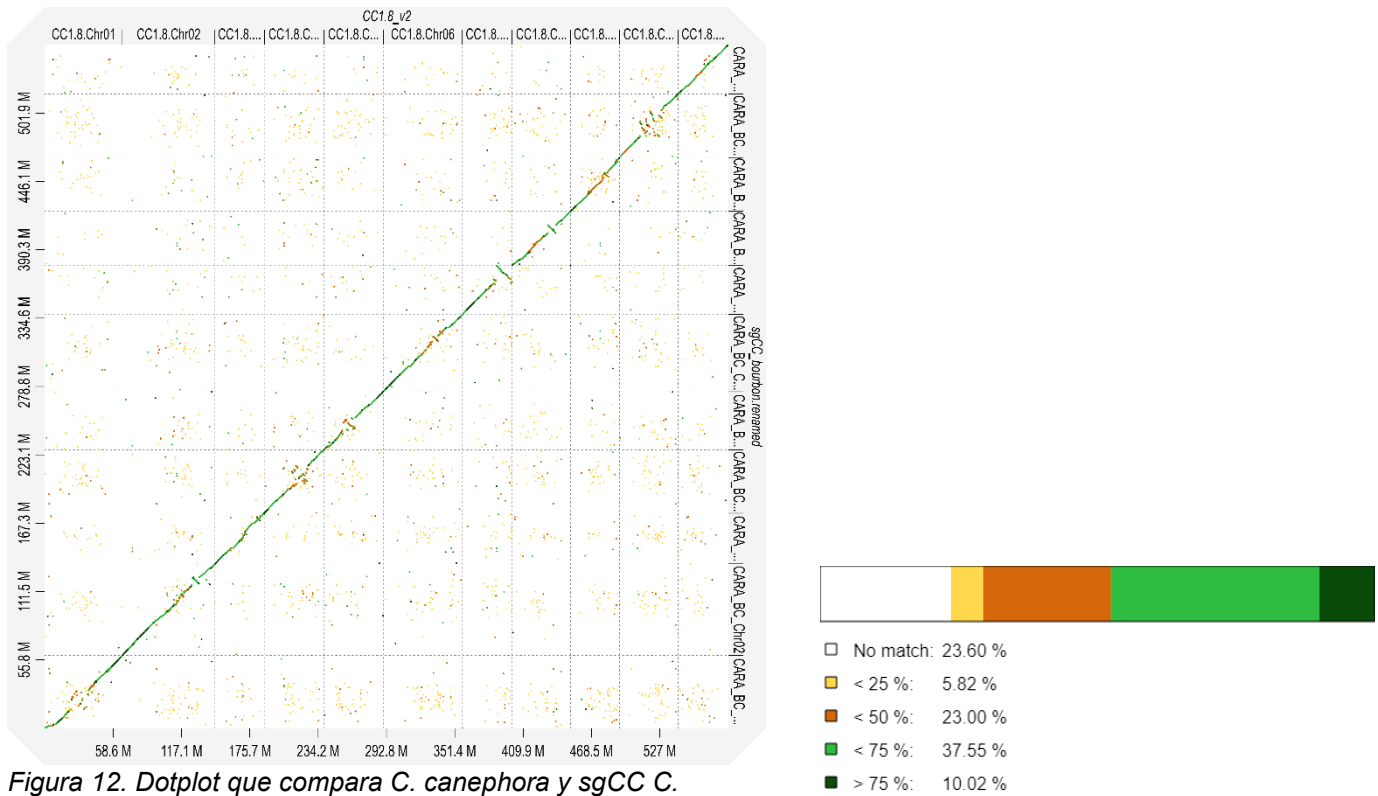


Figura 12. Dotplot que compara *C. canephora* y sgCC *C. arabica* var. bourbon

6.4.6. *C. eugenioides* (*C. eugenioides_BUA*) vs sgEE *C. arabica* var. bourbon (sgEE_bourbon)

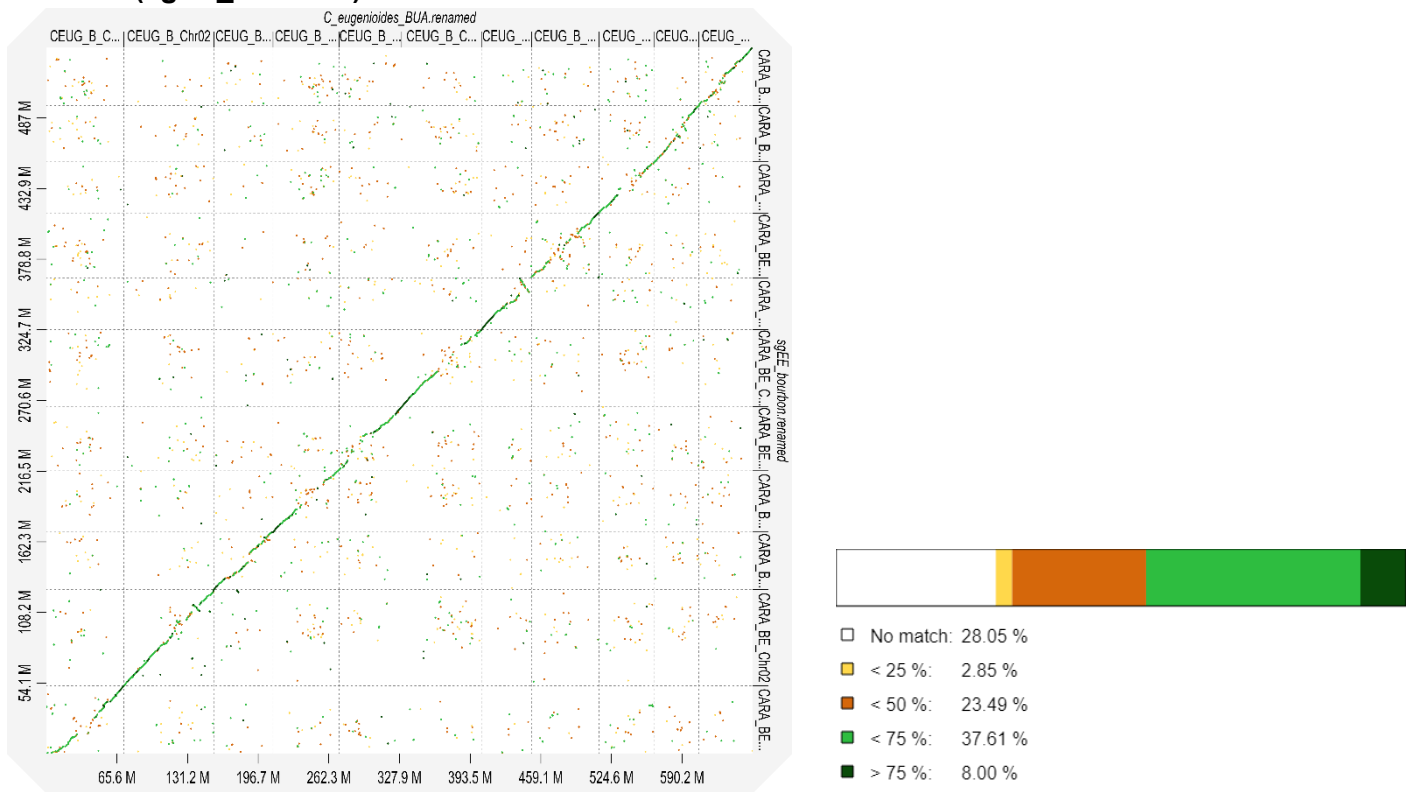


Figura 15. Dotplot *C. eugenioides_BUA* y *sgEE_bourbon*

6.4.7. *C. eugenioides* (*C. eugenioides_BUA*) vs sgEE *C. arabica* var. caturra

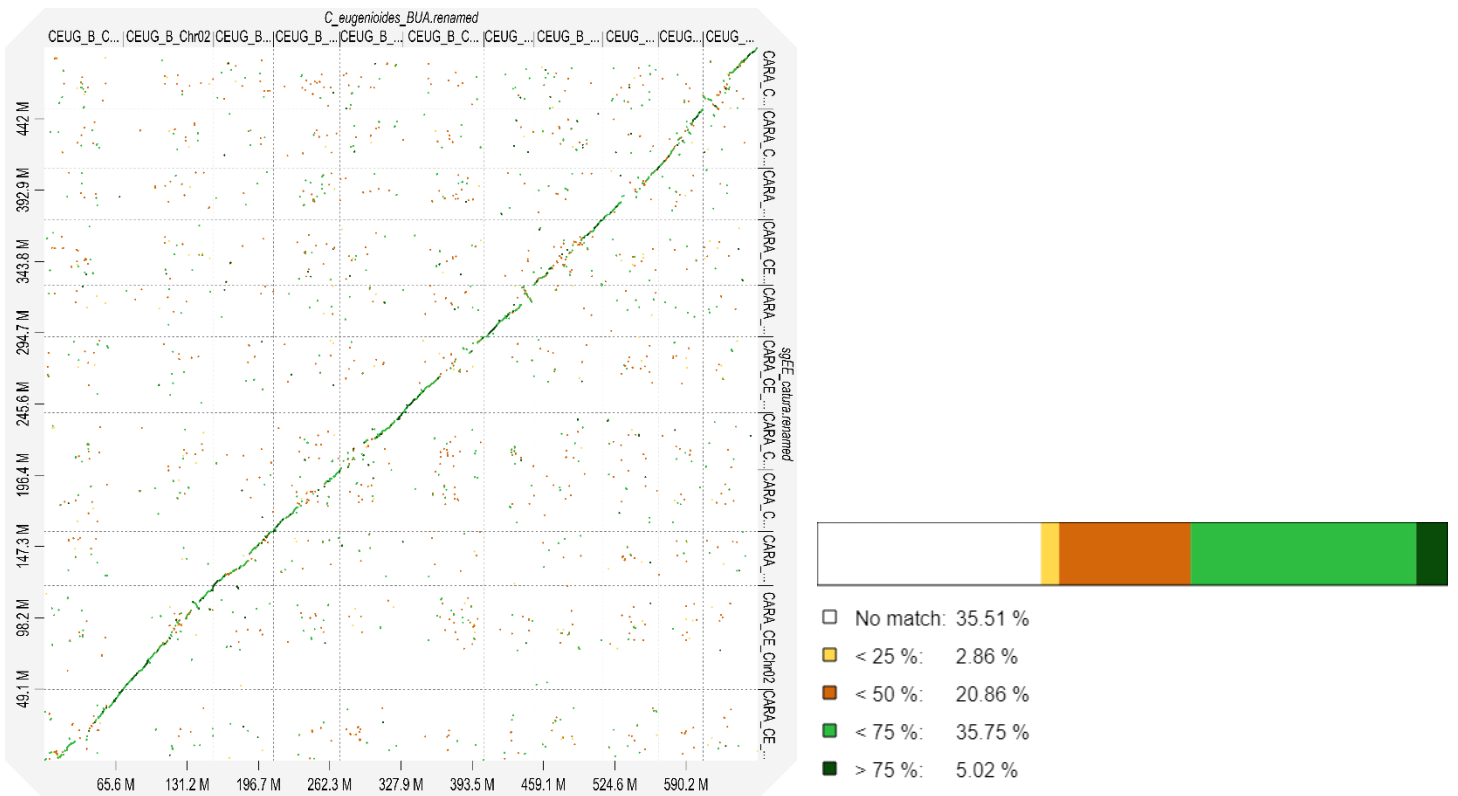


Figura 16. Dotplot *C. eugenioides* y *sgEE_caturra*

6.4.8. *C. eugenioides* (*C. eugenioides_BUA*) vs sgEE *C. arabica* var. *geisha*

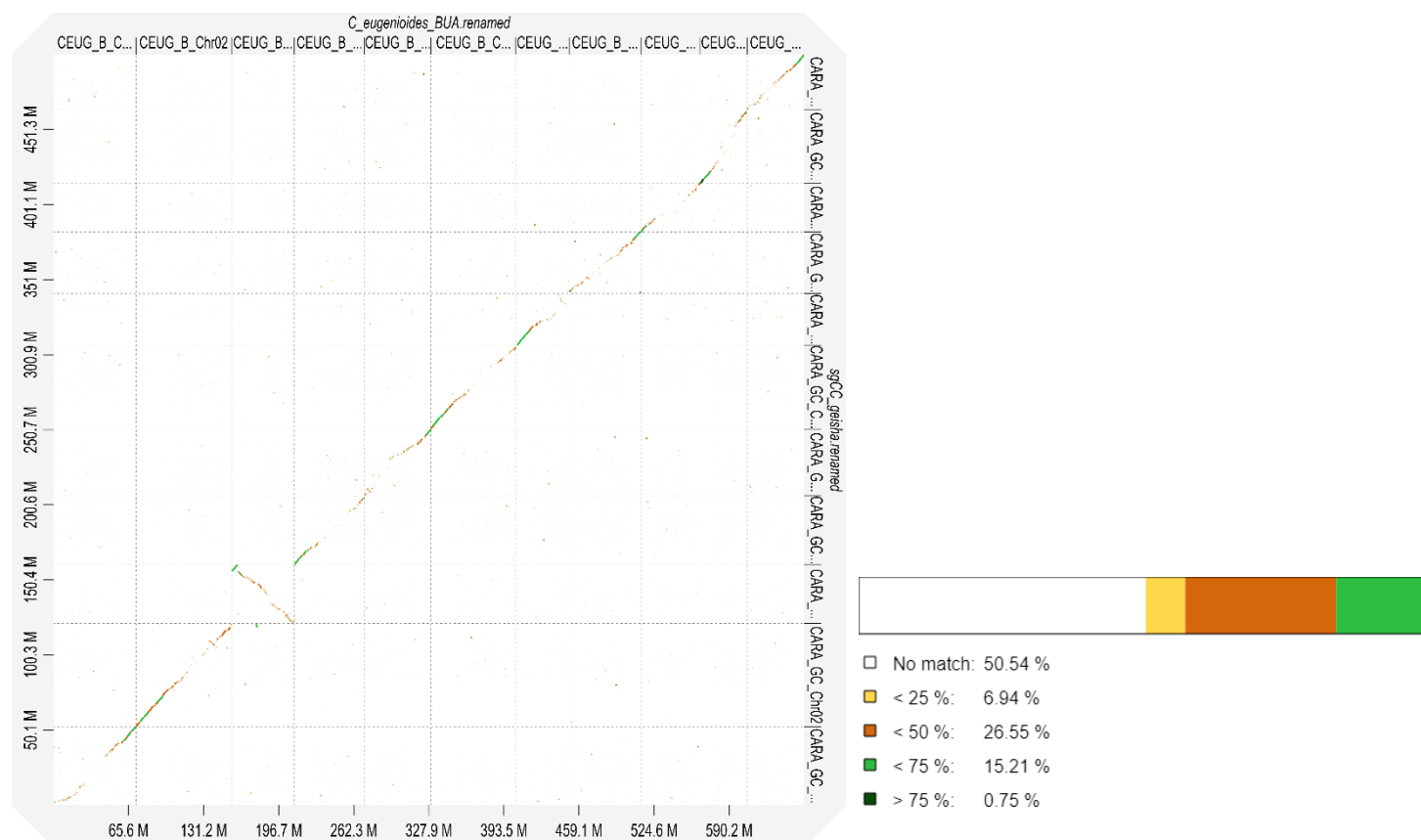


Figura 17. Dotplot *C.eugenioides* y *sgEE_geisha*

6.5. Genes pertenecientes a las familias con mayor número de repeticiones en los genomas evaluados.

Genes Family	Filtered Genes	description
1	Ccan005g011590	Geminivirus AL1 replication-associated protein, catalytic domain
	Ccan007g027000	Geminivirus AL1 replication-associated protein, catalytic domain
	Ccan003g021050	Geminivirus AL1 replication-associated protein, central domain
	Ccan003g021060	Geminivirus AL1 replication-associated protein, catalytic domain
2	Ccan009g013650	Ribonuclease H superfamily
	Ccan005g012060	Ribonuclease H superfamily
	Ccan002g035300	Ribonuclease H-like superfamily
	Ccan005g013680	Ribonuclease H superfamily
	Ccan011g014480	Ribonuclease H superfamily
	Ccan001g013230	Ribonuclease H-like superfamily
	Ccan003g023210	Ribonuclease H superfamily
	Ccan011g008030	Ribonuclease H superfamily
Ccan009g011340	Ribonuclease H superfamily	

Ccan008g017530	Ribonuclease H superfamily
Ccan006g033580	Reverse transcriptase zinc-binding domain
Ccan006g027380	Reverse transcriptase zinc-binding domain
Ccan001g016900	Ribonuclease H superfamily
Ccan005g010850	Ribonuclease H superfamily
Ccan002g035810	Ribonuclease H superfamily
Ccan002g038430	Ribonuclease H superfamily
Ccan008g011470	Reverse transcriptase zinc-binding domain
Ccan001g011600	Ribonuclease H superfamily
Ccan008g013310	Ribonuclease H superfamily
Ccan007g027920	Ribonuclease H-like superfamily
Ccan006g025980	Ribonuclease H superfamily
Ccan011g009780	Reverse transcriptase zinc-binding domain
Ccan011g006080	Reverse transcriptase zinc-binding domain
Ccan008g007350	Ribonuclease H-like superfamily
Ccan004g018450	unnamed product
Ccan002g039650	unnamed product
Ccan008g016570	Ribonuclease H superfamily
Ccan011g014490	Domain of unknown function DUF4283
Ccan003g023200	Domain of unknown function DUF4283
Ccan009g013640	Domain of unknown function DUF4283
Ccan005g010840	Domain of unknown function DUF4283
Ccan009g017980	Domain of unknown function DUF4283
Ccan008g017520	Domain of unknown function DUF4283
Ccan004g020460	Domain of unknown function DUF4283
Ccan007g023780	Protein of unknown function DUF829, TMEM53
Ccan007g014240	Protein of unknown function DUF829, TMEM53
Ccan011g009080	Protein of unknown function DUF829, TMEM53
Ccan001g012810	Ribonuclease H domain
Ccan001g019250	Ribonuclease H-like superfamily
Ccan009g013500	Domain of unknown function DUF4283
Ccan001g015600	Uncharacterized protein At4g02000-like
Ccan004g018860	Domain of unknown function DUF4283
Ccan002g039380	Domain of unknown function DUF4283
Ccan001g015650	Domain of unknown function DUF4283
Ccan002g042210	Domain of unknown function DUF4283
Ccan002g037060	Domain of unknown function DUF4283
Ccan002g042680	Domain of unknown function DUF4283
Ccan006g035020	Domain of unknown function DUF4283
Ccan006g030600	unnamed product
Ccan008g012300	SWIB/MDM2 domain superfamily
Ccan008g020310	unnamed product
Ccan006g028840	Ribonuclease H-like superfamily
Ccan001g017560	unnamed product
Ccan001g017570	Ribonuclease H superfamily
Ccan006g028530	Domain of unknown function DUF4283

Ccan002g038850	Domain of unknown function DUF4283
Ccan010g018040	Domain of unknown function DUF4283
Ccan001g016220	Domain of unknown function DUF4283
Ccan002g038310	Domain of unknown function DUF4283
Ccan002g007350	40S Ribosomal protein S10
Ccan010g018390	40S Ribosomal protein S10
Ccan011g007540	Serine/threonine-protein kinase, active site
Ccan008g013450	40S Ribosomal protein S10
Ccan007g022310	40S Ribosomal protein S10
Ccan004g022590	40S Ribosomal protein S10
Ccan006g026970	Serine/threonine-protein kinase, active site
Ccan009g013440	40S Ribosomal protein S10
Ccan008g015420	Domain of unknown function DUF4283
Ccan010g018920	Domain of unknown function DUF4283
Ccan005g015050	Uncharacterized protein At4g02000-like
Ccan001g012800	Domain of unknown function DUF4283
Ccan003g019490	Domain of unknown function DUF4283
Ccan005g009390	unnamed product
Ccan003g015040	Domain of unknown function DUF4283
Ccan006g025330	Ribonuclease H-like superfamily
Ccan009g014510	Domain of unknown function DUF4283
Ccan005g014620	Domain of unknown function DUF4283
Ccan009g016050	F-box-like domain superfamily
Ccan010g018380	Serine/threonine-protein kinase, active site
Ccan002g006990	Tetratricopeptide-like helical domain superfamily
Ccan008g022750	Serine/threonine-protein kinase, active site
Ccan002g038710	Nucleoside phosphatase GDA1/CD39
Ccan006g028950	Armadillo-like helical
Ccan005g007380	Armadillo-like helical
Ccan001g016540	Domain of unknown function DUF4283
Ccan007g029720	Domain of unknown function DUF4283
Ccan010g020500	Domain of unknown function DUF4283
Ccan009g011480	Domain of unknown function DUF4283
Ccan003g014370	Domain of unknown function DUF4283
Ccan010g016750	Domain of unknown function DUF4283
Ccan006g031650	Domain of unknown function DUF4283
Ccan005g014310	Domain of unknown function DUF4283
Ccan011g010600	unnamed product
Ccan005g007710	Leucine-rich repeat domain superfamily
Ccan001g016620	Ribonuclease H superfamily
Ccan007g029780	Ribonuclease H superfamily
Ccan010g015970	Ribonuclease H superfamily
Ccan006g033120	Domain of unknown function DUF4283
Ccan007g027380	TB2/DP1/HVA22-related protein
Ccan008g010820	Exportin-1, repeat 3
Ccan003g022140	Serine/threonine-protein kinase, active site

Ccan002g023500	Serine/threonine-protein kinase, active site
Ccan007g010420	Serine/threonine-protein kinase, active site
Ccan002g027540	Serine/threonine-protein kinase, active site
Ccan010g011500	Serine/threonine-protein kinase, active site
Ccan007g027020	Serine/threonine-protein kinase, active site
Ccan007g019150	Serine/threonine-protein kinase, active site
Ccan006g022400	Exportin-1, repeat 3
Ccan001g015130	Exportin-1, repeat 3
Ccan002g042150	Armadillo-like helical
Ccan011g005590	Ribosomal protein L4 domain superfamily
Ccan006g028960	Ribosomal protein L4 domain superfamily
Ccan007g028910	Ribosomal protein L4 domain superfamily
Ccan009g014960	Ribosomal protein L4 domain superfamily
Ccan009g016980	Ribosomal protein L4/L1e, eukaryotic/archaeal, conserved site
Ccan007g024290	Ribosomal protein L4/L1e, eukaryotic/archaeal, conserved site
Ccan011g021790	Nucleoside phosphatase GDA1/CD39
Ccan003g023240	Domain of unknown function DUF4283
Ccan008g012740	Endonuclease/exonuclease/phosphatase superfamily
Ccan010g019890	Endonuclease/exonuclease/phosphatase superfamily
Ccan011g009240	Endonuclease/exonuclease/phosphatase superfamily
Ccan010g017150	Ribonuclease H superfamily
Ccan002g041090	Ribonuclease H superfamily
Ccan004g018830	Ribonuclease H superfamily
Ccan004g020160	Reverse transcriptase zinc-binding domain
Ccan010g019400	Ribonuclease H-like superfamily
Ccan009g013470	TB2/DP1/HVA22-related protein
Ccan010g023530	Ribonuclease H superfamily
Ccan004g022900	Ribonuclease H superfamily
Ccan005g013080	Ribonuclease H-like superfamily
Ccan002g036130	Ribonuclease H-like superfamily
Ccan006g025350	Ribonuclease H superfamily
Ccan001g009430	Ribonuclease H superfamily
Ccan007g025690	Ribonuclease H superfamily
Ccan011g011180	Ribonuclease H-like superfamily
Ccan004g019260	Ribonuclease H-like superfamily
Ccan011g008760	unnamed product
Ccan006g025860	Ribonuclease H superfamily
Ccan009g011200	unnamed product
Ccan004g021330	Reverse transcriptase zinc-binding domain
Ccan005g005410	Ribonuclease H-like superfamily
Ccan005g014610	Ribonuclease H superfamily
Ccan005g011510	Ribonuclease H-like superfamily
Ccan001g017040	Ribonuclease H-like superfamily
Ccan008g017380	Ribonuclease H superfamily
Ccan010g010620	Membrane-anchored ubiquitin-fold protein
Ccan004g025460	unnamed product

Ccan006g028910	Ribosomal protein L4 domain superfamily
Ccan011g009060	Small auxin-up RNA
Ccan002g038400	Endonuclease/exonuclease/phosphatase superfamily
Ccan002g039520	Endonuclease/exonuclease/phosphatase superfamily
Ccan006g028340	Reverse transcriptase zinc-binding domain
Ccan003g022480	Ribonuclease H superfamily
Ccan005g014820	Ribonuclease H superfamily
Ccan009g014050	TB2/DP1/HVA22-related protein
Ccan001g017300	TB2/DP1/HVA22-related protein
Ccan009g017940	unnamed product
Ccan006g030410	TB2/DP1/HVA22-related protein
Ccan004g020660	Ribonuclease H-like superfamily
Ccan005g015060	TB2/DP1/HVA22-related protein
Ccan010g017820	Ribonuclease H superfamily
Ccan010g019710	Ribonuclease H superfamily
Ccan008g013470	unnamed product
Ccan005g009290	Reverse transcriptase zinc-binding domain
Ccan003g021680	Reverse transcriptase zinc-binding domain
Ccan003g020720	unnamed product
Ccan003g021120	Ribonuclease H superfamily
Ccan009g015820	Reverse transcriptase zinc-binding domain
Ccan008g009410	Ribonuclease H-like superfamily
Ccan011g008870	unnamed product
Ccan003g022800	unnamed product
Ccan002g021790	Small auxin-up RNA
Ccan011g009020	Small auxin-up RNA
Ccan008g019250	Small auxin-up RNA
Ccan008g019260	Small auxin-up RNA
Ccan002g021780	Small auxin-up RNA
Ccan007g026520	Endonuclease/exonuclease/phosphatase superfamily
Ccan010g017670	Endonuclease/exonuclease/phosphatase superfamily
Ccan004g022580	Reverse transcriptase zinc-binding domain
Ccan004g018950	Endonuclease/exonuclease/phosphatase superfamily
Ccan006g032830	Ribonuclease H superfamily
Ccan006g032950	Ribonuclease H-like superfamily
Ccan002g037030	Ribonuclease H superfamily
Ccan002g032500	Ribonuclease H superfamily
Ccan007g025700	Ribonuclease H superfamily
Ccan007g025270	Ribonuclease H superfamily
Ccan004g018400	Ribonuclease H-like superfamily
Ccan008g013950	Ribonuclease H superfamily
Ccan001g012020	Ribonuclease H-like superfamily
Ccan006g030970	Ribonuclease H superfamily
Ccan007g026400	Ribonuclease H superfamily
Ccan009g018210	Ribonuclease H superfamily
Ccan001g015670	Ribonuclease H superfamily

Ccan004g017690	Ribonuclease H superfamily
Ccan008g021790	Ribonuclease H superfamily
Ccan002g041210	Ribonuclease H superfamily
Ccan006g032970	Ribonuclease H-like superfamily
Ccan011g006090	Ribonuclease H superfamily
Ccan003g027160	Ribonuclease H superfamily
Ccan007g019600	Ribonuclease H-like superfamily
Ccan001g013460	Ribonuclease H superfamily
Ccan003g026490	Ribonuclease H superfamily
Ccan011g009030	Small auxin-up RNA
Ccan002g021770	Small auxin-up RNA
Ccan003g019730	Endonuclease/exonuclease/phosphatase superfamily
Ccan006g028730	Endonuclease/exonuclease/phosphatase superfamily
Ccan006g028720	Endonuclease/exonuclease/phosphatase superfamily
Ccan009g014490	Endonuclease/exonuclease/phosphatase superfamily
Ccan001g019500	Ribonuclease H-like superfamily
Ccan010g018240	Endonuclease/exonuclease/phosphatase superfamily
Ccan005g010980	Endonuclease/exonuclease/phosphatase superfamily
Ccan006g031800	Reverse transcriptase zinc-binding domain
Ccan002g040060	Endonuclease/exonuclease/phosphatase superfamily
Ccan010g018290	Endonuclease/exonuclease/phosphatase superfamily
Ccan001g013930	Endonuclease/exonuclease/phosphatase superfamily
Ccan005g009680	Ribonuclease H superfamily
Ccan010g013710	Ribonuclease H superfamily
Ccan005g015110	Ribonuclease H-like superfamily
Ccan009g013540	Ribonuclease H-like superfamily
Ccan011g008020	Ribonuclease H-like superfamily
Ccan011g008010	Reverse transcriptase zinc-binding domain
Ccan007g027340	Ribonuclease H superfamily
Ccan001g012980	Ribonuclease H superfamily
Ccan005g011380	Ribonuclease H-like superfamily
Ccan006g030610	Ribonuclease H superfamily
Ccan011g009220	Ribonuclease H superfamily
Ccan007g029990	Ribonuclease H superfamily
Ccan001g014100	unnamed product
Ccan006g029560	Reverse transcriptase zinc-binding domain
Ccan005g012830	unnamed product
Ccan009g015910	unnamed product
Ccan006g029220	Endonuclease/exonuclease/phosphatase superfamily
Ccan010g018310	Endonuclease/exonuclease/phosphatase superfamily
Ccan010g018030	Endonuclease/exonuclease/phosphatase superfamily
Ccan002g021800	Tetratricopeptide-like helical domain superfamily
Ccan004g021500	unnamed product
Ccan004g021120	unnamed product
Ccan007g028940	Ribonuclease H-like superfamily
Ccan001g016070	Endonuclease/exonuclease/phosphatase superfamily

	Ccan007g029960	Endonuclease/exonuclease/phosphatase superfamily
	Ccan004g023040	Ribonuclease H-like superfamily
	Ccan001g013000	Ribonuclease H superfamily
	Ccan001g016930	Ribonuclease H superfamily
	Ccan007g029110	Ribonuclease H superfamily
	Ccan003g022700	Ribonuclease H superfamily
	Ccan002g042070	TB2/DP1/HVA22-related protein
	Ccan011g010460	Ribonuclease H-like superfamily
	Ccan002g039260	Ribonuclease H-like superfamily
	Ccan004g021550	Endonuclease/exonuclease/phosphatase superfamily
	Ccan010g018300	Endonuclease/exonuclease/phosphatase superfamily
	Ccan011g009940	Reverse transcriptase zinc-binding domain
	Ccan007g028950	Domain of unknown function DUF4283
	Ccan009g015540	Ribonuclease H superfamily
	Ccan009g015530	unnamed product
	Ccan001g015040	unnamed product
	Ccan001g015050	Ribonuclease H-like superfamily
	Ccan002g042080	unnamed product
	Ccan004g023830	Reverse transcriptase domain
	Ccan006g028370	Endonuclease/exonuclease/phosphatase superfamily
	Ccan011g009700	Endonuclease/exonuclease/phosphatase superfamily
	Ccan002g040210	Endonuclease/exonuclease/phosphatase superfamily
	Ccan002g039310	Endonuclease/exonuclease/phosphatase superfamily
	Ccan003g022670	unnamed product
	Ccan002g041800	Ribonuclease H superfamily
	Ccan003g020630	unnamed product
	Ccan004g021560	unnamed product
	Ccan005g013010	Endonuclease/exonuclease/phosphatase superfamily
	Ccan002g041860	Endonuclease/exonuclease/phosphatase superfamily
	Ccan005g013040	unnamed product
	Ccan005g014110	Endonuclease/exonuclease/phosphatase superfamily
	Ccan001g017460	Kelch-type beta propeller
	Ccan011g008830	Reverse transcriptase zinc-binding domain
	Ccan008g014770	unnamed product
	Ccan006g020630	unnamed product
	Ccan006g031780	unnamed product
	Ccan006g020620	unnamed product
	Ccan001g032460	NADP-dependent oxidoreductase domain superfamily
3	Ccan004g017850	Aminotransferase-like, plant mobile domain
	Ccan007g027630	Aminotransferase-like, plant mobile domain
	Ccan008g020280	Aminotransferase-like, plant mobile domain
	Ccan002g044660	Aminotransferase-like, plant mobile domain
	Ccan011g023650	Aminotransferase-like, plant mobile domain
	Ccan004g016450	Aminotransferase-like, plant mobile domain
	Ccan006g030450	Aminotransferase-like, plant mobile domain
	Ccan002g033960	Aminotransferase-like, plant mobile domain

Ccan008g014670	Aminotransferase-like, plant mobile domain
Ccan004g020530	Aminotransferase-like, plant mobile domain
Ccan005g014030	Aminotransferase-like, plant mobile domain
Ccan008g015840	Aminotransferase-like, plant mobile domain
Ccan002g047630	Aminotransferase-like, plant mobile domain
Ccan005g013560	Aminotransferase-like, plant mobile domain
Ccan008g016950	Aminotransferase-like, plant mobile domain
Ccan004g022750	Aminotransferase-like, plant mobile domain
Ccan002g038020	Aminotransferase-like, plant mobile domain
Ccan007g026220	Aminotransferase-like, plant mobile domain
Ccan008g023510	Aminotransferase-like, plant mobile domain
Ccan011g008470	Aminotransferase-like, plant mobile domain
Ccan011g008490	Aminotransferase-like, plant mobile domain
Ccan003g016740	Aminotransferase-like, plant mobile domain
Ccan009g014140	Aminotransferase-like, plant mobile domain
Ccan009g017250	Aminotransferase-like, plant mobile domain
Ccan006g028750	Aminotransferase-like, plant mobile domain
Ccan006g033610	Aminotransferase-like, plant mobile domain
Ccan003g014750	Aminotransferase-like, plant mobile domain
Ccan005g014810	Aminotransferase-like, plant mobile domain
Ccan007g028080	Aminotransferase-like, plant mobile domain
Ccan010g013810	Aminotransferase-like, plant mobile domain
Ccan002g029330	Aminotransferase-like, plant mobile domain
Ccan005g016200	Aminotransferase-like, plant mobile domain
Ccan007g024820	Aminotransferase-like, plant mobile domain
Ccan009g012580	Aminotransferase-like, plant mobile domain
Ccan006g031020	Aminotransferase-like, plant mobile domain
Ccan002g033790	Aminotransferase-like, plant mobile domain
Ccan011g008480	Aminotransferase-like, plant mobile domain
Ccan011g009770	Aminotransferase-like, plant mobile domain
Ccan002g043110	Aminotransferase-like, plant mobile domain
Ccan002g041130	Aminotransferase-like, plant mobile domain
Ccan009g016270	Aminotransferase-like, plant mobile domain
Ccan010g018410	Aminotransferase-like, plant mobile domain
Ccan010g017620	Aminotransferase-like, plant mobile domain
Ccan007g024710	Aminotransferase-like, plant mobile domain
Ccan003g009310	Aminotransferase-like, plant mobile domain
Ccan002g020390	Aminotransferase-like, plant mobile domain
Ccan002g029340	Aminotransferase-like, plant mobile domain
Ccan002g037890	Aminotransferase-like, plant mobile domain
Ccan007g026180	Aminotransferase-like, plant mobile domain
Ccan006g027470	Aminotransferase-like, plant mobile domain
Ccan006g025910	Aminotransferase-like, plant mobile domain
Ccan009g010770	Aminotransferase-like, plant mobile domain
Ccan003g021010	Aminotransferase-like, plant mobile domain
Ccan009g014540	Aminotransferase-like, plant mobile domain

Ccan005g012170	Aminotransferase-like, plant mobile domain
Ccan010g016100	Aminotransferase-like, plant mobile domain
Ccan011g016940	Aminotransferase-like, plant mobile domain
Ccan008g015830	Aminotransferase-like, plant mobile domain
Ccan003g016100	Aminotransferase-like, plant mobile domain
Ccan008g011660	Aminotransferase-like, plant mobile domain
Ccan011g014060	Aminotransferase-like, plant mobile domain
Ccan005g010380	Aminotransferase-like, plant mobile domain
Ccan002g017280	Aminotransferase-like, plant mobile domain
Ccan004g016550	Aminotransferase-like, plant mobile domain
Ccan007g016070	Aminotransferase-like, plant mobile domain
Ccan002g045040	Aminotransferase-like, plant mobile domain
Ccan009g016920	Aminotransferase-like, plant mobile domain
Ccan003g026820	Aminotransferase-like, plant mobile domain
Ccan011g013940	Aminotransferase-like, plant mobile domain
Ccan001g012490	Aminotransferase-like, plant mobile domain
Ccan002g040510	Aminotransferase-like, plant mobile domain
Ccan008g012400	Aminotransferase-like, plant mobile domain
Ccan006g028650	Aminotransferase-like, plant mobile domain
Ccan009g017540	unnamed product
Ccan003g026810	unnamed product
Ccan001g012690	Aminotransferase-like, plant mobile domain
Ccan011g008430	Aminotransferase-like, plant mobile domain
Ccan001g022030	Aminotransferase-like, plant mobile domain
Ccan002g032540	Aminotransferase-like, plant mobile domain
Ccan005g013630	Aminotransferase-like, plant mobile domain
Ccan001g008090	Aminotransferase-like, plant mobile domain
Ccan010g015580	Wall-associated receptor kinase, galacturonan-binding domain
Ccan010g015560	Wall-associated receptor kinase, galacturonan-binding domain
Ccan003g018020	Aminotransferase-like, plant mobile domain
Ccan008g014430	Aminotransferase-like, plant mobile domain
Ccan010g020490	Aminotransferase-like, plant mobile domain
Ccan003g014330	Aminotransferase-like, plant mobile domain
Ccan010g020700	Aminotransferase-like, plant mobile domain
Ccan011g009820	Aminotransferase-like, plant mobile domain
Ccan010g005060	Aminotransferase-like, plant mobile domain
Ccan006g024680	Aminotransferase-like, plant mobile domain
Ccan003g014710	Aminotransferase-like, plant mobile domain
Ccan004g012960	Aminotransferase-like, plant mobile domain
Ccan002g042610	Aminotransferase-like, plant mobile domain
Ccan011g012550	Aminotransferase-like, plant mobile domain
Ccan004g017160	Aminotransferase-like, plant mobile domain
Ccan002g031570	Aminotransferase-like, plant mobile domain
Ccan002g041530	Aminotransferase-like, plant mobile domain
Ccan009g015860	Aminotransferase-like, plant mobile domain
Ccan009g015760	Aminotransferase-like, plant mobile domain

	Ccan009g017530	unnamed product
	Ccan009g010860	Wall-associated receptor kinase, galacturonan-binding domain
	Ccan011g008440	unnamed product
	Ccan003g014720	unnamed product
	Ccan010g016800	unnamed product
	Ccan005g012100	Aminotransferase-like, plant mobile domain
	Ccan002g024170	Aminotransferase-like, plant mobile domain
	Ccan008g013150	Aminotransferase-like, plant mobile domain
	Ccan009g013250	Aminotransferase-like, plant mobile domain
	Ccan006g034440	Aminotransferase-like, plant mobile domain
	Ccan001g028480	Aminotransferase-like, plant mobile domain
	Ccan009g010830	Wall-associated receptor kinase, galacturonan-binding domain
	Ccan009g010820	Wall-associated receptor kinase, galacturonan-binding domain
	Ccan009g010870	Cytochrome P450 superfamily
	Ccan011g008420	unnamed product
	Ccan003g023820	Thioredoxin-like superfamily
	Ccan002g039210	unnamed product
	Ccan009g014090	unnamed product
	Ccan006g024990	unnamed product
	Ccan006g021980	Aminotransferase-like, plant mobile domain
	Ccan009g015850	unnamed product
	Ccan006g026610	unnamed product
	Ccan002g048270	Aminotransferase-like, plant mobile domain
	Ccan009g014080	Aminotransferase-like, plant mobile domain
	Ccan007g028380	Protein kinase, ATP binding site
	Ccan009g010760	Wall-associated receptor kinase, galacturonan-binding domain
	Ccan009g010790	Serine/threonine-protein kinase, active site
	Ccan009g010800	Serine/threonine-protein kinase, active site
	Ccan009g010810	Serine/threonine-protein kinase, active site
	Ccan002g047820	unnamed product
	Ccan003g011450	unnamed product
	Ccan007g017050	unnamed product
	Ccan006g021970	unnamed product
	Ccan010g014150	Aminotransferase-like, plant mobile domain
	Ccan007g028830	unnamed product
	Ccan006g029960	Aminotransferase-like, plant mobile domain
	Ccan006g026920	unnamed product
	Ccan003g019080	unnamed product
	Ccan009g012860	Serine/threonine-protein kinase, active site
	Ccan009g015750	unnamed product
	Ccan002g047810	Aminotransferase-like, plant mobile domain
	Ccan006g029970	unnamed product
	Ccan009g010780	Wall-associated receptor kinase, galacturonan-binding domain
4	Ccan008g019870	Leucine-rich repeat domain superfamily
	Ccan008g019710	Serine/threonine-protein kinase, active site
5	Ccan008g019490	unnamed product

	Ccan008g019630	unnamed product
	Ccan008g019540	unnamed product
	Ccan008g019570	unnamed product

BIBLIOGRAFÍA

- Alarcó López, A. (2011). *Modelo de gestión productiva para el cultivo de café (Coffea arabica L.) en el Sur del Ecuador* [Proyecto de Fin de Carrera]. Universidad Politécnica de Madrid.
- Anthony, F., Diniz, L. E. C., Combes, M.-C., & Lashermes, P. (2010). Adaptive radiation in *Coffea* subgenus *Coffea* L. (Rubiaceae) in Africa and Madagascar. *Plant Systematics and Evolution*, 285(1-2), 51-64. <https://doi.org/10.1007/s00606-009-0255-8>
- Baroni, L., Sarni, A., & Zuliani, C. (2021). Plant Foods Rich in Antioxidants and Human Cognition: A Systematic Review. Antioxidants. *Multidisciplinary Digital Publishing Institute*, 10, 714.
- Bulatova, N. (2020). Archives: The law of homologous series in variation (N. I. Vavilov). *Comparative Cytogenetics*, 14(3), 329-338. <https://doi.org/10.3897/CompCytogen.v14i3.54511>
- Campa, C., Doulebeau, S., Dussert, S., Hamon, S., & Noirot, M. (2005). Diversity in bean Caffeine content among wild *Coffea* species: Evidence of a discontinuous distribution. *Food Chemistry*, 91(4), 633-637. <https://doi.org/10.1016/j.foodchem.2004.06.032>
- Cardenas, S. (2007). *Caracterización morfológica y agronómica de la colección núcleo de café (Coffea arabica L.) del CATIE* [Tesis de Maestría]. Centro Agronómico Tropical de Investigación y Enseñanza.
- Catacchio, C. R., Maggiolini, F. A. M., D'Addabbo, P., Bitonto, M., Capozzi, O., Signorile, M. L., Miroballo, M., Archidiacono, N., Eichler, E. E., Ventura, M., & Antonacci, F. (2018). Inversion variants in human and primate genomes. *Genome Research*, 28(6), 910-920. <https://doi.org/10.1101/gr.234831.118>
- Cenci, A., Combes, M.-C., & Lashermes, P. (2010). Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome. *Molecular Genetics and Genomics*, 283(5), 493-501. <https://doi.org/10.1007/s00438-010-0534-7>
- Cenci, A., Combes, M.-C., & Lashermes, P. (2012). Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant Molecular Biology*, 78(1-2), 135-145. <https://doi.org/10.1007/s11103-011-9852-3>
- Chen, X., & Tompa, M. (2010). Comparative assessment of methods for aligning multiple genome sequences. *Nature Biotechnology*, 28(6), 567-572. <https://doi.org/10.1038/nbt.1637>
- Chen, Y.-L., Chen, C.-M., Pai, T.-W., Leong, H.-W., & Chong, K.-F. (2013). Homologous synteny block detection based on suffix tree algorithms. *Journal of Bioinformatics and Computational Biology*, 11(06), 1343004. <https://doi.org/10.1142/S021972001343004X>

- Clarindo, W. R., & Carvalho, C. R. (2009). Comparison of the *Coffea canephora* and *C. arabica* karyotype based on chromosomal DNA content. *Plant Cell Reports*, 28(1), 73-81. <https://doi.org/10.1007/s00299-008-0621-y>
- Cotton, J. A., Bennuru, S., Grote, A., Harsha, B., Tracey, A., Beech, R., Doyle, S. R., Dunn, M., Hotopp, J. C. D., Holroyd, N., Kikuchi, T., Lambert, O., Mhashilkar, A., Mutowo, P., Nursimulu, N., Ribeiro, J. M. C., Rogers, M. B., Stanley, E., Swapna, L. S., ... Lustigman, S. (2016). The genome of *Onchocerca volvulus*, agent of river blindness. *Nature Microbiology*, 2(2), 16216. <https://doi.org/10.1038/nmicrobiol.2016.216>
- Davis, A. P. (2011). *Psilanthus mannii*, the type species of *Psilanthus*, transferred to *Coffea*. *Nord.J.Bot*, 29, 471-472.
- Davis, A. P., Chadburn, H., Moat, J., O'Sullivan, R., Hargreaves, S., & Lughadha. (2019). High extinction risk for wild coffee species and implications for coffee sector sustainability. *ScienceAdvances*, 5(1).
- Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., Aury, J.-M., Bento, P., Bernard, M., Bocs, S., Campa, C., Cenci, A., Combes, M.-C., Cruzillat, D., Da Silva, C., ... Lashermes, P. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*, 345(6201), 1181-1184. <https://doi.org/10.1126/science.1255274>
- Espinosa Guzmán, G. L. (2009). *Diversidad genética del café (Coffea arabica L.) distribuida en la Provincia de Loja – Ecuador* [Trabajo de Grado]. Universidad Técnica Particular de Loja.
- Feschotte, C., & Pritham, E. J. (2007). DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics*, 41(1), 331-368. <https://doi.org/10.1146/annurev.genet.40.110405.090448>
- Gordon, D., Huddleston, J., Chaisson, M. J. P., Hill, C. M., Kronenberg, Z. N., Munson, K. M., Malig, M., Raja, A., Fiddes, I., Hillier, L. W., Dunn, C., Baker, C., Armstrong, J., Diekhans, M., Paten, B., Shendure, J., Wilson, R. K., Haussler, D., Chin, C.-S., & Eichler, E. E. (2016). Long-read sequence assembly of the gorilla genome. *Science*, 352(6281), aae0344. <https://doi.org/10.1126/science.aae0344>
- Guyot, R., Lefebvre-Pautigny, F., Tranchant-Dubreuil, C., Rigoreau, M., Hamon, P., Leroy, T., Hamon, S., Poncet, V., Cruzillat, D., & De Kochko, A. (2012). Ancestral synteny shared between distantly-related plant species from the asterid (*Coffea canephora* and *Solanum* Sp.) and rosid (*Vitis vinifera*) clades. *BMC Genomics*, 13(1), 103. <https://doi.org/10.1186/1471-2164-13-103>
- Hardie, D. G. (1999). PLANT PROTEIN SERINE/THREONINE KINASES: Classification and Functions. *Annual Review of Plant Physiology and Plant Molecular Biology*, 50(1), 97-131. <https://doi.org/10.1146/annurev.arplant.50.1.97>
- Hulbert, S. H., Webb, C. A., Smith, S. M., & Sun, Q. (2001). RESISTANCE GENE COMPLEXES: Evolution and Utilization. *Annual Review of Phytopathology*, 39(1), 285-312. <https://doi.org/10.1146/annurev.phyto.39.1.285>

- Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323-329. <https://doi.org/10.1038/nature05286>
- Kleckner, N. (1977). Translocatable elements in procaryotes. *Cell*, *11*(1), 11-23. [https://doi.org/10.1016/0092-8674\(77\)90313-0](https://doi.org/10.1016/0092-8674(77)90313-0)
- Lashermes, P., Andrade, A. C., & Etienne, H. (2008). Genomics of Coffee One of the World's Largest Traded Commodities. En P. H. Moore & R. Ming (Eds.), *Genomics of Tropical Crop Plants* (Vol. 1, pp. 203-226). Springer New York. https://doi.org/10.1007/978-0-387-71219-2_9
- Lashermes, P., Combes, M.-C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F., & Charrier, A. (1999). Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics MGG*, *261*(2), 259-266. <https://doi.org/10.1007/s004380050965>
- Leale, G., Baya, A. E., Milone, D. H., Granitto, P. M., & Stegmayer, G. (2018). Inferring Unknown Biological Function by Integration of GO Annotations and Gene Expression Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *15*(1), 168-180. <https://doi.org/10.1109/TCBB.2016.2615960>
- Liu, D., Hunt, M., & Tsai, I. (2018). Inferring synteny between genome assemblies: A systematic evaluation. *BMC Bioinformatics*, *19*. <https://doi.org/10.1186/s12859-018-2026-4>
- Mahé, L., Combes, M.-C., Várzea, V. M. P., Guilhaumon, C., & Lashermes, P. (2007). Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.). *Molecular Breeding*, *21*(1), 105-113. <https://doi.org/10.1007/s11032-007-9112-z>
- Maluf, M. P., Silva, C. C. D., Oliveira, M. D. P. A. D., Tavares, A. G., Silvarolla, M. B., & Guerreiro Filho, O. (2009). Altered expression of the caffeine synthase gene in a naturally caffeine-free mutant of *Coffea arabica*. *Genetics and Molecular Biology*, *32*(4), 802-810. <https://doi.org/10.1590/S1415-47572009005000090>
- Maurin, O., Davis, A. P., Chester, M., Mvungi, E. F., Jaufeerally-Fakim, Y., & Fay, M. F. (2007). Towards a Phylogeny for *Coffea* (Rubiaceae): Identifying Well-supported Lineages Based on Nuclear and Plastid DNA Sequences. *Annals of Botany*, *100*(7), 1565-1583. <https://doi.org/10.1093/aob/mcm257>
- Mawalagedera, S., Callahan, D., Gaskett, A., Ronsted, N., & Symonds, M. (2019). Combining Evolutionary Inference and Metabolomics to Identify Plants With Medicinal Potential. *Frontiers in Ecology and Evolution*. *Frontiers*. *Media S.A*, 7-267.
- Medrano, J. F., Cantu, D., Minio, A., Dreischer, C., Gibbons, T., Chin, J., Chen, S., Van Deynze, A., & Hulse-Kemp, A. M. (2024). *De novo whole-genome assembly and annotation of a high-quality coffee variety from the primary origin of coffee, Coffea arabica var. Geisha*. <https://doi.org/10.1101/2024.06.21.600137>
- Mende, D. R., Sunagawa, S., Zeller, G., & Bork, P. (2013). Accurate and universal delineation of prokaryotic species. *Nature Methods*, *10*(9), 881-884. <https://doi.org/10.1038/nmeth.2575>
- Meyers, B., Scalabrin, S., & Morgante, M. (2004). Mapping and sequencing complex genomes let's get physical! *Nature Reviews Genetics*, *5*(8).

- Mishra, M. K., Huded, A. K. C., & Jingade, P. (2020). Assessment of the Suitability of Molecular SCoT Markers for Genetic Analysis of Coffee Species. *Botanica*, 26(2), 184-196. <https://doi.org/10.2478/botlit-2020-0019>
- Mishra, M. K., Sandhyarani, N., Suresh, N., Satheesh Kumar, S., Soumya, P. R., Yashodha, M. H., Bhat, A., & Jayarama. (2012). Genetic Diversity Among Indian Coffee Cultivars Determined via Molecular Markers. *Journal of Crop Improvement*, 26(6), 727-750. <https://doi.org/10.1080/15427528.2012.696085>
- Morgante, M., Depaoli, E., & Radovic, S. (2007). Transposable elements and the plant pan-genomes. *Current Opinion in Plant Biology*, 10(2), 149-155. <https://doi.org/10.1016/j.pbi.2007.02.001>
- Nathanson, J. A. (1984). Caffeine and Related Methylxanthines: Possible Naturally Occurring Pesticides. *Science*, 226(4671), 184-187. <https://doi.org/10.1126/science.6207592>
- Noël, L., Moores, T. L., Parker, J. E., & Jones, J. D. G. (1999). Pronounced Intraspecific Haplotype Divergence at the RPP5 Complex Disease Resistance Locus of Arabidopsis. *The Plant Cell*, 11, 2099-2111.
- Noirot, M., Barre, P., Louarn, J., Duperray, C., & Hamon, S. (2002). Consequences of Stoichiometric Error on Nuclear DNA Content Evaluation in Coffea liberica var. Dewevrei using DAPI and Propidium Iodide. *Annals of Botany*, 89(4), 385-389. <https://doi.org/10.1093/aob/mcf056>
- Oliveira, L. N. L. D., Rocha, R. B., Ferreira, F. M., Spinelli, V. M., Ramalho, A. R., & Teixeira, A. L. (2018). Selection of Coffea canephora parents from the botanical varieties Conilon and Robusta for the production of intervarietal hybrids. *Ciência Rural*, 48(4). <https://doi.org/10.1590/0103-8478cr20170444>
- Pacheco, A., Pohlan, J., & Schulz, M. (2008). Allelopathic effects of aromatic species intercropped with coffee: Investigation of their growth stimulation capacity and potential of caffeine uptake in Puebla, Mexico. *Allelopathy Journal*, 21(1).
- Padmanabhan, M., Cournoyer, P., & Dinesh-Kumar, S. P. (2009). The leucine-rich repeat domain in plant innate immunity: A wealth of possibilities. *Cellular Microbiology*, 11(2), 191-198. <https://doi.org/10.1111/j.1462-5822.2008.01260.x>
- Parniske, M., & Jones, J. D. G. (1999). Recombination between diverged clusters of the tomato Cf-9 plant disease resistance gene family. *Proceedings of the National Academy of Sciences*, 96(10), 5850-5855. <https://doi.org/10.1073/pnas.96.10.5850>
- Paterson, A. H., Bowers, J. E., Burow, M. D., Draye, X., Elisk, C. G., Jiang, C.-X., Katsar, C. S., Lan, T.-H., Lin, Y.-R., Ming, R., & Wright, R. J. (2000). Comparative Genomics of Plant Chromosomes. *The Plant Cell*, 12, 1523-1539.
- Pearl, H. M., Nagai, C., Moore, P. H., Steiger, D. L., Osgood, R. V., & Ming, R. (2004). Construction of a genetic map for arabica coffee. *Theoretical and Applied Genetics*, 108(5), 829-835. <https://doi.org/10.1007/s00122-003-1498-3>
- Perrois, C., Strickler, S. R., Mathieu, G., Lepelley, M., Bedon, L., Michaux, S., Husson, J., Mueller, L., & Privat, I. (2015). Differential regulation of caffeine metabolism in Coffea

arabica (Arabica) and *Coffea canephora* (Robusta). *Planta*, 241(1), 179-191. <https://doi.org/10.1007/s00425-014-2170-7>

- Pilozo, W. (2021). *Principales enfermedades causantes de la perdida de rendimientos de los cultivares de café arábigo (Coffea arábica) en la Zona del Sur de Manabí*. [Pregrado]. Universidad Estatal del Sur de Manabí.
- Polturak, G., Liu, Z., & Osbourn, A. (2022). New and emerging concepts in the evolution and function of plant biosynthetic gene clusters. *Current Opinion in Green and Sustainable Chemistry*, 33, 100568. <https://doi.org/10.1016/j.cogsc.2021.100568>
- Prakash, N. S., Marques, D. V., Varzea, V. M. P., Silva, M. C., Combes, M. C., & Lashermes, P. (2004). Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. *Theoretical and Applied Genetics*, 109(6), 1311-1317. <https://doi.org/10.1007/s00122-004-1748-z>
- Purayil, G. P., Saeed, E. E., Mathai, A. M., El-Tarabily, K. A., & AbuQamar, S. F. (2024). A high-quality genome assembly and annotation of *Thielaviopsis punctulata* DSM102798. *Scientific Data*, 11(1), 745. <https://doi.org/10.1038/s41597-024-03458-y>
- Raharimalala, N., Rombauts, S., McCarthy, A., Garavito, A., Orozco-Arias, S., Bellanger, L., Morales-Correa, A. Y., Froger, S., Michaux, S., Berry, V., Metairon, S., Fournier, C., Lepelley, M., Mueller, L., Couturon, E., Hamon, P., Rakotomalala, J.-J., Descombes, P., Guyot, R., & Crouzillat, D. (2021). The absence of the caffeine synthase gene is involved in the naturally decaffeinated status of *Coffea humblotiana*, a wild species from Comoro archipelago. *Scientific Reports*, 11(1), 8119. <https://doi.org/10.1038/s41598-021-87419-0>
- Rivas, L. G. L., & Herrera, G. Z. (2000). *Análisis estadístico de la producción de Café en el Ecuador*. Escuela Superior Politecnica del Litoral.
- Salojärvi, J., Rambani, A., Yu, Z., Guyot, R., Strickler, S., Lepelley, M., Wang, C., Rajaraman, S., Rastas, P., Zheng, C., Muñoz, D. S., Meidanis, J., Paschoal, A. R., Bawin, Y., Krabbenhoft, T. J., Wang, Z. Q., Fleck, S. J., Aussel, R., Bellanger, L., ... Descombes, P. (2024). The genome and population genomics of allopolyploid *Coffea arabica* reveal the diversification history of modern coffee cultivars. *Nature Genetics*, 56(4), 721-731. <https://doi.org/10.1038/s41588-024-01695-w>
- Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Quraishi, U. M., Calcagno, T., Cooke, R., Delseny, M., & Feuillet, C. (2008). Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *The Plant Cell*, 20(1), 11-24. <https://doi.org/10.1105/tpc.107.056309>
- Tettelin, H., Massignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., DeBoy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit Y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., ... Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences*, 102(39), 13950-13955. <https://doi.org/10.1073/pnas.0506758102>

- The French–Italian Public Consortium for Grapevine Genome Characterization. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449(7161), 463-467. <https://doi.org/10.1038/nature06148>
- Uricaru, R., Michotey, C., Chiapello, H., & Rivals, E. (2015). YOC, A new strategy for pairwise alignment of collinear genomes. *BMC Bioinformatics*, 16(1), 111. <https://doi.org/10.1186/s12859-015-0530-3>
- Vergara, I. A., & Chen, N. (2010). Large synteny blocks revealed between *Caenorhabditis elegans* and *Caenorhabditis briggsae* genomes using OrthoCluster. *BMC Genomics*, 11(1), 516. <https://doi.org/10.1186/1471-2164-11-516>
- Villamar, R., Oviedo, B., Torres, Y., Zambrano, C., & Mehdi, S. (2022). Comparative computational genomic analysis between wild and domesticated species of cacao and coffee. *Journal of Pharmaceutical Negative Results*, 13(3). <https://doi.org/10.47750/pnr.2022.13.03.082>
- Wang, Y., & Xu, S. (2024). A high-quality genome assembly of the waterlily aphid *Rhopalosiphum nymphaeae*. *Scientific Data*, 11(1), 194. <https://doi.org/10.1038/s41597-024-03043-3>
- Wolfe, K. (2001). Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics*, 2, 333-341. <https://doi.org/10.1038/35072009>
- Zewdu, A., & Mastewal, G. (2022). Review on comparative genome mapping in crop improvement. *International Journal of Agricultural Science and Food Technology*, 8(3), 218-224. <https://doi.org/10.17352/2455-815X.000167>