



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



FACULTAD DE
CIENCIAS

UDELAR | fcien.edu.uy

UNIVERSIDAD DE LA REPÚBLICA

FACULTAD DE CIENCIAS

Tesis para optar al Título de: Licenciada en Ciencias Biológicas

**CARACTERIZACIÓN Y COMPARACIÓN DE LA DINÁMICA DE LAS
SECUENCIAS REPETIDAS EN LOS GENOMAS DE TRES
ESPECIES DE LA FAMILIA *MYRTACEAE***

Natali Mariel Baz Peña

Tutora: Magdalena Vaio

Co-tutora: Clara Pritsch

**Laboratorio de Evolución y Domesticación de las Plantas
Departamento de Biología Vegetal
Facultad de Agronomía – Universidad de la República
Montevideo, Uruguay**

2023

Dedicatoria y Agradecimientos

Agradecer en primer lugar, a la Facultad de Ciencias, casa de estudios con su excepcional cuerpo docente y por su valiosísimo aporte durante mi formación, al Laboratorio de Biología Vegetal de la Facultad de Agronomía, donde tuvo lugar mi pasantía y donde aprendí muchísimo.

Infinitas gracias a Magdalena Vaio, Clara Pritsch por haber confiado en mí y apoyado cada una de las etapas con esmero y paciencia, me llevo muchísimo para seguir andando. Al equipo del laboratorio de Evolución y Domesticación de las Plantas que han sido de gran inspiración y motivación para seguir adelante. La calidez y la vocación con la que Magdalena me recibió no la olvidaré y siempre estará presente.

Amigos y familia, grandes pilares donde encontré refugio, apoyo y sobre todo aliento para llegar y culminar este ciclo, personas muy importantes sin dudas.

Dedicado especialmente a la memoria de mi mamá y la de mi abuelo, a quienes agradezco la entereza de sus consejos.

Tabla de Contenido

DEDICATORIA Y AGRADECIMIENTOS	2
TABLA DE CONTENIDO	3
RESUMEN	5
1.INTRODUCCIÓN	6
1.1. Secuencias repetidas en tándem: ADN Satélite	7
1.2. Elementos transponibles	8
1.3. <i>Myrtaceae</i>	10
2.HIPÓTESIS	13
3. OBJETIVOS	13
3.1 Objetivo general	13
3.2 Objetivos específicos	13
4. MATERIALES Y MÉTODOS	14
4.1. Material vegetal y Obtención del ADN de <i>Psidium guajava</i>	14
4.2. Secuencias, secuenciación y metodología	15
4.3. Análisis comparativos de secuencias repetidas en <i>Myrtaceae</i>	17
4.4. Análisis filogenéticos de secuencias repetidas	18
5.RESULTADOS	18
5.1. Análisis de las secuencias repetidas de ADN en <i>Psidium guajava</i>	19
5.2. Análisis de las secuencias repetidas de <i>Feijoa sellowiana</i> (<i>syn Acca sellowiana</i>)	21
5.3. Estudio comparativo de las fracciones repetidas de los genomas de <i>F. sellowiana</i> , <i>P. guajava</i> y <i>Eucalyptus grandis</i>	23
5.4. Estudio filogenético de las secuencias <i>Ty1-Copia</i> y <i>Ty3-Gypsy</i>	28
5.5. Secuencias tipo satélite	29
6.DISCUSIÓN	30
6.1. Genomas pequeños pero ricos en secuencias repetidas	30
6.2. Importancia de la anotación del satelitoma	31
6.3. Relaciones evolutivas de linajes <i>Ty1-Copia</i> y <i>Ty3-Gypsy</i>	33
7.CONCLUSIONES	34
8.PERSPECTIVAS	35
9.BIBLIOGRAFÍA	36

RESUMEN

Las secuencias repetidas (repiteoma) juegan un rol relevante en la evolución del genoma de las plantas. El estudio de su composición y dinámica es de sumo interés para la comprensión de procesos evolutivos tales como mecanismos de especiación. El estudio del repitoma en la familia *Myrtaceae* está escasamente desarrollado. Esta familia incluye especies de gran importancia económica, ya sea por sus aptitudes frutales, como es el caso de *Psidium guajava* y *Feijoa sellowiana* (tribu Myrteae), como por ser recursos de biomasa y celulosa, como es el caso de *Eucalyptus grandis* (tribu Eucalypteae). Actualmente, la herramienta RepeatExplorer® facilita la caracterización del repitoma en especies sin genomas de referencia, utilizando secuenciación de baja cobertura y análisis de clustering. Este trabajo propone obtener información de composición y abundancia de elementos transponibles y secuencias repetidas en tándem para cada especie estudiada, sumado a un estudio comparativo entre dos las dos tribus antes mencionadas. Los linajes de secuencias repetidas obtenidos coinciden en las tres especies con diferencias en sus abundancias. Es destacable que, en este trabajo, se genera el primer reporte de cuatro familias de ADN satélite para las mirtáceas. El estudio contribuye al conocimiento de la dinámica de los elementos transponibles en dos tribus de la familia *Myrtaceae*, y a la comprensión de eventos evolutivos en sus genomas.

Palabras clave: Secuencias repetidas, genoma, especies nativas, *Psidium guajava*, *Feijoa sellowiana*, Elementos transponibles, ADN satélite, *Myrtaceae*

1. INTRODUCCIÓN

Entre los eucariotas existe una gran variabilidad en cuanto a los tamaños de los genomas de las especies, desde 2.3×10^{-3} Gb en el hongo *Encephalitozoon intestinalis* (Microsporidia) hasta 148,8 Gb de ADN en la planta *Paris japonica* (Melanthiaceae) (Hidalgo et al. 2017). Como se es muy conocido al día de hoy, la complejidad del organismo no está relacionada con el tamaño del genoma, Se llegó a pensar que la explicación estaba relacionada a lo complejo del organismo, es decir que a organismos más complejos, mayor tamaño del genoma, lo cual no es así, lo complejo del organismo no está estrictamente asociado al tamaño de su genoma, como lo describe el concepto de paradoja del valor C (Thomas CA Jr, 1971). En la actualidad se conoce que gran parte de la diferencia del tamaño de los genomas en eucariotas se debe a la existencia de secuencias repetidas de ADN que puede corresponder hasta un 90% del genoma (Schmidt y Heslop-Harrison, 1998). Al conjunto de las secuencias repetidas de un genoma se lo llama repitoma (Goubert et al., 2015; Jouffroy et al., 2016; Pita et al., 2017; Hannan, 2018) y debido a su dinámica varían en riqueza y abundancia entre las distintas especies (De Koning et al., 2011; Biscotti et al., 2015; Maumus y Quesneville, 2016). En los genomas de las plantas, además de la poliploidía, el tamaño del genoma se ha determinado que varía a causa del incremento o decremento en el número de copias de los distintos tipos de secuencias repetidas y se las ha relacionado, entre otras cosas, con estrategias evolutivas de especiación (Albach y Greilhuber, 2004). Si bien por mucho tiempo esta fracción del genoma fue considerada como ADN basura, recientemente se ha comenzado a determinar que cumplen diferentes funciones en los genomas, ya sea en la recombinación, estabilidad, modificación de la expresión de genes y modulación de la cromatina (Biscotti et al., 2015), así como la relación con la participación en respuesta a diversos tipos de estrés (Negi et al., 2016) y su gran importancia en el despliegue adaptacionista (Casacuberta, 2013). Lo que nos permite inferir que, en los genomas de las plantas, hay secuencias repetidas que son sumamente importantes para la evolución de las especies.

Las secuencias repetidas se clasifican de acuerdo a cómo disponen su organización en el genoma, en dispersas, las predominantes son los llamados elementos transponibles (TE), como transposones y retrotransposones, o en tándem, como es el caso del ADN ribosomal, los satélites de ADN y algunas familias de genes codificantes para proteínas (López-Flores y Garrido-Ramos 2012; Biscotti et al. 2015). Al conjunto de secuencias repetidas se le ha comenzado a llamar repitoma y a su vez al repertorio de secuencias tipo satélite en un genoma se le llama satelitoma (Ruiz-Ruano et al 2016; Kim et al. 2014).

1.1. Secuencias repetidas en tándem: ADN satélite

Las secuencias repetidas en tándem son un gran número unidades de repetición que van de 50 a 1000 pb y que se organizan de manera tal que quedan ampliamente distribuidas en el genoma (Sharma & Raina, 2005). Las secuencias de ADN satélite (ADNsat), son muy útiles en estudios filogenéticos (Gillespie et al., 1982; Widegren et al., 1985). Están dispuestas en tándem en el mismo sentido, (de cabeza a la cola) y generalmente no son codificantes (Garrido-Ramos 2017). Los ADNsat tienen una estructura básica (monómero) de largo variable que se va repitiendo de miles a millones de veces en el genoma dispuestas en series de gran cantidad de monómeros. Los monómeros pueden variar en largo, composición, complejidad y abundancia, clasificándose en diferentes familias según estas características. Las familias pueden estar compartidas entre especies, familias, géneros o ser específicas, ya que es una de las fracciones más dinámicas del genoma (Garrido-Ramos 2015, 2017). Las secuencias satélites son el componente principal de la heterocromatina y se localizan principalmente en las regiones pericentroméricas y subteloméricas de los cromosomas (López-Flores y Garrido-Ramos 2012).

1.2. Elementos transponibles

Los elementos transponibles de ADN son secuencias repetidas dispersas con la capacidad de moverse en los genomas mediante diferentes mecanismos de transposición. En plantas estas secuencias pueden llegar a representar un gran porcentaje del tamaño del genoma, por ejemplo, en maíz llega al 85% (Schnable et al., 2009). Se los clasifica de acuerdo a la forma en que se transponen en los genomas, en dos clases principales (Finnegan, 1992). Los elementos transponibles de clase I, son aquellos que utilizan un intermediario de ARN y mecanismo de transcripción reversa para la transposición, por lo que se les denomina retrotransposones y los elementos de clase II, son aquellos que se transponen de forma directa, sin intermediario de ARN, y son llamados transposones de ADN (Wicker et al., 2007). Los elementos transponibles se han acumulado en los genomas durante la evolución y los moldean mediante sus dinámicas de pérdida y amplificación, siendo los de gran número de copias en el genoma de plantas las secuencias de clase I con respecto a las de clase II que suelen estar en baja copia (Casacuberta & Santiago, 2003).

Las secuencias repetidas de clase I o retrotransposones, están clasificadas en tres órdenes, donde encontramos los retrotransposones tipo LTR (LTR-RT), principalmente *Ty1-Copia* y *Ty3-Gypsy*, los retrotransposones no-LTR (Short Interspersed elements -SINES, Long Interspersed Elements -LINES y Penelope-like elements -PLE's) además de Dictyostelium repetitive sequence -DIRS (Casacuberta, 2013). En estos elementos la diferencia radica en la conformación de su estructura y en el ciclo de vida (Kejnovski, 2015). En plantas los LTR-RT se encuentran en elevado número de copias y su transposición sería la principal causa del aumento del tamaño del genoma (Wicker et al 2018). La estructura de los retrotransposones LTR recuerda a la de los retrovirus (Grandbastien, 2015; Zhang & Gao, 2017), aunque no contienen la porción del gen de envoltura o ésta puede encontrarse como una secuencia no funcional (Eickbush & Jamburuthugoda, 2008; Godinho et al., 2012).

Los elementos de clase II, también conocidos como transposones de ADN, exhiben un mecanismo de transposición que se asemeja a un proceso de "corta y pega", operando sin la intermediación de ARN. La clasificación dentro de esta clase se realiza considerando la presencia o ausencia de TIRs

(repeticiones terminales invertidas), que actúan como sitios de escisión y reintegración reconocidos por una transposasa codificada por el propio elemento (Finnegan, 1989).

Tanto los elementos de clase I como de clase II pueden ser además clasificados en superfamilias o familias de acuerdo a los dominios proteicos que los conforman y su estructura. En los RT-LTR se reconocen dos grandes superfamilias *Ty1-Copia* y *Ty3-Gypsy* que se diferencian entre sí en el orden de sus dominios proteicos. A su vez, pueden ser identificados dentro de cada superfamilia diferentes linajes. Las secuencias de las superfamilias *Ty1-Copia* y *Ty3-Gypsy* pueden ser utilizadas en análisis filogenéticos ya que las mismas cuentan con secuencias proteicas muy conservadas (Neumann et al., 2019).

Los genomas de las plantas son muy dinámicos, esto se ha asociado, entre otros factores, con que sufren una gran exposición a estrés tanto biótico como abiótico por tratarse de organismos con forma de vida sésil, por lo que es necesario que se desarrollen mecanismos capaces de mitigar su efecto sobre el individuo (Casacuberta & González, 2013). Los TEs son activados por elicitores (inductores de procesos de defensa en plantas) internos de manera directa, o externos y de forma indirecta por el efecto que tienen sobre ciertos mecanismos celulares, además, se demostró que éstos tienen la capacidad de vencer el silenciamiento que se les aplica por parte de la célula a la hora de que las plantas intentan mitigar efectos del estrés sufrido. En plantas, se sugirió que los LTR-RT tienen la capacidad de escapar al silenciamiento por parte del huésped con factores anti-silenciadores (Vicent & Casacuberta, 2017). Para evitar un impacto negativo sobre el genoma por parte de los TEs, las plantas aplican diversos mecanismos de control de daños (Giordani et al. , 2016; Zuccolo et al., 2015). Sin embargo, se sabe que los LTR-RT tienen una amplia variabilidad de estructuras, funciones y localización en los genomas.

Las secuencias repetidas también presentan la característica de que las diferentes clases tienen una preferencia de localización en el genoma, por lo que presentan patrones de distribución característicos de la clase en los cromosomas (Li et al., 2017). Un ejemplo son las secuencias de clase I *Ty1-Copia* y *Ty3-Gypsy*, donde las primeras se encuentran distribuidas a lo largo del cromosoma, en la región pericentromérica (Heslop-Harrison et al., 1997), mientras que las segundas, por el contrario, se

encuentran en la región centromérica también (Nagaki et al., 2005). Algunos TEs muestran preferencias por los lugares de inserción por ejemplo Athila, un linaje dentro de los *Ty3-Gypsy*, tiene preferencia por las regiones de heterocromatina (Rafaqat et al 2021). Mientras que algunos linajes dentro de los *Ty1-Copia* se ha visto que están asociados a regiones de eucromatina ricas en genes (Galindo et al 2016).

1.3. Myrtaceae

La familia Myrtaceae Juss. está compuesta por alrededor de 5700 especies organizadas en 19 tribus y 132 géneros (Govaerts et al., 2008). Los frutos de la familia se clasifican en carnosos o capsulares, y basados en estas características las especies estaban divididas en las subfamilias Myrtoideae y Leptospermideae, respectivamente. Sin embargo, los últimos estudios moleculares mostraron que los frutos carnosos evolucionaron independientemente en varios momentos durante la evolución (Thornhill et al. 2015). En la actualidad las especies de la familia se agrupan en dos subfamilias Psiloxylodeae y Myrtoideae, (Wilson et al. 2005). Es una familia de distribución pantropical aunque la mayoría de sus especies se encuentran en las regiones tropicales y subtropicales del hemisferio Sur (Thornhill et al 2015). Las mirtáceas se caracterizan por ser muy estables en su número cromosómico siendo la mayoría diploides con un número de cromosomas básico de 11 (Vijayakumar & Subramanian, 1985, da Costa y Forni-Martins, 2007, Coser et al., 2012). En los géneros con frutos capsulares se ha detectado series disploides ($x = 5, 6, 7, 8, 9, 10, 11, 12$ and 14) mientras que los de frutos carnosos se conocen varios citotipos producto de la poliploidía (ver da Costa et al. 2008). La mayoría de las especies presenta también genomas pequeños sin embargo existe variación en el tamaño desde 234 Mpb (Valor $C= 0,239pg$) en *Myrciaria glazioviana* a 1,1 Gb (Valor $C= 1,1 pg$) en *Melaleuca leucadendra* (Myburg et al 2014, Grattapaglia et al 2012, Heayley et al 2021) representando una variación de 5 veces. Algunos estudios sugieren que el grupo de frutos carnosos tendrían genomas más pequeños a los presentes en las especies de frutos capsulares (da Costa et al., 2008). Dentro de esta familia se encuentran especies de importancia económica, de fruto carnoso como *Psidium guajava* (guayabo brasileiro), *Feijoa*

sellowiana (syn *Acca sellowiana*) (guayabo del país) ambas de la subfamilia Myrtoideae y *Eucalyptus grandis*, de fruto capsular, de la subfamilia Psiloxylloideae.

En las regiones neotropicales la familia Myrtaceae está mayormente representada por la tribu Myrteae (Subfamilia Myrtoideae), de frutos carnosos, que incluye 51 géneros y cerca de 2500 especies lo que representa aproximadamente la mitad de las especies conocidas para la familia (Lucas et al. 2019; Souza Neto et al 2022). Todas las especies de Myrtaceae de América del Sur, menos *Metrosideros stipularis*, pertenecen a *Myrtea* (Wilson et al 2005). Dentro de esta tribu se encuentran especies de gran importancia económica, ya sea por sus frutos carnosos de interés como comestibles, en la agroindustria y como medicinales (Sardi et al 2017) como *Psidium guajava* (guayaba) o *Feijoa sellowiana* (syn *Acca sellowiana*) (guayabo) o por presentar especies con alto poder invasor como *Psidium cattleianum* (Arazá).

Psidium guajava pertenece a un gran género monofilético que es muy importante económicamente a nivel mundial, ya que es muy cultivado entre los trópicos por sus frutos comestibles. Se trata de plantas neotropicales introducidas en Europa junto a la papa y el tabaco, con gran importancia en cuidados de la salud ya que aporta grandes valores de antioxidantes y valores “anti diabetes”, por lo cual se lo llama “hierba multipropósito”. Fue descrita por Linneo en base a materiales de Asia, pero algunos investigadores consideran que la misma sería originaria de América y su distribución original abarcaba desde México a América del Sur (revisado en Feng et al 2021). Recientemente se han compilado las diferentes evidencias que describen el proceso de domesticación en *Psidium* en los últimos 7000 años en América (Arévalo-Marín et al. 2021). Dentro de las mirtáceas es la especie de frutos carnosos más importante económicamente, probablemente por ser económica en sus requerimientos de cultivo, adaptación a distintas condiciones y su riqueza en compuestos vitamínicos (Feng et al. 2021). Las frutas son consumidas frescas, o secas, como mermeladas o aditivos antioxidantes. Con respecto a su genoma, todos los materiales analizados en la especie son diploides con $2n=22$ cromosomas, a pesar de que para el género se conocen especies con distintos niveles de ploidía. Los estudios de contenido de ADN en materiales diploides han mostrado diferencias entre 0,5 a 0,95 pg (valor 2C) pero todos ellos están de acuerdo con un genoma pequeño (Das Graças De Souza et al 2015; Marques et al 2016; da

Costa et al., 2008; da Costa y Forni-Martins 2006, 2007). Recientemente, fue publicado para esta especie un genoma de referencia (Feng et al., 2021). Los resultados de este trabajo muestran 25601 genes y un 43,55% de secuencias repetidas. Las secuencias repetidas fueron anotadas y clasificadas únicamente en grandes grupos, y no se dispone de una caracterización de los linajes de cada familia de repetidos. El mayor porcentaje en la fracción repetida está compuesto por retrotransposones de tipo LTR (Feng et al. 2021).

Feijoa sellowiana (*syn Acca sellowiana*) comúnmente conocida como guayabo del país, es una especie leñosa de fruto carnoso con gran potencial agronómico. Es una especie diploide ($2n=2x=22$), alógama, originaria de Uruguay y Brasil, que presenta un genoma muy pequeño, de unas 245 Mb (da Costa et al., 2008). En Uruguay la especie se encuentra en quebradas y montes ribereños de Rivera, Tacuarembó, Cerro Largo, Artigas, Salto y Treinta y Tres (Brussa y Grela 2007). Esta especie es una gran fuente de nutrientes, vitaminas, minerales y gran aporte de antioxidantes, por lo que agronómicamente reviste muchísima importancia (Seymour et al., 1993). Si bien en Uruguay no hay plantaciones de la especie, es utilizada comercialmente en Nueva Zelanda, Colombia, Estados Unidos e Israel. En el país ha sido modelo de estudio en varios proyectos de investigación, para caracterizar entre otras cosas su diversidad genética y morfológica. Estos estudios permitieron publicar el primer mapa genético por parte de Quezada et al. (2014, 2021) y se está elaborando un genoma de referencia a escala cromosómica (Pritsch et al., datos sin publicar). Todos los materiales analizados para la especie son diploides con 22 cromosomas (ver en da Costa et al 2008) incluyendo materiales colectados en Uruguay (Vázquez 2014). El contenido de ADN, medido por citometría de flujo, publicado por da Costa 2008, para materiales de Brasil tenían $2C = 0,5$ pg mientras que las mediciones realizadas para plantas de Uruguay mostraron un contenido de ADN mayor de $2C = 0,74$ pg (Pritsch et al., 2010).

2.HIPÓTESIS

La diferencia de tamaño de los genomas entre las especies de fruto capsular *Eucalyptus grandis* y las de fruto carnosos, *Psidium guajava* (Pg) y *Feijoa sellowiana* (Fs) es atribuible a la diferente representación de secuencias repetidas de ADN.

Existe mayor riqueza de secuencias repetidas, del tipo RT-LTR entre las especies *Psidium guajava* y *Feijoa sellowiana*, por ser estas las más abundantes en los genomas de plantas.

Las especies *Psidium guajava* y *Feijoa sellowiana* (syn *Acca sellowiana*) más cercanas filogenéticamente, comparten más tipos de secuencias repetidas entre ellas que con *Eucalyptus grandis* (Eg).

3.OBJETIVOS

3.1. Objetivo general:

Determinar la relación entre el contenido de ADN y número de cromosomas con respecto al porcentaje y composición de la fracción repetida del genoma en Myrtaceae.

3.2. Objetivos específicos:

- Describir y caracterizar la composición y abundancia del repitoma de la especie *Feijoa sellowiana*
- Describir y caracterizar la composición y abundancia del repitoma de la especie *Psidium guajava*

- Estudiar de manera comparativa los repitomas de *Feijoa sellowiana* y *Psidium guajava* utilizando el repitoma de *Eucaliptus grandis* como grupo externo, de la familia Myrtaceae
- Determinar las relaciones evolutivas entre los miembros de las superfamilias Ty3-Gypsy y de Copia (retrotransposones de tipo LTR) de *F. sellowiana*, *P. guajava* y *E. grandis* mediante análisis filogenéticos del dominio Retrotranscriptasa (RT).

4.MATERIALES Y MÉTODOS

4.1. Material vegetal y Obtención del ADN de *Psidium guajava*

El ADN se obtuvo a partir de hojas tiernas, jóvenes recolectadas de un árbol adulto de *Psidium guajava* que se mantiene en el predio de la Facultad de Agronomía, Sayago, Montevideo. Para la extracción del ADN se utilizó el kit de extracción Nucleospin Plant™ II (Macherey-Nagel™). En el laboratorio, se pesaron 100 mg de hojas en la balanza de precisión, se maceraron con nitrógeno líquido en un mortero para romper la pared celular pasos previos a la utilización del kit. Las hojas fueron tratadas con SDS para lisis celular, desnaturalización de proteínas y desnaturalización del ADN. Luego fueron utilizados los tampones de lisis PL1 y PL2 cuya utilización depende de los metabolitos secundarios presentes en la muestra siguiendo el protocolo proporcionado por el kit de extracción ([https://www.takarabio.com/assets/documents/User%20Manual/NucleoSpin%20Plant%20II%20Genomic%20DNA%20Purification%20User%20Manual%20\(PT5028-1\)_Rev_09.pdf](https://www.takarabio.com/assets/documents/User%20Manual/NucleoSpin%20Plant%20II%20Genomic%20DNA%20Purification%20User%20Manual%20(PT5028-1)_Rev_09.pdf)).

Se realizaron pruebas para determinar el rendimiento y resultado de la extracción con cada uno de ellos y se seleccionó el tampón PL2 basado en los resultados (concentración, integridad y relaciones A260/A280 y A260/230). Se obtuvieron 6 tubos con muestras de ADN tratadas con el mismo kit. Una vez obtenido el ADN, se verificó la calidad y concentración de las 6 muestras por Nanodrop y mediante

corrida de 5µl de la muestra de ADN en gel de agarosa al 1%. A partir de los resultados obtenidos se seleccionaron 3 de las 6 muestras, que fueron las que presentaron mejores resultados de calidad.

4.2. Secuencias, secuenciación y metodología

Para la secuenciación de baja cobertura de *Psidium guajava*, se utilizó el servicio de secuenciación de BGI (Beijing Genomics Institute), Hong Kong, China. Para el secuenciado se utilizó la plataforma HiSeq Illumina 2000 para obtener secuencias cortas de ADN con la característica de ser un secuenciado paired end de 100 pb ,recomendado para obtener mejor calidad en los análisis de secuencias repetidas con una cobertura 1x (2c= 463,8 Mb) . Las secuencias de *Feijoa sellowiana* fueron obtenidas en un proyecto previo utilizando el mismo protocolo de extracción de ADN, servicio de secuenciación, plataforma y nivel de cobertura. Las secuencias de *E. grandis* fueron obtenidas de la base de datos SRA/NCBI <https://www.ncbi.nlm.nih.gov/sra/?term=eucalyptus+grandis> y las secuencias utilizadas fueron las accesiones SRX7049890 y SRX7049892.

Para las tres especies, una vez obtenidas las secuencias, las mismas se pre procesaron en el Pipeline RepeatExplorer® en el ambiente Galaxy (Novak et al 2021; Novak et al, 2013, <https://repeatexplorer-elixir.cerit-sc.cz/galaxy/>) siguiendo los pasos adecuados para llegar a filtrar las secuencias para obtener las de mejor calidad. Los pasos que se siguieron fueron: cambio del formato original a formato fastq Sanger con la herramienta edición de la plataforma. Luego del cambio de formato se realizó un control de calidad, mediante la herramienta FASTQC chequeando calidad (basado en los valores Q), presencia de adaptadores, contenido GC y presencia de bases sin identificar (N). Luego de esto se filtró por calidad, mediante la herramienta Filter by quality con los parámetros de Phred mayores a 10 en por lo menos 95 % del largo de la secuencia (valor de corte y en el corte respectivamente, CUT VALUE PERCENT y CUT OFF VALUE). Una vez finalizados estos pasos, los archivos resultado del filtrado por calidad se lo transformó a formato .FASTA (herramienta FASTq to FASTA). Los dos archivos resultado de este paso, uno para las secuencias forward (R1) y otro para las secuencias reverse (R2) fueron unidos por medio de la herramienta INTERLACE. De este paso se

generan dos archivos, uno con las secuencias pareadas y otro con las que quedaron sin parear. Las secuencias pareadas fueron muestreadas al azar (read sampling) de manera de tener un número correspondiente a 1x del genoma de cada especie. Para la identificación de las secuencias repetidas se utilizó una metodología basada en agrupamientos, y gráficos (RepeatExplorer®, Novák et al 2010, 2013, 2021) que permite la caracterización e identificación *de novo* de secuencias repetidas y su proporción en un genoma sin necesidad de contar con un genoma de referencia. Para esto se utilizó la herramienta clustering de la plataforma utilizando el pipeline RepeatExplorer® 2. Los clusters resultantes fueron clasificados y anotados automáticamente por búsquedas de similitud contra la base de dominios conservados de elementos transponibles REXdb (Neumann et al 2019) y una base de datos de secuencias de plantas conteniendo los genes ribosomales, secuencias teloméricas, cloroplastos y mitocondrias. Este Pipeline utiliza también la base de datos del GenBank, por lo que por medio de algoritmos de búsqueda y coincidencia contra todas las bases de datos anota automáticamente, clasificando los clusters formados. Los clusters que no fueron anotados automáticamente por el pipeline fueron anotados de forma manual. Para ello se usó el NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>) para buscar similitud en las secuencias no clasificadas anteriormente por el Repeat Explorer® y así ahondar en la clasificación y anotación de las mismas, Una de las herramientas utilizadas para la anotación manual fue BLAST (Basic Local Alignment Search Tool) la cual encuentra regiones de similitud entre secuencias, compara secuencias de nucleótidos o proteínas y mediante un cálculo estadístico muestra la significancia con las secuencias en la base de datos NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>). En el caso de este estudio, se utilizó la herramienta BLASTn, que compara secuencias de nucleótidos, es decir, se realizaron búsquedas de similitud, de los nucleótidos pertenecientes a los clusters con los de la base de datos para poder clasificarlos. Los que no pudieron ser anotados de esa forma, se clasificaron utilizando el BLASTx, que lo que hace es traducir los nucleótidos (de ser posible) y buscar en la base de datos potenciales proteínas codificadas para esos nucleótidos (Figura 1). En un paso previo a finalizar el procesamiento, se excluyeron los clusters que incluyen a las secuencias pertenecientes al ADN mitocondrial y al ADN plastidial con el objetivo de determinar solamente los porcentajes de secuencias

repetidas pertenecientes al ADN nuclear. Para esto se consideró el número de secuencias total analizadas por el pipeline y el número de secuencias para cada cluster. Para la búsqueda de satélites en los dos guayabos, se usó la herramienta TAREAN (Tandem Repeat Analyzer) disponible en el ámbito del Repeat Explorer® (Novak et al., 2017), la cual realiza una búsqueda más exhaustiva de secuencias repetidas en tándem comparado con el análisis de clustering de RepeatExplorer®. Los monómeros detectados se alinearon para determinar su identidad y el número de familias presentes.

4.3. Análisis comparativos de secuencias repetidas en Myrtaceae

Para las especies *P. guajava*, *F. sellowiana* y *E. grandis* se realizaron análisis individuales y comparativos entre ellas. En este último caso se implementó el análisis a partir de un pool de lecturas de *P. guajava*, *F. sellowiana* y *E. grandis*, éste último incluido como grupo externo. Para el análisis comparativo, se tomaron de las tres especies un número significativo de secuencias según el tamaño de genoma, equivalente a 1x del tamaño de cada genoma, es decir, se analizó el mismo porcentaje de secuencias de acuerdo al tamaño de los genomas analizados para mantener la representatividad. Además del muestreo de secuencias a cada conjunto de datos se le agregó un prefijo para identificar la especie y luego los tres fueron unidos en un archivo único (herramienta Concatenate) para ser sometidos al análisis de clustering. Para el análisis de clustering comparativo se seleccionó la opción avanzada *perform comparative analyses*. La anotación de cada cluster y respectiva cuantificación de las secuencias repetidas (en base a número de reads/grupo) fueron realizadas de la misma forma que se mencionó con anterioridad. Los gráficos de porcentaje y relación entre las diferentes secuencias repetidas fueron realizados en el programa Excel.

4.4. Análisis filogenéticos de secuencias repetidas

Para los análisis filogenéticos de los elementos transponibles tipo retrotransposon se utilizó la herramienta DANTE (Domain based Annotation of Transposable Elements) en el ámbito del RepeatExplorer®, la cual permite anotar los elementos transponibles de acuerdo a la base de datos RexDB (Neumann et al 2019). El análisis de DANTE se realizó para recuperar las secuencias de dominio RT de los retroelementos de tipo *Ty3-Gypsy* y *Ty1-Copia* mediante la búsqueda del dominio RT (retrotranscriptasa) en los contigs generados en el RepeatExplorer® en cada análisis individual. Además, se realizó una corrida individual en *E. grandis* de manera de generar el archivo de los contigs necesario para este estudio. El resultado de utilizar esta herramienta es un archivo fasta en el que se encuentran las secuencias de aminoácidos correspondientes al dominio RT. Las secuencias RT para los diferentes linajes fueron alineadas mediante MAFFT en el entorno Geneious. Luego de este paso los alineamientos fueron importados al programa Mega X (<https://www.megasoftware.net/>) para construir una red de Neighbor Joining y determinar el soporte de los clados mediante bootstrap.

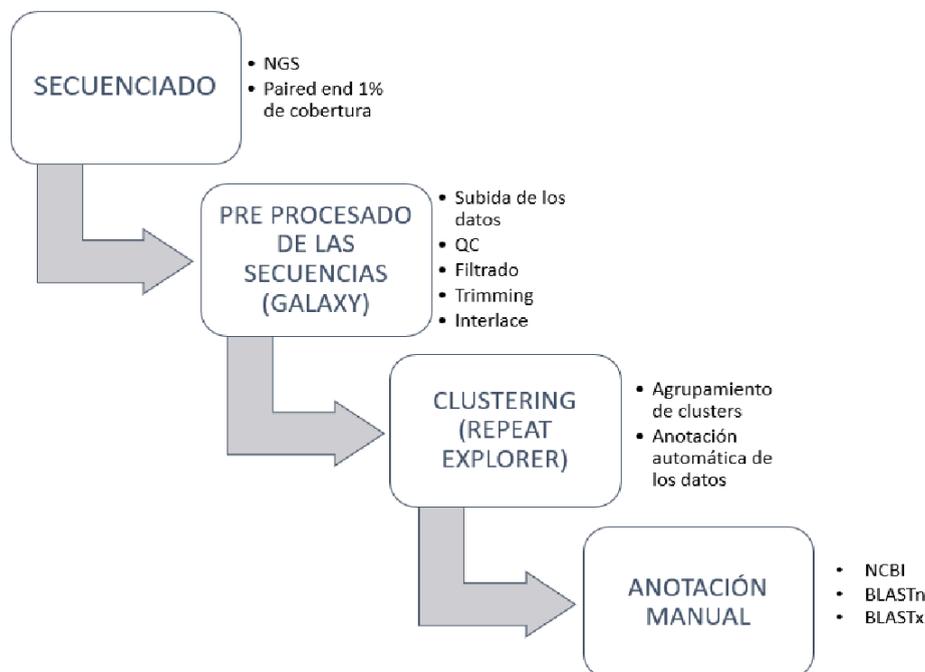


Figura 1. Etapas para el procesamiento de las secuencias de ADN en el estudio de secuencias repetidas en la herramienta Repeat Explorer. Modificado de Novak et al., 2013.

5. RESULTADOS:

5.1. Análisis de las secuencias repetidas de ADN en *Psidium guajava*

Para el análisis de las secuencias repetidas que forman parte del genoma de *Psidium guajava* se realizó un muestreo y selección por calidad de las secuencias originales resultado de la secuenciación. Del total de secuencias muestreadas fueron procesadas en el RepeatExplorer® 2.457.450 que equivalen a una cobertura del 0,83x del genoma de la especie. Las secuencias repetidas fueron agrupadas en 239 clusters y los análisis de la fracción de secuencias repetidas (repetoma) en el genoma de *P. guajava* mostraron que esta equivale a aproximadamente el 43% (Figura 2).

Del total de las secuencias repetidas, un poco más del 80% de las mismas está representado por elementos transponibles clasificados como Retrotransposones tipo LTR Clase I. En el caso de las secuencias en tándem tipo satélite, éstas fueron las menos abundantes con un 0.13% del total. El porcentaje de secuencias repetidas no clasificadas fue de casi 3% (2,84%). Dentro de los Retrotransposones los más abundantes son los del tipo *Ty3-Gypsy*, con casi un 10%, y los *Ty1-Copia* representaron casi un 7% en referencia a su abundancia en el genoma. La familia más representativa dentro de los *Ty3-Gypsy* fue *Chromovirus* con alrededor del 6%, considerando todos los linajes de este grupo: En los *Copia* el linaje más abundante fue *Ikeros* con casi un 2%. Un gran porcentaje (19%) dentro de los retrotransposones tipo LTR no pudieron ser clasificados, que pueden representar variantes no autónomas de los mismos (Figuras 3 y 6). Los repetidos menos frecuentes en el genoma (menos de 4 %) se componen de transposones (1%), LINE (0.4 %), ADNr (2 %).

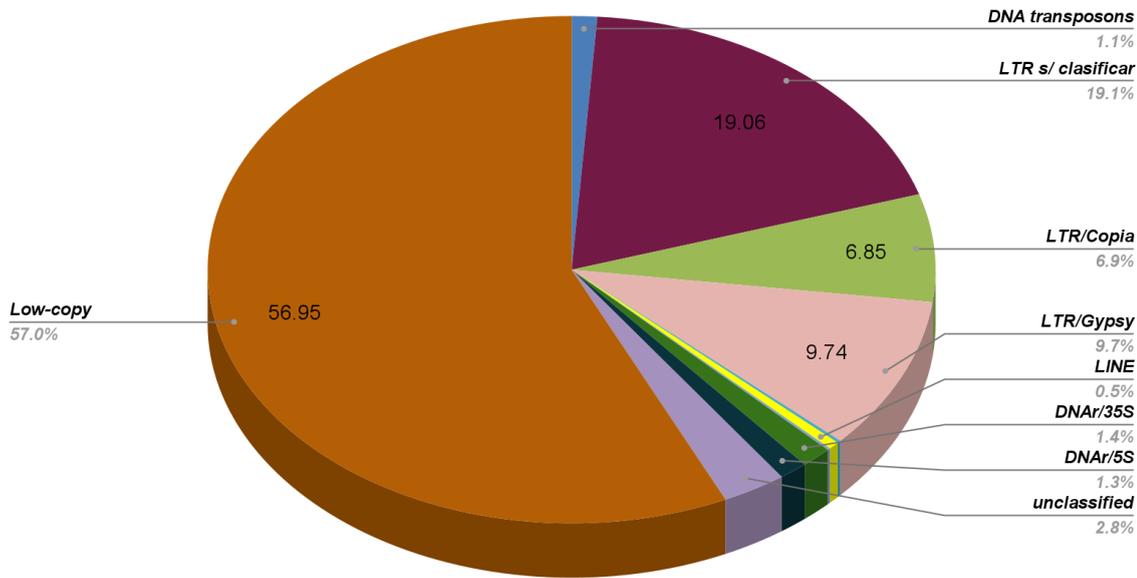


Fig. 2. Composición de las secuencias repetidas en el genoma de *Psidium guajava* mostrando las principales clases y tipos de repetidos y su proporción. La fracción repetida representa un 43% del genoma de esta especie.

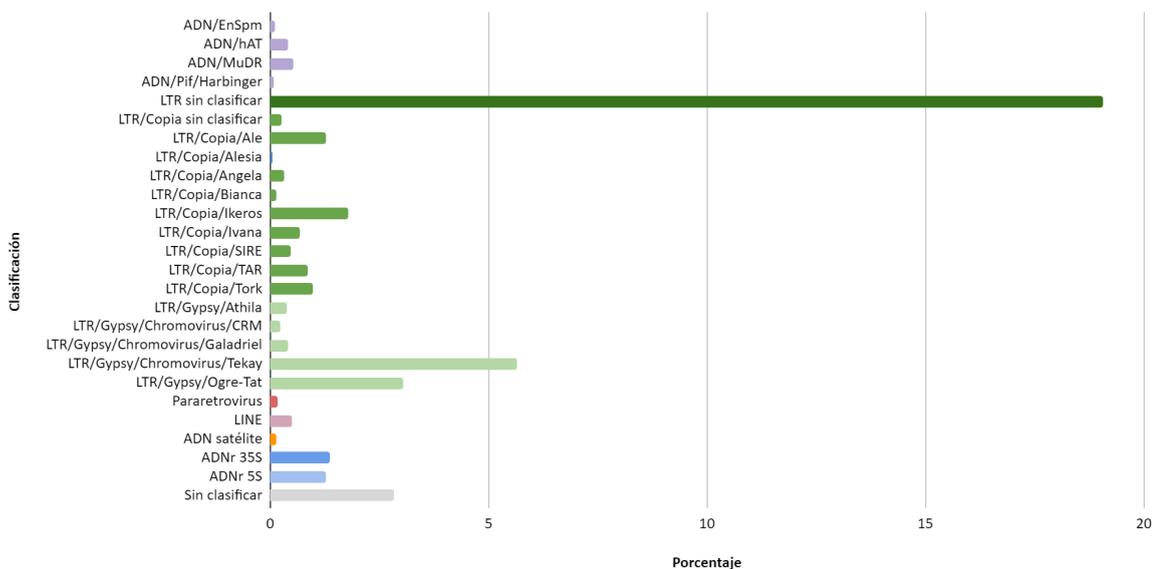


Figura 3. Clasificación de los elementos de ADN repetido en el genoma de *Psidium guajava* y la proporción de cada uno de ellos. Se incluyen transposones (ADN, en Violeta); retrotransposones (tipo LTR, de las superfamilias *Ty1-Copia* y *Ty3-Gypsy*)

otros retrotransposones (pararetrovirus en rojo y LINE en lila); los repetidos en tándem incluyen al ADN satélite (naranja) y genes ADNr. En gris se indica los elementos sin clasificar. En los elementos tipo LTR el nombre indica Orden/Superfamilia/Familia. En los transposones el nombre indica Clase/Superfamilia.

5.2. Análisis de las secuencias repetidas de *Feijoa sellowiana* (syn *Acca sellowiana*)

Para el análisis de las secuencias repetidas que forman parte del genoma de *Feijoa sellowiana* se realizó un muestreo y selección por calidad, al igual que con *P. guajava*, de las secuencias originales. El número de secuencias que fueron procesadas en el RepeatExplorer®, es 2.967.313, que equivalen a una cobertura 1,17x del genoma. Las secuencias repetidas fueron agrupadas en 320 clusters. De este análisis se desprende que un poco más del 34% del genoma se corresponde a la fracción repetida y que las más abundantes son los retrotransposones del tipo LTR. Dentro de estos, los más abundantes fueron tipo *Ty1-Copia* con más del 10% del genoma, mientras que los *Ty3-Gypsy* representaron 6%. Los linajes o familias Ale, Angela, Ikeros, SIRE y Tork representaron todos más del 1%, siendo los dos últimos los más abundantes dentro de *Copia* (2,3% y 2,1% respectivamente). Entre los *Ty3-Gypsy*, los más abundantes fueron los Chromovirus/Tekay y Ogre-Tat con 2,4%. Las secuencias de ADN satélite se encontraron en baja proporción (0,52%). Por otro lado, casi un 4% del genoma está representado por retrotransposones tipo LTR que no lograron ser clasificados y un poco más del 7,5% del genoma son secuencias repetidas que no pudieron ser clasificadas. La abundancia de elementos LINE fue 0,46% y de transposones fue 1,5%.

En cuanto a repetidos en tándem, globalmente las secuencias de ADN satélite representaron un 0,52% del genoma, mientras que las secuencias de ADNr correspondieron a cerca de 2,3%.

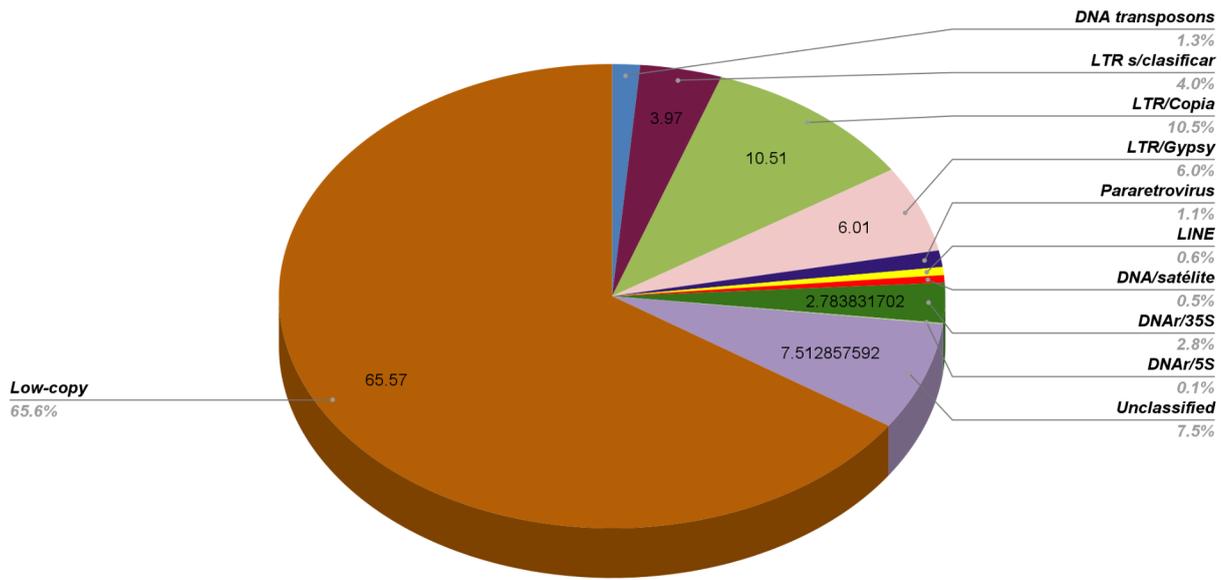


Figura 4. Composición de las secuencias repetidas en el genoma de *Feijoa sellowiana* (syn *Acca sellowiana*), mostrando las clases y proporciones de secuencias repetidas. El porcentaje de secuencias repetidas corresponde al 34% del genoma de la especie.

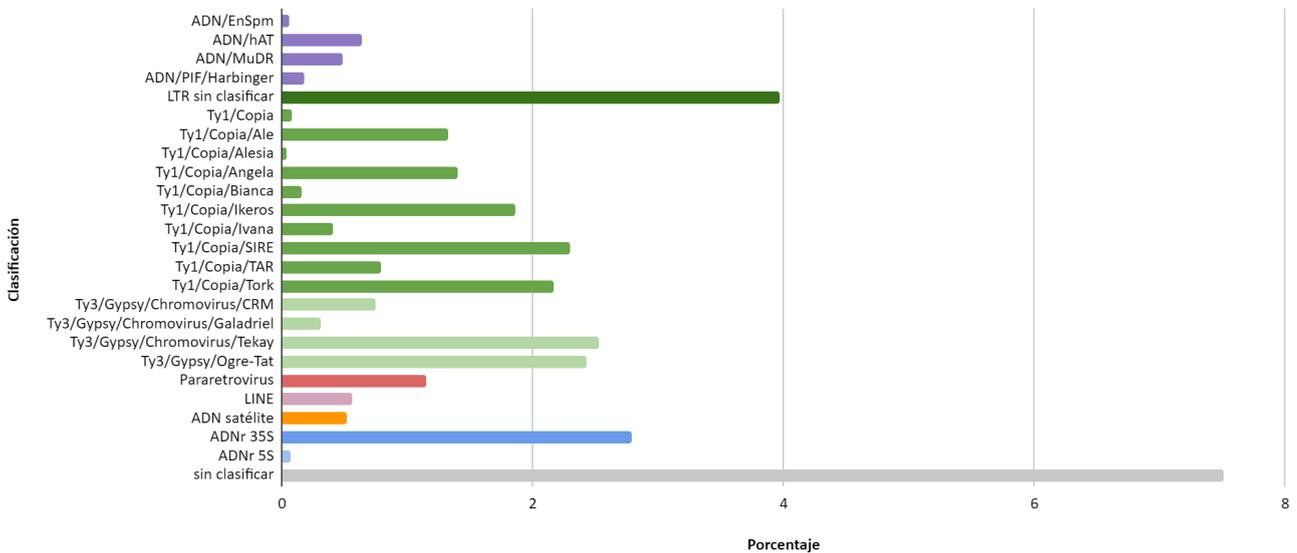


Figura 5. Clasificación de los elementos de ADN repetido en el genoma de *Feijoa sellowiana* y la proporción de cada uno de ellos. Se incluyen transposones (ADN, en violeta); retrotransposones (tipo LTR, de las superfamilias *Ty1-Copia* y *Gypsy*) otros retrotransposones (pararetrovirus en rojo y LINE en lila); los repetidos en tándem incluyen al ADN satélite (naranja) y genes

ADNr. En gris se indica los elementos sin clasificar. En los elementos tipo LTR el nombre indica Orden/Superfamilia/Familia. En los transposones el nombre indica Clase/Superfamilia.

5.3. Estudio comparativo de las fracciones repetidas de los genomas de *F. sellowiana*, *P. guajava* y *Eucalyptus grandis*

Para el análisis se usaron un total de 38000000 secuencias del genoma de *F. sellowiana*, 4.410.000 secuencias del genoma de *P. guajava* y 6.125.000 secuencias del genoma de *E. grandis*, representando el 1x del genoma. Este análisis, que simultáneamente analiza a las tres especies, permite visualizar cuales son los clusters compartidos entre ellas, además de cuantificar las abundancias relativas de cada uno (Tabla 2). En la figura 6 se representan los datos acumulados de las secuencias repetidas por especie. Dentro del orden LTR-retrotransposon, *P. guajava* es más rico en secuencias de la superfamilia *Ty3-Gypsy* (19,7%) y *E. grandis* en superfamilia *Ty1-Copia* (13,5 %), mientras que para *F. sellowiana* *Ty1-Copia* fue levemente superior a *Ty3-Gypsy* con 7,8% y 5,2%, respectivamente. Las familias y superfamilias de elementos son compartidos entre las tres especies, excepto por los transposones de la superfamilia CACTA que fueron detectadas exclusivamente en las especies de frutos carnosos y no fueron detectados en *E. grandis* por lo menos en 0,01% del genoma (Tabla 2). Entre los diferentes tipos de repetidos se observaron variaciones en las abundancias en los linajes (o familias) de cada una de ellas. En términos generales, se conservan los patrones de distribución por linajes observados en los análisis individuales por especie, descritos en la Figs. 3 y 5, para *P. guajava* y *F. sellowiana*, respectivamente. De esta manera ambas especies coinciden en tener como linajes más importantes de *Ty3-Gypsy* a las familias Chromovirus-Tekay y Ogre-Tat. En el caso de *Ty1-Copia*, Ikeros y Ale son los más representados en *P. guajava*, mientras que en *F. sellowiana* a estas dos categorías se agregan las familias SIRE (principal). Tork, y Angela. En *E. grandis*, el linaje SIRE de *Ty1-Copia* fue notoriamente más abundante. En cuanto a los transposones de tipo TIR, las superfamilias Mutator y hAT coinciden en ser las más abundantes para las especies de frutos carnosos.

La abundancia de ADNr 35S en *Eg* fue marcadamente inferior a las otras especies . Por otra parte, el porcentaje de secuencias repetidas sin clasificar fueron similares entre las tres especies.

Es de destacar que los análisis comparativos en el RepeatExplorer® permiten, además, visualizar en cada cluster la distribución de las especies (datos no mostrados). En este trabajo se detectaron clusters compuestos por solo una especie y otros con más de una especie, indicando que algunos clusters representan variantes de familias de repetidos que son especie-específicos, mientras que otros clusters representan variantes de familias compartidos por varias especies.

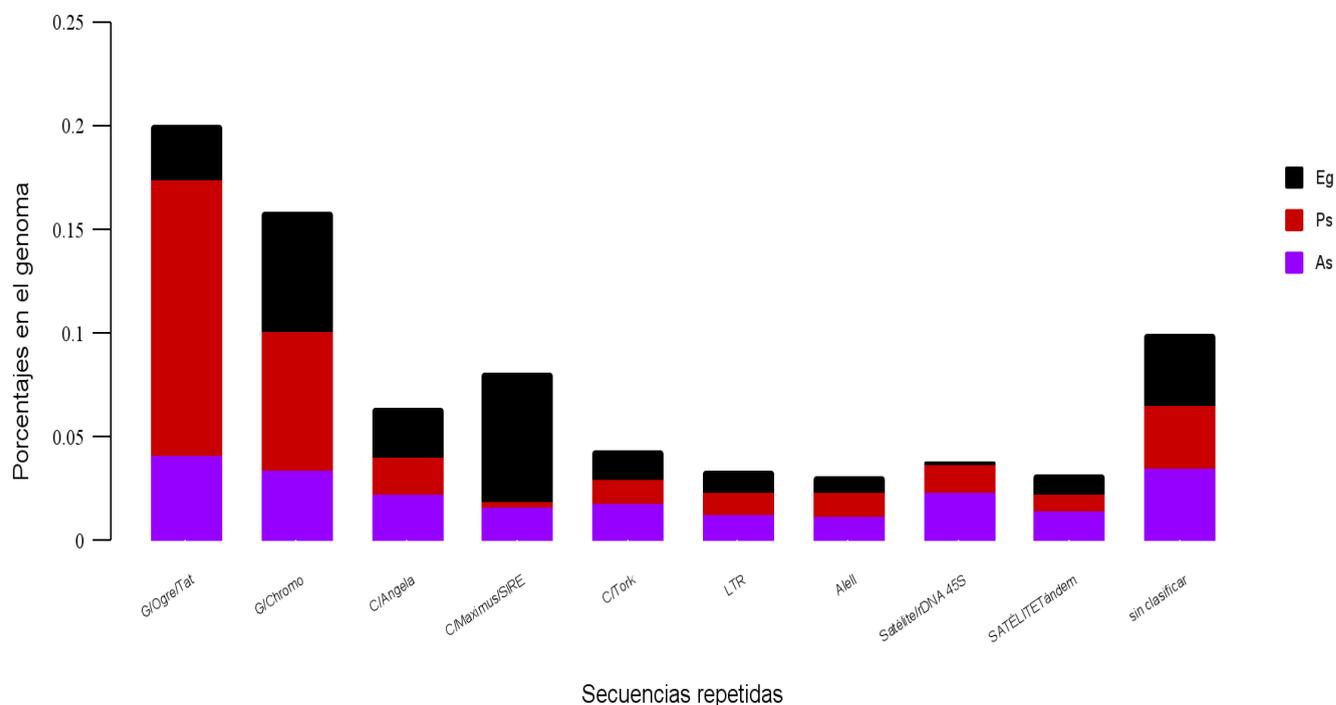


Figura 6. Comparativo de las secuencias repetidas en los genomas de *F. sellowiana*, *P. guajava* y *E. grandis*.

Tabla 2. Estudio comparativo de los elementos repetidos de tres especies de Myrtaceae *Feijoa sellowiana* (Fs), *Psidium guajava* (Ps), *Eucalyptus grandis* (Eg). Los valores representan cantidad de secuencia por genoma, entre paréntesis los porcentajes.

Clasificación	<i>Feijoa sellowiana</i>	<i>Psidium guajava</i>	<i>Eucalyptus grandis</i>
Elementos Clase I			
<i>Retrotransposones tipo LTR</i>			
sin clasificar	3.0942 (0.81)	4.1454 (0.94)	6.6708 (1.09)
<i>Ty1-Copia/ sin clasificar</i>	0.2292 (0.06)	0.6615 (0.15)	0
<i>Ty1-Copia/Ale</i>	3.438 (0.9)	4.1895 (0.95)	7.5276 (1.23)
<i>Ty1-Copia/Alesia</i>	0.1528 (0.04)	0.1323 (0.03)	0.4284 (0.07)
<i>Ty1-Copia/Angela</i>	3.9346 (1.03)	2.2932 (0.52)	1.9584 (0.32)
<i>Ty1-Copia/Bianca</i>	0.4584 (0.12)	0.3087 (0.07)	1.1016 (0.18)
<i>Ty1-Copia/Ikeros</i>	4.775 (1.25)	4.9392 (1.12)	20.1348 (3.29)
<i>Ty1-Copia/Ivana</i>	1.1842 (0.31)	2.6901 (0.61)	3.1824 (0.52)
<i>Ty1-Copia/SIRE</i>	6.7232 (1.76)	1.4994 (0.34)	35.2512 (5.76)
<i>Ty1-Copia/TAR</i>	2.4066 (0.63)	2.9988 (0.68)	4.0392 (0.66)
<i>Ty1-Copia/Tork</i>	6.4558 (1.69)	4.0572 (0.92)	8.6904 (1.42)
<i>Ty3-Gypsy/Chromovirus/CRM</i>	0.955 (0.25)	0.882 (0.2)	2.5092 (0.41)
<i>Ty3-Gypsy/Chromovirus/Galadriel</i>	0.955 (0.25)	1.1907 (0.27)	0.1224 (0.02)
<i>Ty3-Gypsy/Chromovirus/Tekay</i>	5.539 (1.45)	17.5518 (3.98)	7.1604 (1.17)
<i>Ty3-Gypsy/non-chromovirus/OTA/Athila</i>	0.3438 (0.09)	1.1466 (0.26)	1.9584 (0.32)
<i>Ty3-Gypsy/non-chromovirus/OTA/Tat/Ogre</i>	14.5924 (3.82)	66.3705 (15.05)	9.6084 (1.57)
Pararetrovirus	3.0942 (0.81)	0.8379 (0.19)	4.1616 (0.68)
<i>Retrotransposones no-LTR</i>			
LINE	0.1528 (0.04)	1.5435 (0.35)	1.0404 (0.17)
Elementos Clase II			
<i>Transposones de ADN/TIR</i>			
EnSpm_CACTA	0.1146 (0.03)	0.3528 (0.08)	0
hAT	1.4516 (0.38)	1.7199 (0.39)	1.4688 (0.24)
MuDR_Mutator	1.1842 (0.31)	1.2789 (0.29)	1.53 (0.25)
PIF_Harbinger	1.1078 (0.29)	0.8379 (0.19)	3.366 (0.55)
Repetidos en tándem			
ADNr 45S	9.6264 (2.52)	4.6305 (1.05)	0.918 (0.15)
ADNr 5S	0.2292 (0.06)	4.2336 (0.96)	2.754 (0.45)
ADN satélite	0.8022 (0.21)	0.0441 (0.01)	2.8764 (0.47)
Repetidos sin Clasificar	21.3538 (5.59)	17.9928 (4.08)	66.096 (10.8)

5.4. Estudio filogenético de las secuencias *Ty1-Copia* y *Ty3-Gypsy*

Mediante el estudio de filogenias utilizando el método de Neighbour-Joining se analizó la relación evolutiva que existe entre los retroelementos tipo LTR *Ty1-Copia* y *Ty3-Gypsy*, que presentaron dominios RT completos, y presentes en al menos una de las tres especies incluídas en este estudio. Del análisis del árbol obtenido para las secuencias *Ty1-Copia* se desprende que los 9 linajes reportados en la Tabla 2 fueron incluidos en el árbol filogenético. El linaje Ale es el más representado ya que 30 de los 87 variantes *Ty1-Copia* corresponden a este linaje. Siete elementos presentaron clados compuestos por las tres especies. El elemento SIRE solo se asoció a *F. sellowiana* mientras que Alesia a *F. sellowiana* y a *E. grandis*. Es de destacar que la ausencia de elementos SIRE de *E. grandis* llama la atención debido al alto porcentaje detectado de estos elementos (6%) en esta especie. Posiblemente, la no inclusión de elementos de SIRE de *E. grandis*, se deba a que no se hayan detectado secuencias RT completas. De manera similar, la misma explicación podría asociarse a la no inclusión de elementos Alesia de *P. guajava*.

Al analizar el árbol de los elementos del tipo *Ty3-Gypsy*, se observa la menor riqueza de este tipo de secuencias con respecto a las clasificadas de *Ty1-Copia*. Del total de 5 linajes detectados en *Ty3-Gypsy*, los únicos que presentaron dominios RT completos o presentes para realizar estos análisis fueron las Ogre/Tat y los Chromovirus/Tekay. En total se detectaron 39 variantes, 25 correspondientes a Tekay y 14 a Ogre/Tat. Esta distribución no se corresponde con las abundancias relativas reportadas en la Tabla 2 para *F. sellowiana* y *P. guajava*. Para el linaje Chromovirus/Tekay presentan una relación bastante particular donde se observa un clado con secuencias de *F. sellowiana* y *P. guajava* (13 variantes), un clado solo correspondiente a *E. grandis* (6 variantes) y otro con las tres especies (3 variantes). En cambio, para el linaje Ogre/Tat, en general, los clados se componen de las tres especies estudiadas.

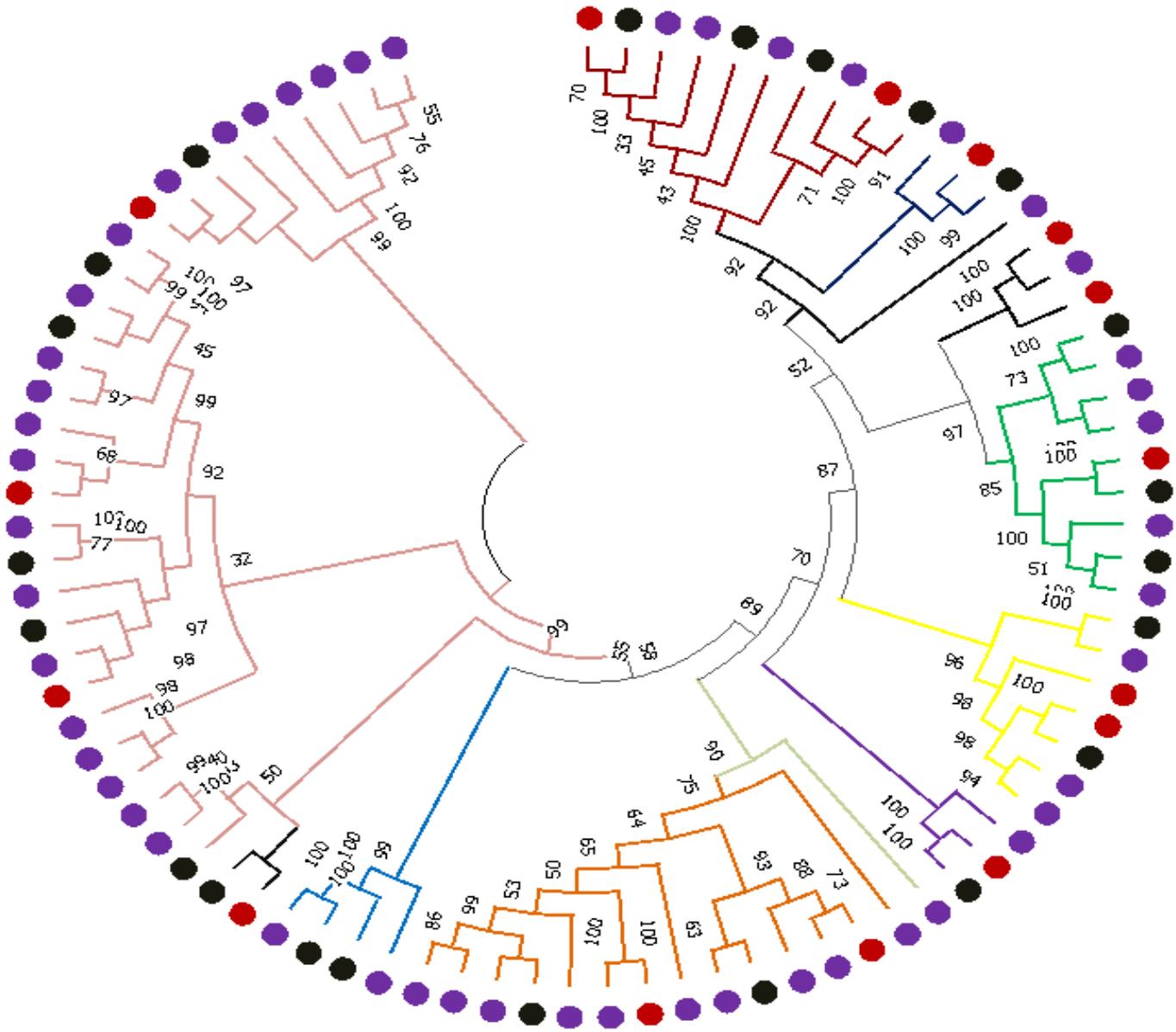


Figura 7. Árbol filogenético de las secuencias Ty1-copia para las tres especies. Negro para *Eucalyptus grandis*, rojo para *Psidium guajava* y violeta para *Feijoa sellowiana* (Ramas- rosa: Ale; celeste: Alesia; naranja: Ivana; verde claro: SIRE; violeta: Bianca; amarillo: TAR; negro: secuencias no clasificadas con similitud; verde oscuro: Tork; azul: Angela; rojo: Ikeros). En los nodos figuran los respectivos números de Bootstrap.

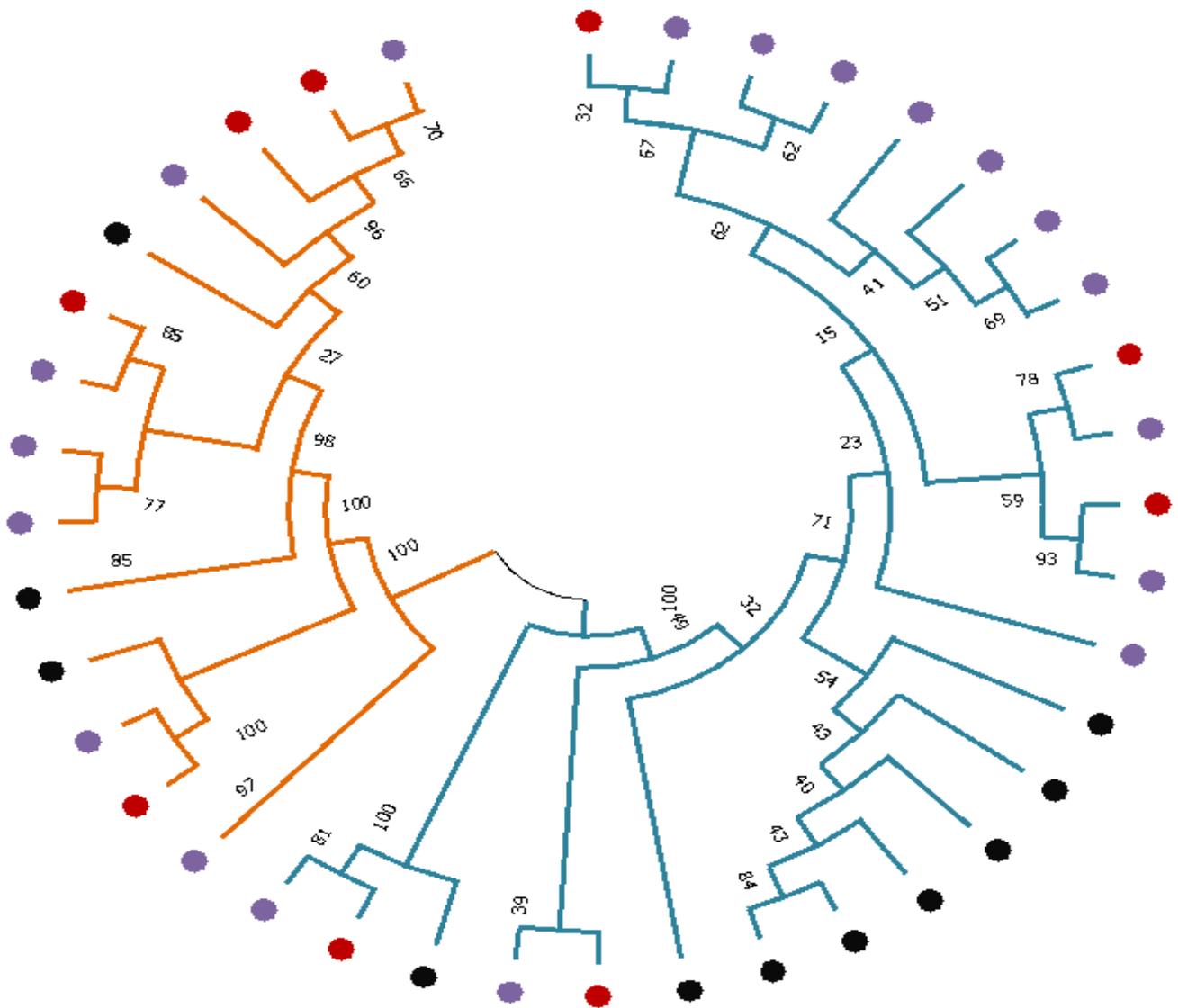


Figura 8. Árbol filogenético para las secuencias del tipo Ty3-Gypsy para las 3 spp. Negro para *Eucalyptus grandis*, rojo para *Psidium guajava* y violeta para *Feijoa sellowiana* (naranja Tat/Ogre; celeste chromovirus/Tekay). En los nodos figuran los respectivos números de Bootstrap.

5.5. Secuencias tipo satélite

Los análisis individuales utilizando TAREAN a partir de lecturas en *F. sellowiana* y *P. guajava*, identificaron dos familias de ADN satélite en cada una de las especies. Los tamaños de los monómeros variaron entre 179 a 712 pb y representan entre 0,06 a 0,4 % del genoma (Tabla 3). Los alineamientos y análisis de identidad mostraron que son cuatro familias diferentes.

Tabla 3. Secuencias repetidas en tándem de ADN de tipo satélite identificadas en las especies *Feijoa sellowiana* y *Psidium guajava*, su porcentaje en el genoma, tamaño del monómero y porcentaje GC.

Identificación de ADN sat	Especie	Tamaño del monómero (pb)	GC (%)	Porcentaje del genoma
Pg_436	<i>Psidium guajava</i>	436	35	0,07
Pg_712	<i>Psidium guajava</i>	712	42	0,44
Fs_179	<i>Feijoa sellowiana</i>	179	37	0,09
Fs_181	<i>Feijoa sellowiana</i>	181	37	0,09

6.DISCUSIÓN

6.1. Genomas pequeños pero ricos en secuencias repetidas

Según los resultados obtenidos en este trabajo, existe consistencia con Heslop- Harrison (2011) y Hanna et al. (2015), donde proponen que el genoma de las plantas superiores está compuesto en gran porcentaje por secuencias repetidas y que este componente es el que explica las grandes diferencias en el tamaño del genoma entre especies, inclusive de la misma familia. En el caso de *Feijoa sellowiana* y *Psidium guajava* los porcentajes de secuencias repetidas son altos, 34% y 43% respectivamente, si consideramos el pequeño tamaño del genoma de ambas especies. Existen diferencias en los datos de tamaño del genoma publicados para las dos especies. Para *P. guajava* los valores van desde 0,5 a 0,95 pg y *F. sellowiana* desde 0,5 a 0,74pg (valor 2C) todos ellos coinciden en un tamaño pequeño (Pritsch et al 2010; Das Graças De Souza et al 2015; Marques et al 2016; da Costa et al., 2008; da Costa y Forni-Martins 2006, 2007). Las diferencias en el contenido de ADN de las muestras estudiadas pueden estar ligadas a la metodología utilizada (tinciones, controles, etc.).

A destacar de este estudio es que, a pesar de tratarse de genomas pequeños, estos tienen alta riqueza de secuencias repetidas. Las diferencias en el tamaño del genoma entre ambas especies se corresponden con un mayor porcentaje de secuencias repetidas en *P guajava*. El tamaño del genoma de mirtáceas a pesar de ser pequeño presenta una variación de 5 veces, desde 234 Mpb (Valor C 0,239pg) en *Myrciaria glazioviana* a 1,1 Gb (Valor C 1,1 pg) en *Melaleuca leucadendra* (Myburg et al 2014, Grattapaglia et al 2012, Heayley et al 2021). Los genomas de algunas especies de Mirtáceas pertenecientes a ambas subfamilias han sido secuenciados o analizados parcialmente (Feng et al. 2021, Heayley et al. 2021, Izuno et al. 2019, Zhang et al., 2023). El porcentaje de secuencias repetidas en *Eucalyptus grandis* (C= 641 Mbp) es de 41%, *E. pauciflora* (C= 594 Mpb) 45% , *Corymbia citriodora* (C= 408 Mpb) 35,8% , *Metrosideros polymorpha* (C= 304 Mpb) 8,06% , *Melaleuca alternifolia* (C= 595 Mpb) 40,43% . Algunas diferencias entre tamaños del genoma no se corresponden con diferencias en la abundancia de secuencias repetidas. Por ejemplo, en *Eucalyptus*, *E grandis* presenta un genoma mayor que *E. pauciflora*, una diferencia de 16%, sin embargo, la segunda especie tiene un mayor porcentaje

de secuencias repetidas. Esto ha sido atribuido a secuencias de copia única o baja copia, distribuidas en el genoma (Wang et al. 2020).

Los datos obtenidos en mi trabajo son comparables a los publicados para el genoma de *Psidium guajava*, en el que el 43,55% (193,2 Mpb) es reconocido como secuencias repetidas (Feng et al., 2021). En mi trabajo el 43% corresponde a secuencias repetidas. Entonces, si bien en el artículo del genoma se llega a ese porcentaje con una alta cobertura del genoma, no da datos de las secuencias LTR que son las que entendemos como las más abundantes en los genomas de plantas, en mi trabajo con solamente 1x de cobertura, fue posible caracterizar no sólo el porcentaje sino también clasificar los diferentes elementos en clases y familias, gracias a la herramienta RepeatExplorer®.

6.2. Importancia de la anotación del satelitoma

Diversos estudios en eucariotas han mostrado que las secuencias de ADN satélite (secuencias repetidas en tándem con monómeros de cientos de pares de bases) pueden ser numerosos llegando a millones de copias (Macas et al., 2002; Plohl et al., 2008). En plantas el porcentaje que representan del genoma varía entre 0,22 en *Passiflora edulis* a más del 50% en *Olea europea* y desde 1 a casi 40 familias diferentes han sido detectadas (revisado en Šatović-Vukšić y Plohl, 2023).

En varias especies, el número de familias de ADN satélite parece no afectar el tamaño del genoma. Algunas familias de ADN satélite podrían estar altamente amplificadas en el genoma y contribuir con una gran parte de éste (por ejemplo, *Olea europea*: Barghini et al., 2014) o una pequeña proporción del mismo podría estar compuesta por muchos ADN satelitales (*Pisum sativum*: Macas et al., 2007; *Luzula*: Heckmann et al., 2013).

La composición de ADN satélite en genomas de especies relacionadas puede ser variable. En algunos casos, el repertorio de ADN satélite puede variar significativamente, como es el caso de especies de *Fritillaria* (Ambrožová et al., 2011) y citotipos de *Prospero autumnale*, (Emadzade et al., 2014) o, ser muy similar como ocurre en el , género *Musa* (Čížková et al., 2013). En el complejo *P. autumnale*, el

satélite PaB6 se localiza predominantemente en regiones peri-centroméricas de todos los citotipos pero su número de copias y por lo tanto proporción genómica varía ampliamente (de 0,1% a 10%; Emadzade et al., 2014). En otros trabajos como el de Gaiero et al. (2019) sobre secuencias repetidas en solanáceas, se reportó que las secuencias satélites entre tomates y papas, (fuertemente emparentadas) no fueron coincidentes. En varios trabajos recientes de genomas completos de mirtáceas tales como *Metrosideros polymorpha* y *P. guajava* no se reporta el detalle de la composición de ADN satélite ni de los loci de ADNr (Izuno et al. 2019, Feng et al 2021). Los abordajes de anotación empleados suelen estar más dirigidos a las secuencias génicas que codifican proteínas, mientras que las secuencias repetidas se describen de manera más general.

El abordaje utilizado en este trabajo basado en el paquete RepeatExplorer®, ha sido una herramienta valiosa para la caracterización de diferentes elementos dentro del repitoma de *F. sellowiana* y *P. guajava*. Este trabajo identificó y caracterizó por primera vez para la familia Myrtaceae cuatro familias diferentes de ADN satélite, dos en *F. sellowiana* y dos en *P. guajava*. Mientras que ambas especies mostraron niveles similares de representación del ADN satélite en el genoma (0,3-0,4 %), tanto las secuencias como el tamaño del monómero de las diferentes familias no fueron coincidentes. La falta de coincidencia en los resultados de *F. sellowiana* y *P. guajava* podría ser interpretado bajo la luz de la *teoría de la biblioteca* postulada por Fry y Salser (1977). Esta teoría indica que las posibles diferencias de composición de ADN satélite observadas entre especies emparentadas, podrían ser resultado de fluctuaciones especie-específicas en el número de copias de diferentes miembros del repertorio (biblioteca) de secuencias satélites compartidos. De esta manera, se puede explicar tanto la ausencia como la sobre-representación de familias de ADN satélites, entre especies.

6.3. Relaciones evolutivas de linajes *Ty1-Copia* y *Ty3-Gypsy*

La diversificación del genoma entre especies puede estar explicada por las dinámicas de amplificación y diversificación de elementos transponibles. Los elementos transponibles más importantes en plantas son los retroelementos LTR *Ty3-Gypsy* y *Ty1-Copia*. Este trabajo encontró los mismos resultados para las tres especies mirtáceas estudiadas; además se identificaron 9 linajes (familias) dentro de *Ty1-Copia* y 5 dentro de *Ty3-Gypsy*. Estos resultados son novedosos ya que la caracterización por linajes de los RT-LTR *Ty1-Copia* y *Ty3-Gypsy* está poco descrita en mirtáceas, tal como se discutió más arriba con respecto a ADN satélite y ADN ribosomal.

El estudio de las relaciones evolutivas entre los linajes de los RT LTR *Ty1-Copia* y las de tipo *Ty3-Gypsy* mediante estudios filogenéticos permite detectar aquellos elementos heredados desde ancestros comunes (ampliamente compartidos) de aquellos novedosos, posiblemente de origen más recientes con una distribución especie-específica. En base al análisis del dominio RT, la mayoría de los RT LTR *Ty1-Copia* (7 de 9 linajes) y el linaje Ogre/TAt del RT LTR *Ty3-Gypsy* eran compartidos por *F. sellowiana*, *P. guajava* y *E. grandis*, indicando que dichos linajes estarían presentes en una especie ancestral común. En cambio, el linaje SIRE (*Ty1-Copia*) solamente fue representado por *F. sellowiana*, posiblemente debido a que las secuencias SIRE de *P. guajava* y *E. grandis* no eran completas. De manera similar, el clado correspondiente al linaje *Ty1-Copia* Alesia, solamente se integró *F. sellowiana* y *E. grandis*. Únicamente los linajes Chromovirus/Tekay y Ogre/Tat pudieron ser analizados filogenéticamente, en base a la disponibilidad de secuencias completas RT. Para Chromovirus/Tekay, con alta representación en las tres especies, fue posible distinguir de manera general un clado compuesto de representantes de las tres especies, otro compuesto únicamente por *E. grandis* (fruto capsular) y un tercero de *F. sellowiana* y *P. guajava* (frutos carnosos). Estos resultados podrían estar indicando que estos linajes estarán desarrollando procesos evolutivos diferentes entre las tribus Myrteae y Eucalypteae. Estudios biogeográficos y de datación de las distintas tribus de la familia, mostraron que la separación entre las tribus Eucalypteae y Myrteae ocurrió aproximadamente hace 70

mya. Los resultados sugieren que la mayoría de las variantes de los linajes de la superfamilia *Ty1-Copia* y *Ty3-Gypsy*, se conservaron durante ese tiempo evolutivo desde un ancestro común. Para el caso de Chromovirus/Tekay en cambio, algunas variantes habrían surgido de manera independiente en las tribu Myrteae y Eucalypteae, después de la divergencia entre estos taxones. Myrteae surge hace 55 mya y el nodo que separa ambos géneros de guayabos está datado ca. 35 mya. (Thornhill et al. 2015).

Mi trabajo también sugiere que un gran número de los retrotransposones sobre todo dentro de los *Ty3-Gypsy* podrían estar incompletos indicando, que podrían tratarse de elementos antiguos, que estarían mostrando una importante acumulación de mutaciones generadas durante un extenso período desde su inserción.

7.CONCLUSIONES

El repitoma de *F. sellowiana* (34%) y *P. guajava* (43%) se compone básicamente de elementos Clase I y en menor medida Clase II.

Considerando *F. sellowiana*, *P.guajava* y *E. grandis*, los retroelementos LTR-RT tipo *Ty1-Copia* y *Ty3-Gypsy* son los más abundantes. La composición de linajes dentro de *Ty1-Copia* y *Ty3-Gypsy* fue similar, pero se observaron diferencias en sus respectivas abundancias relativas. En el linaje Chromovirus/Tekay, de alta representación, se identificaron elementos tribu-específicos.

Por primera vez para la familia Myrtaceae, se identificaron cuatro familias diferentes de ADN satélite en mirtáceas, dos en *F. sellowiana* y dos en *P. guajava*.

El programa RepeatExplorer® es una herramienta poderosa para la identificación y clasificación detallada de un amplio rango de secuencias repetidas, independiente de la disponibilidad de un genoma

de referencia, apoyado en una base de datos con actualización permanente. El procedimiento de identificación basado en análisis de grafos permite la posible detección de elementos novedosos.

No subestimar los tamaños de los genomas, los pequeños pueden esconder alta riqueza de secuencias repetidas.

8. PERSPECTIVAS

Para complementar la información obtenida en este trabajo, sugiero ampliar el análisis del paisaje de los elementos transponibles Ty1-copia y Ty3-Gypsy y del ADN satélite y, caracterizar la ubicación de estos elementos en los genomas de *F. sellowiana* y *P. guajava*. La localización de dichas secuencias: se podría implementar mediante: i) hibridación in situ de cromosomas metafásicos con sondas fluorescentes o ii) mediante análisis in silico de genomas ensamblados.

En este trabajo, se detectó que algunas variantes de Chromovirus/Tekay serían específicas de *Myrteae*, mientras que otras serían específicas de *Eucalypteae*. Sin embargo, dado que el número de especies analizadas es bajo, propondría ampliar el número de especies de cada tribu en futuros análisis.

9. BIBLIOGRAFÍA

- Albach, D. C., & Greilhuber, J. (2004). Genome size variation and evolution in *Veronica*. *Annals of Botany*, 94(6), 897-911.
- Ambrožová, K., Mandáková, T., Bureš, P., Neumann, P., Leitch, I. J., Koblížková, A., ... & Lysak, M. A. (2011). Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. *Annals of Botany*, 107(2), 255-268.
- Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, 23(3), 415-420.
- Biscotti, M.A., Olmo, E. & Heslop-Harrison, J.S.(. Repetitive DNA in eukaryotic genomes. *Chromosome Res* 23, 415–420 (2015).
- Brussa, C. A., & Grela, I. (2007). *Flora Arbórea del Uruguay. Con énfasis en las especies de Rivera y Tacuarembó*. COFUSA. Mosca. Montevideo, Uruguay, 543.
- Casacuberta, E., & González, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular ecology*, 22(6), 1503-1517.
- Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 371(6494), 215-220.
- Čížková J, Hřibová E, Humplíková L, Christelová P, Suchánková P, Doležel J (2013) Análisis molecular y organización genómica de los principales satélites de ADN en banano (*Musa* spp.). *PLoS ONE* 8(1): E54808. <https://doi.org/10.1371/journal.pone.0054808>

Coser, S. M., da Silva Ferreira, M. F., Ferreira, A., Mitre, L. K., Carvalho, C. R., & Clarindo, W. R. (2012). Assessment of genetic diversity in *Psidium guajava* L. using different approaches. *Scientia Horticulturae*, 148, 223-229.

Itayguara Ribeiro da Costa & Eliana Regina Forni-Martins (2006) Chromosome studies in Brazilian species of *Campomanesia* Ruiz & Pávon and *Psidium* L. (Myrtaceae Juss.), *Caryologia*, 59:1, 7-13, DOI: 10.1080/00087114.2006.10797891

ITAYGUARA RIBEIRO DA COSTA , ELIANA R. FORNI-MARTINS, Karyotype analysis in South American species of Myrtaceae, *Botanical Journal of the Linnean Society*, Volume 155, Issue 4, December 2007, Pages 571–580, <https://doi.org/10.1111/j.1095-8339.2007.00704.x>

da Costa, I. R., Dornelas, M. C., and Forni-Martins, E. R. (2008). Nuclear genome size variation in fleshy-fruited Neotropical Myrtaceae. *Plant Syst. Evol.* 276, 209–217.

Das Graças De Souza, A., Resende, L.V., De Lima, I.P. *et al.* Chromosome number and nuclear DNA amount in *Psidium* spp. resistant and susceptible to *Meloidogyne enterolobii* and its relation with compatibility between rootstocks and commercial varieties of guava tree. *Plant Syst Evol* 301, 231–237 (2015). <https://doi.org/10.1007/s00606-014-1068-y>

de Koning, A. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS genetics*, 7(12), e1002384.

De Souza Neto, J. D., Dos Santos, E. K., Lucas, E., Vetö, N. M., Barrientos-Diaz, O., Staggemeier, V. G., ... & Turchetto-Zolet, A. C. (2022). Advances and perspectives on the evolutionary history and diversification of Neotropical Myrteae (Myrtaceae). *Botanical Journal of the Linnean Society*, 199(1), 173-195. doi: 10.1007/s00606-008-0088-x

Eickbush, T. H., & Jamburuthugoda, V. K. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus research*, 134(1-2), 221-234.

Emadzade, K., Jang, T. S., Macas, J., Kovařík, A., Novák, P., Parker, J., & Weiss-Schneeweiss, H. (2014). Differential amplification of satellite PaB6 in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). *Annals of Botany*, 114(8), 1597-1608.

Feng, C., Feng, C., Lin, X., Liu, S., Li, Y., and Kang, M. (2021). A chromosome level genome assembly provides insights into ascorbic acid accumulation and fruit softening in guava (*Psidium guajava*). *Plant Biotechnol. J.* 19, 717–730. doi: 10.1111/pbi.13498

Finnegan, D. J. (1992). Transposable elements. *Current opinion in genetics & development*, 2(6), 861-867.

Fry K , Salzer W. (1977). Nucleotide sequences of HS-alpha satellite DNA form kangaroo rat *Dypodomis ordii* and characterization of similar sequences in other rodents. *Cell* 12, 1069-1084.

Gaiero, P., Vaio, M., Peters, S. A., Schranz, M. E., de Jong, H., & Speranza, P. R. (2019). Comparative analysis of repetitive sequences among species from the potato and the tomato clades. *Annals of Botany*, 123(3), 521-532.

Galindo-González, L., Mhiri, C., Grandbastien, MA. et al. Ty1-copia elements reveal diverse insertion sites linked to polymorphisms among flax (*Linum usitatissimum* L.) accessions. *BMC Genomics* 17, 1002 (2016).
<https://doi.org/10.1186/s12864-016-3337-3>

Garrido-Ramos, M. A. (2015). Satellite DNA in plants: more than just rubbish. *Cytogenetic and Genome Research*, 146(2), 153-170.vv

Garrido-Ramos, M. A. (2017). Satellite DNA: an evolving topic. *Genes*, 8(9), 230.

Gillespie, D., Adams, J. W., Costanzi, C., & Caranfa, M. J. (1982). New orientations of ancestral, "long interspersed repeated sequences"(LINES) in human DNA. *Gene*, 20(3), 409-414.

- Giordani, T., Cossu, R. M., Mascagni, F., Marroni, F., Morgante, M., Cavallini, A., & Natali, L. (2016). Genome-wide analysis of LTR-retrotransposon expression in leaves of *Populus× canadensis* water-deprived plants. *Tree Genetics & Genomes*, 12(4), 1-14.
- Godinho, S.; Paulo, O.S.; Morais-Cecilio, L.; Rocheta, M. A new gypsy-like retroelement family in *Vitis Vinifera*. *VITIS* 2012, 51, 65–72.
- Goubert C, Modolo L, Vieira C, Valiente Moro C, Mavingui P, Boulesteix M (2015) De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biol Evol* 7:
- Govaerts, R., Sobral, M., Ashton, P., Barrie, F., Holst, B. K., Landrum, L. L., ... & Lucas, E. (2008). World checklist of Myrtaceae. Royal Botanic Gardens.
- Grandbastien, M. A. (2015). LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1849(4), 403-416.
- Grattapaglia, D., Vaillancourt, R. E., Shepherd, M., Thumma, B. R., Foley, W., Külheim, C., ... & Myburg, A. A. (2012). Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus. *Tree Genetics & Genomes*, 8, 463-508.
- Hannan, A. J. (2018). Tandem repeats and repeatomes: delving deeper into the 'Dark Matter' of genomes. *EBioMedicine*, 31, 3-4.
- Healey, A. L., Shepherd, M., King, G. J., Butler, J. B., Freeman, J. S., Lee, D. J., ... & Henry, R. J. (2021). Pests, diseases, and aridity have shaped the genome of *Corymbia citriodora*. *Communications biology*, 4(1), 537.
- Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., ... & Houben, A. (2013). The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. *The Plant Journal*, 73(4), 555-565.

- Hemleben, V., Kovarik, A., Torres-Ruiz, R. A., Volkov, R. A., & Beridze, T. (2007). Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. *Systematics and Biodiversity*, 5(3), 277-289.
- Henikoff, S., Ahmad, K., & Malik, H. S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, 293(5532), 1098-1102.
- Heslop-Harrison, J. S., Brandes, A., Taketa, S., Schmidt, T., Vershinin, A. V., Alkhimova, E. G., ... & Harrison, G. E. (1997). The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica*, 100(1), 197-204.
- Heslop-Harrison, J. S., & Schwarzacher, T. (2011). Organization of the plant genome in chromosomes. *The Plant Journal*, 66(1), 18-33.
- Izuno, A., Wicker, T., Hatakeyama, M., Copetti, D., & Shimizu, K. K. (2019). Updated Genome Assembly and Annotation for *Metrosideros polymorpha*, an Emerging Model Tree Species of Ecological Divergence. *G3: Genes, Genomes, Genetics*, 9(11), 3513-3520.
- Jouffroy, O., Saha, S., Mueller, L., Quesneville, H., & Maumus, F. (2016). Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. *BMC genomics*, 17(1), 1-15.
- Kejnovsky, E., Hawkins, J. S., & Feschotte, C. (2012). *Plant transposable elements: biology and evolution* (pp. 17-34). Springer Vienna.
- Kejnovsky, E., Tokan, V., & Lexa, M. (2015). Transposable elements and G-quadruplexes. *Chromosome Research*, 23(3), 615-623.
- Kumar, A., & Bennetzen, J. L. (1999). Plant retrotransposons. *Annual review of genetics*, 33(1), 479-532.

- Li, S. F., Su, T., Cheng, G. Q., Wang, B. X., Li, X., Deng, C. L., & Gao, W. J. (2017). Chromosome evolution in connection with repetitive sequences and epigenetics in plants. *Genes*, 8(10), 290.
- López-Flores, I., & Garrido-Ramos, M. A. (2012). The repetitive DNA content of eukaryotic genomes. *Repetitive DNA*, 7, 1-28.
- Lucas, E. J., Holst, B., Sobral, M., Mazine, F. F., Nic Lughadha, E. M., Barnes Proença, C. E., ... & Vasconcelos, T. N. (2019). A new subtribal classification of tribe Myrteae (Myrtaceae). *Systematic Botany*, 44(3), 560-569.
- Macas, J., Meszaros, T., & Nouzova, M. (2002). PlantSat: a specialized database for plant satellite repeats. *Bioinformatics*, 18(1), 28-35.
- Macas, J., Neumann, P., & Navrátilová, A. (2007). Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC genomics*, 8(1), 1-16.
- Marques, A., Tuler, A. C., Carvalho, C. R., Carrijo, T., da Silva Ferreira, M. F., & Clarindo, W. (2016). Refinement of the karyological aspects of *Psidium guineense* (Swartz, 1788): a comparison with *Psidium guajava* (Linnaeus, 1753). *Comparative Cytogenetics*, 10(1), 117-128.
- Maumus, F., & Quesneville, H. (2016). Impact and insights from ancient repetitive elements in plant genomes. *Current opinion in plant biology*, 30, 41-46.
- McClintock B. 1950 The origin and behavior of mutable loci in maize. *Proc. Natl Acad. Sci. USA* 36, 344–355.
- Myburg A, Grattapaglia, D, Tuskan G, et al. (2014). The genome of *Eucalyptus grandis*. *Nature*. 510: 352-362.
- Nimisha S, Kherwar D, Ajay KM, Singh B, et al. (2013). Molecular breeding to improve guava (*Psidium guajava* L.): Current status and future prospective. *Sci. Hortic.* 164: 578-588.
- Nagaki, K., Kashihara, K., & Murata, M. (2005). Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant *Luzula nivea*. *The Plant Cell*, 17(7), 1886-1893.

Negi, P.; Rai, A.N.; Suprasanna, P. Moving through the stressed genome: Emerging regulatory roles for transposons in plant stress response. *Front. Plant Sci.* 2016, 7.

Neumann, P., Novák, P., Hošťáková, N., & Macas, J. (2019). Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA*, 10(1), 1-17.

Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., & Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic acids research*, 45(12), e111-e111.

Novák, P., Guignard, M. S., Neumann, P., Kelly, L. J., Mlinarec, J., Koblížková, A., ... & Leitch, A. R. (2020). Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nature Plants*, 6(11), 1325-1329.

Novák, P., Neumann, P., & Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC bioinformatics*, 11, 1-12.

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., & Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, 29(6), 792-793.

Oriane Hidalgo, Jaume Pellicer, Maarten J. M. Christenhusz, Harald Schneider, Ilia J. Leitch, Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*, *Botanical Journal of the Linnean Society*, Volume 183, Issue 4, 1 April 2017, Pages 509–514

Pita, S., Panzera, F., Mora, P., Vela, J., Cuadrado, Á., Sánchez, A., ... & Lorite, P. (2017). Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease. *PLoS One*, 12(7), e0181635.

Plohl, M., Luchetti, A., Meštrović, N., & Mantovani, B. (2008). Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. *Gene*, 409(1-2), 72-82.

Pritsch C., Mazzella C., Vaio M., Quezada M., Da Cruz I., Lombardo P., Ccabrera D. & Vignale B. 2010. Hacia la caracterización de la estructura del genoma de la especie frutal nativa *Acca sellowiana* (Berg.) Burret: abordajes genéticos, citológicos y moleculares. 5o Encuentro Nacional sobre Frutos Nativos. Salto, Uruguay. Serie Actividades de Difusión 602. INIA - FAGRO - MGAP. 56 p.

Quezada, M., Amadeu, R. R., Vignale, B., Cabrera, D., Pritsch, C., & Garcia, A. A. F. (2021). Construction of a high-density genetic map of *Acca sellowiana* (Berg.) burret, an outcrossing species, based on two connected mapping populations. *Frontiers in Plant Science*, 12, 626811.

Rafaqat A. Gill, Federico Scossa, Graham J. King, Agnieszka A. Golicz, Chaobo Tong, Rod J. Snowdon, Alisdair R. Fernie & Shengyi Liu (2021) On the Role of Transposable Elements in the Regulation of Gene Expression and Subgenomic Interactions in Crop Genomes, *Critical Reviews in Plant Sciences*, 40:2, 157-189, DOI: 10.1080/07352689.2021.1920731

Ruiz-Ruano, F.J.; López-León, M.D.; Cabrero, J.; Camacho, J.P.M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* 2016, 6, 28333. [CrossRef] [PubMed]

25. Kim, Y.B.; Oh, J.H.; McIver, L.J.; Rashkovetsky, E.; Michalak, K.; Garner, H.R.; Kang, L.; Nevo, E.; Korol, A.B.; Michalak, P. Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel. *Proc. Natl. Acad. Sci. USA* 2014, 111, 10630–10635. [CrossRef]

Sardi J de CO, Freires IA, Lazarini JG, Infante J, de Alencar SM, Rosalen PL. 2017. Unexplored endemic fruit species from Brazil: antibiofilm properties, insights into mode of action, and systemic toxicity of four *Eugenia* spp. *Microbial Pathogenesis* 105: 280–287.

Šatović-Vukšić, E., & Plohl, M. (2023). Satellite DNAs—From Localized to Highly Dispersed Genome Components. *Genes*, 14(3), 742. <https://doi.org/10.3390/genes14030742>

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., ... & Presting, G. G. (2009). The B73 maize genome: complexity, diversity, and dynamics. *science*, 326(5956), 1112-1115.

Seymour, G. B., Taylor, J. E., Tucker, G. A. (1993). *Biochemistry of fruit ripening*. London: Chapman & Hall.

Sharma, S., & Raina, S. N. (2005). Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. *Cytogenetic and genome research*, 109(1-3), 15-26.

Thomas CA (1971). The genetic organization of chromosomes. *Annual Review of Genetics* 5:237-256

Thornhill, A. H., Ho, S. Y., Külheim, C., & Crisp, M. D. (2015). Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. *Molecular Phylogenetics and Evolution*, 93, 29-43.

Ugarković, Đ., & Plohl, M. (2002). Variation in satellite DNA profiles—causes and effects. *The EMBO journal*, 21(22), 5955-5959.

Vázquez Medina, S. N. (2014). *Psidium cattleianum* Sabine y *Acca sellowiana* (Berg.) Burret (Myrtaceae): caracterización cromosómica y cariotípica en poblaciones silvestres y genotipos seleccionados en programas nacionales de mejoramiento.

Vijayakumar, N., & Subramanian, D. (1985). *Cytotaxonomical studies in South Indian Myrtaceae*. *CYTOLOGIA*, 50(3), 513–520. doi:10.1508/cytologia.50.513

Vicient, C. M., & Casacuberta, J. M. (2017). Impact of transposable elements on polyploid plant genomes. *Annals of Botany*.

Weiss-Schneeweiss, H., Leitch, A. R., McCann, J., Jang, T. S., & Macas, J. (2015). Employing next generation sequencing to explore the repeat landscape of the plant genome. *Next generation sequencing in plant systematics*. *Regnum Vegetabile*, 157, 155-179.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., ... & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973-982.

Wicker, T.; Gundlach, H.; Spannagl, M.; Uauy, C.; Borrill, P.; Ramirez-Gonzalez, R.H.; De Oliveira, R.; Mayer, K.F.; Paux, E.; Choulet, F. Impact of Transposable Elements on Genome Structure and Evolution in Bread Wheat. *Genome Biol.* 2018, 19, 103.

Widegren B, Arnason U, Akusj/irvi (1985) Characteristics of a conserved 1,579-bp highly repetitive component in the killer whale, *Orcinus orca*. *Mol Biol Evol* 2: 411-419

Wilson, P. G., O'Brien, M. M., Heslewood, M. M., & Quinn, C. J. (2005). Relationships within Myrtaceae sensu lato based on a matK phylogeny. *Plant Systematics and evolution*, 251(1), 3-19.

Zhang X, Chen S, Zhang Y, Xiao Y, Qin Y, Li Q, Liu L, Liu B, Chai L, Yang H, Liu H. Draft genome of the medicinal tea tree *Melaleuca alternifolia*. *Mol Biol Rep.* 2023 Feb;50(2):1545-1552. doi: 10.1007/s11033-022-08157-8. Epub 2022 Dec 13. PMID: 36513867.

Zhang, Q. J., & Gao, L. Z. (2017). Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* species. *G3: Genes, Genomes, Genetics*, 7(6), 1875-1885.

Zuccolo, A., Scofield, D. G., De Paoli, E., & Morgante, M. (2015). The Ty1-copia LTR retroelement family PARTC is highly conserved in conifers over 200 MY of evolution. *Gene*, 568(1), 89-99.