



AGENCIA NACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN

Informe final publicable de proyecto

Caracterización de las duplicaciones génicas en eucaliptos.

Código de proyecto ANII: FCE_3_2022_1_172481

Fecha de cierre de proyecto: 01/12/2025

GIORELLO, Facundo (Responsable Técnico - Científico)

FARIAS COLMAN, Joaquina (Co-Responsable Técnico-Científico)

UNIVERSIDAD DE LA REPÚBLICA (Institución Proponente) \ \

UNIVERSIDAD DE LA REPÚBLICA. CENTRO UNIVERSITARIO REGIONAL NORESTE

Resumen del proyecto

El género *Eucalyptus* agrupa a más de 800 especies de árboles de gran importancia económica que representan un claro ejemplo de rápida adaptación a diversos climas y nichos ecológicos a lo largo de millones de años. Sorprendentemente, a pesar de su enorme diversidad, sus genomas son muy similares. La hipótesis principal es que su rápida evolución está impulsada por las duplicaciones génicas en tándem, es decir, que ciertos genes se copian repetidamente en una región del ADN. Como esta idea se ha estudiado solo en unas pocas especies, nuestro proyecto se propuso determinar si esta característica es común a más miembros del género. Para ello, secuenciamos y analizamos los genomas de 11 especies de *Eucalyptus*. Los resultados de esta investigación son útiles para comprender el rol de estas duplicaciones en la evolución y adaptación del género, y la información obtenida puede ser aplicable en programas de mejoramiento genético de eucaliptos.

Ciencias Naturales y Exactas / Ciencias Biológicas / Genética y Herencia (Genética Médica va en "Ciencias Médicas y de la Salud") / Genética Evolutiva

Palabras clave: genómica comparada / duplicaciones génicas / /

Antecedentes, problema de investigación, objetivos y justificación.

El género *Eucalyptus* es uno de los linajes más diversos de árboles. Perteneciente a la familia Myrtaceae, este grupo diverso se compone de más de 800 especies que, tras una radiación adaptativa que se remonta al Cenozoico medio (hace 25 a 10 millones de años), se han adaptado a una amplia gama de hábitats, desde regiones áridas y de baja fertilidad hasta ambientes alpinos con heladas extremas. Su capacidad para prosperar en biomas propensos a incendios y su extensa producción de metabolitos secundarios (como terpenoides y alcaloides) reflejan su adaptabilidad ecológica y química. Además de su relevancia biológica como ejemplo de evolución rápida, los eucaliptos son económicamente trascendentes a nivel global, siendo la base de importantes industrias forestales.

El enorme potencial adaptativo que ha demostrado el género de *Eucalyptus* contrasta en gran medida con lo hallado para sus genomas. A pesar de la vasta diversidad fenotípica y adaptativa del género *Eucalyptus*, sus genomas muestran un grado de conservación remarcable (1,2). Típicamente, todas las especies presentan los mismos 11 cromosomas, con tamaños promedio de unos 600 Mb, y con niveles de reordenamientos cromosómicos llamativamente bajos, incluso al comparar géneros cercanos como *Eucalyptus* y *Corymbia* (2).

Estudios pioneros, principalmente en *E. grandis* (cuyo genoma fue secuenciado y anotado con gran detalle), identificaron una característica muy particular: aproximadamente el 34% de los genes codificantes de proteínas se encuentran en duplicaciones en tándem (1). Esta es la proporción más grande reportada hasta la fecha en genomas de plantas. Las duplicaciones en tándem, que consisten en la repetición de un gen de manera adyacente en el cromosoma, son un mecanismo dinámico que permite a la especie retener copias de genes originales mientras las copias duplicadas están libres para mutar y adquirir nuevas funciones. Estas duplicaciones, a menudo asociadas a genes de respuesta al estrés, podrían explicar la capacidad adaptativa de los eucaliptos. Myburg y colaboradores (2014), de hecho, encuentran que dentro del género *Eucalyptus* las duplicaciones en tándem son dinámicas; por ejemplo, para un clúster de factores de transcripción MYB, *E. globulus* carece de 4 de las 9 duplicaciones en tándem encontradas para *E. grandis* (2).

Justificación

Dada la relevancia del género *Eucalyptus*, resulta interesante determinar hasta qué punto esta característica se extiende a otras especies de *Eucalyptus* e identificar los genes involucrados en tal proceso.

Objetivo General:

Caracterizar las duplicaciones génicas en varias especies de *Eucalyptus*.

Objetivos Específicos:

Optimizar la caracterización de CNVs: Definir y validar una estrategia bioinformática que permita identificar CNVs interespecíficas con precisión.

Estudiar la extensión de la dinámica de las duplicaciones en tándem: Determinar si todas las especies de *Eucalyptus* estudiadas presentan un número significativo y similar de duplicaciones en tándem en relación a *E. grandis*.

Determinar el número de duplicaciones por gen: Comparar el número de copias por familia génica entre las distintas especies de *Eucalyptus*.

Analizar los genes asociados a este fenómeno evolutivo: Identificar los genes localizados en las regiones duplicadas y determinar su función biológica, esperando que estén relacionados con la respuesta al estrés biótico y abiótico.

Las Hipótesis principales contrastadas fueron:

Las especies de *Eucalyptus* estudiadas presentan una cantidad significativa de genes duplicados en tándem en comparación con otras especies vegetales.

Las distintas especies tendrán un número distinto de copias para las distintas familias génicas.

Los genes asociados a las regiones duplicadas están relacionados a la respuesta al estrés biótico y abiótico.

Metodología/Diseño del estudio

1. Obtención de Datos y Secuenciación

* Muestras: El material genético provino de especies facilitadas por INIA Tacuarembó. Gracias al alto rendimiento obtenido en las extracciones de ADN, se logró incluir exitosamente una especie adicional, finalizando con 11 especies analizadas: *E. grandis*, *E. tereticornis*, *E. dunnii*, *E. smithii*, *E. maidenii*, *E. amplifolia*, *E. sideroxylon*, *E. macarthurii*, *E. globulus*, *E. saligna* y *E. nitens*.

* Preparación y Secuenciación: El ADN fue extraído utilizando el protocolo CTAB modificado para obtener ADN de alto peso molecular en especies vegetales complejas, como los eucaliptos (3). Las muestras se enviaron a un servicio externo (Macrogen Inc.) para secuenciación masiva en la plataforma Illumina, garantizando la profundidad de x30 requerida, y de PacBio Hifi.

2. Optimización y Definición del Pipeline Bioinformático (Objetivo 1)

Evaluamos y consideramos más de una decena de programas entre alineadores y detectores de CNV en pos de detectar las duplicaciones, tanto a partir de reads cortos (con e.g. *cn.mops*, *LUMPY*), ensamblados de reads cortos con megahit (con e.g. *SVIM-asm*, *cuteSV*), y genomas completos (trasladando las anotaciones con *LiftOn*, *GEMOMA*).

La evaluación se hizo utilizando los datos genómicos y genomas de *E. grandis* y *Corymbia citriodora* adquiridos en la base de datos del NCBI.

3. Identificación y Caracterización de Duplicaciones Génicas (Objetivos 2 y 3)

* Identificación de Duplicaciones en Tándem: Mediante scripts propios (que analizan el archivo gtf generado por LiftOn), se identificó la presencia de genes en tándem en las nuevas especies. Básicamente el script busca genes duplicados separados por menos de 20 genes (4).

* Análisis de Ortología: Para comparar el número de copias génicas entre especies (Objetivo 3), se establecieron previamente las relaciones de ortología entre *E. grandis* y otras especies relacionadas (*Rhodamnia argentea*, *Corymbia citriodora*). Esto permitió identificar grupos de ortólogos que eran de copia única en las especies de control pero duplicados (en tándem o no) en las nuevas especies.

4. Análisis Funcional y Validación Experimental (Objetivo 4)

Análisis Ontológico: Los genes identificados en las regiones duplicadas fueron analizados mediante análisis ontológico (utilizando la herramienta DAVID) para determinar si existe un enriquecimiento de ciertas funciones biológicas en comparación con el total de genes del genoma.

Resultados, análisis y discusión

1. Ensamblado de genomas y pipeline.

* Calidad del ensamblado: Para varias de las 11 especies analizadas logramos ensamblados genómicos de excelente calidad (e.g., N90 = 21 Mb) en parte debido a un exitoso procedimiento de extracción de ADN genómico de alto peso molecular.

* Pipeline: Definimos un pipeline eficiente para encontrar las duplicaciones que consistió directamente en ensamblar los genomas y hacer las anotaciones cruzadas con LiftOn. Esta estrategia fue la más rápida y precisa, sin embargo no es la más económica dado que requiere secuenciación mediante PacBio Hifi. En breve, los programas para reads cortos que utilizamos (DELLY, LUMPY, cn.MOPS, smooove, TARDIS y svabas) arrojaron resultados disímiles en relación a las duplicaciones. Básicamente lo mismo sucedió cuando se utilizaron los programas cuteSV, SVIM-asm, SVIM, Sniffle2 para reads largos, genomas y/o ensamblados de los reads cortos con megahit. Sin embargo, con el traslado de la anotación utilizando el LiftOn entre especies (*E. grandis* y *C. citriodora*), el 86% de los genes anotados se recuperan, donde la gran mayoría se ubica en los grupos ortólogos correctos (especialmente si estos grupos de genes ortólogos estaban compuestos por pocos genes). BRAKER, en cambio, sobrestimaba notoriamente el número de genes (seteado por defecto utilizando secuencias proteicas como referencia). Por otro lado, los ensamblados preliminares que corrimos para evaluar Oxford Nanopore y Pacbio Hifi, arrojaron que los ensamblados de esta última plataforma eran notoriamente superiores.

2. Duplicación Génica

El análisis del número de duplicaciones en tándem reveló una tendencia clara que apoya la hipótesis general del proyecto:

* **Extensión de la Característica:** Encontramos que las 11 especies de eucalipto analizadas tienen un número significativo y similar de duplicaciones en tándem. En particular, observamos que más del 60% de los genes duplicados (en relación a *E. grandis*) están duplicados en tándem en las especies analizadas. Este número es más alto que lo reportado en otras especies vegetales (5). Esto demuestra que la alta tasa de duplicación génica en tándem parece ser un rasgo evolutivo generalizado en varias especies del género *Eucalyptus*. Ninguna especie mostró ser particularmente estable en términos de duplicaciones en comparación con *E. grandis*.

* **Dinámica Intra- e Interespecífica:** Un hallazgo particularmente interesante fue que el nivel de duplicaciones se mostró similar a nivel interespecífico (entre especies) e intraespecífico (entre individuos de *E. grandis*). Si bien es posible que las duplicaciones intraespecíficas en *E. grandis* hayan sido sobreestimadas por la metodología (la transferencia de anotación con LiftOn puede anotar pseudogenes como genes en genomas muy similares), creemos que este alto dinamismo intraespecífico puede ser un fenómeno biológico real. Análisis futuros más profundos se requieren para confirmar este hallazgo.

* **Duplicaciones Únicas y Compartidas:** Si bien encontramos genes duplicados en común entre todas las especies (algo esperado dado las relaciones filogenéticas de las especies), también identificamos un gran número de duplicaciones que son únicas para cada una de las especies estudiadas.

3. Implicaciones Funcionales y Adaptativas

* **Compuestos Secundarios:** Encontramos un claro enriquecimiento de funciones biológicas relacionadas con la producción de compuestos secundarios, específicamente diterpenos. Estos compuestos son esenciales para la interacción de la planta con el medio ambiente, actuando como defensas químicas, atrayentes de polinizadores o mediadores de señalización. La duplicación de los genes que controlan estas rutas metabólicas permite una mayor producción y/o la evolución de nuevas variantes químicas, lo que explicaría la diversidad fitoquímica observada entre las especies de *Eucalyptus* y su capacidad para adaptarse a distintos nichos ecológicos.

* **Resistencia a Patógenos:** Los resultados también mostraron un fuerte enriquecimiento en genes relacionados con la resistencia a patógenos (e.g., respuesta de defensa a bacterias, oomycetes y otros organismos). La presión constante de enfermedades fúngicas y bacterianas en el hábitat natural de los eucaliptos (Australia) puede explicar las copias génicas adicionales en las rutas de defensa. La retención diferencial de estas duplicaciones podría explicar las distintas capacidades de resistencia observadas en las especies.

Estos hallazgos confirman la hipótesis de que los genes asociados a las duplicaciones están relacionados con la respuesta al estrés biótico y abiótico, y apoyan la idea de que las duplicaciones en tándem son el principal mecanismo que permite a *Eucalyptus* generar la variación necesaria para la adaptación.

Conclusiones y recomendaciones

Del proyecto nacen dos conclusiones importantes, una metodológica y otra sobre el papel de las duplicaciones en eucaliptos. Con respecto a la primera, pudimos definir un pipeline que en cuestión de pocas horas obtiene qué genes están duplicados en la especie analizada para grupos de genes ortólogos previamente establecidos (en nuestro caso *E. grandis*, *C. citriodora* y *Rhodamnia argentea*). Esta aproximación, aunque resultó ser mejor que las otras analizadas que partían de reads cortos o evitaban el ensamblado de las lecturas largas, se puede mejorar en los siguientes aspectos. Por ejemplo, muchas

veces los haplotipos que nacían del ensamblado podían tener distinto número de copias. Esto es un problema inherente al ensamblado de organismos con heterocigosidad alta. Una forma rápida de remediarlo para evitar falsos positivos es exigiendo que ambos haplotipos presenten el mismo número de copias. Por otro lado, cuando anotamos los nuevos genomas en base a *E. grandis*, asumimos que esos genes identificados en la nueva especie son parte del mismo grupo ortólogo, pero esto obviamente es una aproximación. Para remediar esto sería necesario correr nuevamente Orthofinder para reestablecer de manera estricta la ortología, lo que inevitablemente agrega un tiempo considerable al pipeline. Más allá de estas debilidades puntuales y esperadas, creemos que la estrategia tiene un excelente equilibrio entre rapidez y precisión. Puede ser especialmente útil para escanear rápidamente por diferencias en las duplicaciones para cuando se tienen muchas especies o muchos individuos. Sin embargo, se necesita tener genomas bien ensamblados.

En relación a las duplicaciones, encontramos que un número importante de ellas están en tándem y, tanto éstas como las duplicaciones en general, parecen estar reflejando distintos procesos adaptativos en las distintas especies. Un resultado interesante es la dinámica evolutiva intraespecífica que estuvo a la par de las interespecíficas. Aunque este resultado merece un análisis más exhaustivo para confirmarlo, en caso de que efectivamente sea así, puede ser de utilidad para programas de mejoramiento. En particular puede ser interesante vincular ciertos genes duplicados con ciertas características fenotípicas a nivel poblacional.

Referencias bibliográficas

1. Myburg, A. A. et al. The genome of *Eucalyptus grandis*. *Nature* 510, 356–362 (2014).
2. Butler, J. B. et al. Comparative genomics of *Eucalyptus* and *Corymbia* reveals low rates of genome structural rearrangement. *BMC Genomics* 18, 397 (2017).
3. Inglis, P. W., Pappas, M. C. R., Resende, L. V. & Grattapaglia, D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One* 13, e0206085 (2018).
4. Noel, B. et al. Pervasive tandem duplications and convergent evolution shape coral genomes. *Genome Biol.* 24, 123 (2023).
5. Huang, Y. et al. The impact of tandem duplication on gene evolution in *Solanaceae* species. *J. Integr. Agric.* 21, 1004–1014 (2022).

Licenciamiento

Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional. (CC BY-NC-ND)