

RESUMEN

El cacao (*Theobroma bicolor*) es una especie neotropical del género *Theobroma* con potencial agroalimentario y de importante valor para estudios de evolución y mejoramiento genético. Su estrecha relación filogenética con *Theobroma cacao*, la especie de mayor relevancia económica del género, la convierte en un modelo de interés para análisis genómicos comparativos. Sin embargo, los recursos genómicos disponibles para *T. bicolor* son aún limitados. En este estudio se desarrolló y optimizó un flujo de trabajo integral para la generación de recursos genómicos de alta calidad utilizando tecnologías de secuenciación de lecturas largas de Oxford Nanopore Technologies (ONT). Se optimizó un protocolo de extracción de ADN de alto peso molecular basado en CTAB para maximizar la integridad, pureza y rendimiento del ADN, y posteriormente se ajustó la preparación de librerías mediante el kit Native Barcoding (SQK-NBD114.24) y enzimas de New England Biolabs con el fin de mejorar la obtención de lecturas largas de alta calidad. Los datos obtenidos permitieron generar un ensamblaje genómico de alta calidad, el cual fue comparado con el genoma de referencia de *T. cacao* para evaluar el grado de conservación genómica y explorar posibles diferencias estructurales entre ambas especies. Este trabajo establece un protocolo robusto para la obtención y análisis de datos genómicos de *T. bicolor*, proporcionando una base para futuros estudios de diversidad genética, evolución y mejoramiento dentro del género *Theobroma*.

INTRODUCCIÓN

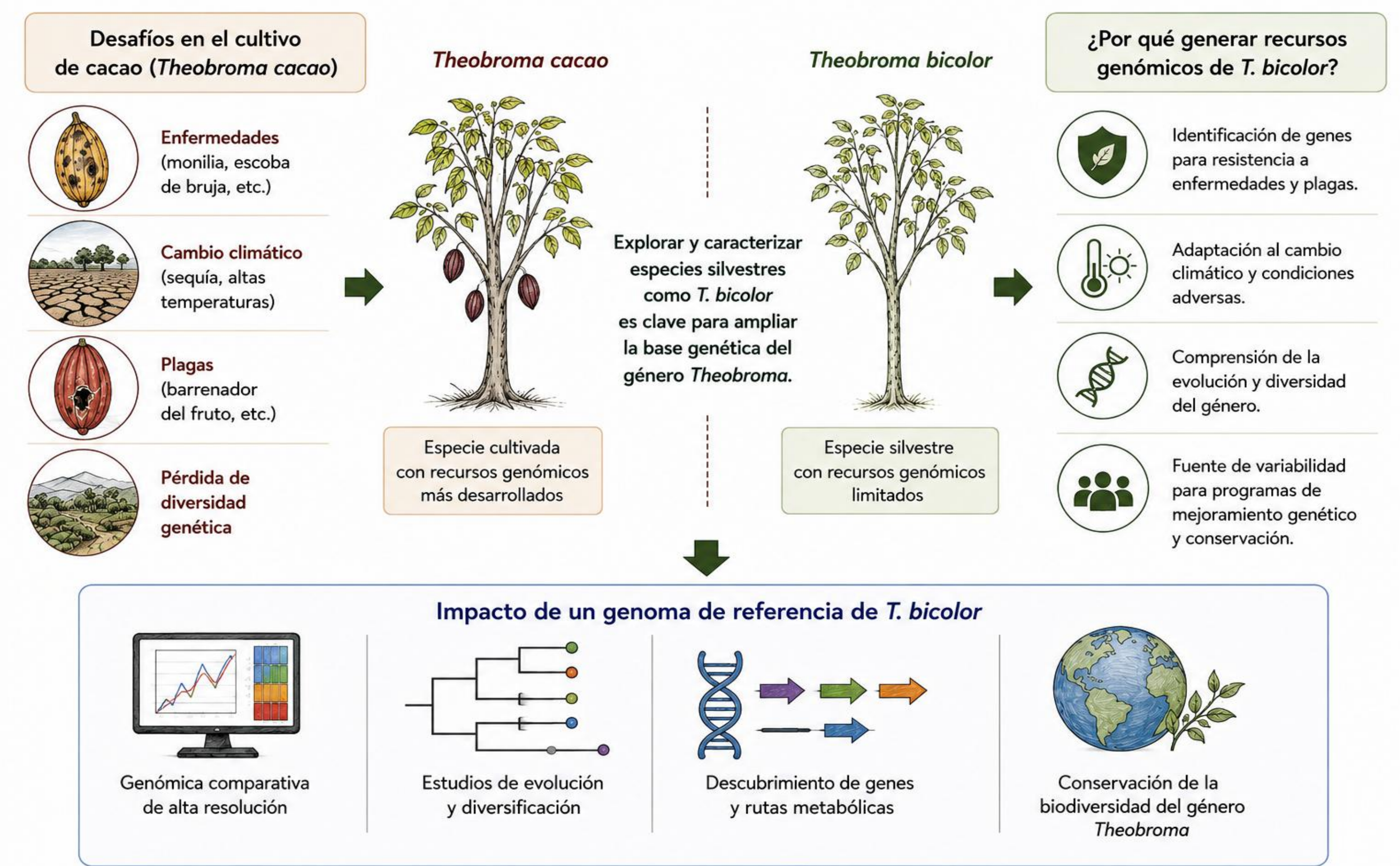


Fig 1. Relevancia de *Theobroma bicolor* como fuente de recursos genómicos para estudios de evolución, diversidad genética y genómica comparativa del género *Theobroma*

MATERIALES Y MÉTODOS

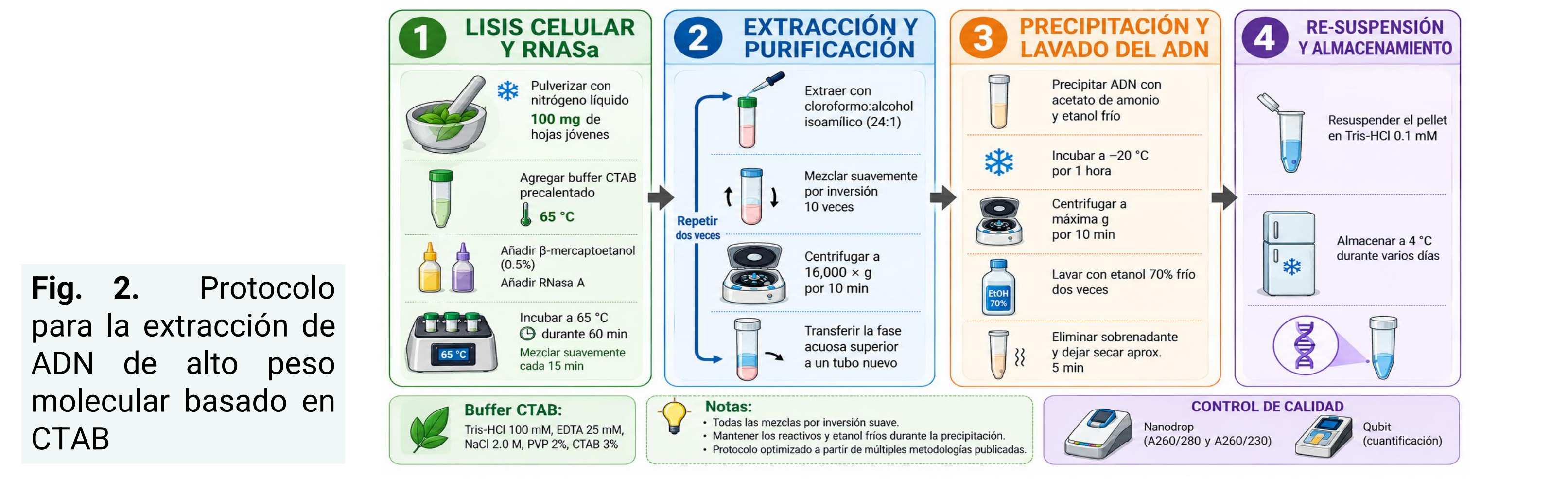


Fig. 2. Protocolo para la extracción de ADN de alto peso molecular basado en CTAB

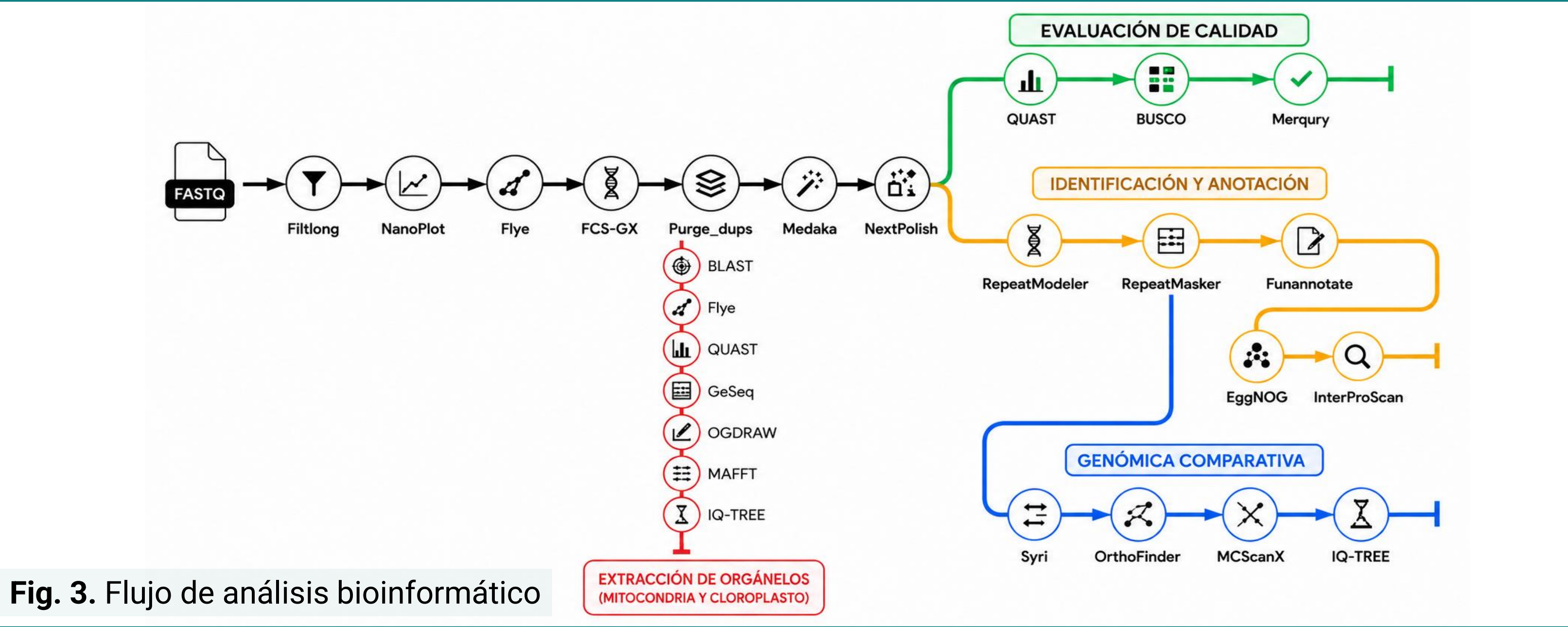


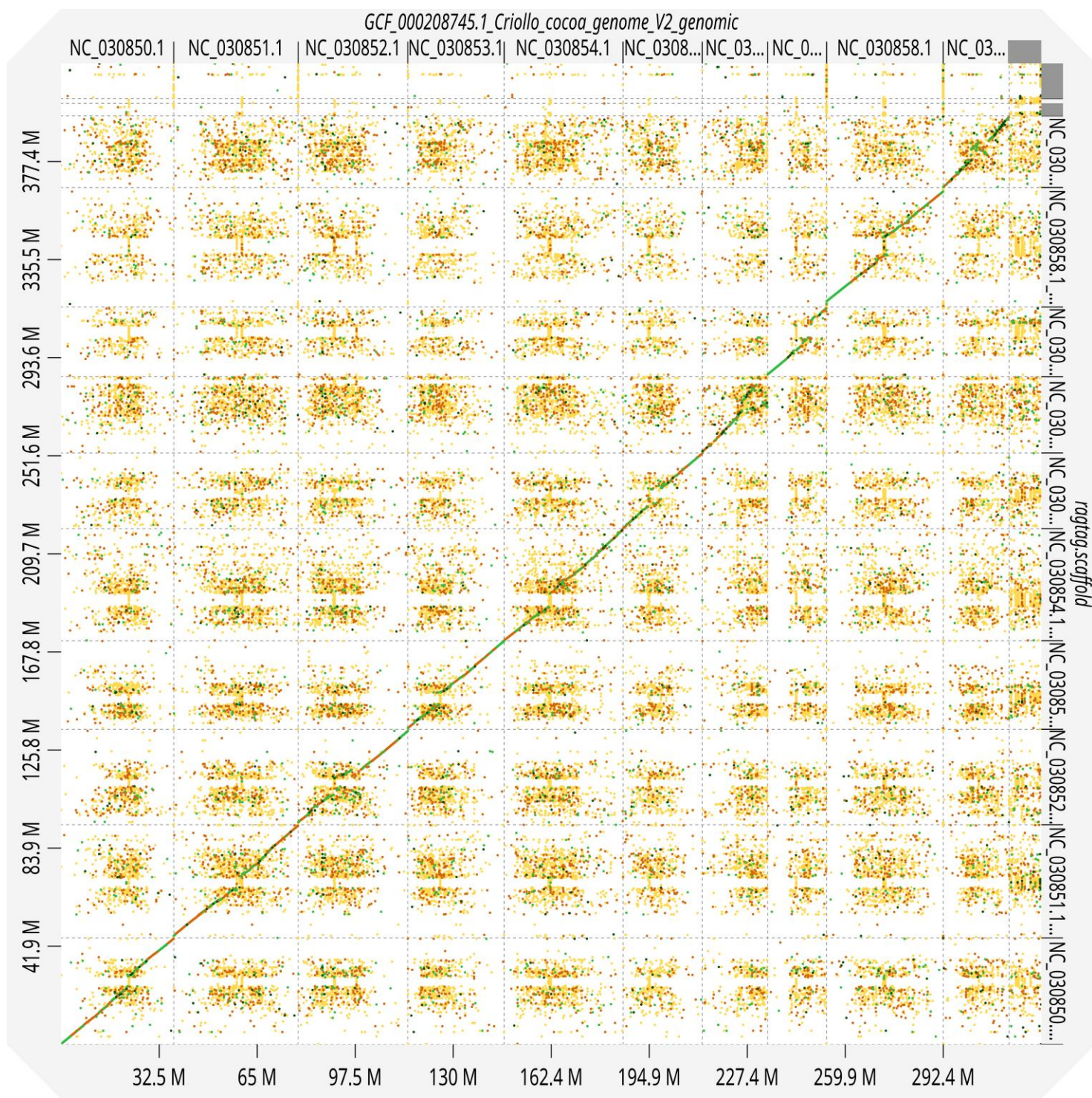
Fig. 3. Flujo de análisis bioinformático

RESULTADOS

Evaluación de la calidad y completitud del ensamblaje genómico

NANOPLOT	DATOS CRUDOS	DATOS FILTRADOS
Longitud media de lectura	6,901	17,303
Calidad media de lectura (Q-score)	17	19
Longitud mediana de lectura	2,865	10,251
Calidad mediana de lectura (Q-score)	19	21
Número de lecturas	3,526,711	1,155,851
N50 de lecturas	19,477	26,520
Bases totales secuenciadas	24,339,038,172	20,000,002,468

ESTADÍSTICAS	ENSAMBLAJE FLYE	POLISH MEDAKA	POLISH NEXTPOLISH
# contigs	310	300	300
# contigs (>= 0 bp)	312	302	302
# contigs (>= 1000 bp)	307	297	297
# contigs (>= 5000 bp)	294	284	284
# contigs (>= 10000 bp)	242	232	232
# contigs (>= 25000 bp)	183	175	175
# contigs (>= 50000 bp)	148	144	144
Contig más largo	45,302,138	45,307,160	45,307,281
Tamaño total	419,907,246	419,383,812	419,376,607
N50	18,273,095	18,275,069	18,274,479
N90	2,072,254	2,072,440	2,072,591
L50		8	
L90		30	
GC (%)		33	



BUSCO
Completos (C): 99.8%
Únicos (S): 98.6%
Duplicados (D): 1.1%
Fragmentados (F): 0.1%
Ausentes (M): 0.2%
Grupos evaluados (n): 1614

Fig 4. Comparación genómica entre el ensamblaje de *Theobroma bicolor* y el genoma de referencia de *Theobroma cacao* (variedad Criollo) mediante D-Genies. La visualización muestra la correspondencia entre regiones cromosómicas, permitiendo evaluar la conservación de la estructura genómica y la relación sinténica entre ambas especies.

Análisis de genómica comparativa de sintenia y ortogrupos en especies relacionadas

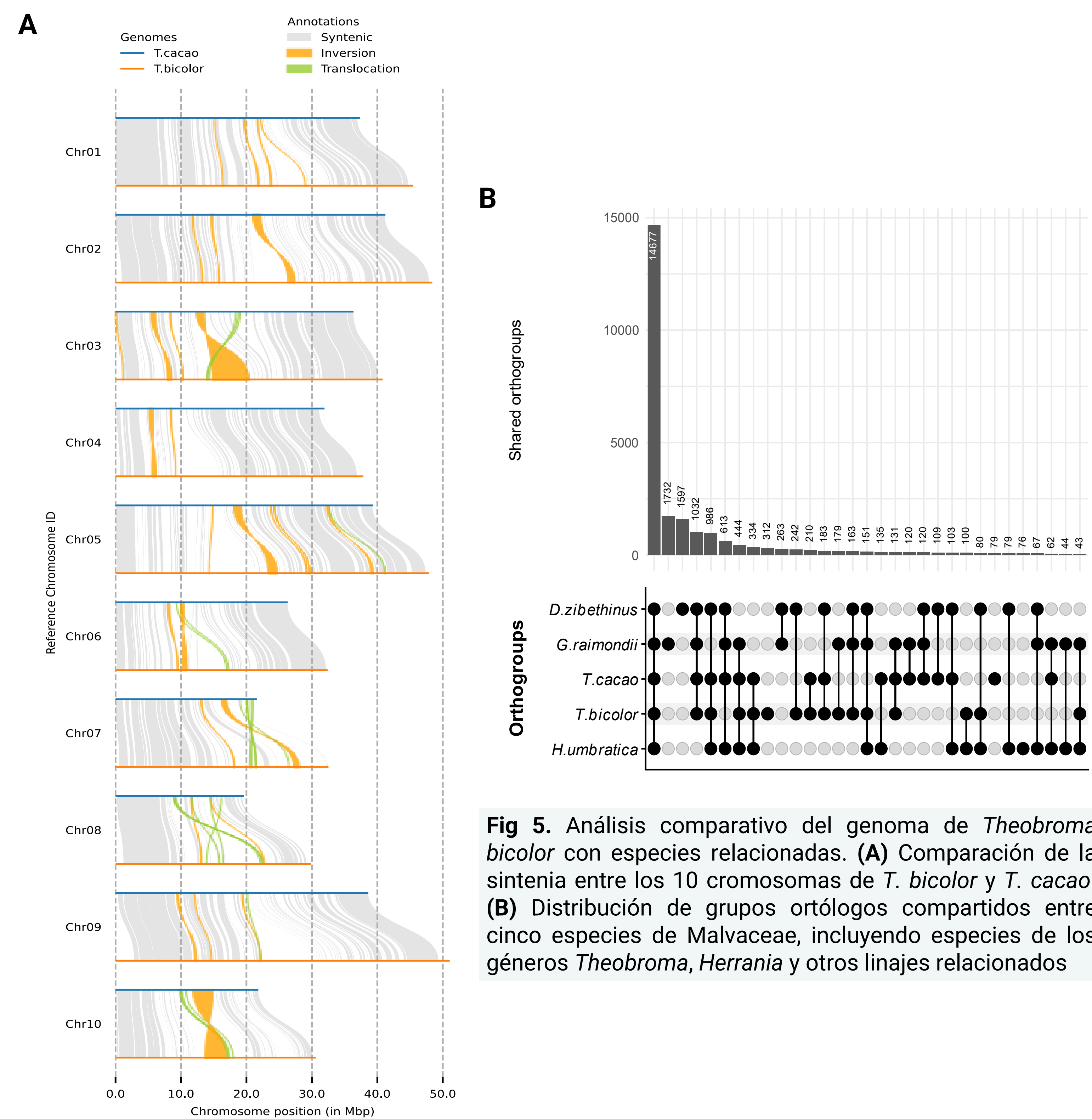


Fig 5. Análisis comparativo del genoma de *Theobroma bicolor* con especies relacionadas. (A) Comparación de la sintenia entre los 10 cromosomas de *T. bicolor* y *T. cacao*. (B) Distribución de grupos ortólogos compartidos entre cinco especies de Malvaceae, incluyendo especies de los géneros *Theobroma*, *Herrania* y otros linajes relacionados

CONCLUSIONES

- La optimización del protocolo de extracción de ADN basado en CTAB y del flujo bioinformático estableció una estrategia robusta para la generación de recursos genómicos en *Theobroma bicolor*.
- El ensamblaje generado constituye un recurso genómico de referencia para *Theobroma bicolor* y facilita estudios comparativos con *Theobroma cacao*.
- Este trabajo amplía los recursos genómicos disponibles para el género *Theobroma* y sienta las bases para futuras investigaciones en evolución, diversidad genética, conservación y mejoramiento.

- Argout, X., Salse, J., Aury, J. M., Guitinan, M. J., Droc, G., Gouzy, J., Allegre, M., Ruiz, M., Abrouk, M., Murat, F., & Lanaud, C. (2011). *The genome of Theobroma cacao*. **Nature Genetics**, **43**, 101–108.
- Tineo, D., Ananco, O., et al. (2024). *Characterization and phylogenetic analysis of the complete plastid genome of Theobroma bicolor (Malvaceae) from Peru*. **Mitochondrial DNA Part B: Resources**, **9**(2), 227–232.