**Título**

Genoma de *Dalbulus maidis*, la chicharrita del maíz, versión actualizable v0.1, INTA Argentina

**Resumen**

1. Introducción

Las plantaciones de maíz en el Norte del país y recientemente en la zona núcleo han enfrentado una amenaza creciente por la proliferación de la chicharrita del Maíz, cuya propagación ha sido exacerbado por el cambio climático. Esta situación ha puesto en riesgo la seguridad alimentaria y la economía del sector agrícola, con estimaciones que señalan pérdidas económicas de más de US$2200 millones para el país debido al último brote de esta plaga y las enfermedades asociadas solo en la última campaña. Ante este escenario crítico, era imperativo avanzar en la generación de información estratégica que complementara las respuestas de manejo inmediatas con enfoques a largo plazo basados en tecnologías de vanguardia. En este contexto, el INTA ha liderado un proyecto en curso de secuenciación, ensamblado y anotación del genoma de *Dalbulus maidis*, desde el CIAP, un centro histórico de referencia en el estudio de este insecto. Aquí presentamos una versión borrador inicial ensamblada del genoma correspondiente a las primeras secuencias masivas generadas y una cronología del proceso:

1.1 Colonia de insectos

Los ejemplares de *Dalbulus maidis* utilizados en este análisis se obtuvieron a partir de una colonia sana propagada en invernadero bajo condiciones controladas de temperatura (25°C) y humedad (60%) en el Instituto de Patología Vegetal (CIAP-INTA), Córdoba, Argentina por Mariana Ferrer y Karina Torrico.

1.2 Extracción de ácidos nucleicos

Para la extracción de ADN genómico (ADNg) de alto peso molecular se utilizó el kit comercial Monarch® HMW DNA Extraction Kit for Tissue (NEB, USA) siguiendo las especificaciones recomendadas por el fabricante. En este trabajo se utilizaron 20 ejemplares machos de *D. maidis* homogenizados en nitrógeno líquido utilizando un mortero y pistilo. La calidad y cantidad de ADNg se evaluó mediante espectrofotometría (Nanodrop-1000, Thermoscientific), flurometría (Quantus, Promega) y mediante electroforesis en geles de agarosa al 1% teñidos con GelRed® Nucleic Acid Gel Stain (Biotium, USA).

1.3 Secuenciación empleando una estrategia híbrida ONT + Illumina

Para la primera etapa de secuenciación se utilizaron 1,2 μg de ADNg para la construcción de librerías en la plataforma ONT (long reads) empleando el kit Ligation sequencing gDNA (SQK-LSK109) siguiendo las instrucciones del fabricante. La librería final fue cuantificada mediante flurometria (Quantus, Promega) y secuenciada en un dispositivo Minion 1Kb y flowcell con química R9.4.1 en el IPAVE-CIAP, Córdoba, Argentina. Los datos crudos fueron procesados (basecall) empleando dorado version 0.7.1 (https://github.com/nanoporetech/dorado) en modo sup. En una segunda etapa se secuenciarán 1ug del mismo ADNg en la plataforma NovaSeq de Illumina (150PE,Truseq DNA Nano 350bp library ).

1.4 Ensamblado *de novo*

Las lecturas provenientes de la primer etapa de secuenciación (ONT) fueron trimmeadas para eliminar secuencias de adaptadores usando Porechop versión 0.2.4 (https://github.com/rrwick/Porechop) y filtradas por calidad (--min\_mean\_q 15) y tamaño (--min\_length 1000) empleando Filtlong v0.2.1 ( https://github.com/rrwick/Filtlong). Las principales estadísticas de las lecturas (pre y post trimming) fueron evaluadas mediante NanoComp versión 1.19.1 (https://github.com/wdecoster/nanocomp). Las lecturas resultantes fueron ensambladas de novo empleando Flye v2.9.1 (https://github.com/fenderglass/Flye) en el servidor del CIAP-INTA. Para el ensamblado se emplearon los siguientes parametros: --nano-hq fastq --genome-size 1g -t 20 --read-error 0.03 --meta -i 4 . El tamaño del genoma se estimó en base a reportes previos de cicadélidos mientras que la opción --meta se utilizó para permitir el ensamble de la mitocondria y posibles organismos simbiontes (bacterias y/o virus). El ensamblado final fue pulido empleando medaka versión 1.12.0 (https://github.com/nanoporetech/medaka) con el modelo r941\_min\_sup\_g507. Las métricas generales del ensamblado se evaluaron empleado Quast versión 5.2.0 en la plataforma https://usegalaxy.eu/. Por otro lado, se evaluó la completitud del ensamblado mediante BUSCO (Mani et al, 2021).

2. Resultados preliminares

2.1 Calidad del ADNg de alto peso molecular

El DNAg extraído cumplió con los requisitos de calidad en de relación 260/280 = 1,97 medidos por Nanodrop y por la visualización en geles de agarosa de una banda intensa sin degradación correspondiente a DNA de alto peso molecular. El rendimiento final medido por Quantus fue de 116 ng/ul.

2.2 Secuenciación y ensamblado

Se realizó una corrida de 48hs totales en la que se ejecutó una recarga (aproximadamente a las 24 horas) de librería para aumentar el rendimiento de la flowcell. En total se obtuvieron 2,397,865 (Total bases: 6,690,474,495) de lecturas crudas con N50 de 5,823 bp y un valor de calidad promedio de 14,1. Luego del procesamiento se obtuvieron 1,398,710 de lecturas (Total bases: 6,151,663,624) con un N50 de 6.782 bp y un valor de calidad promedio de 15,1. El ensamblado (versión Dalbulus\_CIAPv0.1) consistió en 22107 contigs que acumularon un tamaño final de 584.127.104 bases (aproximadamente 580Mb) con un N50 de 50453bp (Tabla 1). En relación al nivel de completitud el análisis llevado a cabo con BUSCO (mode genome) arrojó un valor del 68.6% cuando se utilizó el linaje insecta\_odb10 (number of genomes: 75, number of BUSCOs: 1367) en las comparaciones.

3. Perspectivas

La emergencia en el sector agropecuario debe atenderse no solo con acciones de manejo inmediatas sino también con una mirada estratégica a mediano plazo que favorezca la generación de respuestas novedosas para atender problemáticas. El genoma de este insecto es un recurso invaluable, vacante y necesario para comprender y afrontar este patosistema asociado a la cadena de maíz de gran impacto en el sector. Es importante destacar que este es un proceso de mejora continua. El genoma draft v0.1 de *Dalbulus maidis* depositado en este conjunto de datos se actualizará regularmente como un "genoma viviente", conforme se generen nuevos datos a través de procesos de secuenciación adicionales que se encuentran en desarrollo y los datos de secuenciación cruda serán depositados a medida que se vayan generando en el Sequence Read Archive (SRA) de NCBI dentro del Bioproject específico de este proyecto PRJNA1120083 (BioSample accession SAMN41674197). De esta manera, el INTA reafirma su compromiso con la transparencia y la accesibilidad de los datos científicos para facilitar análisis secundarios y generación de conocimiento comunitaria a través de la liberación inmediata de esta información.

**Autores**

Debat, Humberto Julio1,2; Fernández, Franco1,2

**Filiación**

1Instituto de Patología Vegetal – Centro de Investigaciones Agropecuarias – Instituto Nacional de Tecnología Agropecuaria (IPAVE-CIAP-INTA), Camino 60 Cuadras Km 5,5 (X5020ICA), Córdoba, Argentina.

2Unidad de Fitopatología y Modelización Agrícola – Consejo Nacional de Investigaciones Científicas y Técnicas (UFYMA-CONICET)., Camino 60 Cuadras Km 5,5 (X5020ICA), Córdoba, Argentina.

**Fecha**

2024-6-4

**Cobertura espacial y temporal**

Córdoba, Argentina; mayo-junio 2024

**Proyectos**

Fondos específicos asignados por Dirección Nacional INTA

**Editorial**

Instituto de Patología Vegetal (IPAVE-CIAP-INTA)

**Formato**

FASTA (secuencias)

**Tipo de documento**

Conjunto de datos

**Documentos Relacionados**

https://www.ncbi.nlm.nih.gov/bioproject/1120083

**Palabras Claves**

Dalbulus maidis, Chicharrita, Maíz, Achaparramiento, Genoma, Argentina

**Derechos de acceso**

Abierto